

REGULATION MECHANISM OF CHIKUNGUNYA VIRUS CELL ENTRY PROTEINS

By

BIBEKANANDA SAHOO

LIFE11201304002

**National Institute of Science Education and Research (NISER)
Bhubaneswar**

*A thesis submitted to the
Board of Studies in Life Sciences*

In partial fulfillment of requirements

For the Degree of

DOCTOR OF PHILOSOPHY

Of

HOMI BHABHA NATIONAL INSTITUTE



May, 2020

STATEMENT BY AUTHOR

This dissertation has been submitted in partial fulfillment of requirements for an advanced degree at Homi Bhabha National Institute (HBNI) and is deposited in the Library to be made available to borrowers under rules of the HBNI.

Brief quotations from this dissertation are allowable without special permission, provided that accurate acknowledgement of source is made. Requests for permission for extended quotation from or reproduction of this manuscript in whole or in part may be granted by the Competent Authority of HBNI when in his or her judgment the proposed use of the material is in the interests of scholarship. In all other instances, however, permission must be obtained from the author.

Bibekananda Sahoo

Bibekananda Sahoo

DECLARATION

I, hereby declare that the investigation presented in the thesis has been carried out by me. The work is original and has not been submitted earlier as a whole or in part for a degree / diploma at this or any other Institution /University.

Bibekananda Sahoo
Bibekananda Sahoo

List of Publications arising from the thesis

Journal

1. Conformational changes in Chikungunya virus E2 protein upon heparan sulfate receptor binding explain mechanism of E2-E1 dissociation during viral entry. **Sahoo B** and Chowdary TK, *Bioscience Reports*, **2019**, 39(6), BSR20191077 (**published**)
2. Acidic pH triggered conformational changes in Chikungunya virus fusion protein, E1 – a spring twisted region in domain I-III linker acts as a hinge point for swivelling motion of domains. **Sahoo B**, Gudigamolla NK, and Chowdary TK (**In review**)
3. Bacterially expressed Chikungunya virus cell entry proteins, E1, E2 and E3, are structurally and functionally similar to insect cell expressed or virion surface ‘cleaved-off’ proteins. **Sahoo B** and Chowdary TK (**In preparation**)

Chapters in books and lectures notes

None

Conferences

1. Characterization of Chikungunya virus E2 - receptor interactions. **Sahoo B**, Chowdary TK, (**Jan, 2019**), *The 14th International conference on vectors and vector borne diseases*, ICMR – Regional Medical Research Center, Bhubaneswar, India (**Oral presentation**)
2. Regulation mechanism in Chikungunya virus cell entry proteins. **Sahoo B**, Gautam A, Satyamurthy K, Chowdary TK, (**Dec, 2016**), *National workshop on Practical aspects of membrane protein crystallization*, IISc Bangalore, India (**Poster presentation**)
3. Protein-protein interactions and Structural mechanism in Chikungunya virus cell entry. **Sahoo B**, Gautam A, Chowdary TK, (**Oct, 2014**), *Symposium-cum-Workshop entitled “Frontiers of Structural Biology: New Advances in X-ray diffraction and Cryo-electron Microscopy”*, INSA New Delhi and RCB Gurgaon, India (**Poster presentation**)
4. Protein-protein interactions in Chikungunya virus cell entry. **Sahoo B**, Gautam A, Chowdary TK, (**Aug, 2014**), *International conference and workshop on “Recent Advances in Structural Biology and Drug Discovery”*, IIT Roorkee, India (**Poster presentation**)

Others

None


Bibekananda Sahoo

DEDICATIONS

I dedicate this thesis to my family for their unconditional love and support for my success in life.

ACKNOWLEDGEMENTS

I sincerely thank my mentor Dr. Tirumala Kumar Chowdary for his continuous guidance and motivation during my Ph.D. thesis research. He has been an inspiration and being around him has immensely helped me to grow and develop as a better individual both personally and professionally. I offer my deepest gratitude to him for all his help, support, encouragement and critiques that have brought the best out of me and helped me to reach my goal.

I express my gratitude to our Director for providing the infrastructure and working facility for research at NISER. I am also thankful to my doctoral committee members: Dr. Palok Aich, Dr. Abdur Rahman, Dr. Rudresh Acharya, and Dr. Dileep Vasudevan for their valuable inputs and suggestions during progress monitoring meetings. Further, I thank PGCS convener and Chairperson of school of biological sciences, Dean academics and all other faculties and staff members of NISER for their timely help during my Ph.D.

I would like to thank Prof. Scott Weaver of University of Texas Medical Branch, Galveston for providing us the PCR amplicons of Chikungunya virus (LR 2006 strain) E1 and E2 coding sequences, which I have used in my research work. I am really thankful to Council of Scientific and Industrial Research (CSIR) for financial assistance through JRF and SRF fellowships during my Ph.D. I thank NISER and my mentor for providing financial assistance during the academic extension period.

I am thankful to all my lab mates and friends who have made my stay really joyful at NISER. List is really long, but few names are worth mentioning. My Ph.D. seniors Dr. Uday Singh, Dr. Santosh Kumar, Dr. Anup Narayan Pillai and my friends Omprakash Singh, Khursid-ul Hassan, and Barsa Tripathy are some of my best buddies at NISER. I thank all my lab mates: Mr. Satyamurthy Kundarapu, Mr. Naresh Kumar G, Mr. A Simanchal Dora, Ms. Anupriya Gautam, Mr. Debashis Panda and Mr. Deepak Ranjan Behera, Mr. Sapuru Vinay Kumar, Mr. Subham Kumar Tripathy and Ms. Priti Chanda Behera for all their help and support. Special thanks to Anupriya for her help and company during the beginning days when lab started. Big thanks to Naresh for his time and help towards completion of my thesis work over last one year. Thanks to Satya for his help, helpful discussions on work and also for his time on and off the lab.

I truly thank my wife Dr. Amita Rani Sahoo for her help and moral support during this period. My appreciation goes to my brother, my sister and all other family members for their motivation and extensive support at every step during my Ph.D. Last but not the least, I am very much thankful to my parents for their blessings and persistent moral support towards my success during Ph.D.

I thank 'God', above all, for giving me the strength to achieve this difficult milestone.

Homi Bhabha National Institute

Recommendations of the Viva Voce Committee

As members of the Viva Voce Committee, we certify that we have read the dissertation prepared by **Bibekananda Sahoo** entitled "**Regulation mechanism of Chikungunya virus cell entry proteins**" and recommend that it may be accepted as fulfilling the dissertation requirement for the Degree of Doctor of Philosophy.

Chair - Dr. Palok Aich

Date: 15/05/2020

Palok Aich

Guide/Convener - Dr. Tirumala Kumar Chowdary

Date: 15/05/2020

Ch. Tirumala Kumar Chowdary

Examiner - Prof. Balasubramanian Gopal

Date: 15/05/2020

Balasubramanian Gopal

Member 1 - Dr. Rudresh Acharya

Date: 15/05/2020

Rudresh Acharya

Member 2 - Dr. Abdur Rahaman

Date: 15/05/2020

Abdur Rahaman

Member 3 - Dr. Dileep Vasudevan

Date: 15/05/2020

Dileep Vasudevan

Final approval and acceptance of this thesis is contingent upon the candidate's submission of the final copies of the thesis to HBNI.

I hereby certify that I have read this thesis prepared under my direction and recommend that it may be accepted as fulfilling the thesis requirement.

Date: 15/05/2020

Place: NISER

Ch. Tirumala Kumar Chowdary
Guide

CONTENTS

	Page No.
SUMMARY	I
LIST OF FIGURES	III
LIST OF TABLES	VIII
ABBREVIATIONS	IX
GENERAL INTRODUCTION	1
Thesis research hypothesis and objectives	46
<i>CHAPTER 1 EXPRESSION AND PURIFICATION OF CHIKV CELL ENTRY PROTEINS FOR STRUCTURE-FUNCTION STUDIES</i>	
1.1 INTRODUCTION	47
1.2 MATERIALS AND METHODS	52
1.2.1 Chikungunya virus envelope protein sequence	52
1.2.2 Chikungunya virus E3-E2-E1 structure coordinates used for analyses	52
1.2.3 Preparation of constructs for mammalian, insect cells and <i>E. coli</i> expression	52
1.2.4 Mammalian cell expression and purification	55
1.2.5 Insect cell expression and purification	55
1.2.6 Bacterial expression and purification	57
1.2.7 Size exclusion chromatography	61
1.2.8 Anti-E1 and anti-E3-E2-epitope specific pAb generation	61
1.2.9 CD spectroscopy	62
1.2.10 Fluorescence Spectroscopy	62
1.2.11 Glutaraldehyde cross-linking assay	63
1.2.12 Lipid co-floatation assay	64
1.2.13 Pyrene based lipid association assay	65
1.2.14 Fluorescein based liposome content mixing assay	66
1.2.15 Negative stain electron microscopy	68
1.3 RESULTS	69
1.3.1 Mammalian cell expression of CHIKV cell entry proteins	69

1.3.2	Insect cell expression and purification of CHIKV cell entry proteins	70
1.3.3	Bacterial expression and purification of CHIKV cell entry proteins	72
1.3.4	Structural characterization of <i>E. coli</i> purified CHIKV cell entry proteins	75
1.3.5	Functional characterization of <i>E. coli</i> purified CHIKV envelope proteins	79
1.3.6	Testing E3-E2 and E1 heterodimer formation with <i>E. coli</i> expressed CHIKV cell entry proteins	91
1.4	DISCUSSION	93
CHAPTER 2 CHARACTERIZATION OF CHIKV E2 - HS INTERACTIONS TO EXPLAIN RECEPTOR BINDING REGULATION OF ENTRY		
2.1	INTRODUCTION	97
2.2	MATERIALS AND METHODS	100
2.2.1	Molecular docking studies	100
2.2.2	Search for HBDs on E2 sequence of alphaviruses	101
2.2.3	Cloning, expression and purification of CHIKV E2, E3-E2 and E3-E2 mutant proteins	103
2.2.4	HS/HEP binding assays	103
2.2.5	Microscale thermophoresis assay	104
2.2.6	Molecular Dynamics simulations of E2 in presence or absence of HS	105
2.2.7	Far-UV CD, and intrinsic tryptophan fluorescence spectroscopy	106
2.2.8	Fluorescence resonance energy transfer (FRET) to monitor domain movements in E3-E2 upon HS binding	107
2.3	RESULTS	108
2.3.1	Heparan sulfate binds to a positively charged pocket on CHIKV E2 at the interface between domain A and arch 1 of β -ribbon connector	108
2.3.2	HS binds to CHIKV E2 at predicted HBD	113
2.3.3	HS binding to E2 results in domain C movement in, towards E1	120
2.4	DISCUSSION	129
CHAPTER 3 CHARACTERIZATION OF CHIKV E1 DI-DIII INTERFACE MUTANTS TO EXPLAIN ACIDIC PH TRIGGERED CONFORMATIONS DURING ENTRY		
3.1	INTRODUCTION	135

3.2	MATERIALS AND METHODS	140
3.2.1	Sequence and structure analysis of CHIKV E1	140
3.2.2	Constant pH molecular dynamics (MD) simulation on E1 at pH 5.5	140
3.2.3	Cloning, expression and purification of CHIKV E1-WT and mutant proteins	141
3.2.4	Far-UV circular dichroism and intrinsic tryptophan fluorescence spectroscopy	141
3.2.5	Size exclusion chromatography	141
3.2.6	Lipid co-floatation assay	141
3.2.7	Pyrene excimer fluorescence assay for studying E1-membrane association	142
3.2.8	Liposome content mixing assay	142
3.3	RESULTS	143
3.3.1	A conserved arginine-aspartic acid/glutamic acid pair in alphavirus E1 dI-III linker	143
3.3.2	A conserved spring-twisted structure in the dI-III linker is untwisted in pre- to post-fusion conformation change	145
3.3.3	MD simulations of E1 at acidic pH show dI-II rotation on dIII over the linker spring-twisted hinge	146
3.3.4	Purified mutant ectodomain proteins are functionally active in membrane association and membrane fusion assays	150
3.3.5	Domain I and III separation at acidic pH is essential for fusion activity of E1, and domain I-III interface breaking is not triggered at the conserved His331	157
3.4	DISCUSSION	161

CHAPTER 4 *CHARACTERIZATION OF CHIKV E3 - MEMBRANE INTERACTIONS TO EXPLAIN ROLE OF E3 IN REGULATION OF ENTRY*

4.1	INTRODUCTION	167
4.2	MATERIALS AND METHODS	170
4.2.1	E3 sequence and structure analysis	170
4.2.2	Cloning, expression and purification of CHIKV E3 F15W mutant protein	170
4.2.3	SDS-PAGE, Chloroform staining, and Western blotting	170

4.2.4	Far-UV CD and intrinsic tryptophan fluorescence analysis of E3 structural changes upon interactions with liposomes	171
4.2.5	Size exclusion chromatography for studying E3-liposome interactions	172
4.2.6	Lipid co-floatation assay for studying E3-liposome interactions	172
4.2.7	Liposome content release assay	172
4.3	RESULTS	173
4.3.1	E3 sequence and structure analysis reveal key interactions at E3-E2 interface	173
4.3.2	Interaction of E3 with membranes	174
4.3.3	E3 does not interact with liposome membrane	178
4.3.4	E3's N-terminal hydrophobic patch neither inserts into lipid bilayer nor does it undergo structural transition in presence of LUVs	184
4.4	DISCUSSION	187
CHAPTER 5 DEVELOPMENT OF A BACULOVIRUS PSEUDOVIRION SYSTEM FOR USE IN FUNCTIONAL ENTRY ASSAYS		
5.1	INTRODUCTION	189
5.2	MATERIALS AND METHODS	196
5.2.1	Lentiviral packaging of CHIKV envelope proteins	196
5.2.2	Construction of 'gp64-start+Cm-GFP+gp64-end' cassette for recombination	196
5.2.3	Maintenance of pRED/ET recombination plasmid inside <i>E. coli</i> DH10Bac cells	198
5.2.4	Preparation of electro-competent pRED/ET bearing <i>E. coli</i> DH10Bac cells	200
5.2.5	Homologous recombination of "gp64-start+Cm-GFP+gp64-end" cassette with baculovirus genome	200
5.2.6	Preparation of viral envelope protein coding constructs and their transposition into Δ gp64-Cm-GFP bacmid genome	201
5.2.7	Transfection of purified recombinant bacmids into sf9 cells for pseudo-virus production	202
5.2.8	Syncytia assay	202
5.2.9	Pseudovirion transduction/cell entry assay	203
5.2.10	Immunofluorescence assay to check surface expression of CHIKV cell entry proteins in sf9 cells using CHIKV E1-pAb	204

5.3	RESULTS	205
5.3.1	Production of CHIKV pseudo viruses using lentiviral packaging system	205
5.3.2	Development of gp64 knockout and Cm-GFP knock in bacmid genome	206
5.3.3	Generation of pseudo viruses with BV gp64 and VSV G proteins	208
5.3.4	Testing of BV gp64 and VSV G pseudotyped virions in functional assays	211
5.3.5	Checking surface expression of CHIKV envelope proteins in sf9 cells	213
5.3.6	Optimization of BV pseudo typing with CHIKV envelope proteins	214
5.4	DISCUSSION	216
	CONCLUDING REMARKS	219
	APPENDICES	224
	REFERENCES	229

LIST OF FIGURES

Name	Title	Page No.
Figure 0.1	Enveloped virus entry by receptor binding followed by virus-host cell membrane fusion.	3
Figure 0.2	Influenza virus HA1 and HA2 envelope proteins are made from the HA0 precursor.	4
Figure 0.3	HIV gp120 and gp41 envelope proteins are made from the gp160 precursor.	4
Figure 0.4	Schematic representation of virus-cell membrane fusion mediated by viral envelope glycoproteins	8
Figure 0.5	A hypothetical energy diagram (free energy vs. fusion progress graph) of membrane fusion process.	10
Figure 0.6	Structural organization of class I, II and III viral fusion proteins.	11
Figure 0.7	Receptor binding regulates fusion protein complex conformational changes and activity during cell entry of HIV1.	15
Figure 0.8	Regulation mechanism of IAV HA structural mechanism during cell entry of IAV.	18
Figure 0.9	CHIKV is a mosquito-transmitted virus, which causes CHIKV fever and is spread globally.	21
Figure 0.10	CHIKV virus particle structure and genome organization. (A) Virus particle structure with diameter in nanometers is shown.	23
Figure 0.11	Scheme of alphavirus structural polyprotein production from sub-genomic RNA followed by processing through secretory pathway to heterodimer conformation.	26
Figure 0.12	Comparison of the immature (pE2) and mature (E3-E2) structures.	28
Figure 0.13	Domain organization of CHIKV E1 protein. The structure is extracted	29

from CHIKV E3-E2-E1 heterodimer structure in pre-fusion state (PDB: 3N42).

Figure 0.14	Domain organization of mature CHIKV E2 protein.	31
Figure 0.15	Pre-fusion conformation of CHIKV E3-E2-E1 heterodimer and trimer of E2-E1 heterodimer in CHIKV surface spike.	33
Figure 0.16	SFV E1-homotrimeric structure in post-fusion state from SFV (PDB: 1RER) and CHIKV (model structure).	34
Figure 0.17	Acidic pH induced fusion of SINV-liposome fusion	36
Figure 0.18	CHIKV envelope proteins pre- to post-fusion conformation switch.	40
Figure 0.19	Interactions known to affect HT formation.	44
Figure 1.1	Mammalian-cell expression construct of full-length CHIKV envelope proteins.	69
Figure 1.2	Insect cell expression and purification of CHIKV envelope proteins.	71
Figure 1.3	Bacterial expressions of CHIKV envelope proteins in <i>E. coli</i> BL21 cells	73
Figure 1.4	Bacterial expression and purification of CHIKV envelope proteins.	74
Figure 1.5	Far-UV CD spectra analysis on <i>E. coli</i> purified CHIKV envelope ectodomain proteins.	76
Figure 1.6	Intrinsic tryptophan fluorescence spectra analysis on <i>E. coli</i> purified CHIKV envelope proteins.	78
Figure 1.7	Quaternary structure analyses of <i>E. coli</i> purified CHIKV E1 WT ectodomain protein.	80
Figure 1.8	Analysis of liposome integrity and E1 secondary structure in presence of SDS.	81
Figure 1.9	Crosslinking of <i>E. coli</i> purified CHIKV E1 WT ectodomain protein.	83
Figure 1.10	Lipid-floatation analysis of <i>E. coli</i> purified CHIKV E1 WT	85

ectodomain protein.

Figure 1.11	Lipid association analysis of <i>E. coli</i> purified CHIKV E1 WT ectodomain protein using Pyrene based fluorescence assay.	86
Figure 1.12	Membrane fusion analysis of <i>E. coli</i> purified CHIKV E1 WT ectodomain protein using fluorescein based liposome fusion assay.	87
Figure 1.13	TEM analyses on negatively stained <i>E. coli</i> purified CHIKV E1 WT ectodomain protein.	89
Figure 1.14	Heterodimer formation analysis on <i>E. coli</i> expressed and purified CHIKV E3-E2 and E1 ectodomain proteins.	92
Figure 2.1	Docking of HS onto CHIKV E2-E1 trimeric spike using AutoDock Vina.	109
Figure 2.2	Search for Alphavirus E2-HS binding consensus motif using MEME.	111
Figure 2.3	Structural analysis of HS-binding motif (HBDs) regions on E2 structure from CHIKV.	112
Figure 2.4	Docking of protein receptors and heparin tetra-saccharide onto CHIKV E2-E1 trimeric spike using ClusPro server.	113
Figure 2.5	Characterization of E2-HS/HEP interactions.	115
Figure 2.6	Characterization of E3-E2-HEP interactions.	116
Figure 2.7	Characterization of CHIKV E3-E2 docking site-HBD mutant proteins.	117
Figure 2.8	Characterization of HEP binding property of CHIKV E3-E2 mutant proteins.	118
Figure 2.9	Analysis of HEP binding property of CHIKV E3-E2 mutant proteins.	118
Figure 2.10	E3-E2 and HS binding affinity measurement through Microscale thermophoresis assay.	119
Figure 2.11	Analysis of MD simulation trajectory of E2-HS complex.	121

Figure 2.12	RMSD and RMSF analysis of MD simulation trajectory of E2-HS complex.	122
Figure 2.13	Tertiary structure analysis on HS bound E2 and E3-E2.	123
Figure 2.14	Secondary structure analysis on HS bound E2 and E3-E2.	124
Figure 2.15	Analysis of E3-E2 structure for FRET pair introduction.	125
Figure 2.16	Characterization of CHIKV E3-E2 double cysteine mutant protein.	126
Figure 2.17	Characterization of HEP binding property of CHIKV E3-E2 double cysteine mutant protein.	127
Figure 2.18	FRET analysis on HS bound E2 and E3-E2 structure.	128
Figure 2.19	Comparison of reported MXRA8 and HS binding sites on alphavirus spike to HS docked site on CHIKV E2 from this study.	131
Figure 2.20	Mechanism for alphavirus E2-E1 dissociation upon receptor binding.	134
Figure 3.1	CHIKV envelope proteins pre- to post-fusion conformation switch.	136
Figure 3.2	Analysis of alphavirus E1 dI-III linker region in pre- and post-fusion state.	144
Figure 3.3	Structure analysis of pre-fusion (PDB: 3N42).	148
Figure 3.4	Purification and characterization of E1 dI-dIII interface mutants.	151
Figure 3.5	Functional characterization of E1 WT and cystine staple mutant in liposome co-floatation and pyrene excimer fluorescence assays.	154
Figure 3.6	Analyses of membrane interaction properties of CHIKV E1 WT and H331D/E mutant proteins in lipid-floatation, and pyrene excimer fluorescence assay.	155
Figure 3.7	Comparison of CHIKV E1 percent secondary structural contents predicted from far-UV CD data and calculated from crystal structure.	157
Figure 3.8	Liposome fusion assay for E1.	159
Figure 3.9	Analyses of membrane fusion by CHIKV E1 WT and H331D/E	161

mutant proteins in lipid-mixing assay.

Figure 4.1	CHIKV E3 sequence and structure analysis.	175
Figure 4.2	CHIKV E3 sequence analysis for presence of membrane interactive region(s).	176
Figure 4.3	CHIKV E3 structure analysis.	177
Figure 4.4	Purification and characterization of CHIKV E3F15W mutant.	179
Figure 4.5	Intrinsic tryptophan fluorescence spectra analysis of E3-liposome interactions.	180
Figure 4.6	Size-exclusion chromatography analysis of E3-liposome interactions.	182
Figure 4.7	Lipid-floatation analysis of E3-liposome interactions.	183
Figure 4.8	Characterization of E3-N-terminus insertion into liposomes.	185
Figure 5.1	Scheme showing overview of baculovirus (BV) pseudovirion development plan.	195
Figure 5.2	Homologous recombination cassette.	197
Figure 5.3	Lentiviral pseudo-typing of full-length CHIKV envelope proteins.	206
Figure 5.4	Steps for generation of recombinant bacmids through homologous recombination approach.	207
Figure 5.5	Homologous recombination verification.	208
Figure 5.6	Transposition reaction procedure.	209
Figure 5.7	Verification of transposition reaction.	210
Figure 5.8	Syncytia assay.	212
Figure 5.9	Cell entry assay.	212
Figure 5.10	Immunofluorescence assay using E1-pAb on CHIKV envelope protein expressing sf9 cells.	214
Figure 5.11	Testing of CHIKV pseudo type in syncytia and transduction assay.	215

LIST OF TABLES

Name	Title	Page No.
Table 1.1	Optimum conditions for expression and solubilization of CHIKV cell entry proteins.	59
Table 1.2	Comparison of percent secondary structural content from far-UV data and crystal structure.	77
Table 1.3	Intrinsic fluorescence spectra analysis on <i>E. coli</i> purified CHIKV cell entry proteins.	78
Table 2.1	412nm absorbance values of BSA, E3-E2 WT and S154C+S296C mutant proteins and the buffer control from the Ellman's assay.	127
Table 3.1	Comparison of dihedral angles of "RVVD" residues from the dI-dIII linker region between $t=0$ ns and $t=23.37$ ns structures from MD simulation studies.	149
Table 3.2	Ellman's assay results with purified CHIKV E1WT and cystine staple mutant proteins.	153
Table 4.1	Key interactions at E3-E2 interface involve charged residues.	173
Table a1.1	List of overlapping primers for E3-gene synthesis.	225
Table a1.2	List of overlapping primers for 6K-gene synthesis.	226
Table a1.3	<i>E. coli</i> , insect cell and mammalian cell expression construct details.	227
Table a1.4	Homologous recombination cassette and heterologous envelope protein coding construct details.	229

ABBREVIATIONS

AF	Alexa Fluor
Amp	Ampicillin
APS	Ammonium Persulfate
ATP	Adenosine triphosphate
ATPS β	ATP Synthase β
AURAV	Aura virus
BFV	Barmah forest virus
Bp	Base pair
BSA	Bovine Serum Albumin
BV	Baculovirus
CBB	Coomassie Brilliant Blue
CCR5	C-C chemokine receptor type 5
CD	Circular dichroism
CD4	Cluster of Differentiation 4
CD64	Cluster of Differentiation 64
CDS	cDNA coding sequence
CHIKV	Chikungunya virus
Chol	Cholesterol
Cm	Chloramphenicol
Cryo-EM	Cryo-Electron microscopy
CXCR4	C-X-C chemokine receptor type 4
DAPI	4',6-diamidino-2-phenylindole
DC-SIGN	Dendritic Cell-Specific Intercellular adhesion molecule-3-Grabbing Non-integrin
DENV	Dengue virus
DGS-NTA	1,2-dioleoyl-sn-glycero-3-[(N-(5-amino-1-carboxypentyl)iminodiacetic acid)succinyl] (nickel salt)
DMEM	Dulbecco's Modified Eagle Medium

DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
DPBS	Dulbecco's phosphate-buffered saline
d.p.i	Days post infection
d.p.t	Days post transfection
<i>E. coli</i>	Escherichia coli
ECL	Enhanced chemiluminescence
ECM	Extracellular Matrix
ecto	Ectodomain
EDTA	Ethylene diamine tetra acetic acid
EEEV	Eastern equine encephalitis virus
EM	Electron microscopy
ER	Endoplasmic reticulum
FBS	Fetal bovine serum
FGF	Fibroblast growth factors
fl	Full-length
FL	Fusion loop
FP	Forward primer
FRET	Förster resonance energy transfer
GA	Glutaraldehyde
GAGs	Glycosaminoglycans
Gent	Gentamycin
GETV	Getah virus
GFP	Green Fluorescence Protein
Gm	Gram (mass unit)
Gp/gp/G	Glycoprotein
GTP	Guanosine-5'-triphosphate
GUI	Graphical user interface
HA	Hemagglutinin

Hairpin	Post-fusion homotrimer of viral envelope fusion protein
HBD	HS/Heparin binding domain/motif
HEP	Heparin
HIV	Human immunodeficiency virus
Hr	Hour
HR	Heptad repeats
HRP	Horseradish peroxidase
HS	Heparan sulfate
HT	Homotrimer
IAV	Influenza A virus
IF	Immunofluorescence
IgG	Immunoglobulin G
IPTG	Isopropyl- β -D-thiogalactoside
Kan	Kanamycin
Kb	Kilo base
KDa	Kilo Dalton
LB	Luria-Bertani
L-SIGN	Liver/lymph node-specific intercellular adhesion molecule-3-grabbing integrin
Ltr	Liter
LUV	Large, Unilamellar Vesicles
M	Molar
mAb	Monoclonal antibody
MAYV	Mayaro virus
MFP	Membrane fusion protein
Mg	Milligram
MIDDV	Middelburg virus
ml	Milliliter
mM	Millimolar

MOI	Multiplicity of infection
MRE	Mean residue ellipticity
MSA	Multiple sequence alignment
MST	Microscale thermophoresis
NaAc	Sodium Acetate
NaCl	Sodium Chloride
NCM	Nitrocellulose membrane
NCP1-C	Niemann-Pick type C1
NHS	N-Hydroxysuccinimide
nm	Nano meter
NTP	Nucleotide triphosphate
OD	Optical Density
ONNV	O'nyong-nyong virus
ORF	Open reading frame
POPA	1-palmitoyl-2-oleoyl-sn-glycero-3-phosphate
pAb	Polyclonal antibody
PCR	Polymerase chain reaction
POPC	1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine
POPE	1-palmitoyl-2-oleoyl-sn-glycero-3-phosphoethanolamine
PFA	Paraformaldehyde
PHB	Prohibitin
PMSF	Phenylmethylsulfonyl fluoride
PS	Penicillin, Streptomycin
PSA	Penicillin, Streptomycin, Amphotericin B
R0	Förster distance
RBP	Receptor binding protein
RIPA	Radioimmunoprecipitation assay
RMSD	Root mean square deviation
RMSF	Root mean square fluctuation

RP	Reverse primer
RPM	Rotation per minute
RRV	Ross river virus
RT	Room temperature
SAGV	Sagiyama virus
SDS-PAGE	Sodium dodecyl sulfate – poly acrylamide gel electrophoresis
SDV	Sleeping disease virus
<i>Sf9</i>	Spodoptera frugiperda 9
SFM	Serum free media
SFV	Semliki forest virus
SINV	Sindbis virus
SiRNA	Short interfering RNA
SNARE	Soluble NSF attachment protein receptor
SPDV	Salmon pancreas disease virus
SUV	Small, Unilamellar Vesicles
T20	Enfuvirtide
TAE	Tris-acetate-EDTA
TBEV	Tick-borne encephalitis virus
TBS	Tris buffered saline
TBST	Tris buffered saline plus tween20
TEM	Transmission Electron microscopy
TEMED	Tetramethylethylenediamine
TIM1	T-cell immunoglobulin and mucin domain 1
TM	Trans membrane
TN	Tris-NaCl
TNE	Tris-NaCl-EDTA (1mM)
TNEP	Tris-NaCl-EDTA (1mM)-PMSF (100µM)
TX-100	Triton X 100
UV	Ultraviolet

VEEV	Venezuela eastern equine encephalitis virus
VLP	Virus like particles
VSV	Vesicular stomatitis virus
w.r.t.	With respect to
WB	Western blot
WEEV	Venezuela eastern equine encephalitis virus
X-gal	5-bromo-4-chloro-3-indolyl- β -D-galactopyranoside
λ	Wavelength
λ_{em}	Emission wavelength
λ_{ex}	Excitation wavelength

CHAPTER
CONCLUDING REMARKS

Concluding remarks

Viruses are intriguing biological entities. Viral molecular machines evolved to accomplish biological processes using simple organization and without energy requirement – many of these processes require multiple proteins and chemical energy in cellular realm of life. Enveloped viral proteins involved in cell entry are an epitome of the intricate design of viral molecular machines. In cellular form of life membrane-membrane fusion and its regulation involves several proteins and cellular signals. Viruses achieve membrane fusion, and in a very regulated manner using one or two proteins from cell surface. Understanding structure-function relationships of viral entry proteins and explaining its mechanism has potential for designing therapeutic intervention.

Despite several years of studies, structure-function relationships and understanding on mechanism of viral proteins is far from complete – especially in case of several reemerging viruses such as Chikungunya virus and Dengue virus. My Ph.D. thesis work focused on understanding the regulation of structural mechanism of CHIKV cell entry proteins. The surface proteins (E3-E2-E1 complex) of the enveloped virus undergo dynamic conformational changes to mediate fusion of virus and host cell membrane during cell entry. Cellular regulatory conditions such as receptor binding, acidic pH, membrane interactions etc. are known to regulate the conformational dynamics of each of these proteins that in turn govern their coordinated functioning during cell entry. I aimed to explain the regulatory mechanisms of CHIKV entry proteins by studying the mechanism of receptor binding-induced, and acidic pH-induced conformational dynamics of the envelope proteins.

I took biochemical and structural biology approach for my studies. I characterized viral protein - receptor interactions using biochemical functional assays, and studied conformational dynamics of the entry proteins using purified recombinant

Concluding remarks

proteins and different biophysical methods. A primary requirement for these studies is to express and purify these proteins in conditions that would allow folding of these proteins into their near native conformation. I prepared DNA clones for over-expression of CHIKV E1, E2, E3, E3-E2-E1 complex and full-length structural protein polypeptide in different expression systems (insect cells for E3-E2-E1 complex, bacterial system for E1, E2, E3-E2 and E3 proteins individually, and mammalian expression system for function studies). Results of these studies are discussed in Chapter 1 of the thesis. I optimized a bacterial expression protocol for expression of E1, E2, E3 and E3-E2 with good protein yield, required for biochemical studies. Proteins, thus purified, are folded to their native conformation (characterized using circular dichroism spectroscopy, gel-filtration, fluorescence spectroscopy and single particle negative stain electron microscopy) and are functional (studied using heparan sulfate binding studies, liposome-based membrane fusion assays and pseudo-typed baculovirus infections in insect cells). However, bacterially expressed E2 and E1 did not form E2-E1 heterodimers, as those would do when expressed in insect cells or in viral infected cells. A probable reason for this could be, that expression of E1, E2 and E3 as a single polypeptide (along with capsid protein) and posttranslational modifications may be a requirement for E2-E1 heterodimer formation. As explained in chapter 1, even viral surface cleaved off SFV E1 protein cannot associate with E2 *in vitro*. Further, despite several optimization attempts, I did not get good protein yields from insect cell expression using a recombinant baculovirus expression method.

In Chapter 2, results from characterization of CHIKV E2-HS interactions and concomitant conformational changes in E2 upon HS binding are presented. Using *in silico* docking, motif search and *in vitro* biochemical assays we identified a novel HBD with XBXXBX pattern in E2 sequence of alphaviruses and confirmed by mutagenesis

Concluding remarks

HS binding to this region on CHIKV E2. Receptor binding induced changes in E2 conformational dynamics were predicted using MD simulations and then validated using *in vitro* biochemical assays. My results suggest allosteric domain movements in E2 may lead to E2-E1 dissociation. Further studies using heterodimer conformation of E3-E2-E1 proteins (insect cell expression systems are shown to produce these complex structures) can be tested for dissociation upon heparan sulfate binding using different assay systems. Prior to my study there was no detailed knowledge on receptor-E2 interactions for any alphaviruses. Earlier studies depended on mutational analysis (with whole virion particles) to characterize E2-receptor interactions (96, 103). Hence, no precise mapping of the receptor binding site was possible. I fine mapped receptor binding site on E2 after thorough biochemical characterizations. Notably, many sites of receptor interactions on E2 mapped by others (using mutant viruses) overlapped with the site I proposed from my study, validating my observations. I also explained a possible mechanism of E2-E1 dissociation, upon receptor binding to E2.

In Chapter 3, results from characterization of E1 conformational changes under acidic pH condition are presented. We generated and characterized a cysteine-stapling mutant and other negatively charged residues mutants at dI/dIII interface of E1 to understand E1 dI swiveling upon dIII over a flexible hinge that connects dI-dIII. We used *in vitro* lipid based assay systems for studying membrane interaction and fusion properties of these mutants and compared those to similar properties of the WT protein. Our results suggest that E1 indeed undergoes conformational changes over the dI-dIII hinge under acidic pH condition for membrane fusion function. Structural studies using other alphaviruses have reported E1 swiveling movements upon acidic pH trigger. Another study also supports E1 conformational change by suggesting role of the flexible dI-dIII hinge towards acidic pH triggered E1 post-fusion conformation. More

data on CHIKV E1 protein intermediate conformations at acidic pH condition may help in better understanding the nature of E1 conformational changes that drive membrane fusion during entry.

In Chapter 4, results from characterizations on mechanism of E3 release from E2-E1 complex to understand its regulation on structure-activity of E2-E1 proteins are presented. We tested two possible mechanisms by which E3-E2 dissociation may be triggered: E3 interacting with membrane, and acidic pH of the endosome as a trigger for the dissociation. Using *in silico* predictions we identified a putative membrane interacting region at N-terminus of E3. But our results from *in vitro* biochemical assays suggest E3 does not interact with membrane. However further characterization of E3-membrane interaction with E3-E2 protein or with different lipid composition liposomes is required to make a conclusion. Based on sequence and structure analysis we proposed a conserved ‘GYG’ motif in E3 from which the Y48 residue make interactions with an acid sensing HIS (H256) residue from E2, as a regulatory switch for E3-E2 dissociation. We hypothesize that E3 Y48 – E2 H256 interaction would break in the acidic pH of the endosome, initiating E3-E2 dissociation. Studies to test this hypothesis are going on in the laboratory.

In Chapter 5, results from the development of a baculovirus pseudo virion system for use in functional entry assays are presented. Using the pRED/ET homologous recombination approach we successfully engineered the baculovirus genome as Δ gp64 – Cm + GFP for making pseudo viruses. Using transposition method, we incorporated envelope protein coding sequences of BV gp64, VSV G and CHIKV E proteins into the engineered Δ gp64 bacmid. Using these bacmids we purified pseudo viruses from sf9 cells. Through syncytia and cell transduction assays, we show successful pseudo-tying with BV gp64 and VSV G proteins. This proved the

Concluding remarks

functionality of the system. Further using immunofluorescence assay, we successfully checked the surface expression of CHIKV E proteins in sf9 cells. However, our trials of pseudo-typing with CHIKV E proteins were not successful so far. This could be because, while alphaviruses bud out of interactions between capsid protein and cytoplasmic domain of envelope proteins is required for packaging. In baculoviruses capsid protein may be required for pseudo-typing CHIKV envelope proteins onto baculovirus. In the future, optimizations on pseudo-typing of a class II or even a class I fusion protein complex, using this baculovirus system would be interesting to make it useful as a universal system for pseudo-typing of viral envelope proteins from all three classes.

Overall, my thesis research work contributes towards understanding the regulation of structural mechanism of CHIKV cell entry proteins. Identification of the putative receptor binding site on CHIKV E2 protein has great potential in terms of therapeutic development by using receptor binding site blockers. Such approaches involving use of soluble heparan sulfate as blockers of receptor binding to E2 protein have been reported (102). Approaches leading to use of E1 domain regions (especially E1-dIII) that block E1 conformational changes have been tested (139). However, more structural information on the nature of extended intermediate conformation of E1 is expected to improve development of other similar strategies. To add to these, we developed thoroughly optimized protocols for utilizing the simple and convenient *E. coli* expression-purification methods to purify these difficult to express viral envelope proteins for structure-function studies. Also, our baculovirus pseudo virion development strategy can be improved in future as a universal system for pseudo-typing of different classes of viral envelope proteins.

CHAPTER

GENERAL INTRODUCTION

General introduction

Viruses as human, animal and plant pathogens have huge impact on public health and welfare. In recent past, several viruses have re-emerged in epidemic proportions: 2006-2010 chikungunya virus epidemic, 2009 swine flu pandemic, 2013-2016 Ebola virus epidemic, and more recently Zika virus, Nipah, Hendra and Corona virus epidemics are glaring examples of the global health threat that viruses pose. Viruses are ubiquitous and infect multiple life forms including animals, plants and even microorganisms like bacteria and archaea.

Viruses are quite distinct from other life forms with respect to structure—lack cellular architecture. In its most simplistic form, a virus particle is a mere supramolecular assembly of a nucleic acid (genome) and proteins (as a protective layer for nucleic acid, referred to as capsid). In some viruses, a lipid bilayer envelops nucleic acid and capsid assembly, as the outermost layer. Most of the animal viruses are enveloped. Envelope is acquired from host cell membrane, when new virus particles bud out of an infected cell. Such viruses with a lipid bilayer envelope are called enveloped viruses (1). As a result of this simple, non-cellular organization, a virus is an obligate parasite of the host cell, exclusively dependent on host ribosomes for protein synthesis, membranes and other metabolites. Hence, entry into the right host cell (a permissible cell type of a specific virus) becomes the most critical first step in virus infection biology.

Of particular interest to the current study is entry of enveloped viruses into a cell. Several human pathogens, such as HIV, Influenza (flu), SARS, Ebola, Zika, Dengue and Chikungunya virus are all enveloped viruses. In case of both enveloped and non-enveloped viruses, cellular membrane is the first barrier that the virus has to overcome in order for the viral nucleic acid to gain access to cellular cytoplasm and ribosomes. An enveloped virus fuses its envelope with cell membrane, which

generates a lipid pore through which nucleic acid (and capsid and associated proteins in certain cases) can escape into cytoplasm. However, viral and cell membrane fusion cannot be a spontaneous event, as spontaneous fusion with an off-target cell (say, a human pathogen fusing with an amoeba membrane) would be futile. Instead, membrane fusion (viral cell entry, in general) is a highly regulated process. Enveloped viruses code for a specialized (both in terms of their structure and function) class of proteins that are anchored into viral membrane and presented on virion surface—hence, called envelope proteins (for example Chikungunya virus E1, E2 proteins). Envelope proteins facilitate binding of an enveloped virus to a specific target cell, and viral-cell membrane fusion. Thus, viral entry is the primary function of envelope proteins. In an infected cell, envelope proteins coded by viral genome, mature through secretory pathway, and are presented on cell surface. When new viral particles bud out of an infected cell from plasma membrane, envelope (lipid membrane) along with envelope proteins is added to virion structure. While these proteins traverse through the secretory pathway of the infected cell, those get glycosylated and processed by ER and Golgi proteases. Almost all viral envelope proteins known so far are glycosylated; hence these are also referred to as viral envelope glycosylated proteins, gp (for example HIV gp120 and gp41).

Enveloped viruses - envelope proteins mediate cell entry

Enveloped virus cell entry follows two sequential steps (Figure 0.1): (i) attachment to a specific cell type (viral cell-tropism) by binding to a cell surface receptor followed by (ii) fusion of virus and host cell membranes to form a fusion-pore through which the nucleocapsid can be released into a host cell (2–4). Viral envelope-anchored glycoproteins (one or more envelope proteins or E proteins) perform both these functions during viral cell entry (2–5).

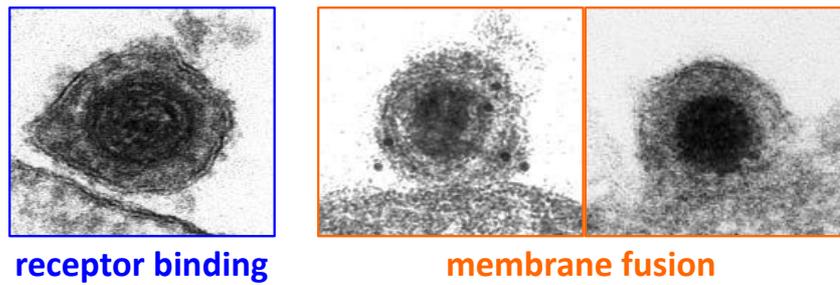


Figure 0.1 Enveloped virus entry by receptor binding followed by virus-host cell membrane fusion. This figure is adapted from *Granzow et al, J. Virol., 2005 (7)*.

In several cases, such as influenza (flu), HIV and Dengue virus, a single protein performs both functions. The viral envelope protein in these cases has different domains/regions from a single polypeptide – one for receptor binding and the other for membrane fusion. In several other viruses (for example herpesviruses), multiple proteins (for example herpes simplex virus gB, gH/gL, and gD) are employed - few for receptor binding and others for membrane fusion function (6). Receptor binding protein and membrane fusion protein are associated by either covalent interactions or non-covalent interactions. In chikungunya virus (and other alphaviruses as well), though the envelope proteins (E1 and E2) are a result of proteolytic processing of a single polypeptide of structural proteins, these are referred to as receptor binding protein (E2) and fusion protein (E1). To maintain consistency with the literature, receptor binding and fusion domains/regions will be referred to as receptor binding protein and fusion protein, respectively, for the rest of the thesis.

In case of influenza, envelope protein hemagglutinin (HA) has two polypeptides: HA1, the receptor binding part, and HA2 that is responsible for viral-cell membrane fusion. HA1 and HA2 are proteolysis products from the precursor protein HA0 (8). After proteolytic processing, HA1 and HA2 (three of each) are held together in a single mushroom like structure on virion surface by both covalent (disulfide bridges) and non-covalent interactions (Figure 0.2). The three HA1 units

become ‘mushroom’ head, whereas HA2 trimer becomes the stalk of the ‘mushroom’. Similarly, in HIV, the receptor binding protein, gp120, and the membrane fusion protein, gp41, are proteolytic products of precursor protein, gp160 (9, 10). Polypeptides gp120 and gp41 are in a tight complex on virion surface, held together by covalent and non-covalent interactions (Figure 0.3).

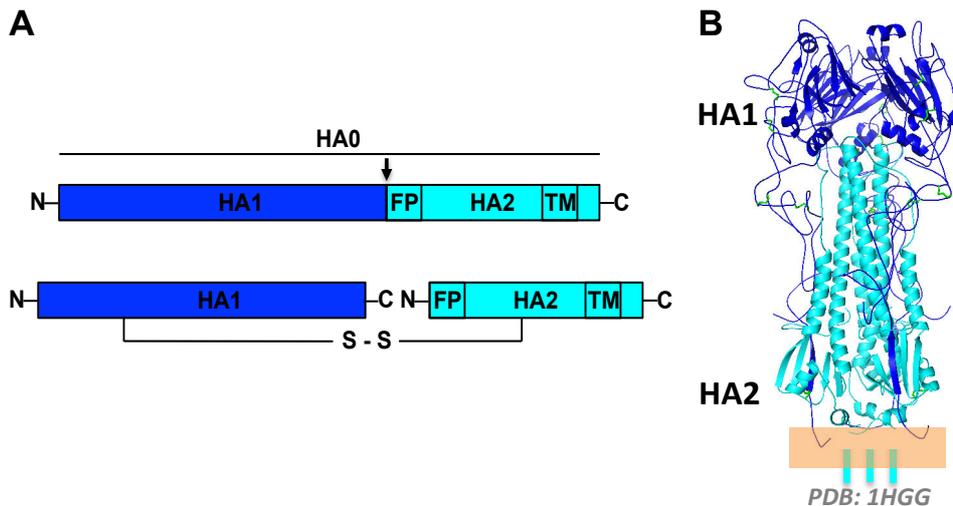


Figure 0.2 Influenza virus HA1 and HA2 envelope proteins are made from the HA0 precursor. (A) Schematic of HA0 cleavage into disulfide linked HA1 and HA2 subunits. The down facing arrow marks the protease cleavage site. FP is fusion peptide and TM is transmembrane region. (B) Crystal structure of complex of trimers of HA1 (blue color) and HA2 (cyan color) (PDB: 1HGG) (8). TM region is shown as vertical lines in cyan color and the membrane schematic is shown as orange colored horizontal bar.

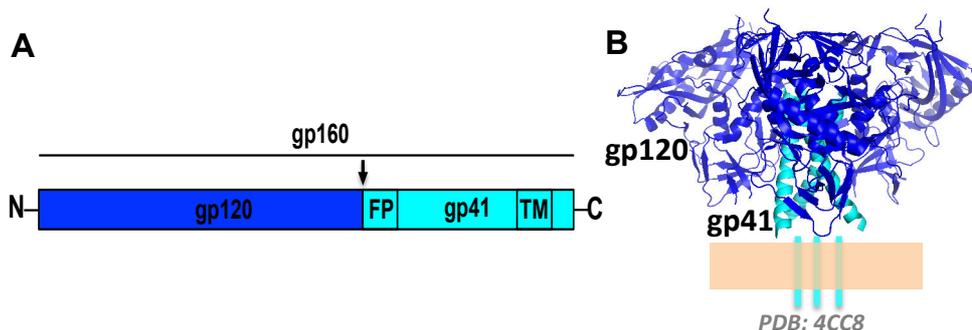


Figure 0.3 HIV gp120 and gp41 envelope proteins are made from the gp160 precursor. (A) Schematic of gp160 precursor. Cleavage site for production of gp120 and gp41 subunits are marked with a down facing arrow. FP is fusion peptide and TM is transmembrane region. (B) Crystal structure of complex of trimers of gp120 (blue color) and gp41 (cyan color) (PDB: 4CC8) (9). TM region is shown as vertical lines in cyan color and the membrane schematic is shown as orange colored horizontal bar.

General introduction

A striking structural feature of all viral fusion proteins, studied so far, is that receptor binding domain is in tight association with membrane fusion protein. This association masks the region of the fusion protein responsible for membrane fusion (either a fusion peptide or fusion loop). Unless the fusion protein is unmasked, fusion protein cannot perform membrane-to-membrane fusion. Thus, receptor binding domain/protein and fusion protein association is a structural feature that has an important consequence in regulating viral fusion protein function, till virus binds to the right target cell. Virus binding to cellular receptor (for viruses that enter cell at plasma membrane) as well as the acidic pH of the cellular endosomal compartment (for viruses that take endocytic entry route) trigger conformational rearrangements in the complex structure of receptor binding protein and the membrane fusion protein. These changes allow exposure of a fusogenic region on the membrane fusion protein that is responsible for membrane insertion and initiation of virus to cell membrane fusion. After viral-cell membrane fusion, fusion protein and receptor binding protein association is not seen (4, 5, 11).

Why fusion protein function has to be regulated? Virus-host cell membrane fusion like any other important cellular membrane fusion events involves exchange of contents between two membrane bound compartments without the loss of membrane integrity and leakage of trapped contents. This process requires bending and generation of local curvatures in lipid bilayers of fusing membranous compartments around the point of contact for subsequent lipid mixing followed by fusion pore formation. This makes the overall membrane fusion process an energy-dependent step (4, 12, 13). In eukaryotes a conserved family of proteins called SNAREs along with other accessory regulatory proteins perform this process, which involves external energy source in form of ATP and GTP hydrolysis. Moreover several proteins,

associate with SNARE proteins in a context dependent manner and in the presence of specific cellular signals, which would make membrane-membrane fusion a tightly regulated process in a cell.

However, if such fusion protein activity were not regulated inside a cell, it may result in complete de-compartmentalization of the cell and loss of identity. In contrast to cellular membrane fusion, viral fusion proteins do not use chemical energy (as there is no access to chemical energy source until virus enters a cell). Instead, viral fusion proteins are ‘designed’ to bring about membrane-membrane fusion by spontaneous changes in their conformation. Viral fusion proteins, on virion surface are ‘locked’ in a metastable conformation. This is also called the pre-fusion conformation as the fusogenic region (either a fusion peptide or a fusion loop) that is present in the fusion protein structure is masked in this conformation by a domain of the receptor binding protein. Receptor binding protein association with the fusion protein is one factor that keeps the fusion protein in this metastable conformation (14, 15). Thus, energy trapped in the viral envelope protein structure is proposed to drive the virus-cell membrane fusion reaction in the forward direction (4, 12, 13).

Site of membrane fusion for viruses following endocytic pathway

For viruses that follow endocytic route of entry, membrane fusion happens in the maturing endocytic vesicle. Endocytosis of virus starts upon signaling from the receptor, bound to viral protein. In viruses that follow endocytic pathway, membrane fusion protein is sprang into action only in the endosome. What triggers fusion protein in endosome? In most cases, acidic pH of endosome alone is enough to ‘trigger’ fusion-protein activation and conformational changes (4, 5, 11, 16). This proposal is supported by the observation that crystal structures of fusion proteins in acidic pH show a different conformation than the structure at neutral pH (pre-fusion structure).

Thus, for viruses that take endosomal pathway for entry, endosome acidification acts as another level of regulation. Based on viral fusion protein complex structures in pre-fusion and post-fusion conformations it is proposed that receptor binding protein, in most cases, dissociates from membrane fusion protein as a pre-requisite for activation of the later protein in endosomes.

Fusion protein activation and membrane fusion may take place in any one of the endosomal compartments. Among different endosomal compartments, early endosomes (Rab5 and ESCRT positive) have a pH range of ~6.0-6.5, maturing endosomes (Rab5, Rab7 and ESCRT positive) have a pH of ~6.0, late endosomes (Rab7, Rab9 and LAMP1 positive) have a pH range of ~5.0-6.0. Most of the acidic pH dependent viral fusion proteins function in the pH range of 5.0-6.5 (16–19). The acidic pH of these endosomes are maintained by function of the vacuolar ATP dependent proton pumps (19).

Universal structural mechanism of viral membrane fusion protein

All viral fusion proteins studied so far seem to follow a universal mechanism to facilitate membrane fusion (Figure 0.4) (4, 11, 20). Receptor binding brings the viral and cell membrane (endosome membrane, if the virus enters through endocytic route) to close proximity. The two apposing membranes fuse with each other, forming a fusion pore by the end of the membrane fusion process, through an unstable hemifusion intermediate (also referred to as hourglass intermediate) (12). In the hourglass intermediate of membrane fusion, only outer leaflets of both the membranes fuse (Figure 0.4).

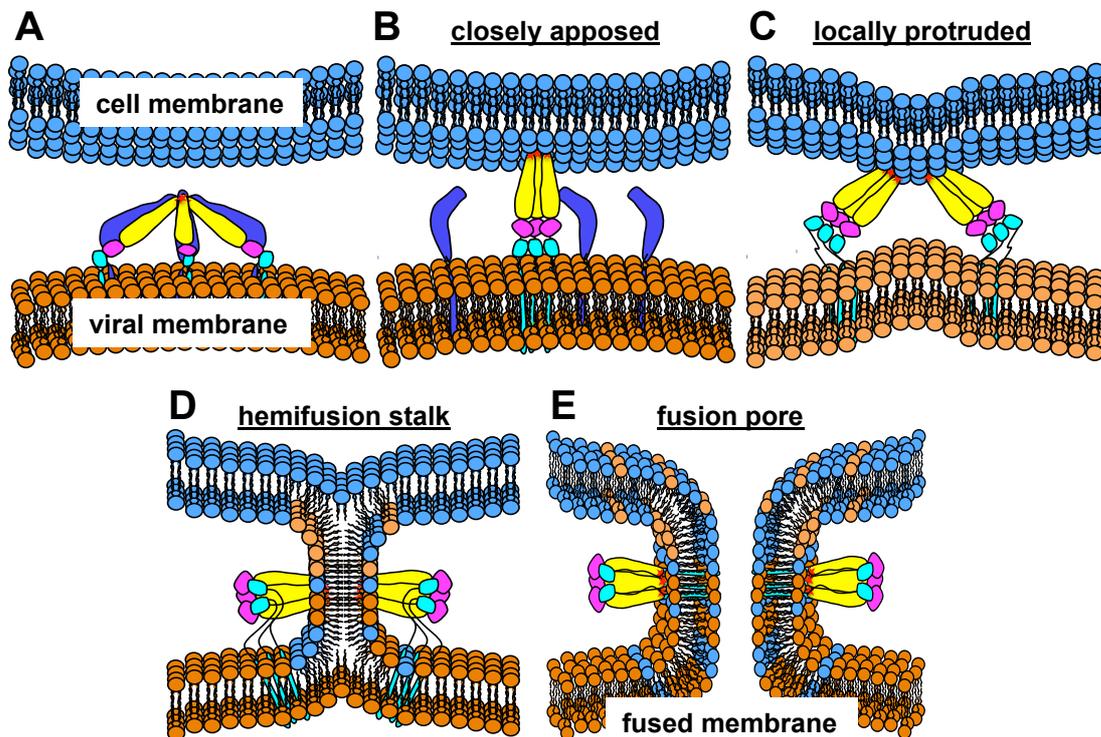


Figure 0.4 Schematic representation of virus-cell membrane fusion mediated by viral envelope glycoproteins. Viral membrane (orange colored bilayer) and cell membrane (sky blue colored bilayer) are labeled. Trimers of receptor binding protein (blue colored elongated structure) and membrane fusion protein (three domains of the single protein are shown as cyan, magenta, and yellow colored oval shaped structures, and the fusion peptide or loop at tip of the yellow domain is marked with a red colored star) heterodimers on viral membrane are shown. The fusion of the two lipid bilayers through the hemifusion pathway is represented in A through E. (A) Virus approaches a cell (viral envelope glycoproteins are present in their pre-fusion state), (B) Closely apposed viral and cell membranes (receptor binding and or acidic pH triggers fusion protein extension to reach target membrane), (C) Local protrusion in both bilayers are induced because of minimization of hydration (acidic pH triggered membrane inserted fusion protein conformational change pulls both membrane close to each other), (D) Hemifusion state with mixing of lipids from outer leaflets of both membranes while inner leaflet remains to be mixed (significant rearrangement in fusion protein structure towards a low energy conformation), (E) Fusion of both leaflets leads to formation of a fusion pore for content exchange (fusion protein attains the stable post-fusion hairpin conformation that is a trimer)

During the process of membrane fusion, viral fusion protein undergoes a conformational transition from pre-fusion state to post-fusion state. The pre-fusion conformation of fusion protein is triggered by receptor binding, which leads to

receptor binding protein dissociation from the fusion protein. Further, when the fusion protein is exposed to acidic pH, it adapts an extended intermediate conformation. In this conformation the fusion protein extends to insert either fusion peptides/fusion loops present in its structure into the outer leaflet of the cell membrane. This extended intermediate conformation of the fusion protein is an unstable structure and so it folds back onto itself to form the post-fusion hairpin like conformation. As a result, the fusion peptide anchored into the outer leaflet of the cell membrane, pulls the cell membrane towards the viral membrane. This results in formation of the hemifusion intermediate in membrane fusion. Pre-fusion conformation for some viral fusion proteins is a trimer (for example, influenza HA), whereas in few others (for example Dengue virus E) it is a dimer. Post fusion, all fusion proteins (studied so far) are homotrimers, irrespective of their pre-fusion conformation. Conformational change from pre-fusion to extended intermediate to post-fusion hairpin structure is proposed as a universal mechanism for all viral fusion proteins known so far (11, 21). This mechanism is proposed based on observation of pre- and post-fusion structures of several viral fusion proteins. However, extended intermediate structure is characterized only for a few viral fusion proteins (for example SINV (22), HA (23)). An estimate from cryo-electron microscopy studies shows an extension of 100Å-150Å for HA fusion protein (22, 23).

Membrane fusion is not a spontaneous process (13). There is energy requirement to induce membrane curvature as well as to overcome the hydration barrier (that resists two bilayers coming into proximity, beyond 10-20Å distance) for initiating fusion pore formation. In viral-cell membrane fusion, energy cost of membrane-fusion is provided by conformational rearrangement of the fusion protein. Energy released upon viral fusion protein conformational changes supply the energy

and act as catalyst to cross the barrier and facilitates to bring two membranes close together — viral fusion proteins lower the kinetic energy barrier of membrane fusion (~40-50 kcal/mol) that leads hemifusion state to fused membrane state (Figure 0.5) (4, 24–26). Further, Co-operative interactions between post-fusion protein homotrimers (two or more) may supply the energy required for initiation and expansion of the fusion pore. Cooperative interactions are reported for influenza virus HA where three or four HA trimers facilitate fusion pore expansion (27), lateral interactions among 5-6 trimers forming ring-like structures are reported to facilitate fusion pore expansion for SFV (28).

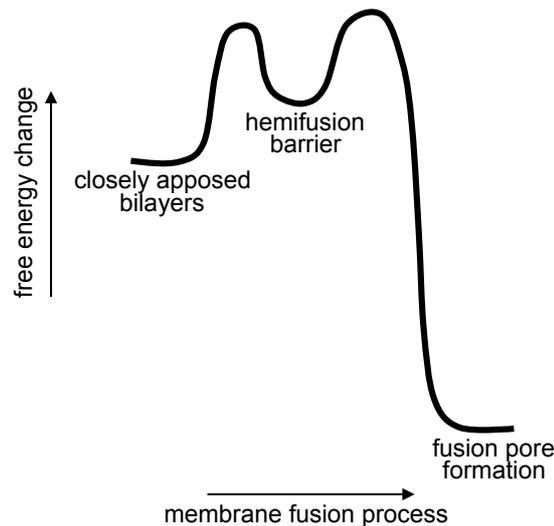


Figure 0.5 A hypothetical energy diagram (free energy vs. fusion progress graph) of membrane fusion process. The free energy barriers are depicted arbitrarily. Energetically favorable conformational changes in fusion proteins help overcome these barriers during virus-cell membrane fusion. This figure is adapted from *Harrison et al, Nat Struct Mol Biol., 2008* (4).

Structural architecture of viral fusion proteins

Viral fusion proteins are grouped into three different classes based on their structural properties, class I, II and III (Figure 0.6) (21, 29, 30).

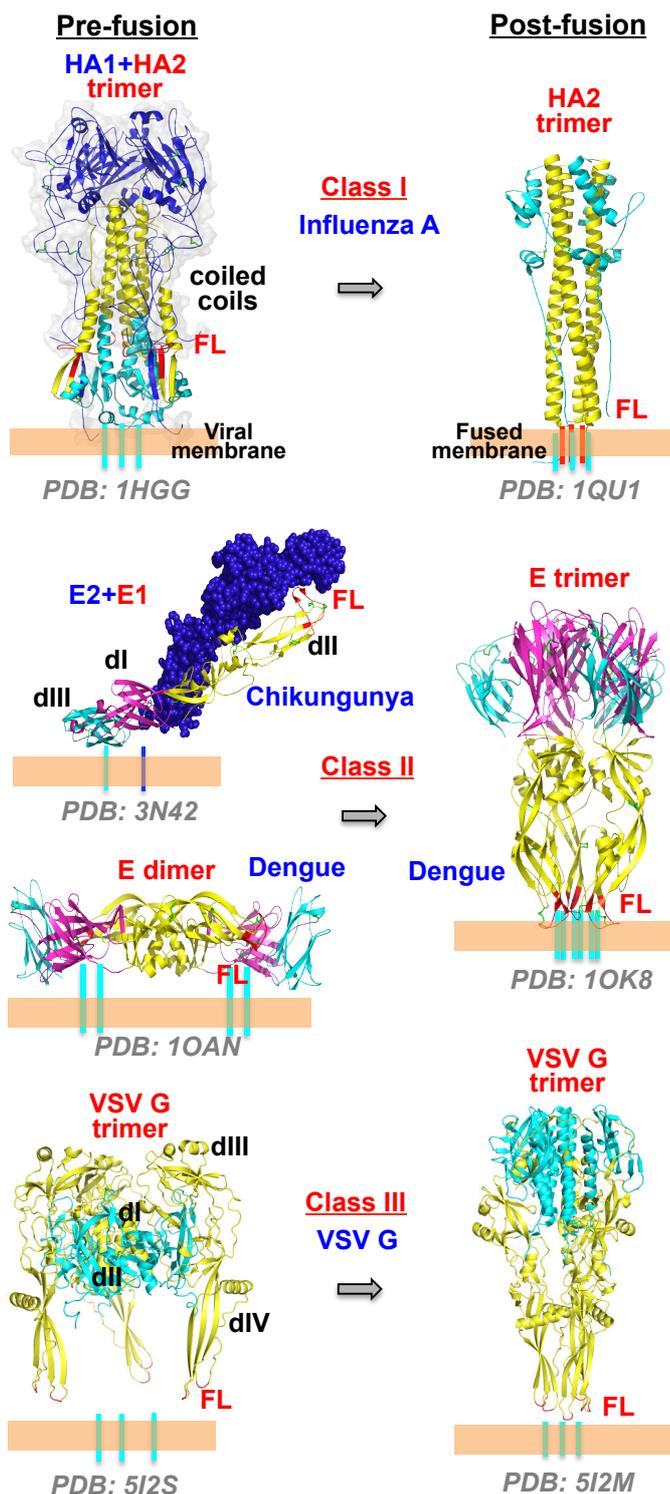


Figure 0.6 Structural organization of class I, II and III viral fusion proteins. From top to bottom – viral fusion proteins from class I, II, and III are shown in their pre-fusion (to the left) and post-fusion conformation (to the right). For class I, and III, fusion proteins are colored as cyan (the N-terminal region), yellow (the C-terminal region). For class II, cyan, magenta and yellow colored regions represent domains III, I, and II respectively. In all fusion proteins, FL is fusion loop (colored in red), and receptor binding protein (wherever present) is represented in blue color.

General introduction

Characteristic features of a class I fusion proteins are: i. Central long coiled-coil α -helices formed out of heptad repeats (HR) (typically two HR sequences, an N-terminal coil and a C-terminal coil), ii. A short hydrophobic stretch at the N-terminus (preceding the N-terminal HR coil) called fusion peptide that would insert into the membrane during fusion, iii. Form homotrimers (both in pre-fusion state as well as in post-fusion state), and iv. Fusion peptides from three of the protomers, in the pre-fusion trimer structure are buried in the trimer interface. Few examples of viral fusion proteins from class-I group are Influenza HA2 and HIV-gp41.

The characteristics of class-II viral fusion proteins are: i. Have predominantly β -sheets, ii. Organized into three domains (named as domain I, II and III), iii. On virion surface (pre-fusion conformation) are either heterodimers (in complex with the receptor binding protein, as in alphaviruses E2-E1 heterodimer) or homodimers (as dimers of the fusion protein, as in Dengue virus E), iv. A loop connecting two strands in the domain II inserts into the cell membrane during membrane fusion (hence, called fusion loop), v. Form homotrimers in the post-fusion state (compared to their pre-fusion heterodimer or homodimer conformations), and vi. The internal fusion loops in the pre-fusion structure is masked by the receptor binding protein/domain. Fusion proteins from alphaviruses (for example SFV, SINV or CHIKV E1 protein), and flaviviruses (for example JEV or DENV E protein) belong to class II group.

The class-III group (for example VSV G, Baculovirus gp64) fusion proteins have features of both class-I and class-II: i. Have coiled-coils as central domains and also have β -sheet rich domains, ii. Have internal fusion loops similar to class II fusion proteins, and iii. Form trimers in both pre-fusion and post-fusion conformations, similar to class I fusion proteins. The structural features of class I, II and III viral fusion proteins are reviewed in references (21, 29, 30). In fusion proteins from all

three classes, post-fusion conformations are referred to as hairpin conformations because of their resemblances to a hairpin (N-terminal and C-terminal regions of the fusion proteins fold back onto each other, where both the regions resemble to both arms of a hairpin). Comparison of fusion protein structures in their pre-fusion and post-fusion conformations reveal that both the conformations share overall structural similarity but with altered domain reorganizations. The post-fusion hairpin lies perpendicular to the target membrane, where both the fusion loops and the trans-membrane helices are present at the same end of the trimeric structure at the fused target membrane.

Regulation of fusion activity of viral fusion proteins:

Unregulated fusion activity of the fusion protein would be futile. Fusion activity should be triggered, only when the virus is attached to the host cell. Different viruses use different ways of regulating the fusion activity of their fusion proteins. Receptor binding protein (or domain) binding to a cell surface receptor and dissociating from the fusion protein (or domain) is seen as a regulation mechanism in many viruses. For example, in HIV the receptor binding domain, gp120, dissociates from the fusion domain, gp41, as a first step in viral entry (31, 32). Unless, gp120 dissociates from gp41, membrane fusion does not go forward. On the contrary, in herpes simplex virus (HSV), the receptor binding protein, gD interacts with the fusion protein, gB, after binding to cognate receptor, triggering membrane fusion (6).

In viruses that take endocytic route of entry, acidic pH (pH 4.5-6.5) of the maturing endosome acts as a 'trigger' for fusion activity of the fusion protein. Acidic pH triggers conformational changes in the fusion protein, required for fusion function.

Receptor binding as a regulation mechanism

As early events in viral entry, receptor binding to the receptor binding protein triggers conformational rearrangements leading to dissociation of receptor binding protein and membrane fusion protein. Specifically, upon binding to a cellular receptor allosteric structural rearrangements are observed in the receptor binding protein. These conformational changes in turn, either open up a conserved buried structure on the receptor binding protein for binding to another cell surface molecule (may be another receptor or a co-receptor) or induce concomitant conformational changes in the fusion protein. These receptor binding-induced conformational changes in receptor binding protein ultimately leads to rearrangement of key interactions between receptor binding protein and membrane fusion protein, leading to dissociation of membrane fusion protein from receptor binding protein and thus unmasking the membrane fusion protein structure (31, 33–37).

Receptor binding determines viral tissue/cell tropism. Also, in several cases provides the initial trigger for cell entry by directly affecting receptor binding protein and membrane fusion protein conformational changes, and thereby regulating fusion protein activity. Significance of receptor binding triggering conformational changes in receptor binding protein, leading to unmasking of fusion protein structure is well established for several viruses. For example in HIV-1, binding to receptor (CD4) and co-receptor (CCR5 or CXCR4) induces conformational changes in receptor binding domain (gp120) (Figure 0.7) (32).

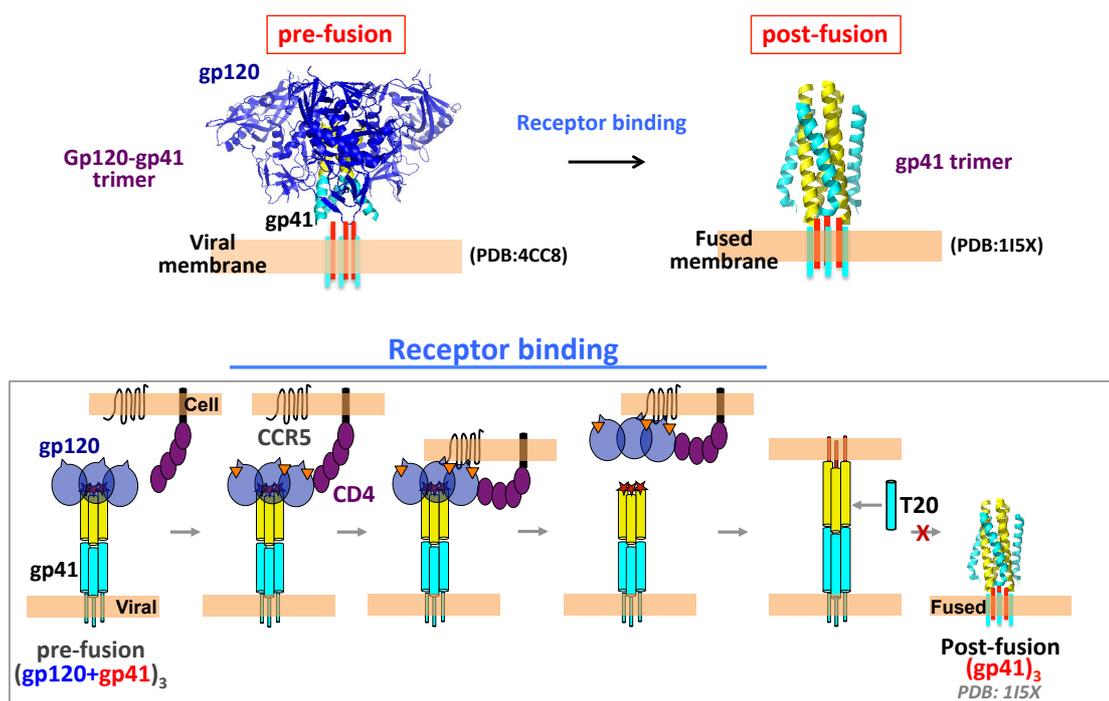


Figure 0.7 Receptor binding regulates fusion protein complex conformational changes and activity during cell entry of HIV1. Membrane fusion protein post-dissociation from receptor binding protein, are triggered to form the homotrimeric post-fusion hairpin conformation essential for performing virus-cell membrane fusion. The top panel shows the HIV1 gp41+120 complex structure in pre-fusion and post-fusion conformations. Bottom panel shows in schematic the conformation changes in these proteins and their regulation by receptor and co-receptor binding. Gp120 is shown in blue and gp41 in yellow (N-terminal half)+cyan (C-terminal half) in both crystal structure and the schematic. The orange colored triangle represents CCR5 binding site and red colored star represents fusion loop. T-20 shown as a cyan colored helix in bottom schematic is an inhibitor to gp41 conformational transition.

Interaction of gp120 with CD4 receptor (relatively weak interactions) and CCR5 or CXCR4 co-receptor (this interaction stabilizes the gp120 bound CD4+CCR5/CXCR4 conformation) leads to allosteric conformational changes in N- and C- terminal region as well as in the V3 loop region on gp120, which in turn affects relative positioning of helical regions on gp41. As a result of this the fusion peptide region on gp41 move to a closer distance to host membrane from where

membrane fusion takes place (32).

In measles virus, binding of receptor binding protein (H) to its cognate receptor (CD46) results in movement of its head domains relative to each other, to trigger fusion activity of fusion protein (F) (36). Binding of CD46 on to H protein takes place at a site opposite to the H-dimer interface and this binding pattern induces domain movements in the head region of the H protein in relative orientation to each other. These resultant movements when translated onto the fusion protein (F) are able to trigger irreversible refolding of F-trimer and thereby membrane fusion by F (36). Another example is Ebola virus, where binding of an endosomal receptor (NPC1-C) to receptor binding protein (GP1) induces conformational changes that lead to change in the state of the fusion loop on the fusion protein (GP2), triggering membrane fusion (38). In case of Ebola virus, upon triggering through the endocytic pathway of entry, by binding to low affinity cell surface receptors (DC-SIGN, L-SIGN etc.), the virus is proteolytically primed by cathepsin L and B proteins. It is in this form the primed receptor binding protein (GP1) binds to its endosomal receptor NCP1-C, which leads to conformational changes in a short helical region that moves upward in GP1 allowing exposure of the fusion peptide region on GP2, which in turn triggers membrane fusion (38). All these observations point to the fact that receptor binding is the first regulatory feature in viral cell entry. Receptor binding triggers dissociation of receptor binding protein and membrane fusion protein, unmasking the fusion protein.

So, the specific triggers that induce dissociation of receptor binding protein from membrane fusion protein, and membrane fusion protein conformational rearrangements vary from virus to virus with some conservation. In paramyxoviruses (for example in Newcastle disease virus, parainfluenza virus 5), receptor binding in some cases (when HN and F are produced and transported to cell surface as a

complex) leads to dissociation of receptor binding protein (HN) from fusion protein (F) at the cell surface and in some other cases (when HN and F are produced as individual subunits and transported to cell surface individually) receptor binding by HN leads to transient association between HN and F, which triggers the later for fusion function (and eventually the HN protein dissociates from F protein) (39, 40). In HIV, receptor and co-receptor binding leads to dissociation of gp120 from gp41, and this is a key regulatory step for membrane fusion (31). In influenza virus (41) and avian sarcoma leukosis virus (42), receptor binding triggers entry through endocytic pathway where acidic pH triggers fusion protein activation for membrane fusion.

In addition, role for a thiol exchange reaction post receptor binding-induced conformational changes is employed in certain viruses that have disulfide-linked receptor binding and membrane fusion protein subunits. Cysteines from motifs in receptor binding protein CXXC and membrane fusion protein CX₆CC, where C is a cysteine and X is any residue form a disulfide bond. Upon receptor binding-induced conformational changes CXXC motif in the receptor binding protein gets exposed and undergoes a disulfide exchange reaction that leads to breaking of the covalent interaction and then dissociation of the receptor binding protein from the membrane fusion protein. Such mechanisms are studied for Moloney murine leukemia virus (43), HIV and SARS coronavirus (44), and Avian sarcoma leukosis virus as well (45). In Ebola virus, proteolytic priming at endosomal pH is essential before to binding to the endosomal receptor, which in turn triggers receptor binding protein membrane fusion protein conformation rearrangements for fusion activation of the later (33).

Regulation by acidic pH trigger

Once the fusion protein is unmasked, it then extends (in most cases being triggered by the acidic pH of the endosome in the endocytic entry pathway) to reach the cell membrane (possibly by forming an extended intermediate structure), where it inserts its hydrophobic fusion peptide or loop into the outer leaflet of the host cell membrane (22, 23, 29, 46). For example, in case of influenza virus, the proteolytically processed membrane fusion protein remains in the metastable form by association with accompanying receptor binding protein until the virus is triggered by receptor binding for entry through the endocytic pathway. Once inside the endosome, acidic pH of the endosomal compartment triggers the HA2 fusion protein to undergo dramatic conformational changes to facilitate membrane fusion (47). The acidic pH regulations of HA2 conformational changes are represented in Figure 0.8.

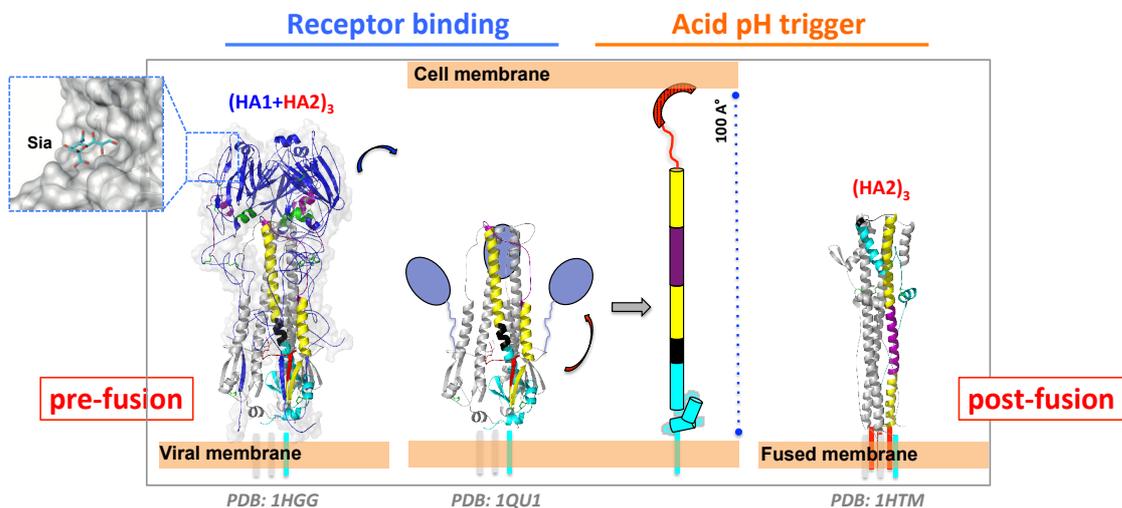


Figure 0.8 Regulation mechanism of IAV HA structural mechanism during cell entry of IAV. Sialic acid binding by HA1 initiates HA1 dissociation from HA2, followed by acid pH triggered HA2 conformational rearrangements from pre-hairpin to hairpin conformation, driving membrane fusion during cell entry. ‘Sia’ is sialic acid.

Several ‘acid sensing’ residues present in the viral fusion protein are likely to contribute to conformational change of the fusion protein once the protein is exposed to acidic pH environment (for example to endosomal acidic pH). Residues such as HIS (pKa 6.6), ASP (pKa 4.0) as well as GLU (pKa 4.4) is likely to play important role in induction of viral membrane fusion protein conformational rearrangements (21, 48). Specifically, HIS residues have been reported to be critical to induce conformational changes (relevant for membrane fusion) for different classes of viral fusion proteins. Role of HIS protonation change triggering of viral fusion protein conformational changes can be seen as two interesting events: At pH 7.0, in the pre-fusion conformation of the viral fusion protein, these HIS residues make charge interactions with positively charged residues (primarily as hydrogen bond donors) that are present near to these HIS residues within an interacting distance (21, 48). However, at endosomal acidic pH, many of the histidines get bi-protonated (as pKa of HIS side chain is 6.0). HIS residues protonation at acidic pH destabilizes the pre-fusion interactions (repulses nearby positively charged residues and by doing so induces structural changes), and lead to formation of new favorable interactions (hydrogen bonds or salt bridges) with negatively charged residues that stabilize the protonated form in post-fusion conformation of viral fusion proteins (21, 48).

Importance of studying viral fusion protein regulation mechanisms

Understanding the fusion protein regulation (by receptor binding and by acidic pH) is important for possible entry-inhibition kind of antiviral therapeutic strategy development. Knowledge on viral fusion protein regulation mechanisms of IAV HA and HIV gp (49), two of the well-studied examples in the field, contributed to development of therapeutics that target multiple steps of the viral cell entry process (50). For example, specific monoclonal antibodies that stabilize or prevent necessary

viral surface protein rearrangements can block virus-cell membrane fusion and entry. Blocking of Influenza virus HA binding to sialic acid receptor is a key antiviral strategy and liposomes coated with penta-saccharide of α 2, 6-linked sialic acid is an example of a multivalent and safe drug against IAV that competes with receptor binding and prevents attachment to host cells (51). Receptor and co-receptor binding blockers (CD4 domain linked to HS molecule competes with both receptor and co-receptor binding sites) have been shown to be successful as viral entry inhibitors for HIV-1 and developing such strategies for other class of viruses are of interest in the field at this moment (52). Several small molecule inhibitors are also developed that target conserved hydrophobic sites and prevent folding back of fusion protein to hairpin conformation, and thereby block fusion and entry (53–56). Selected peptides to prevent formation of the post-fusion hairpin structure have been used as well (57–60). One example is the T-20 peptide that blocks HIV gp41 hairpin formation by binding and stabilizing an intermediate structure, which is now used as a licensed anti-retroviral drug (50). Viral cell entry inhibitors summarized above constitute one of the important aspects of antiviral strategy and interestingly similar strategies are currently being explored for development of therapeutics for the less studied groups of enveloped entry protein complexes of enveloped viruses from alphaviruses and flaviviruses as well.

To that end, there are several other enveloped viruses (specifically several critical pathogenic animal viruses from the alphavirus and flavivirus genus that mainly constitute the less explored class-II enveloped virus fusion protein group) for which the regulation mechanism that governs the viral entry protein complex conformational rearrangements remains poorly understood (even though both receptor binding and acidic pH trigger are believed to play a role in viral membrane fusion

triggering process (61, 62) of these viruses). Further exploration of these important aspects of class-II fusion protein complex may serve as useful for development of antiviral strategies including viral entry inhibitors for enveloped viruses from these groups.

Chikungunya virus - an enveloped alphavirus

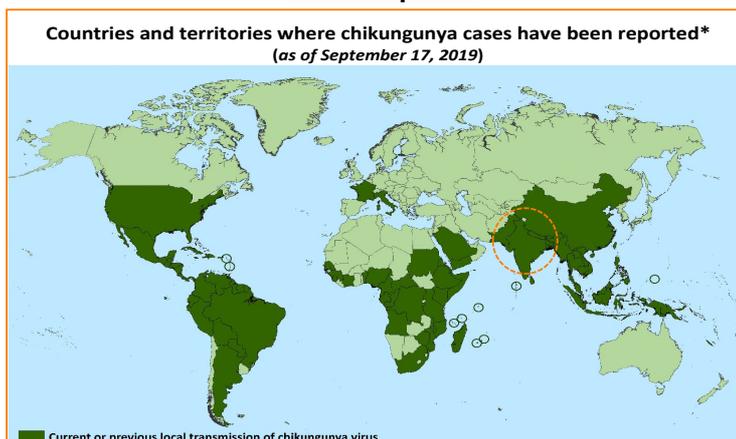
Chikungunya virus (CHIKV) is a mosquito-transmitted virus that has re-emerged as a critical human pathogen of global importance (Figure 0.9).

Transmitted primarily by aedes mosquitoes and causes CHIKV fever



Photo by: Luz Sosa, PAHO/WHO

Global spread



<http://www.cdc.gov/chikungunya>
<http://www.who.int/mediacentre/factsheets>

Figure 0.9 CHIKV is a mosquito-transmitted virus, which causes CHIKV fever and is spread globally.

General introduction

CHIKV is primarily transmitted by aedes mosquitoes, with global spread of the virus attributed to adaptation of the Indian Ocean isolates of the virus to a new vector species, *Aedes albopictus* (63, 64). There are four lineages of Chikungunya virus – West African, East/Central/South African (ECSA), Asian and the Indian Ocean Lineage; the later emerging from the ECSA lineage. Since its re-emergence, CHIKV has caused epidemics in several South East Asian countries in recent past, including millions of infections in India alone during the 2006-2010 epidemic. From then, CHIKV has spread to many different parts of Europe and Americas, where it had never been seen before (64–66).

CHIKV infection leads to onset of fever in ~95% of the infected individuals. The acute symptoms of CHIKV infection include joint pain, rash, high fever, nausea, myalgia etc. and the chronic symptoms can be serious (seen in ~12% of infected individuals), which includes persistent joint pain (arthritis) that can last for months to years (65, 67). The disease effects are severe amongst younger and older populations (68). In neonates the disease is associated with encephalitis (69). In older individuals (>65 years) the disease mortality rates are 5 times higher compared to those under 45 years of age (70). Poly-arthralgia is seen in 30-90% of the cases of infection and in chronic phases joint pain is debilitating (65, 67, 71, 72). In rare cases the disease symptoms are shown to be associated with ophthalmic diseases (73), encephalitis (74) and even circulatory symptoms (75). In current scenario no vaccines or antiviral treatments are available for CHIKV infection. Antiviral strategy that includes inhibition of viral entry is a potential and proven approach and can be utilized to target CHIKV cell entry process.

CHIKV is an alphavirus. It is grouped in Togaviridae family, which includes a number of viruses (~31 different viruses) that are critical human pathogens. Few of

the viruses from alphavirus genus other than CHIKV are Semliki forest virus (SFV), Sindbis virus (SINV), Venezuelan equine encephalitis virus (VEEV), Eastern equine encephalitis virus (EEEV), Ross river virus (RRV) etc. Alphaviruses are generally ~70 nm in diameter sized virus particles. These viruses have positive sense single stranded RNA genome (~11.8 Kb) (Figure 0.10). The alphavirus genome is transcribed and translated to produce polyproteins that are processed later to individual non-structural (nSP1-nSP4) and structural (Capsid-E2-E1-6K-E1) proteins (Figure 0.10). The non-structural proteins are mainly responsible for viral genome replication and structural proteins are mainly responsible for the structural makeup of the virus as well as play roles during exit and entry of the virus particle from/into a host-cell.

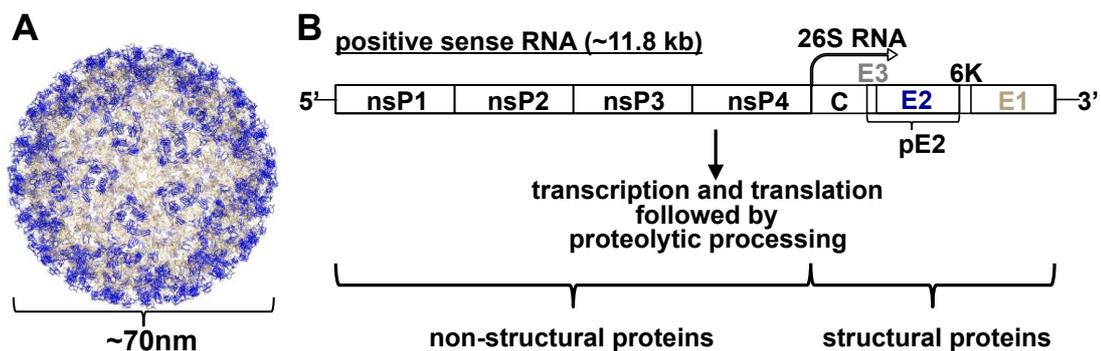


Figure 0.10 CHIKV virus particle structure and genome organization. (A) Virus particle structure with diameter in nanometers is shown. (B) The positive sense RNA genome of CHIKV is shown in a schematic with relative positions of non-structural and structural protein coding regions. The 26S sub-genomic RNA transcription start site is marked with a forward arrow at the beginning of capsid coding region. The path to non-structural and structural polyproteins production is shown.

Among the structural proteins, the alphavirus membrane fusion protein E1 is an example of the less explored class-II fusion protein group, which also includes flavivirus E (for example DENV E), rubivirus E1 (rubella virus), and phlebovirus Gc proteins (76). All these different proteins are synthesized initially as part of the

structural polyprotein complex, which subsequently undergoes proteolytic maturation. The maturation process produces a companion protein (the regulatory protein/receptor binding protein, E2), which chaperones the folding of the fusion protein and remains associated with it on virion surface, masking and stabilizing the fusion protein metastable pre-fusion structure.

Alphaviruses follow the endocytic entry pathway for cell entry. Binding or early attachment to a cell-surface receptor triggers these viruses to follow clathrin coated endocytic pathway, where either in early endosomes (pH = ~6.0 – 6.5) or in late endosomes (pH = ~5.5 – 6.0), the acidic pH triggers fusion protein conformational rearrangements essential for virus-cell membrane fusion during entry (62, 76, 77). While the role of acidic pH in triggering dissociation of receptor binding protein and membrane-fusion protein as well as acidic pH triggered conformational changes in the alphavirus membrane-fusion proteins are studied to some extent, several questions remain to be answered with respect to understanding the role of receptor binding and triggering of receptor binding protein and membrane-fusion protein dissociation because of receptor binding trigger as well as the exact nature and path of alphavirus fusion protein conformational rearrangements upon acidic pH trigger. CHIKV constitutes a good model to address and advance our understanding of the cell entry protein regulations, of CHIKV and in general of class-II fusion protein containing group of viruses as well, by both ‘receptor binding’ and ‘acidic pH’ triggers during cell entry.

Cell entry of CHIKV – role of cell entry proteins

Based on our current understanding of viral entry protein structures from different alphaviruses, a fairly informative path for alphavirus cell entry process is known (62). Cell entry of CHIKV, like other alphaviruses, follows two sequential

steps: 1. Binding to cell surface receptor – triggering entry through endocytic pathway, and 2. Fusion of virus-host cell membrane – leading to opening a fusion pore for nucleocapsid release. Envelope glycoproteins E1 (the membrane fusion protein) and E2 (the receptor binding protein) perform these functions during viral cell entry (62, 77). The regulatory role of the third smaller protein E3, which is present through its interactions on E2, in cell entry, is not known.

Biogenesis of CHIKV E3-E2-E1 protein complex:

In an infected cell, structural proteins of Chikungunya virus – capsid, E3, E2, 6K and E1 that make up the structure of the virus are produced by transcription and translation of a sub-genomic RNA (Figure 0.11). The 26S sub-genomic RNA is transcribed from an internal promoter present preceding the structural protein coding region in the 3'-end of the genome (77). This sub-genomic transcript codes for a single polypeptide (Figure 0.10). On the polypeptide of structural proteins, capsid, E3, E2, 6K and E1 are arranged in N-terminus to C-terminus order. The polypeptide is proteolytically processed (by viral and host proteases) into individual components in a complex maturation process (78). In the first step of the process, auto-proteolytic activity of the capsid protein clips itself from the rest of the structural polyprotein. The remaining polypeptide, guided by E3 secretory signal sequence, translocate into the ER lumen (79). E3 and transmembrane regions of E1 and E2 are likely to arrange in the ER membrane, as shown in the schematic in figure 0.11 (80). Subsequent to capsid protein cleavage from the polypeptide, the new N-terminus of the remaining polypeptide (N-terminus of E3) is glycosylated at an early stage of transport through the secretory pathway (81). Based on this observation, Garoff et al proposed that E3 would retract out of the ER membrane soon after the remaining polypeptide translocate into ER. E3 would become the new N-terminus protein of the remaining

polypeptide. In the crystal structure (PDB: 3N42) using drosophila expressed protein complex two glycosyl moieties could be seen (Figure 0.11). However, extent and pattern of glycosylation in host cells is not known.

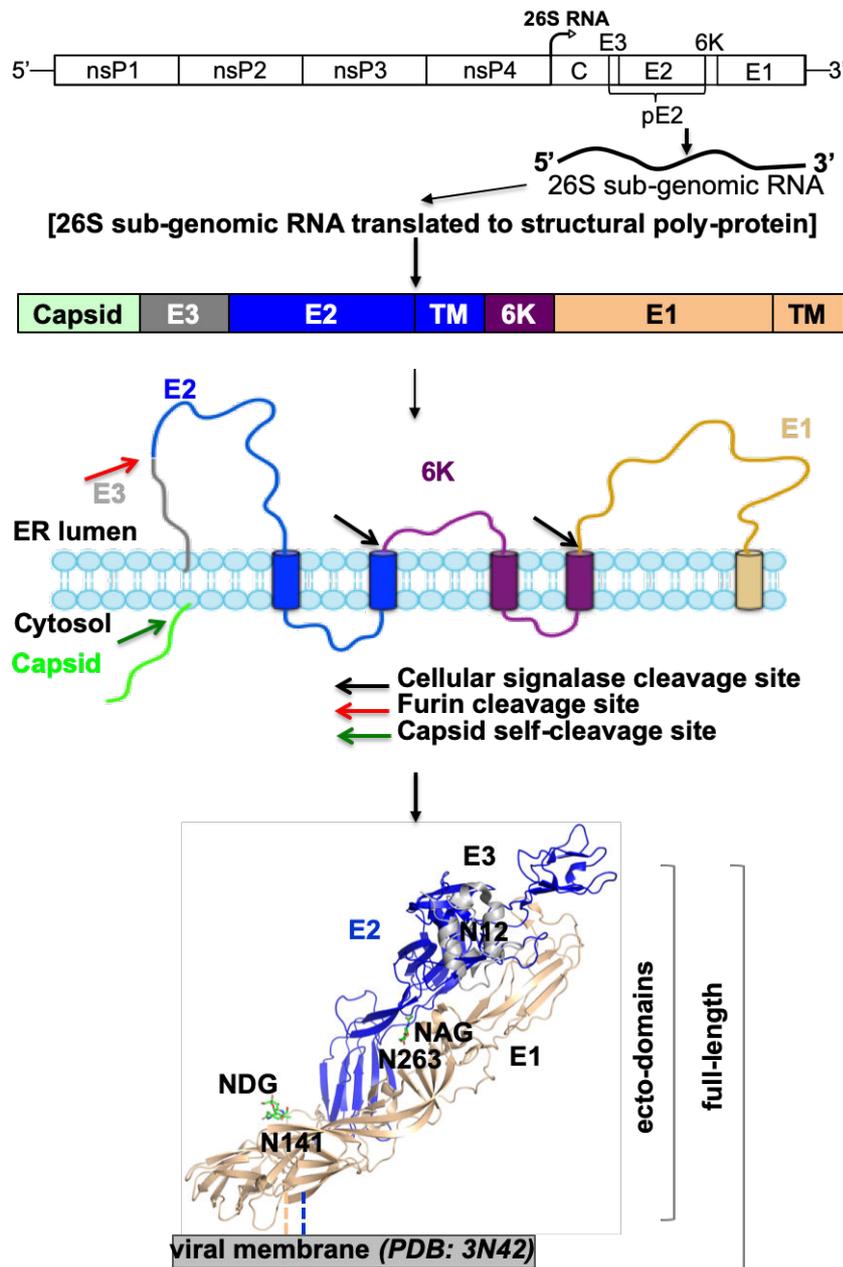


Figure 0.11 Scheme of alphavirus structural polyprotein production from sub-genomic RNA followed by processing through secretory pathway to heterodimer conformation. End product is an envelope anchored folded E3-E2-E1 heterodimer conformation (functional ectodomains protrude outside from viral envelope bilayer) (PDB: 3N42) (82). N-glycosylation sites are shown in sticks and labelled. NAG - N-ACETYL-D-GLUCOSAMINE, NDG - 2-(ACETYLAMINO)-2-DEOXY-A-D-GLUCOPYRANOSE.

The polypeptide is further cleaved into p62 and E1 by cellular signalases, (please refer to the schematic in figure 0.11). The p62 protein is an immature precursor form of E2 where N-terminal E3 is not proteolytically processed. The p62 and E1 traverse the ER and the Golgi as a heterodimer complex, where p62 plays a chaperone role for E1 during Golgi processing and presentation onto cell surface, as indicated by the following reports. When E3 is replaced with a cleavable signal sequence (no E3 in the premature E2-E1 complex) in SFV, E1 is retained in the ER and not presented onto the infected cell surface (79). In another study, Uchime et al (83) noted that when E3 and E2 interaction is disturbed by specific mutations of the residues at the E3-E2 interface (in p62), then E1 is not presented onto the plasma membrane. Similar observations were also made by Fields and Kielian (84). Both these studies suggested that E3 (as part of p62) prevents E2-E1 dissociation and premature triggering of E1 fusion function in the acidic pH environment of trans-Golgi (in the secretory pathway). Trans-Golgi resident furin protease cleaves E3 from p62 (refer to the schematic Figure 0.11). Post E3-E2 cleavage, E3 is retained on the E2-E1 heterodimer, through weak interactions with E2. A comparison between the immature (p62, furin un-cleaved form) and the mature (E3 and E2, furin cleaved) protein structures of E3-E2 are presented in figure 0.12. Both the structures show overall similarity, except the structural details of few residues from either side of the furin cleavage site remain missing from the furin cleaved mature E3 and E2 structure (Figure 0.12). Further, the furin cleaved E3 that is held on E2 through weak interactions is eventually released from the E2-E1 heterodimer in few alphaviruses (for example SFV, and SINV) (85–87), probably because of altered interactions between E3 and E2 at neutral pH of the cell surface. One mechanism that is studied for E3 release from E2-E1 complex at neutral pH is by breaking of interactions of

conserved tyrosine residues from a mildly conserved “GY₄₇Y₄₈” motif on E3, at the E3-E2 interface. These tyrosine residues from E3, Y47 that forms cation- π interaction with K254 of E2, and E3 Y48 interactions with E2 H256 and E166 stabilize the un-cleaved E3-E2 interface in p62, and thereby prevent dissociation of p62/E1 heterodimer in acidic pH of the trans-Golgi during biogenesis (83, 84). The transition to the extracellular neutral pH environment is proposed to change the protonation states of the conserved K254 or H256 residues on E2 impacting the E3-E2 interface stability towards E3 release from E2 (83, 84).

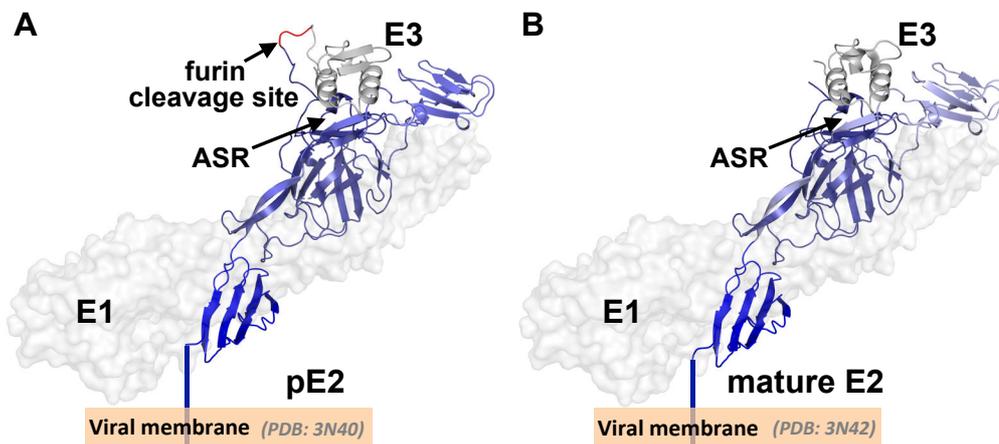


Figure 0.12 Comparison of the immature (pE2) and mature (E3-E2) structures. The E2 precursor structure pE2 (extracted from PDB: 3N40) (82) and the mature E2 structure (extracted from PDB: 3N42) (82) are represented as cartoons, E3 is shown in gray and E2 is shown in blue color. The position of the furin cleavage site (top arrow points to this region) at the E3-E2 junction is marked in red. ASR stands for acid sensitive region (bottom arrow points to this) in E2 upon which E3 is situated.

Structure and organisation of CHIKV cell entry proteins

Structures of E2-E1 (or E3-E2-E1) from few alphaviruses are published. This section describes the domain organization and different structural features seen in CHIKV E1, E2 and E3. Pre-fusion and post-fusion conformation changes in the protein complex are also discussed. Post-fusion homotrimer conformation structure of

CHIKV E1 is not available. Hence, post-fusion (acidic pH structures) conformation of SFV E1 (88) is discussed.

Membrane fusion protein, E1

E1 domains (from viral membrane proximal to distal; dIII-dI-dII) constitute an elongated E1 molecule (Figure 0.13). The E1-dI lies in between dIII and dII and connects to each of these domains through flexible hinge regions. Two long insertions (E1-residues 38-130 form insertion 1, and E1-residues 169-273 form insertion 2) that arise from dI form the elongated dII. Two loops connect β -stands at the tip of each of these insertions - the cd-loop and the ij loop.

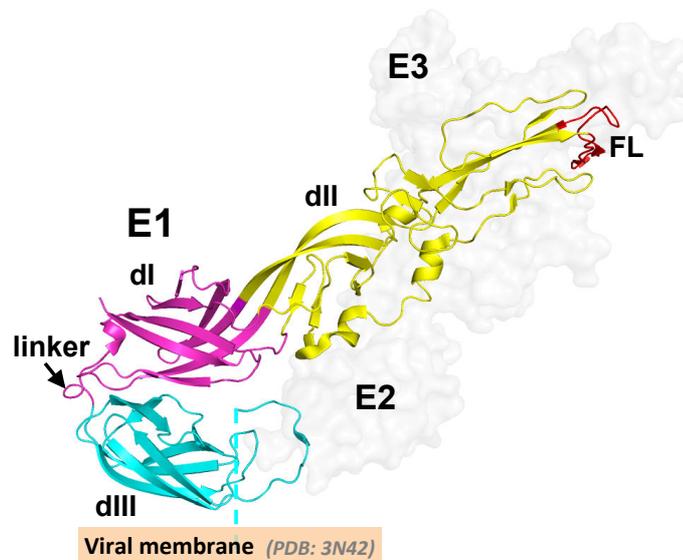


Figure 0.13 Domain organization of CHIKV E1 protein. The structure is extracted from CHIKV E3-E2-E1 heterodimer structure in pre-fusion state (PDB: 3N42). Domain organization and position of dI-dIII linker and fusion loop (FL) are shown as cartoon representation in different colors. The dIII connects to the stem, and transmembrane (TM) region in the viral membrane.

The cd-loop contains the hydrophobic fusion loop essential for target membrane insertion and virus-cell membrane fusion by E1. The C-terminus of E1-dIII

connects to the E1-stem region, which in turn, connects to the E1 trans-membrane anchor. Domain III is arranged, bending back onto domain I. E1 dI-dIII interactions, primarily through charge interactions hold the bent conformation. The interaction at dI-dIII and the dI-dIII connecting region called 'linker' hold this bent conformation of E1.

E1 domain I - domain III linker

An important structural feature, essential for dIII bending back onto domain I, and forming hairpin structure, is a linker connecting domains I, and III (dI-dIII linker). The linker does not have any regular secondary structures, either in pre-fusion conformation or post-fusion structure (as seen in structures of alphavirus E1 proteins) (Figure 0.13). Zheng et al (89) studied importance of dI-dIII linker in E1 pre- to post-fusion conformation change and found that the linker region plays critical role during acidic pH trimerization of E1 in SFV.

Receptor binding protein, E3-E2

E3 remains associated with E2 through non-covalent interactions in the mature E3-E2 protein (Figure 0.14). The mature E2 protein has three-immunoglobulin like domains (from viral membrane proximal to distal; dC-dA-dB) (Figure 0.14). E2-dA is sandwiched between E2-dC and E2-dB. An elongated β -ribbon connector is also part of the E2 structure, and runs alongside E2-dA connecting E2-dB with E2-dC. The β -ribbon connector has three distinct structural parts (arch1 at the interface between dA and dC, arch2 at the interface between dA and dB, and the central arch that has an acid sensitive region (ASR), which runs alongside dA and makes most of the contacts with E3 on side and with E1-dII on the other side). Finally, the E2-dC connects to E2-stem region, which is linked to the E2 trans-membrane anchor.

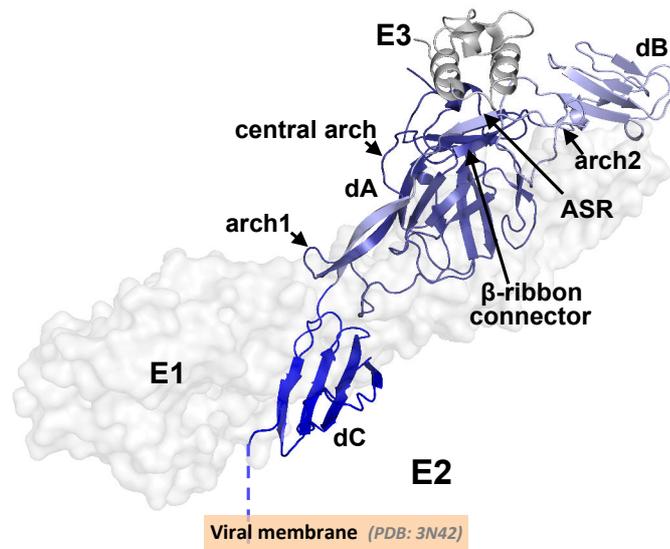


Figure 0.14 Domain organization of mature CHIKV E2 protein. The structure is extracted from CHIKV E3-E2-E1 heterodimer structure in pre-fusion state (PDB: 3N42). Domain organization and β -ribbon connectors are shown as cartoon representation in different shades of blue and are labeled. The dC region connects to stem, and transmembrane (TM) region in the viral membrane.

The smaller E3 protein

The first structure of alphavirus E3 protein came from the chikungunya virus E3-E2-E1 complex (PDB: 3N40 and 3N42) (82). In this complex structure, E3 is positioned on E2 in the E3-E2-E1 complex, covering domain A and B of E2, which in turn covers fusion loops of E1 (Figure 0.15). The core of E3 structure is like a horseshoe of three α helices, and all three helices of the core make several contacts with E2. E3 interactions with E2 is at an acid sensitive region (ASR), a region of E2 β -ribbon arch that has rigid conformation with regular structure in neutral pH, but becomes disordered at acidic pH. E1 interacts on the opposite side of this ASR.

Pre-fusion conformation

E3-E2-E1 heterodimer

The elongated E2 molecule remains tightly associated with E1 on virion surface through non-covalent interactions throughout the E2-E1 interface (Figure 0.15A). E1 is positioned in oblique orientation to viral membrane and E2 is positioned in such a manner that it masks most of the E1 structure. Major interactions that stabilize E2-E1 structure are between E2 domain C, β -ribbon connector and E1-dII. E2-dB also makes key interactions around the E1-fusion loops at dII tip. This E2-E1 arrangement stabilizes E1 in its metastable pre-fusion conformation. Importantly, E2 masks the hydrophobic fusion loop present on E1, at the tip of its dII, by forming a clamp around it by its dA and dB. From the crystal structure of CHIKV E3-E2-E1, E3 presence on E2 forms a brace that stabilizes E2 dA-dB clamping around E1-FL. Since, in chikungunya virus E3 is retained on virion surface through its tight interaction with E2, the name E3-E2 will be used for the rest of the chapter, as the CHIKV receptor binding protein. At the E3-E2 interface, E3 Y47 forms a hydrogen bond with E2/K254; E3 Y48 forms a hydrogen bond with E2 E166 and a π - π interaction with E2 H256.

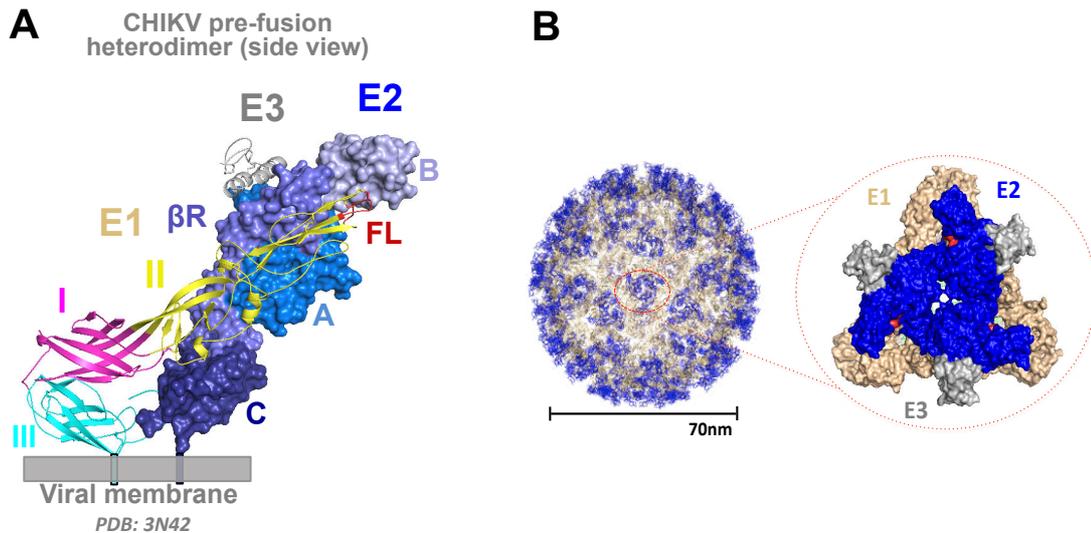


Figure 0.15 Pre-fusion conformation of CHIKV E3-E2-E1 heterodimer and trimer of E2-E1 heterodimer in CHIKV surface spike. (A) Domain organization of E3-E2 and E1 structures are shown. (B) CHIKV trimeric spike organization from viral surface is zoomed and shown. The trimeric spike structure is prepared in PyMol using the PDB: 6NK7, which is a cryo-EM structure of CHIKV in complex with a CHIKV receptor, MXRA8 (MXRA8 is not shown here for clarity) (90).

Virion surface ‘spike’, trimer of E3-E2-E1 heterodimers

Cryo-EM reconstructions of the alphavirus envelope with E2-E1 (available for CHIKV, SINV, SFV, VEEV etc.) show that heterodimers of E2-E1 arrange into spike like structures, where trimers of E2-E1 heterodimers are present in a manner, wrapping around a hypothetical central trimeric spike axis (80, 82, 88, 91, 92). CHIKV and other alphaviruses follow a T=4 icosahedral shape of the capsid (because of E2-capsid interactions) and hence, 80 of these spikes on the entire virion surface (240 copies each of E2-E1 heterodimers). In the spike, three peripherally positioned E1 molecules (in their pre-fusion bent conformation) surround three centrally positioned E2 molecules (Figure 0.15B). E2-dA interactions among three E2 molecules in the spike further stabilize the intra-spike interactions, while inter-spike interactions are mediated through E1. The smaller E3 protein is held on to E2 through

weak interactions at the ASR region on E2. This positioning of E3 in the virion surface is at the apex of the spike, in a way protruding outwards from spike center.

Post-fusion conformation – E1 homotrimer

The post-fusion hairpin conformation of alphaviruses is a homotrimer of fusion protein E1. Post-fusion structure is known for SFV E1 (88) (Figure 0.16). For CHIKV and other alphaviruses the post fusion structure is not available.

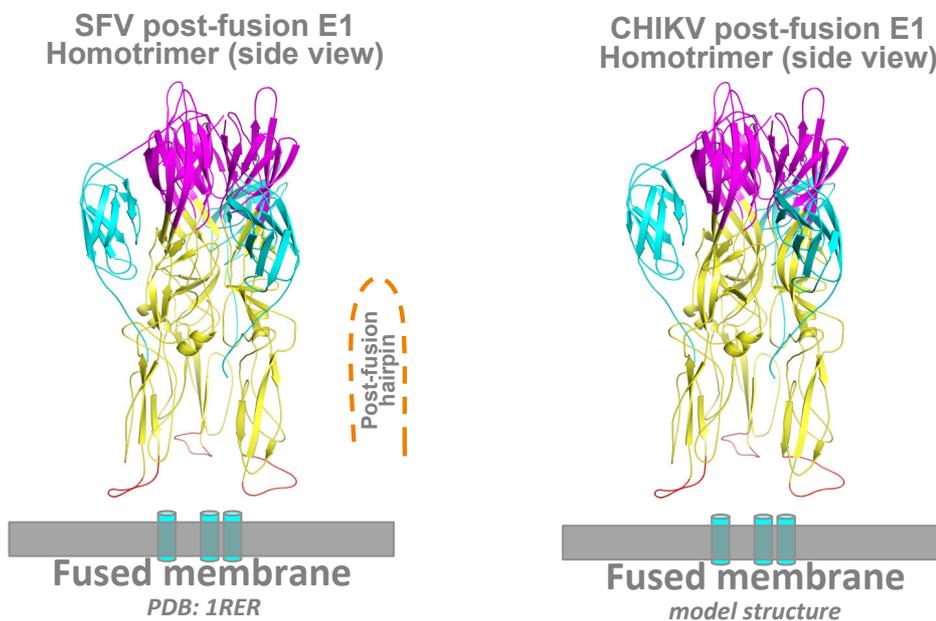


Figure 0.16 SFV E1-homotrimeric structure in post-fusion state from SFV (PDB: 1RER) and CHIKV (model structure). E1-domains follow the same color-coding as it is shown for E1 in the pre-fusion state (Figure 0.11). The post-fusion hairpin schematic refers to the hairpin conformation of E1 in post-fusion state.

We modeled the post-fusion structure of CHIKV E1 using homology modeling using the SFV post-fusion E1 homotrimer structure (PDB: 1RER) as template. The modeled structure of CHIKV E1 homotrimer (HT) showed good overlap with the template structure (homology modeled structure is discussed in detail in chapter 3). The post-fusion trimer reveals hairpin conformation of E1 with E1-dI and E1-dII

constituting the central trimeric part (majorly contribute to interactions that stabilize the trimer interface), and E1-dIII and the connecting stem region forming the outer trimeric part (E1-dIII+stem pack into the groove formed between adjacent E1-dI-dII rods from the central trimeric structure) (figure 0.16).

In the post-fusion conformation, all three domains of E1 maintain their original fold. Major change is in interactions between E1-domains and overall domain positions. With respect to the pre-fusion conformation, in this structure E1-dII rotates by $\sim 15^\circ$ around the dI-dII connecting hinge, and forms a straight continuous rod shape structure composed of E1-dI-dII. Also, E1-dIII moves from its original position in the pre-fusion state by $\sim 37 \text{ \AA}$ towards the fusion loop in the post-fusion state. This lead to breaking of existing interactions at domain I-dIII interface from the pre-fusion structure to formation of new interactions in the post-fusion structure. A highly conserved histidine, H331, in dIII interacts with K16 (also conserved) in dI of E1 at the dI-dIII interface, in the pre-fusion state. In post-fusion conformation this interaction is not present. In post-fusion conformation, H331 from one E1 protomer interacts with N149 from another protomer in the E1 trimer structure. Another conserved histidine, H3 is in dI (not at dI-dIII interface), which also has altered interactions between pre-fusion and post-fusion conformations.

Extended intermediate structure

Very recently, a low-resolution ($\sim 14.5 \text{ \AA}$) extended intermediate structure (after acidic pH trigger) of alphavirus E1 was described. Cao et al (22) in a cryo-EM based study analysed the structural details of extended bridge kind of connections between SINV envelope and liposomes (Figure 0.17). At pH of 6.4, density for three of the E2 molecules at the center of the trimeric spike remain unaltered. However, densities for three of the E1 molecules at the periphery of the trimeric spike were missing. The

authors have discussed and proposed a model that the extended bridge kind of connections seen between the fused SINV and liposomes at the point of contact are extended conformations of the E1 protein. From their model (Figure 0.17), three E1 molecules arrange in a domain-linearized manner (dIII-dI-dII in a linear extended form), where they extend to insert the fusion loop (present at dII tip) into the liposome membrane and are present in their membrane bound forms. The distance calculations between the fused virion and liposomes are measured to be $\sim 150\text{\AA}$, which is similar to the distance of alphaviruses E1 while all three domains are present in a linearized form.

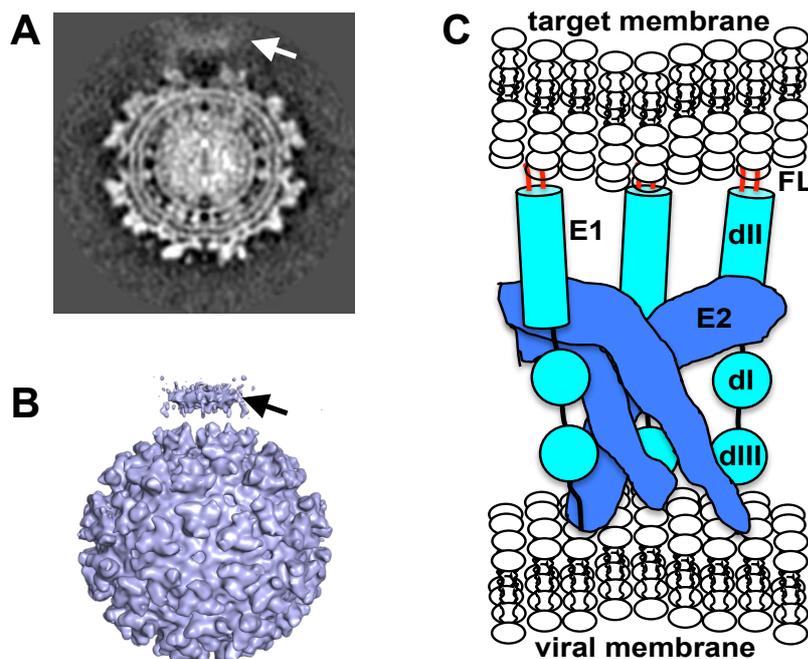


Figure 0.17 Acidic pH induced fusion of SINV-liposome fusion. (A) A cross-section of the EM map shown in B-panel. (B) Surface rendering of the EM reconstruction map (EMD-2374) of SINV fused with a target liposome membrane. The arrow in both A and B panel points to the position of the liposomal target membrane, which is calculated to be $\sim 150\text{\AA}$ away from the outer leaflet of the viral membrane. (C) Model proposed by the authors of the study for E2-E1 arrangement at the point of fusion between SINV and liposome. The figures are adapted from the study by *Cao et al, PNAS, 2013 (22)*.

Regulation of alphavirus entry proteins during cell entry

Receptor binding – Alphaviruses receptors known

Alphaviruses are mosquito-transmitted pathogens and infect wide range of species like insects, birds, and mammals. The wide host range may be attributed to binding of different receptors on different cell-types. Earlier studies have shown that alphaviruses bind to cell surface, binding is saturable and through E2 protein, and bound viruses can be removed by proteolytic digestion of cell surface molecules without affecting virus surface proteins. Several cell surface molecules have been proposed as alphavirus attachment factor/receptors that mediate cell entry of these viruses. Some of these molecules are high-affinity laminin receptor, integrin, heparan sulfate (HS), DC-SIGN and L-SIGN (34). All the molecules reported to function as receptors for CHIKV and other alphavirus are detailed below.

Reported receptors for CHIKV

Like other alphaviruses, CHIKV shows broad tissue tropism and infects multiple tissue and cell types (77). Examples of such cell types include skin keratinocytes, muscle and joint fibroblasts, liver endothelial cells, macrophages and dendritic cells from the lymphoid tissue, brain epithelial and endothelial cells etc. (65, 77). In absence of a cognate receptor of CHIKV, several cell surface molecules are proposed as CHIKV entry partners. Cell surface proteins, Prohibitin (93), T-cell immunoglobulin mucin domain 1 (94), ATP synthase β -subunit (95), matrix remodeling-associated protein 8 (MXRA8) (96), and cell surface glycosaminoglycans such as heparan sulfate (HS) molecules are reported to mediate cell entry of CHIKV. Wintachai et al (93) through two dimensional virus overlay assay showed that Prohibitin binds to CHIKV in microglial cells. Through co-localization, co-

immunoprecipitation, infection blocking using antibody and siRNA-mediated approach the author confirmed a role for Prohibitin in CHIKV entry (93). Moller-Tank et al (94) through pseudo virion binding and transduction assays showed that TIM-1 may play a role in entry of alphaviruses including CHIKV. Using exactly similar approach as that used for identification of Prohibitin (through virus overlay protein binding assays), C. Fongsaran et al suggested role of a 50kDa protein, named ATP synthase β subunit to be a cell entry mediator of CHIKV in insect cells (95). Through co-immunoprecipitation and co-localization studies they showed CHIKV and ATPS- β co-localization both at cell-surface and intracellular regions. Using targeted antibody mediated inhibition and siRNA-mediated down regulation of ATPS- β , they showed significantly reduced viral cell entry (95). However, none of the above studies characterized the exact receptor binding site/sites on E2 or studied the nature of conformational changes on viral envelope upon receptor binding. In contrast, Zhang R et al (96) recently identified MXRA8 as a receptor for multiple alphaviruses. Through mutational studies they showed that multiple residues on surface of E2-dA and E2-dB regions are responsible for MXRA8 receptor binding. They also showed inhibition of CHIKV infection in different cell types (including primary human synovial fibroblasts, chondrocytes osteoblasts, and skeletal muscle cells), when anti-Mxra8 monoclonal antibodies were used for blocking. In a very recent study (while this thesis was in preparation) by Song et al (97) and Basore et al (90), cryo-EM data on MXRA8 in complex with CHIKV trimeric spike shows the binding site of MXRA8 to be at the groove between two adjacent E2 protomers in the trimeric spike, where residues from E2-dB and E2-dA mediate most of the contacts with MXRA8. This is the first-ever structural data of a receptor with any alphavirus envelope so far and was published very recently in parallel to our work.

Heparan Sulfate as receptor of CHIKV

Of all the receptor molecules (proteins as well as glycosaminoglycans), HS is widely used as an attachment factor or receptor for many different viruses (98). This is mainly because of its presence on almost all the cell types as part of heparan sulfate proteoglycans of the extracellular matrix. The nature of HS/Heparin-protein interactions are studied and reported earlier. Studies with HS/Heparin binding proteins has lead to identification of consensus HBDs in these proteins through which they interact with HS/heparin molecules. Among few of the consensus HBD motifs proposed so far, i. XBBBXXBX and ii. XBBXBX are two of them, where X is any residue and B is a basic residue (R/K) (99), among others XBBBXXBBBXXBBXBX and TXXBXXTBXXXTB (T defines a turn, B is a basic residue, X is a hydrophatic residue) were proposed for Von Willebrand factor, and α - and β -fibroblast growth factors (α FGF, β FGF), respectively (100). These studies not only identified such key sequence factors for interactions with HS/heparin, but also have given insight into the nature and mode of such interactions between proteins and HS/heparin molecules. Similarly, viruses bind to these molecules at cell surface and utilize them for attachment and entry into host cells. Several Arboviruses (including the alphavirus, CHIKV (101–103)) among many other viruses enter into cells using HS/Heparin (104). Multiple reports suggest role of HS in cell entry of alphaviruses. For example SINV, SFV, and EEEV can use heparan sulfate to trigger cell entry (78, 101–103, 105).

Receptor binding induced conformational changes in E3-E2-E1 is the earliest event in alphavirus cell entry

E3-E2-E1 dissociation is an essential step for alphavirus membrane fusion (as it is for cell entry of many other enveloped viruses) is evident from comparison of

pre- and post-fusion structures of fusion proteins from alphaviruses (80, 82, 88, 92)
(Figure 0.18).

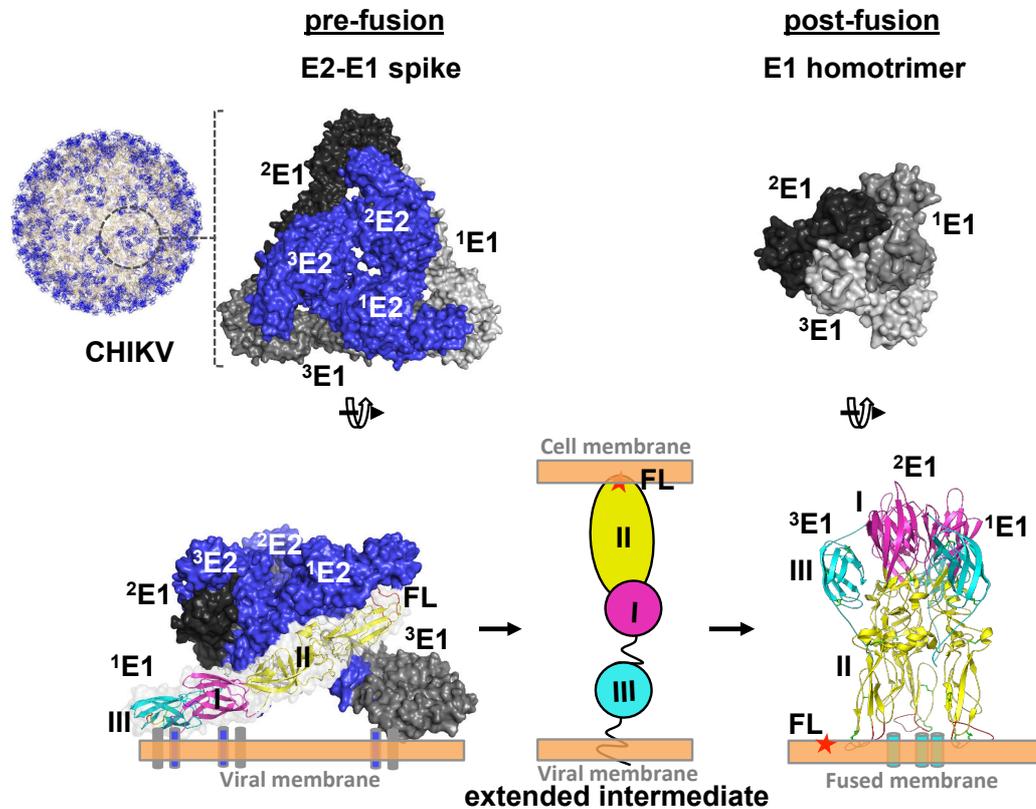


Figure 0.18 CHIKV envelope proteins pre- to post-fusion conformation switch. Top (left to right) - Image from 3D reconstruction of CHIKV E1-E2 into CHIKV VLPs (EMDB id: 5577), surface view (labelled CHIKV), is shown. One trimeric spike structure is shown in circle. Zoomed in image of one E2-E1 trimeric spike is shown to the left (PDB: 3J2W). Top view of E1 (different shades of gray) and E2 (blue) are shown in surface representation. Three protomers of E1 and E2 are labelled, ¹E1, ²E1 and ³E1, and ¹E2, ²E2 and ³E2, respectively. A single monomer of E1 (from PDB: 3N42), in cartoon representation, is shown to explain the pre-fusion hairpin conformation, with the three domains (dI, dII and dIII) and fusion loops (FL) labelled. dI-III linker is marked with a left black arrow. Top view of the post-fusion E1 homo-trimer (PDB: 1RER) is shown to the right. Bottom: shows side view of the trimeric spike in the left, schematic of a possible extended intermediate in the middle (dIII, dI, dII and fusion loops depicted in cyan, magenta and yellow ovals and a red star, respectively), and post-fusion E1 homo-trimer conformation in the right. VM, CM and FM refer to viral membrane, cell membrane and fused membrane, respectively. Dashed lines connecting protein structure figures with VM, represent stem and TM region that anchor the proteins into VM or FM.

Similar to what is observed with many other viruses, receptor binding protein and fusion protein dissociation in alphaviruses may be initiated with the virus binding to the cell surface receptor. This proposition came from the work of Flynn et al (35). They observed that the monoclonal antibody epitopes on E1 and E2 are altered on the virion surface, while the virus is still on the plasma membrane surface. The mAb epitopes, that are not accessible on virion surface, are accessible for mAb binding after the virus attached to the cell surface, but not endocytosed. One possible explanation for this, as explained by the authors, is that the E3-E2-E1 undergoes conformational changes upon receptor binding, and possibly by dissociating from each other or a change in trimeric spike conformation. In another study by Abell et al (106), the authors suggest that SINV membrane fusion can be achieved by reduction of exposed disulfide bridges at cell surface, and this exposure could be triggered by receptor binding induced conformational changes in viral surface proteins. Also, Voss et al (82) have proposed role of receptor binding in induction of E2-dB movements that may lead to fusion loop exposure on E1 at neutral pH. Li et al (80) have speculated E2-dC rearrangements as a possible mechanism for E2-E1 conformational changes and dissociation. They discussed that both receptor binding and/or acidic pH could play a role in this process.

Acidic pH-induced conformational changes in E3-E2-E1

Alphaviruses enter cell through endocytic pathway (16, 77, 78). Alphaviruses follow the clathrin-mediated endocytic pathway (107) and membrane fusion is triggered either in early endosomes (108) or late endosomes. SFV membrane fusion is in early endosome, at ~pH 6.2 (109), whereas SINV fusion is at ~pH 5.6 in late endosome (110). For CHIKV membrane fusion trigger is at ~pH range 5.6-6.2 (111). Alphaviruses are endocytosed immediately post cell surface attachment (112).

Endosomal acidification blockers such as inhibitors of vacuolar ATPase (110, 113), weak bases (114–116) and proton ionophores (109, 117) prevented nucleocapsid release, possibly by preventing membrane fusion by viral fusion protein.

Acidic pH of the endosome leads to E3-E2-E1 dissociation, and triggering of E1 for fusion. Crystal structure of E2-E1 heterodimer of SINV at acidic pH (~pH 5.6) condition, show partial E2-dB dissociation from E1 and disordering of ASR of E2- β -ribbon connector (80). Several cryo-EM based studies with alphaviruses at acidic pH have been reported (22, 28, 88). These studies reveal that interactions between virus and target membranes occur only at acidic pH condition. In a study by Cao et al (22) the authors have studied fusion between SINV and liposomes at a mildly acidic pH (pH 6.4) and based on their study proposed an extended intermediate conformation of E1 that has dissociated from E2 in the trimeric spike (refer to details under extended intermediate structure description, and Figure 0.17). However, since these studies were done without considering the receptor binding event, the contributions from receptor binding and moving away of the three E2 molecules from the trimeric spike center (where these are still present as shown in the same study by Cao et al (22)) to allow moving in of all three E1 molecules to the center of spike to form the E1 homotrimeric post-fusion structure (PDB: 1RER) (88), is not explained in these studies. Gibbons et al (88) proposed that the extended conformation of E1 observed at acidic pH could possibly arise by swiveling of dI-dII around the connecting dI-dIII linker region, or by rotation around the dI-dII hinge, or by changes in stem region or by combinations of all these events. Fuller et al (118) suggested that E1, through centripetal movement, would move to the center of the spike (while E2, through centrifugal movement may escape from the center of the spike), upon triggering by acidic pH. Voss et al (82) suggested that E1, post acid sensing, would swivel over the

dI-dIII linker to the center of the trimeric spike. Further, this extended structure may subsequently proceed to the folded back hairpin conformation (post-fusion conformation) that is essential for driving virus-cell membrane fusion. Like other enveloped viruses, alphavirus fusion proceeds through a hemifusion state. Subsequent formation and expansion of fusion pore for content exchange between effector and target cells has been shown to occur specifically at acidic pH with simultaneous conformational rearrangements in viral envelope fusion protein E1 (28, 119). E1-homotrimers show cooperative associations among each other (either through dI-dII, dIII or FL interactions) at acidic pH and form rosettes of E1-homotrimers, simultaneously driving virus-cell membrane fusion (28, 119, 120).

Role of conserved histidines as sensors of acidic pH in alphavirus fusion protein

Fusion protein E1 in alphaviruses facilitates viral–endosome membrane fusion in the maturing endosome. As endosome gets matured in the cytoplasm, vacuolar proton pumps drive H⁺ ions into the vesicle leading to acidification. When endosome acidification reaches to a threshold pH, E1 undergoes changes in conformation. Following the universal mechanism of membrane fusion by viral fusion proteins, E1 changes from pre-fusion hairpin (monomer), to an extended intermediate, to post-fusion hairpin homotrimer conformation as the membrane fusion proceeds (Figure 0.18, bottom panel).

Several key charged residue protonation/interaction changes are likely to regulate alphaviruses E1 conformational rearrangements (Figure 0.19). Several important interactions (for example H3, H125, H230, D188, G91 etc.) including the interactions of conserved histidines that affect membrane fusion pH threshold (requires more acidic pH for fusion) and stability of the E1 post-fusion trimer structure are reported (89, 121, 122). These residues possess highly conserved

interaction partners (different interaction partners observed in both pre-fusion and post-fusion conformations) are likely to contribute to the stability of both these structures (the pre-fusion state seen at neutral pH, and the post-fusion state seen at acidic pH). Change in protonation of the histidine residues, in general, has been shown to trigger viral fusion protein conformational changes.

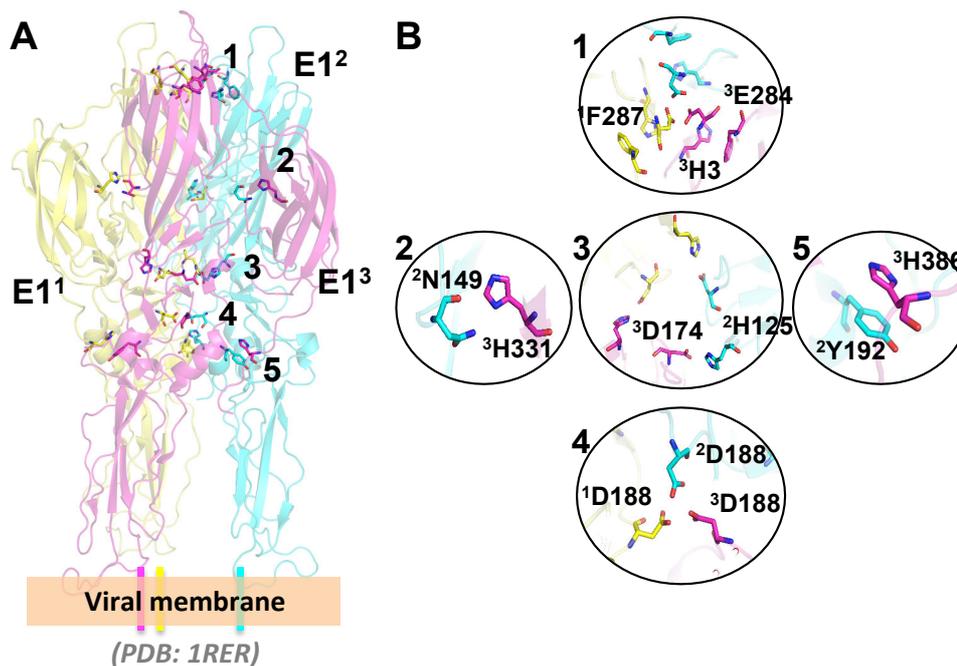


Figure 0.19 Interactions known to affect HT formation. (A) Post-fusion homotrimeric E1 structure from SFV (PDB: 1RER). The three E1 monomers in the trimer structure are colored distinctly and labeled. The interaction networks of the important residues that affect homotrimer formation and stability are shown (represented as sticks and each interaction network is assigned with a number). (B) Each of the interaction networks numbered in panel-A are zoomed and labeled. Interactions among the residues from different E1 protomers are shown by labeling the name of the residue with a starting super scribed number that corresponds to the E1 protomer in panel A. 1, 3, 4 are top views and 2, and 5 are side views of position in panel A.

Class I fusion proteins can be triggered to post-fusion trimeric conformations *in vitro* by factors such as temperature or by urea. Such factors are not shown to trigger a class II fusion protein from SFV into a post-fusion trimer conformation. However, role of other such factors such as a detergent or presence of a lipid bilayer

membrane even at neutral pH could be tested for triggering of class II fusion protein into its post-fusion trimeric conformation. These studies can significantly contribute to our understanding of time, location and the very role of E1 trimerization in the membrane fusion process during cell entry.

E3 as a brace to ‘lock’ E2 dA-B clamp over the E1 fusion loop – a possible fusion regulation mechanism

The first structure of E3 is known from the crystal structure of chikungunya virus E3-E2-E1 (PDB: 3N42) (Figure 0.15) (82). In this structure, E3 protein is positioned like a brace on E2 at the apex of the E2-E1 complex structure. In this way, E3 keeps the domain A and B of E2 in place that clamp around the E1 fusion loops. Implying from this, Voss et al (82), suggested that E3 might, act as a ‘brace’ to keep domain B of E2 in position, covering fusion loops of E1 until E2-E1 dissociation at acidic pH of endosome takes place. As described earlier, Lobigs et al (79) proposed that E3 has N-terminal signal sequence for translocation of structural polyproteins into ER. Garoff et al (81) proposed that E3 would anchor and then retract out of the ER membrane soon after the structural polypeptide translocate into ER. Based on this, it can be hypothesized that the free N-terminus of E3 on the virion surface spike may interact with a target membrane, to which virus approaches for attachment and cell entry.

Unanswered questions:

Because of lack of structural data on alphaviruses E2 at different stages of entry pathway, extensive rearrangements of the E2 protein leading to E2-E1 dissociation remains to be visualized. In addition to this, further heterodimer rearrangements should occur in order for E1-homotrimer formation, but the triggers of

these rearrangements and their exact nature remains elusive.

Thesis research hypothesis:

For understanding the regulation mechanism of CHIKV cell entry proteins, I propose the following hypothesis based on the available information from published literature and based on our current understanding of the subject;

- Receptor binding to CHIKV E2 may lead to its dissociation from E1
- Acidic pH triggers CHIKV E1 conformational changes, dI-dIII linker region is the hinge for swiveling of E1
- E3 has a regulator function in viral entry

Objectives of thesis research work:

I propose the following set of objectives based on my thesis hypothesis to be characterized, for addressing the key questions with respect to understanding of the CHIKV entry protein complex regulation mechanism, as part my thesis research work;

- Optimization of over-expression of CHIKV E1, E2, E3 proteins using insect cell and *E. coli* expression system
- Characterization of CHIKV E2-HS interactions to understand receptor binding trigger of membrane fusion
- Characterization of E1-dI/dIII cysteine stapling mutant to understand acidic pH triggered conformational changes in E1.
- Characterization of E3-membrane interactions to explain a plausible mechanism for E3 release from E2.
- Development of a functional pseudo virion system for entry assay.

CHAPTER 1

EXPRESSION AND PURIFICATION OF CHIKV CELL ENTRY PROTEINS FOR STRUCTURE-FUNCTION STUDIES

Bacterially expressed Chikungunya virus cell entry proteins, E1, E2 and E3, are structurally and functionally similar to insect cell expressed or virion surface 'cleaved-off' proteins. **Sahoo B** and Chowdary TK (**In preparation**)

1.1 INTRODUCTION:

Viral envelope proteins undergo complex maturation steps through the secretory pathway of the infected cell. In general, it is observed that several viral envelope proteins depend on a chaperone for proper maturation. In majority cases, receptor binding protein or a product of proteolytic processing of pre-mature envelope protein, acts as a chaperone. However, in few cases fusion protein and receptor binding protein chaperone each other. In such cases, co-expression of both the proteins is required. For example in flaviviruses, prM protein chaperones E, the fusion protein, for proper folding, assembly and surface presentation (123). Another example of co-expression of envelope protein along with its chaperone protein is that of HSV gH/gL. In a cell expression study, when gL is not co-expressed along with gH, antigenic conformation and surface representation of the gH protein is affected and the protein is retained inside the ER (124). In case of HIV-1, co-expression of the envelope glycoprotein subunits, gp120 and gp41, has been reported to be necessary for generation of antigenic conformation (125) and normal fusogenic function (126). Similarly, in case of Ebola virus, surface co-expression of GP1 and GP2 subunits of the GP protein has been reported to be essential and the level of their expression is reported to regulate production and infectivity of virus particles (127). The antigenic properties of Influenza A and B viruses were retained and were of similar levels as that of the WT virus when envelope glycoprotein subunits - HA1 and HA2 or recombinant HA were expressed and tested in insect as well as mammalian expression systems (128).

Necessity of co-expression of a companion protein, which in many cases functions as a chaperone for proper folding of the viral fusion protein makes expression and purification of viral fusion proteins for high-resolution structure-function studies very

challenging. In fact, despite several attempts by various research groups to obtain practically usable protein yields, HIV gp41 and gp120 recombinant expression proved to be very difficult (129). Moreover, complex post-translational processing, such as glycosylation and Golgi resident protease cleavage of viral envelope proteins necessitates use of insect or mammalian expression hosts for recombinant viral envelope protein production.

Bacterial expression systems have also been used for viral envelope protein expression and purification. Functional form of HIV-1 envelope glycoproteins in gp160 (gp120 plus gp41 subunits) form were expressed and characterized in *E. coli* cells (130) as well. *E. coli* expressed influenza HA1 and HA2 proteins have been tested in functional assays successfully (131, 132). Bacterial expression systems for recombinant protein production are simple, cost effective and produce better protein yields. However, bacterial expression systems lack the post-translational modification machinery (such as glycosylation and proteolytic maturation), required for proper maturation and native folding of the viral envelope proteins. Hence, bacterial expressed recombinant viral cell entry proteins have to be thoroughly characterized for their folding and functional activity.

So far, structure-function studies on alphavirus entry proteins (including CHIKV) are limited to use of either the entire virion particle or proteins cleaved off, proteolytically, of viral surface (22, 28, 88, 133, 134). However, there are major caveats in using these proteins for high-resolution structure studies. For example, with the proteins cleaved-off of viral surface or of cells expressing these proteins on their surface after viral infection, the yield is too low that they can be studied using biochemical and biophysical approaches. Also, in a previous study by Kielian et al (135), the authors have shown that viral surface

E2-E1 proteins (the ectodomains) when proteolytically cleaved and detergent isolated, they dissociated from each other even at neutral pH and did not retain the E2-E1 heterodimer conformation that is seen on the virion surface.

Several studies also reported recombinant over expression of alphavirus E1, E2 and E3 proteins in insect and mammalian cells. Voss et al (82) used S2 drosophila cells for over-expression of CHIKV E3-E2-E1 soluble ectodomain protein complex. The proteins were expressed as secreted out proteins using a native secretory signal. The first crystal structure of E3-E2-E1 complex was solved using this purified protein. Crystal structure (PDB: 3N42) revealed that the protein complex, expressed and purified using S2 cells, folded to its native conformation and processed correctly – formed a heterodimer complex of E2-E1, and E3 was held through interactions with E2. On the contrary, SINV E2-E1 complex purified by Li et al (80) in a similar manner (using S2 cells and expressing only soluble ectodomains) formed a trimer (three E2-E1 molecules surrounding a central axis, similar to trimeric spike organization on virion surface). Moreover, in this case, E2-E1 complex were not dissociable in a buffer solution with pH range of 5.5-9.5 (even below the pH threshold of fusion, pH 6.0~6.5, which is believed to trigger E2-E1 dissociation in alphaviruses). It is possible that the difference in structural organization of recombinant-expressed and purified proteins could be because of differences in sequences between CHIKV and SINV. Most importantly, in both these reports (CHIKV E3-E2-E1 expression and SINV E2-E1 expression), authors have discussed about the difficulties with purification, and have mentioned requirement of multiple purification steps (further to affinity purification, use of double strep-tag/6His-tag for CHIKV and SINV proteins, respectively). Other insect cell expression trials in sf9 cells (using the baculovirus

expression system) were done but limited to small scale expression studies only (136). For structure-function studies using biochemical assays, typically, more amount of protein than what is possible with the methods used in the literature cited above is required. And, it is desirable to have a method to express these proteins independently, if one wishes to characterize conformational dynamics in receptor binding protein and membrane fusion protein. Studying such dynamics for each protein in a complex would be difficult.

Liao et al (60), expressed SFV E1 protein domain III in *E. coli* expression system, purified the protein from inclusion bodies and used the refolded protein domain in membrane fusion assays. The refolded protein domain inhibited SFV membrane fusion with liposomes, implying that the bacterial expressed E1 domain III is functionally active. In another study by Weber et al (102), *E. coli* expressed CHIKV E2 protein (or its domains), competed with CHIKV virus particles for HS binding on cell surface. Bacterially expressed alphavirus E1 and E2 are used for serodignosis (137) (138) and viral-host cell interaction studies (102, 139). Sánchez et al (139) have used the *E. coli* expression-purification system to produce different domain regions of SFV E1 protein and have successfully studied trimer formation using these purified proteins in *in vitro* biochemical assays. Thus, *E. coli* purified proteins can be utilized for structure-function studies.

Advantages of using bacterial expression system are: easy adaptability and generally, higher protein yields (required for biochemical assays) compared to insect cell expression system or mammalian cell expression system. I intended to develop protocols for expression of E2-E1 complex using insect cell expression system, and in parallel, optimize protocols for bacterial expression of E2 and E1 proteins, independently. My rationale for adapting both the expression method is, E2-E1 complex expressed in insect

cells, as reported in the earlier literature, would fold into a native conformation. But, if I fail to get good protein yield (as has been the case in earlier studies from others), bacterial expression system should serve as an alternative source.

In this chapter, I described cloning, expression, and purification protocols for expression of E2, E1 and E3 in insect cells, mammalian cells and *E. coli* expression systems. Briefly, I optimized over-expression and purification protocol for CHIKV E3, E3-E2, E2 and E1 proteins using *E. coli* expression system. I did a thorough characterization of these proteins for their structure and function using different biochemical and biophysical approaches. I also developed antibodies against E2 and E1 peptides, which served as good resources for optimizing the protein expression trials. My attempts to purify E3-E2-E1 complex using Sf9 insect cells using recombinant baculovirus method, did not give desirable results – very poor protein yield and purity. I was not successful with mammalian expression system as well. At the end, I discuss results of expression and purification trials in comparison with expression attempts of others.

1.2 MATERIALS AND METHODS:

1.2.1 Chikungunya virus envelope protein sequence

CHIKV envelope protein coding sequences used are from LR2006_OPY1 strain of the virus and details of the protein sequences (protein_id="ABD95938.1") are available at <https://www.ncbi.nlm.nih.gov/protein/90654094/>. The ectodomain coding regions were confirmed using the virus pathogen resource database at <https://www.viprbrc.org/brc/viprStrainDetails.spg?ncbiAccession=KT449801&decorator=toga>. PCR amplicons of CHIKV E2 and E1 full-length coding sequences were a kind gift from Prof. Scott C. Weaver, UTMB-Galveston.

1.2.2 Chikungunya virus E3-E2-E1 structure coordinates used for analyses

We analysed the crystal structure of CHIKV E3-E2-E1 furin cleaved mature protein complex (PDB ID: 3N42) for percent secondary structure calculations and crystal structure of CHIKV E3-E2-E1 immature protein complex (PDB ID: 3N40) for identification of a conformation specific epitope at E3-E2 junction for polyclonal antibody development.

1.2.3 Preparation of constructs for mammalian, insect cells and *E. coli* expression

The constructs for recombinant protein expression of CHIKV envelope structural proteins were generated by PCR amplifying the coding sequences of E3, E2, 6K, and E1 full-length and ectodomain regions (full-length without the trans-membrane part) for cloning. For cloning into pCDNA3.1 for surface expression in mammalian cells, first E3 and 6K coding regions were synthesized by using multiple small oligonucleotide fragments (a set of 17 primers for E3 and 12 primers for 6K, Appendices Tables, a1.1 and a1.2 respectively, were generated using E3 sequence from DRDE-07 strain of CHIKV,

GenBank: EU372006.1 and 6K sequence from LR_2006_OPY1 strain of CHIKV, GenBank: ABD95938.1 as inputs in Gene2Oligo server) in an overlapping PCR method, using a published protocol (140). The coding regions for full-length E2 and full-length E1 are PCR amplified from pUC19+E1-fl/pUC19+E2-fl constructs. The final clone (pcDNA3.1 + E3-E2-fl-6K-E1-fl) was prepared in two sequential steps. First, E3 and E2 coding regions were cloned in a three-point ligation into pcDNA3.1 vector. This was followed by in frame cloning of 6K and E1 coding regions into the previously cloned pcDNA3.1 + E3-E2 construct to prepare the final mammalian cell expression construct of full-length forms of CHIKV envelope structural proteins. We used this construct for cell-surface expression of CHIKV envelope structural proteins for packaging into pseudo lentiviruses using a 3rd generation lentivirus packaging system (discussed in detail in chapter 5).

For making the insect cell expression construct, we PCR amplified the E3-E2-ectodomain, E2-ectodomain, and E1-ectodomain regions for cloning. The coding regions (from LR_2006_OPY1 strain of CHIKV, GenBank sequence accession no. DQ443544) for E2-ectodomain (CDS: 8542-9810) and E1-ectodomain (CDS: 9811-9993) sequences were PCR amplified from CHIKV-E2-fulllength and CHIKV-E1-fulllength sequences from an intermediate pUC19 based clone (pUC19+E1-fl and pUC19+E2-fl clones) as explained previously. The CDS for E3-E2-ectodomain was generated by cloning E2-ectodomain coding region in frame with E3 coding region (CDS: 8350-8541). The insect cell expression construct (Mel-E3-E2-GS-E1-ectodomain complex) was prepared in a multistep process using multi-intermediate clones. The final ectodomain expression construct has Honeybee melittin (at N-terminus), E3, E2-ectodomain, Glycine-Serine (GS) linker

(‘GSGSHRQRRSTGSGS’ - replaces the E2-TM and E1-TM regions and connects end of E2-ectodomain with beginning of E1-ectodomain), E1-ectodomain and a C-terminal 6X-His tag. A similar GS-linker (size and composition) was used by Voss et al (82) for expressing E3-E2-E1 in drosophila cells and determining the crystal structure. All the coding regions for expression of these proteins as a complex were cloned in the order of sequences mentioned above (from left to right) into pFastBac1 insect cell expression vector (please refer to results section). I used these clones to generate recombinant bacmids in *E. coli* DH10Bac cells following a standard Bac-to-Bac™ protocol. These recombinant bacmids were purified and used for transfections into *Spodoptera frugiperda* (sf9) cells for recombinant baculovirus generation for protein expression and purification studies.

Coding sequences for each of the CHIKV envelope structural proteins: E3, E3-E2-ectodomain, E2-ectodomain, and E1-ectodomain regions were PCR amplified and cloned into pET24b expression vector with a C-terminal 6His-tag for *E. coli* expression-purification purposes. Since bacterially expressed E3-E2 can not be proteolytically processed, as it would be in insect cell expression or mammalian cell expression, E2 ectodomain alone is also cloned into pET24b. For dual expression in *E. coli* the dual expression vector pET-Duet1 was used. E3-E2-ectodomain coding region was cloned into MCS1 and E1-ectodomain coding region was cloned into MCS2 of this vector for dual expression purpose. All the final expression constructs are sequence verified and the details about primers and restriction sites used for each of the cloning’s are listed in Appendices Tables, a1.1, a1.2 and a1.3.

1.2.4 Mammalian cell expression and purification

We used the pcDNA3.1 based pBS002 construct (pcDNA3.1 + E3-E2-fl-6K-E1-fl) for cell-surface expression of full-length structural envelope proteins of CHIKV in HEK293T, and Hela cells. Cells were grown in DMEM media with 10% FBS and 1X PSA antibiotic at 37°C inside a humidified temperature incubator with 5% constant supply of CO₂. Cells were grown as adherent cultures and trypsin-EDTA (0.05%) solution was used for cell detachment during sub-culturing. 12 hrs prior to transfection, cells were seeded at $\sim 0.5 \times 10^6$ cells (with >95% viability)/well of a 6 well plate so as to have $\sim 70\%$ confluency. Cells were transfected with expression plasmids using Lipofectamine 3000 reagent following a standard suppliers protocol (in DMEM media without any antibiotic and FBS). Transfection mixture used was removed from the treated cells ~ 6 hrs post transfection and replaced with 10% FBS and 1X PSA antibiotic containing fresh DMEM media. For expression analysis, cells transfected pBS002 expression construct was lysed in 1X RIPA buffer and whole cell lysates were run in a 12% SDS-PAGE gel for further analysis by WB using E1 polyclonal antibody.

1.2.5 Insect cell expression and purification

Recombinant bacmids used for transfections into sf9 cells for expression studies were prepared by transposing (in between the Tn7L and Tn7R flanking sites) the CHIKV ectodomain expression ORF from pBS001 vector (pFB1 + Mel-E3-E2-GS-E1-6XHis ectodomain complex) inside *E. coli* DH10Bac cells. Insect (sf9) cells were grown in sf900II SFM with 1X PS antibiotic at RT both as semi-adherent and suspension cultures. Cells were transfected with recombinant bacmids using Cellfectin II reagent (purchased from Invitrogen™) following a standard supplier's protocol. Post transfection, recombinant

baculoviruses positive for expression of CHIKV envelope structural proteins were collected from the supernatant fraction of transfected cells. P1 viral stock was collected 6 d.p.t, and P2 viral stocks were collected 6 d.p.i, with P1-viral stocks. P3 viral stocks were generated in suspension cultures after infection with P2 viral stocks. These P3 stocks are collected at 6 d.p.i, and used for infection of sf9 cells (at a cell density of $\sim 2-3 \times 10^6$ cells/ml with $>95\%$ viability) at 10-20ml of P3 stock/1ltr of suspension culture for protein expression-purification purpose.

For checking expression and purification of CHIKV Mel-E3-E2-GS-E1 protein complex, sf9 cells transfected with recombinant bacmids were assessed (both cellular and media fractions were checked at 6 d.p.t.) by SDS-PAGE and WB analysis. Once expression was confirmed, 1liter batch cultures were set for expression-purification purposes. At optimum expression conditions (72hrs post infection with P3), culture medium was collected from recombinant baculovirus infected sf9 cells. These media containing the secreted-out proteins were concentrated and buffer exchanged to TN (20mM Tris pH8.0, 150mM NaCl) buffer using Akta Flux - tangential flow concentrator with a 5kDa MWCO membrane filter. The concentrated fraction was loaded onto the Ni-NTA column for His-tag affinity purification. Specifically, bound proteins (after washing with 10 column volumes of wash buffer, containing 25mM imidazole in TN buffer) were eluted at 150mM and 300mM imidazole concentrations in TN buffer along with some contaminants. The elution fractions were pulled together and used in another round of purification step using anion exchange chromatography (Mono-Q column used was purchased from GE Healthcare) to remove contaminants. The different elution peaks from this purification step containing protein of interest were concentrated together. As a final purification step, size

exclusion chromatography (Superdex 200 10/300 GL column used was purchased from GE Healthcare) was used. Aliquots at different elution volumes of size exclusion chromatography were concentrated using 10kDa MWCO amicon ultra centrifugal filter units and stored at 4°C (with cOmplete™ protease inhibitor cocktail from Roche purchased from Sigma) till further use in western blotting with anti-His antibodies.

Concentrations of the purified proteins were estimated, by taking absorbance at 280 nm with protein specific molecular weight and molar extinction coefficient parameters as inputs, using NanoDrop machine. We also used SDS-PAGE and WB methods for qualitative analysis of the purified CHIKV envelope protein complex.

1.2.6 Bacterial expression and purification

Bacterial expression-purification of each of the recombinant envelope proteins of CHIKV E3, E3-E2-ectodomain, E2-ectodomain and E1-ectodomain were optimized using BL21 (DE3) strain of *E. coli*. For this, expression plasmid transformed cells were grown as primary cultures in LB medium supplemented with 50 µg/ml of kanamycin at 37°C overnight with shaking at 220 rpm. The secondary cultures (inoculated with 1-2% of primary culture) were induced at 0.8 OD₆₀₀ with 1mM IPTG for over-expression of these proteins. Optimum expression conditions for each of the proteins are summarized in Table 1.1.

At optimum expression condition, recombinant protein expressing cells were harvested by centrifugation at 4°C for 5 minutes at 7000 rpm. We performed expression analysis by running the samples in either 12% or 15% SDS-PAGE gels in SDS-loading dye. For purification studies, cells were first lysed in lysis buffer (TN buffer + 0.2mg/ml lysozyme and 100µM PMSF) followed by sonication at 40% amplitude with 10sec on,

10secs off pulse cycle for 15 minutes. Following cell lysis, the total bacterial proteins were fractionated into soluble and insoluble fractions by centrifugation at 4°C for 40 minutes at 12,000 rpm. Both cytoplasmic and cell-membrane fractions were analysed by SDS-PAGE and WB to check for the presence of over-expressed viral envelope proteins in soluble (supernatant) and/or insoluble fraction (pellet). All the proteins found to be in the insoluble inclusion bodies fraction after cell lysis were then solubilized using optimum urea concentration for a period of 3hrs-5hrs at RT. The expression conditions were optimized by varying the temperature of the growth, IPTG concentration, IPTG induction time. Similarly, protein extraction from the inclusion bodies, and subsequent purification steps were optimized. Different urea concentrations, and incubation time in with urea were tested. Optimal condition for expression and purification were determined based on the final protein yield and purity. Optimal conditions of expression and solubilization, for E3, E3-E2, E2 and E1, are listed in Table 1.1.

Table 1.1: Optimum conditions for *E. coli* expression and solubilization of CHIKV cell entry proteins.

Protein	Secondary reaches OD₆₀₀=0.8	Secondary growth condition in <i>E. coli</i> BI21 (DE3)	Solubilization buffer
E3	Inoculate with 1% primary for 3 ^{1/2} hrs	1mM IPTG induction for 3hrs at 16°C, 220 rpm,	TN pH7.6+6M urea
E3-E2	Inoculate with 2% primary for 5 ^{1/2} hrs	1mM IPTG for 4hrs at 37°C, 250 rpm	TN pH8.0+6M urea or TN pH10.0+2M urea (equally efficient)
E2	Inoculate with 1% primary for 2 ^{1/2} hrs	1mM IPTG for 3hrs at 25°C, 220 rpm	TN pH8.0+8M urea
E1	Inoculate with 1% primary for 3 hrs	1mM IPTG for 3hrs at 37°C, 220 rpm	TN pH7.6+4M urea

The urea solubilized protein fractions were subjected to batch purification under denaturing conditions using 6His-tag affinity chromatography. Briefly, 10 ml of urea solubilized His-tag fusion protein containing supernatants were loaded onto 4ml of the buffer (TN+6-8M urea buffer) equilibrated Ni-NTA resin in a gravity flow column and incubated for 30 minutes at RT. The column was washed with 10-column volumes (CV) of wash buffer (TN + 6-8M urea + 25mM-35mM imidazole). Specifically, bound protein was then eluted with 1 CV of elution buffer (TN + 4-8M urea and 300mM imidazole). The eluted protein fractions were analysed by SDS-PAGE for assessing the purity of the purified proteins.

The fractions that are reasonably pure, from affinity purification, were mixed

together and then the protein was refolded by slow removal of urea. We used the slow dilution method in ice cold refolding buffer (TN pH 8.0 + 1mM EDTA and 100 μ M PMSF) for refolding of each of these proteins. Proteins were refolded by first diluting the samples into ice-cold refolding buffer so as to have a final 1M urea concentration. This process was done rapidly by adding the protein samples into ice cold refolding buffer under constant stirring on a magnetic stirrer. Followed by this a 30 minutes incubation of the diluted protein samples (diluted to \sim 50 μ g/ml of protein) was done on ice.

As final step of purification, the protein was exchanged into ice cold refolding buffer (TN pH 8.0 + 1mM EDTA and 100 μ M PMSF) with several buffer exchanges, and concentrated using ultrafiltration. Slow refolding by removing the urea at a very slow rate is critical for the proper refolding of the protein, as rapid exchange into buffers without urea resulted in precipitation of the protein. Care was taken that the protein does not precipitate out of solution during the buffer exchange and concentration step. In final step, post concentration, protein samples were spun at 14000 rpm for 30 minutes at 4 $^{\circ}$ C to remove any aggregated proteins from the concentrated protein samples. These protein samples were then stored either at -80 $^{\circ}$ C or 4 $^{\circ}$ C till further use. The final protein samples (after concentration step) did not contain any urea (TN pH 8.0 + 1mM EDTA and 100 μ M PMSF).

Concentration of the purified proteins were estimated, by taking absorbance at 280 nm for each of the individual proteins with protein specific molecular weight and molar extinction coefficient parameters as inputs, using NanoDropTM. All these parameters were obtained by using the protparam program of ExPASy available at <https://web.expasy.org/protparam/>. SDS-PAGE and WB analysis was performed for

qualitative analysis of the purified proteins.

1.2.7 Size exclusion chromatography

We performed size exclusion chromatography on purified E1 protein for structural analysis. We loaded ~200µg of purified E1 protein per run in TN buffer (with 150mM NaCl) by loading the samples onto buffer pre-equilibrated FPLC column (Superdex 200 10/300 GL). At 0.5ml/min flow rate, protein elution was detected using 280nm detection wavelength. The peak fractions for purified proteins were then verified by SDS-PAGE and WB analysis. Gel filtration standard markers (Alcohol dehydrogenase – 200kDa, Beta amylase – 150kDa, Bovine serum albumin – 66kDa, Carbonic anhydrase – 29 kDa, Cytochrome C – 12 kDa) ran in Superdex 200 column under similar parameters were used for generating a standard curve for accurate molecular weight calculations of purified proteins. Molecular weight calculations were performed using the formula $K_{av}(B) = \{V_e(B) - V_0\} / (V_c - V_0)$, where $V_e(B)$ is the retention volume of unknown samples, V_c is the total column volume (23.56ml), and V_0 is the void volume (8.34ml).

1.2.8 Anti-E1 and anti-E3-E2-epitope specific pAb generation

We used the commercial service of Abgenex for polyclonal antibody development. The protocol is as follows: *E. coli* purified E1 protein and an epitope of E3-E2 protein were used for generation of custom polyclonal antibodies. For anti-E1 pAb development *E. coli* purified, refolded and concentrated recombinant E1-ectodomain protein was used as an immunogen. Similarly, for E3-E2-ectodomain junction specific pAb development, a 15 amino acid peptide – “CSPHRQRRSTKDNFN” from the E3-E2 connecting region was chosen as an epitope and synthesized with N-terminal Cysteine residue for KLH conjugation for immunization. 6-month-old New Zealand white rabbits were used for

immunization. At intervals of 15 days, three booster doses were given with respective antigens. 3 days after the third booster, a final intravenous booster dose was also given. Animals were bled for 3 days after the final booster. Serum was separated; purified antibody concentration was measured using Bradford assay. Specificity of the antibodies was assessed by WB analysis on ~100ngs of purified E1-ectodomain and E3-E2-ectodomain protein samples. Optimum dilutions as standardized for E1-pAb are 1:250000 and 1:10000 for WB and immunofluorescence analysis, respectively. Similarly, optimum dilutions as standardized for E3-E2 epitope specific pAb are 1:25000 and 1:1000 for WB and immunofluorescence analysis, respectively.

1.2.9 CD spectroscopy

Far-UV CD spectra for each of the *E. coli* purified protein samples were recorded with 0.2mg/ml protein concentration in TNE buffer at RT. A 0.1 cm path length quartz cuvette was used for taking CD spectral recordings. The representative spectra in figures are average of three readings each/sample, and blank-corrected. The percentage secondary structural content calculations from far-UV CD spectra were performed using CDPro available at <https://sites.bmb.colostate.edu/sreeram/CDPro/> and also using K2D3 software available at <http://cbdm-01.zdv.uni-mainz.de/~andrade/k2d3/>.

1.2.10 Fluorescence Spectroscopy

Intrinsic tryptophan fluorescence spectra for each of the proteins were recorded with 0.2mg/ml protein concentration in TNE buffer at RT. A 1cm path length quartz cuvette was used for taking fluorescence spectral recordings. The representative spectra in figures are average of three readings each/sample, and blank-corrected. The spectra recorded for

unfolded proteins were taken in presence of 8M urea in TNE buffer. The excitation wavelength was set to 295nm and emission spectra were recorded from 300nm to 450nm. The excitation and emission slit widths were set to 5nm. The highest intensity points in the emission spectra were used to select the λ -emission max values for each of the proteins, either in presence or absence of urea, for comparison.

1.2.11 Glutaraldehyde cross-linking assay

We performed glutaraldehyde (GA) crosslinking (25% glutaraldehyde solution in dH₂O was purchased from Sigma, USA) to capture E1 oligomeric conformations as well as E3-E2-ectodomain plus E1-ectodomain heterodimeric conformation. 0.01% GA was used for cross-linking of purified protein samples. For E1 cross-linking ~5 μ g of total protein was used (in a 25 μ l reaction volume). Crosslinking was performed in sodium phosphate buffer pH 7.4 and pre-calibrated volume of sodium acetate was used to bring down the pH to 5.5 during acidic pH triggering. Crosslinking reaction was stopped, by adding SDS-PAGE loading buffer. Samples (without boiling) were loaded onto 7-10% SDS-PAGE gels for analysis. For E3-E2-E1 heterodimer conformation crosslinking, both E3-E2 and E1 samples were used at 0.2 mg/ml concentration to have ~2-4 μ g of total protein in a sample reaction buffer of 25 μ l. Both the proteins were mixed at 1:1 molar ratios for 1 minute at RT. Crosslinking was performed in sodium phosphate buffer pH 7.4. Crosslinking reaction was stopped by adding SDS-PAGE loading buffer and heating the samples to 95°C for 5 minutes before cooling and loading the samples onto 7-10% SDS-PAGE gels for analysis.

1.2.12 Lipid co-floatation assay

We used lipid-protein co-floatation assay for studying E1-membrane interactions. LUVs were prepared by adding individual lipids at PC: PE: PA: Cholesterol = 1: 1: 0.3: 2 molar ratios. 4 μ moles of total lipid were dissolved in 1ml of chloroform and vacuum evaporated in a round bottom glass container. The flask was kept under continuous rotation to allow for thin layer film formation. 4hrs of further vacuum drying was performed to remove any residual chloroform left. Then overnight solvation was performed in TN buffer pH 7.4. The solvated cloudy mixture was extruded through 0.1 μ m filter at least 11 times or till the mixture become almost transparent. LUV samples prepared were used immediately at 0.5 mM working concentration for co-floatation assay. For the same assay, \sim 100 μ g of E1 protein was used in TN + 0.01% SDS buffer, either at pH 7.4 or pH 5.5. We verified that even in presence of 0.01% SDS LUVs/SUVs structures were intact using EM. Also, we verified using Far-UV CD analysis that 0.01% SDS did not alter the secondary structure of the E1 protein. Samples (\sim 100 μ g of E1 protein plus 0.5mM LUVs in TNE buffer, either at pH 7.4 or pH 5.5) for the experiment were layered over 40% sucrose cushion in a step gradient formed by 4ml each of 40% sucrose, 20% sucrose, and 5% sucrose (in order from bottom to top). The sucrose solutions for the experiment were also prepared in the sample buffer with appropriate pH values. The gradient layering's including the protein samples were done inside a transparent 12 ml ultra-centrifugation tube. All the tubes were properly balanced up to three digits beyond the decimal point. Samples were run using SW 41 Ti rotor for 3 hrs at 35000 RPM (\sim 2 lakh X g), 4°C in a Beckman ultracentrifuge. Post centrifugation samples from respective tubes were collected carefully without disturbing the gradient. Three equal volume fractions corresponding to top, middle and bottom

fractions of the tube were collected and analysed by WB. Percentage of E1 floating along with liposomes was calculated by quantifying band intensities of proteins in top fractions over band intensities of total amount of protein (top+bottom fractions) using ImageJ software.

1.2.13 Pyrene based lipid association assay

We studied E1-membrane destabilization properties using pyrene-based fluorescence assay. In this case, the LUVs prepared were labelled from outside with 40 μ M pyrene (purchased from Sigma-Aldrich, USA). The labelled LUVs were separated from free dyes using a PD10 desalting column (purchased from GE Healthcare, USA). Final experiments were performed with freshly prepared pyrene-labelled LUVs (at 0.5mM working concentration) and bacterially purified recombinant E1 WT protein (at 0.2mg/ml concentration). TN buffers were used for assays either at pH 7.4 or pH 5.5 (pre-calibrated volumes of sodium acetate pH 4.6 were added to bring down the pH to 5.5).

For the experiment, we took spectral readings of pyrene labelled LUVs in presence or absence of E1 WT protein. The excitation wavelength was set to 340nm and emission spectra were recorded from 350-550nm. The excitation and emission slit widths were kept to 5nm. The emission spectra from different samples were obtained, as accumulations of three spectral readings each. All recorded spectra were blank corrected and normalized by setting pyrene emission monomer peak at 397nm as constant. Effect on lipid membrane as percent dilution of lipids were indirectly calculated by measuring the drop in pyrene excimer peak values at \sim 475nm, in presence of E1 WT protein. The percent lipid dilution calculations were performed by considering the difference between the pyrene excimer peak values of pyrene-labelled LUVs in absence of any protein (taken as 0%) or in absence

of any protein plus 0.5% TX-100 (taken as 100%). The effects of presence of E1 WT protein towards percent lipid dilutions were calculated over this range. The experiment was repeated thrice with similar trends observed in each case. Representative spectra and percent lipid dilutions calculated from a single experiment are plotted and shown in the results section.

1.2.14 Fluorescein based liposome content mixing assay

We studied E1-membrane fusion properties using fluorescein based fluorescence assay. In this case, the LUVs prepared were labelled from inside with quenching concentrations (100mM) of carboxyfluorescein (purchased from Thermo Scientific™, USA). The labelled LUVs were separated from free dyes using a PD10 desalting column (purchased from GE Healthcare, USA). Final experiments were performed with freshly prepared fluorescein-loaded LUVs (at 0.5mM working concentration) and bacterially purified recombinant E1 WT protein (at 0.2mg/ml concentration). TN buffers were used for assays either at pH 7.4 or pH 5.5 (pre-calibrated volumes of sodium acetate pH 4.6 were added to bring down the pH to 5.5).

For the assay, first E1 WT protein labelled DGS-LUVs (LUVs were prepared as mentioned above, in this case with an extra component - DGS-NTA at 0.4 molar ratio) were prepared by incubating E1 WT protein at 0.2mg/ml concentration for 30 minutes with 0.5 mM DGS-LUVs at RT. Then the fluorescein-labelled LUVs (loaded from inside) were added to the E1 WT protein labelled DGS-LUVs. Following this, we took fluorescein emission fluorescence recordings by setting the excitation maximum for the experiment to 490nm and emission maximum to 520nm (emission maxima of fluorescein). The excitation and emission slit widths were set to 5nm. The emission spectral readings in this case were

collected as kinetic readings for the fusion assay. All readings with different samples were taken as: 15 minutes at pH 7.4, followed by 15 minutes at pH 5.5 (by addition of pre-calibrated volumes of sodium acetate pH 4.6), and finally 10 minutes in presence of 0.5% TX-100 detergent. The total duration of each reading was kept to ~30 minutes (including the steps where halts were taken for addition of acidic pH buffer and TX-100 detergent). Each individual spectrum represented in figures is accumulations of three spectral readings each and blank corrected. Normalization of effect of acidic pH on fluorescein fluorescence was done in the following manner. The drop-in fluorescein emission fluorescence value at acidic pH was normalized to the level of neutral pH values. Similar normalization was also performed when TX-100 was used, and the fluorescein emission fluorescence values at acidic pH+TX-100 were normalized to the level of fluorescein emission fluorescence values at neutral pH+TX-100. The fluorescein emission fluorescence values for all other samples in presence of E1 WT protein at acidic pH was normalized accordingly. The experiment was repeated thrice with similar trends observed in each of the cases. Representative spectra from one single experiment are plotted as final kinetic spectral readings and shown in the results section. The fluctuations in spectral readings for all the samples were smoothed similarly by using the moving average functions in excel before final representation.

Percent membrane fusion as an indirect measure of increase in fluorescein fluorescence because of its dilution below quenching concentration was calculated in presence of E1 WT protein sample. The average of last few data points from only LUV sample towards the end of pH 5.5 reading was considered as base line of fusion (considered as 0% liposome content dilution). The average of last few data points towards the end of

0.5% TX-100 reading (post-acidic pH treatment) of the same sample (only LUV sample) was considered as end point dilution (considered as 100% liposome content dilution). Effect of LUV anchored E1 WT protein on inducing membrane fusion/liposome content dilution was calculated as percent value (percent liposome content dilution) in the range of above-mentioned values obtained with the control samples (between 0%-100%).

1.2.15 Negative stain electron microscopy

We analysed the liposome samples alone in TN+0.01% SDS buffer and *E. coli* purified E1-ectodomain protein (0.1mg/ml) along with LUVs (at 0.5 mM concentration) in TN+0.01% SDS buffer triggered to pH 5.5. The sample was incubated at RT for 30 minutes before it was applied to glow-discharged continuous carbon-coated grids. Staining was performed using UranylLess EM stain (cat. no. 22409) from electron microscopy sciences. The grids were imaged on TEM microscope at 40K-45K magnifications. Images collected were analysed using EMAN2.1 software.

1.3 RESULTS:

1.3.1 Mammalian cell expression of CHIKV cell entry proteins

We tested expression of full-length forms of CHIKV envelope proteins for cell surface expression in mammalian cells. Our construct included E3, full length E2 and E1 proteins including the trans-membrane regions and also included the native 6K sequence that joined E2 end and E1 beginning coding sequences (Figure 1.1).

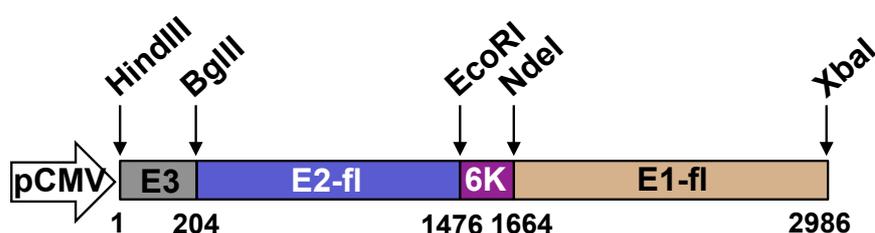


Figure 1.1 Mammalian-cell expression construct of full-length CHIKV envelope proteins. Restriction enzyme sites used at the junction between two protein-coding sequences are shown with down faced black arrows. At the bottom, the base numbering of the expression ORF is shown.

This entire ORF was kept in frame with the cytomegalovirus (CMV) promoter for expression in mammalian cells. We transfected the CHIKV envelope protein expression construct into both HEK293T cells as well as Hela cells for checking expression of these proteins in these mammalian cell lines. We performed expression analyses on both the transfected cell lines using western blotting method as detailed in the materials and methods section. I did not find expression of E3, E2, E1 in the conditions tested, neither in HEK293T nor in Hela cells. Though CHIKV envelope proteins are used earlier for production of VLPs (141), lentiviral packaging of CHIKV envelope proteins that requires expression in mammalian cells was not established thoroughly. In a recent study, CHIKV envelope proteins were used towards packaging into lentiviral system (102), where the authors have

reported use of complex and multiple constructs with IRES sites for successful expression in mammalian cells for packaging of CHIKV envelope proteins into lentiviruses.

1.3.2 Insect cell expression and purification of CHIKV cell entry proteins

We used the Bac to BacTM protocol for insect cell expression and purification of CHIKV envelope entry protein complex ectodomain regions only. Initial expression check following transfection with recombinant bacmids (prepared by transposing the expression construct shown in Figure 1.2a into baculovirus bacmid genome) was done at different time points, and 72 h.p.t was found to be optimum for CHIKV envelope protein complex expression. Sf9 cells grown to $2-3 \times 10^6$ cells/ml in suspension culture with >95% viability was infected initially with different volumes (10/15/20 ml) of P3 viral stock for expression and purification studies. However, we did not observe any significant increase in the level of protein expression at optimal expression conditions when varied volumes of P3 viral stocks were used for infection for expression-purification of secreted out CHIKV envelope ectodomain protein complex in the media fraction. We followed the Bac-to-BacTM manual for generation of p3 stock. Assuming the titer to be $\sim 1 \times 10^8$ pfu/mL as suggested in the manual, we tested the above-mentioned volumes of p3 stock for protein expression trials keeping in mind that more volume of p3 may add more number of defective viral particles and may hamper expression.

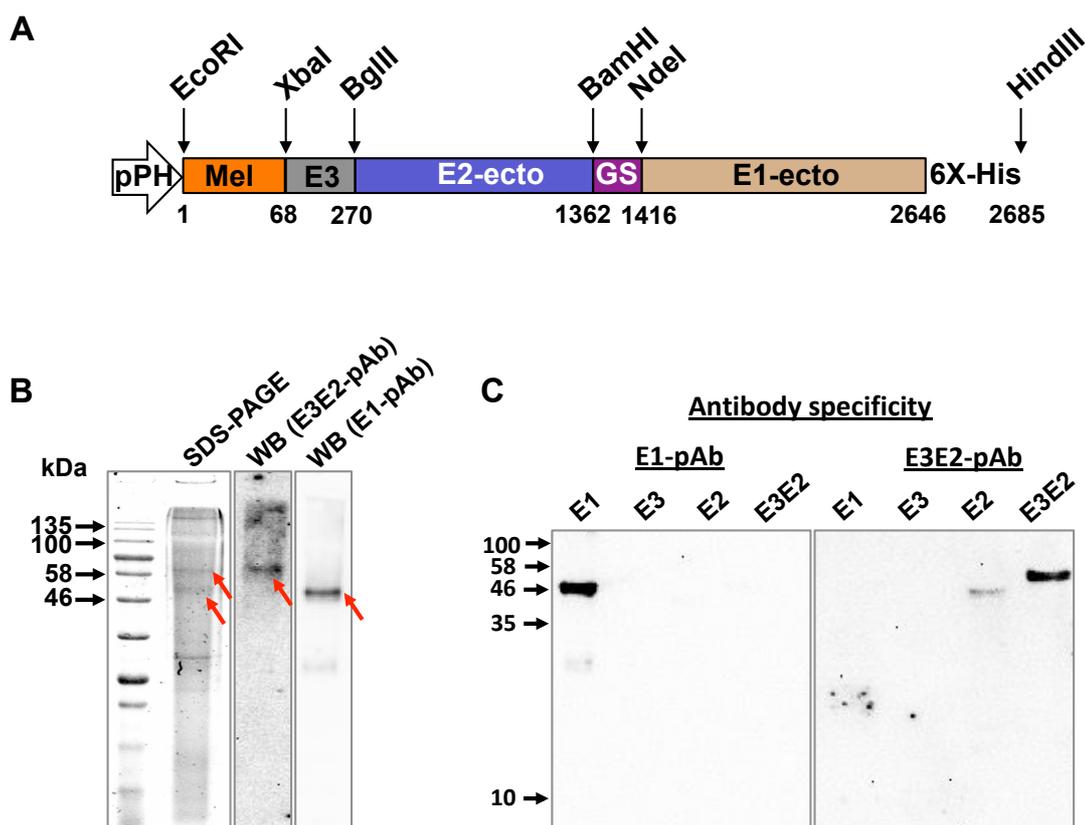


Figure 1.2 Insect cell expression and purification of CHIKV envelope proteins. (A) Baculovirus expression construct. Restriction enzyme sites used at the junction between two protein-coding sequences are shown with down faced black arrows. At the bottom, the base numbering of the expression ORF is shown. (B) SDS-PAGE and WB analysis (both E3-E2 and E1 pAb were used for testing) on insect (sf9) cell purified E3-E2-GS-E1 complex are shown. (C) Specificity of antibodies (developed in this study, refer to materials and methods section) used to detect insect cell purified E3-E2-GS-E1 complex proteins are shown. Bacterially purified individual E3, E2, E3-E2, and E1 ectodomain proteins were used for testing specificity of these antibodies.

The secreted out protein from media were buffer exchanged and concentrated. Purification attempts using Ni-NTA affinity chromatography, followed by ion exchange chromatography (using Mono-Q column), followed by a gel filtration step did not yield the pure fractions of the protein, as desired. SDS PAGE followed by Coomassie brilliant blue staining and western blotting with anti E1 polyclonal antibodies showed that the protein is

expressed, but in small amounts that the protein purity could not be achieved (Figure 1.2b). We used the polyclonal antibodies (both E1-pAb or E3-E2-junction specific pAb), which were developed in this study and have specific activity against E3-E2 and E1 proteins for our WB analysis (Figure 1.2c).

Using the insect cell expression approach, the yield (~50-100ug per litre of culture after final-purification step) and purity (~70% based on visual observation from SDS-PAGE analyses) that we observed are extremely poor for further characterizations.

1.3.3 Bacterial expression and purification of CHIKV cell entry proteins

Using the sequence verified constructs, we optimized over-expression of CHIKV E3, E3-E2-ectodomain, E2-ectodomain, and E1-ectodomain proteins in *E. coli* BL21 (DE3) cells (similar or low level of expression were also observed in Rosetta™ (DE3) and BL21 (DE3) pLysS strains of *E. coli*). Expression profile analyses for each of the proteins expressed as C-terminal 6X-His fusion proteins were performed by SDS-PAGE. Since, proteins were found to be expressing in the insoluble fractions (Figure 1.3), we used urea for solubilization and also for purification under denaturing conditions. After purification proteins were refolded by slow buffer exchange to remove urea. Optimized expression and solubilization conditions for each of the proteins and are summarized in Table 1.1.

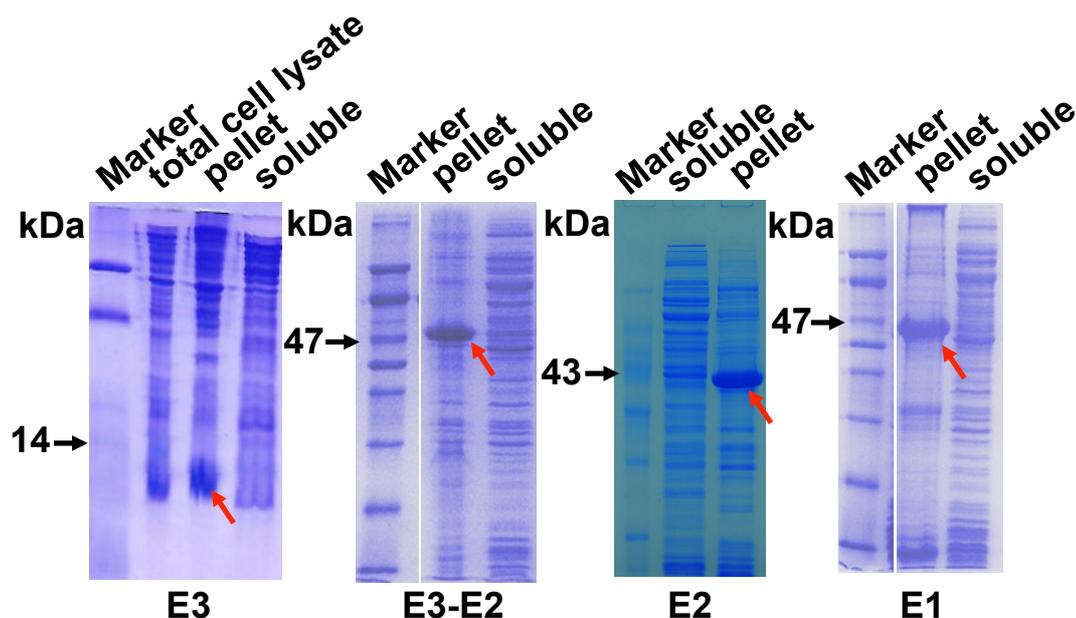


Figure 1.3 Bacterial expressions of CHIKV envelope proteins in *E. coli* BL21 cells. Left to right – expression profiles after lysis of *E. coli* cells expressing E3, E3-E2, E2, and E1 respectively are shown. All four proteins expressing in insoluble fractions (labeled as pellet) are shown. 15% SDS-PAGE gel was run in case of E3, all other samples were run in 12% SDS-PAGE gels. Molecular weights of these proteins are shown in Figure 1.4.

All the ectodomain proteins from sequence verified constructs were affinity purified using the C-terminal 6X-His tag (Figure 1.4a). Briefly, the urea solubilized protein fractions were affinity purified using Ni-NTA resin under denaturing condition. The pure fractions (as assessed by SDS PAGE) from this step were used for refolding. Proteins were refolded using a slow dilution method as detailed in the experimental section. The purity of each of the proteins was assessed by SDS-PAGE, which shows that each of the proteins showed >95% purity after a single step affinity purification process (Figures 1.4b and 1.4c). All the purified proteins except E3-E2-ectodomain protein could be concentrated up to ~1mg/ml and are stable at this concentration at 4°C.

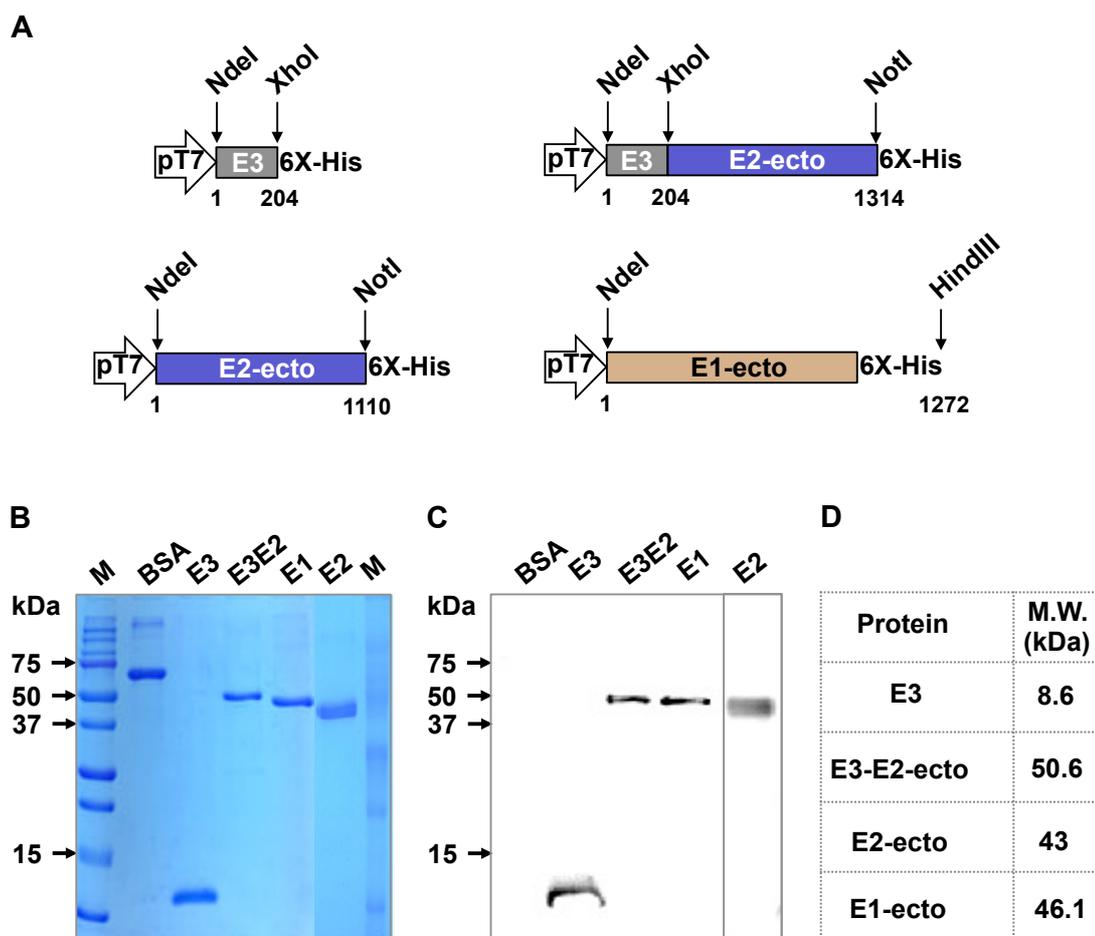


Figure 1.4 Bacterial expression and purification of CHIKV envelope proteins. (A) *E. coli* expression constructs for individual envelope protein ectodomains. Restriction enzyme sites used at the junction between two protein-coding sequences are shown with down faced black arrows. At the bottom, the base numbering of the expression ORF is shown. (B) SDS-PAGE and (C) WB analysis (penta-HIS-HRP antibody was used for WB testing) on *E. coli* purified E3, E3-E2, E2, and E1 ectodomain proteins are shown. (D) Molecular weights of each of the purified proteins as predicted using ProtParam and as observed in our SDS-PAGE and WB analysis are reported.

The E3-E2-ectodomain could be concentrated up to ~ 0.8 mg/ml at maximum. We calculated the theoretical ‘ ϵ ’ (molar extinction coefficient) values for each of the proteins using ExPASy ProtParam program (molar extinction coefficient values for ectodomain proteins of E1 is $52800 \text{ M}^{-1}\text{cm}^{-1}$, for E3-E2 is $58425 \text{ M}^{-1}\text{cm}^{-1}$, for E2 is $53580 \text{ M}^{-1}\text{cm}^{-1}$, and

for E3 is $4845 \text{ M}^{-1}\text{cm}^{-1}$) and used it along with their molecular weight values (Figure 1.4d) for protein concentration measurements using UV-visible absorption spectroscopy. E3 does not contain any tryptophan residues, hence the low molar extinction coefficient value. All the proteins were used freshly for further characterizations after purification.

While the *E. coli* expression and purification conditions are simple and robust, the yield ($\sim 3\text{-}5\text{mg}$ per litre of culture) and purity ($>95\%$ based on visual observation from SDS-PAGE analyses) are better in comparison to insect cell expression and purification of CHIKV E3-E2-GS-E1 complex. To validate that these proteins are usable, we further characterized these proteins for their structure and function, and compared it to insect cell expressed and viral surface cleaved proteins from CHIKV and other alphaviruses.

1.3.4 Structural characterization of *E. coli* purified CHIKV cell entry proteins

I recorded far-UV CD spectra for *E. coli* purified proteins and then analysed the CD spectrum for each of the proteins using both CDPPro as well as K2D3 servers for prediction of percentage of secondary structural content from far-UV CD data (Figures 1.5a-1.5c, and Table 1.2).

All four structural protein forms of CHIKV, as E3, E3-E2-ectodomain, E2-ectodomain, and E1-ectodomain proteins that were purified from *E. coli* expression system were characterized for secondary structure. All these proteins showed similarity in terms of their percent α -helix and β -sheet content predicted from far-UV CD data, when compared to the percentage of α -helix and β -sheet content calculated using the CHIKV E3-E2-E1 crystal structure (PDB: 3N42, this crystal structure was obtained with S2 insect cell line expressed protein complex) data. The detailed values for secondary structure estimation and comparison thereof are listed in Table 1.2.

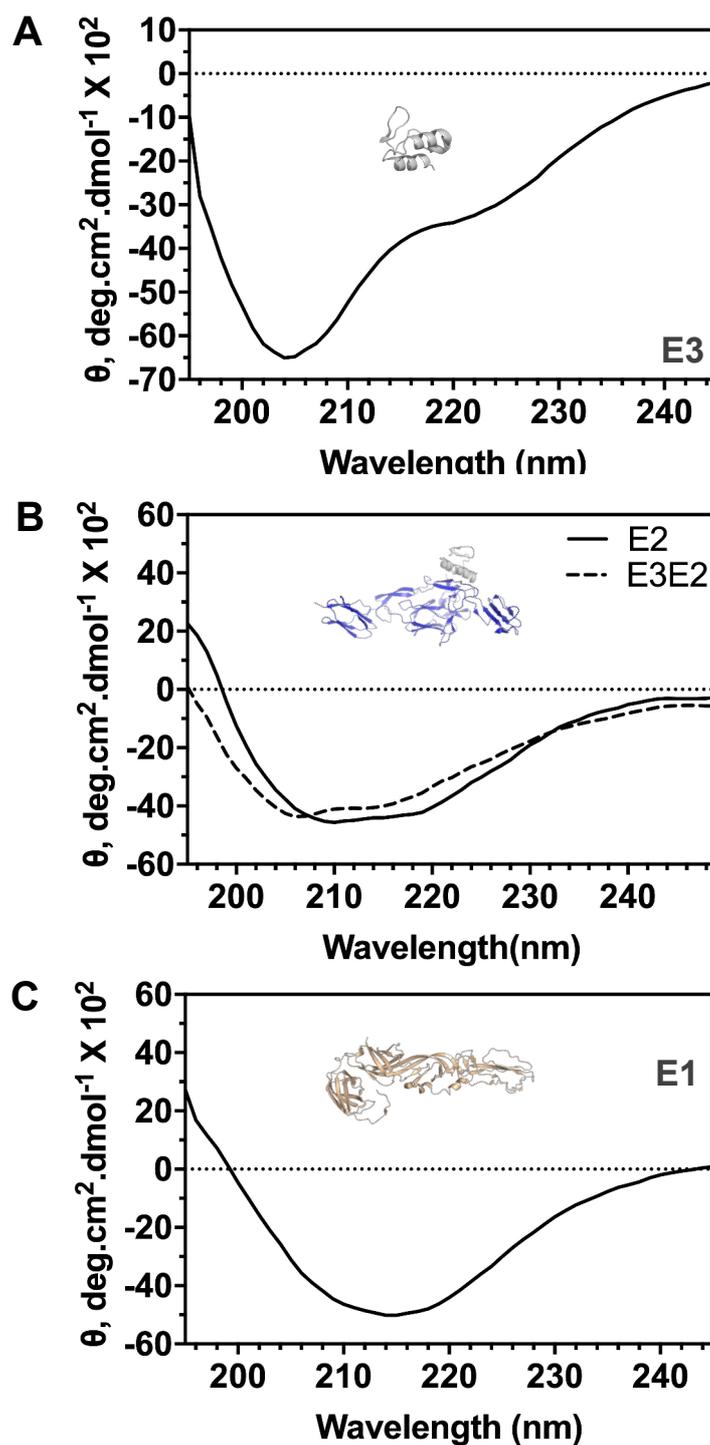


Figure 1.5 Far-UV CD spectra analysis on *E. coli* purified CHIKV envelope ectodomain proteins. Experimentally recorded far-UV CD spectra of CHIKV envelope protein with corresponding structure as inset are shown. ‘ θ ’ is mean residue ellipticity. Far-UV CD spectra of (A) E3 (B) E3-E2, E2 and (C) E1 ectodomain proteins are shown.

Table 1.2: Comparison of percent secondary structural content from far-UV data and crystal structure.

	Secondary structure	E3	E3-E2	E2	E1
Far-UV CD data	α -helix	38	3.09	1.76	6.0
	β -sheet	13.9	36.41	47.42	40.96
PDB: 3N42 structure	α -helix	40	7.07	2.36	4.62
	β -sheet	13.3	37.70	47.9	45.74

As expected from the crystal structure, E3-E2-ectodomain, E2-ectodomain and E1-ectodomain proteins showed CD signature of predominantly β -sheet structure, while E3 protein showed predominantly α -helical structure. This indicated that right amount of secondary structures are present in these *E. coli* purified and refolded CHIKV envelope structural proteins.

Intrinsic fluorescence spectra of E3, E3-E2, E2 and E1 are presented in Figure 1.6. Our results from intrinsic tryptophan fluorescence spectra analyses revealed that these proteins have λ -emission maximum in the range of \sim 333nm-334nm for E3-E2-ectodomain, E2-ectodomain and E1-ectodomain proteins (E3 does not have a tryptophan residue). This suggests that all these proteins have their tryptophan residues, present in buried environments, which is indicative of folded conformations of these proteins. When we recorded intrinsic fluorescence spectra in presence of 8M urea, the λ -emission maximum observed was in the range of \sim 351nm-352nm for these proteins. This large red shift seen

in presence of a denaturant indicates transition from folded to unfolded forms of these proteins (Table 1.3) further validating proper folding of *E. coli* expressed and urea unfolded and refolded proteins.

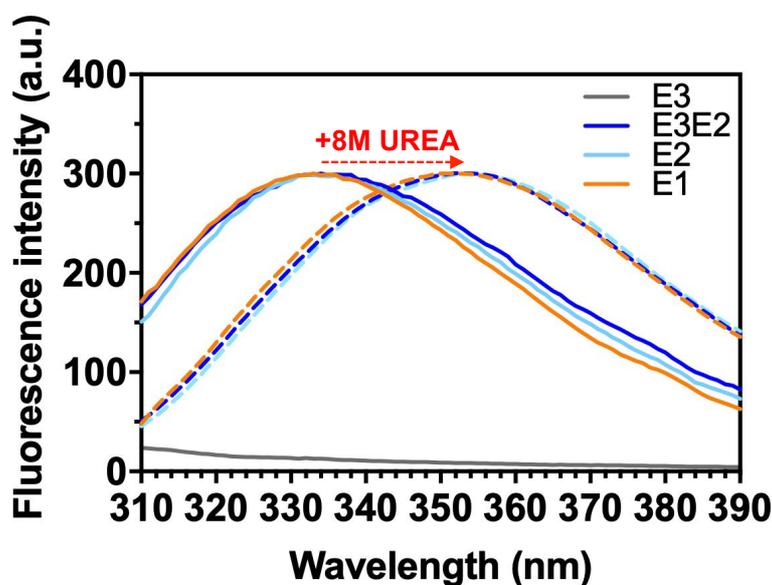


Figure 1.6 Intrinsic Tryptophan fluorescence spectra analysis of *E. coli* purified CHIKV envelope ectodomain proteins. Experimentally recorded fluorescence spectra of folded proteins (colour coded solid lines) and urea-unfolded proteins (colour coded dashed lines) are shown.

Table 1.3: Intrinsic fluorescence spectra analysis on *E. coli* purified CHIKV cell entry proteins.

Intrinsic Trp fluorescence (λ - excitation max = 295 nm)	λ - emission max (nm)	E3	E3-E2	E2	E1
	Refolded	No Trp	334	334	333
	+8M urea	No Trp	352	352	351

All together these results strongly suggest that these proteins, which have the right amount of secondary structural content, have folded properly into their native tertiary conformations that is essential for their function. More results on structural and functional characterizations are explained in respective chapters on structure-function studies of each of these proteins.

1.3.5 Functional characterization of *E. coli* purified CHIKV envelope proteins

We studied CHIKV envelope structural proteins, the receptor binding protein - E2 or E3-E2, in *in vitro* receptor binding assay to show that these *E. coli* purified proteins can bind to their cellular receptor heparan sulfate/heparin. Detailed characterizations on this are presented in the next chapter, chapter 2. For structure function studies or characterization of CHIKV E1-ectodomain protein, we first purified the protein using *E. coli* expression-purification approach. None of the alphavirus E1 proteins in their full ectodomain forms were previously purified using bacterial expression-purification approach. We observed high amounts of purified proteins (~3-5mg/1ltr of culture) from the bacterial expression-purification approach. To make sure that the *E. coli* purified E1 protein possesses similar structural and functional properties as that of its native counterparts, we characterized its structure and function as detailed further.

We did size exclusion chromatography experiment using the *E. coli* purified CHIKV E1 protein for analysing its quaternary structure (Figure 1.7). The gel filtration profile of the E1 protein alone showed a large peak near the void volume (in addition to smaller peaks near regions that would correspond to monomer, dimer, and trimer populations of E1) suggestive of higher order oligomeric forms of the protein. Because of the inherent nature of the protein for oligomerization, it is expected that E1 could form

oligomeric conformations when it is free in solution even at neutral pH, and especially when triggered with low pH. Similar results (smaller and larger oligomeric populations were observed even at neutral pH) were obtained with viral surface cleaved and detergent isolated (0.5mM Triton-X100) SFV E1 protein (135).

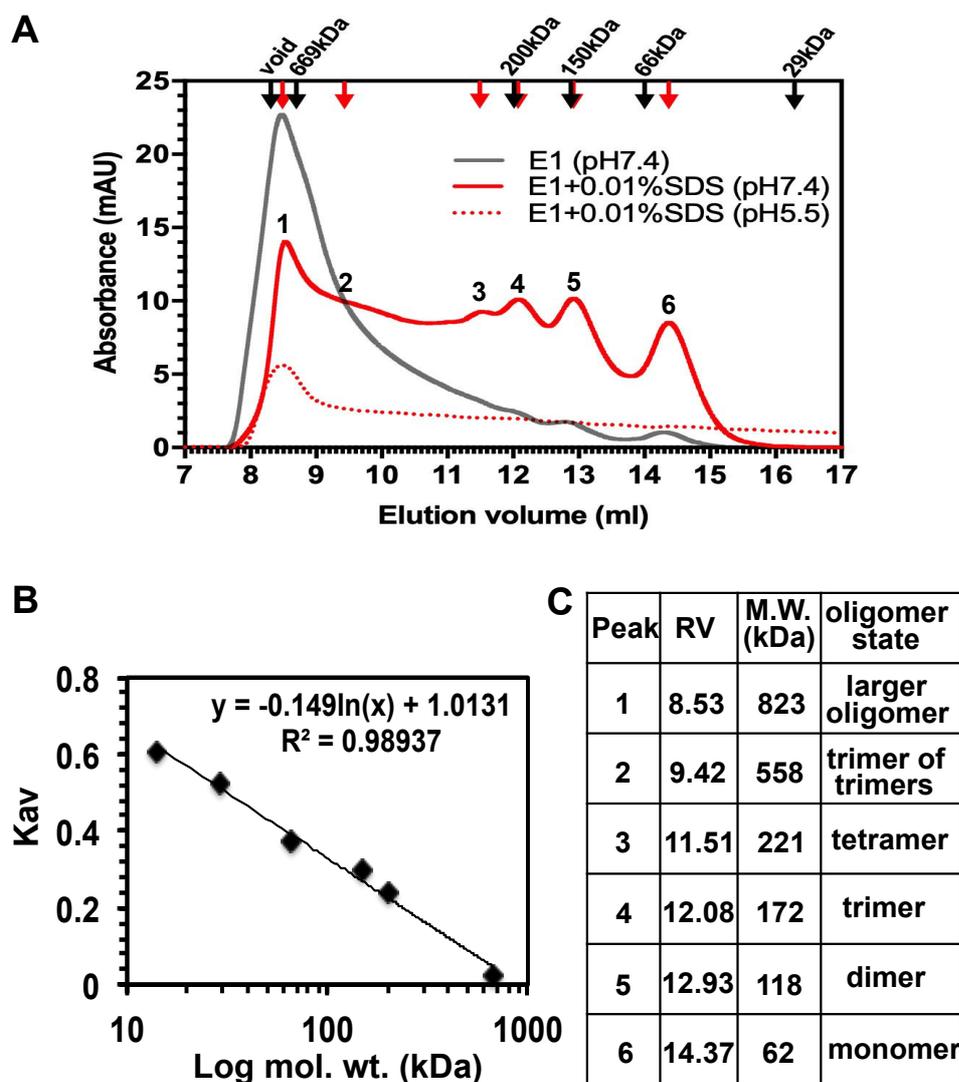


Figure 1.7 Quaternary structure analyses of *E. coli* purified CHIKV E1 WT ectodomain protein. (A) Gel filtration profile. Positions of the gel filtration standards are marked with down faced black arrows above the chromatogram, red arrows mark the position peaks from E1 gel-filtration profile, which are also marked with numbers 1-6. (B) Gel filtration standard curve prepared using known molecular weight markers. (C) Estimated mass of E1 from each peak fraction and respective oligomeric states are mentioned. RV – retention volume, M.W. – molecular weight.

We further checked the gel filtration profile of E1 in presence of 0.01% SDS to see if this mild detergent could affect the larger oligomeric populations of E1 protein. In fact, smaller amounts of detergents are earlier reported to stabilize alphavirus fusion protein E1 conformations in solution (142) without affecting its biological activity (143). We also tested by far-UV CD spectra analysis that 0.01% SDS does not affect the LUVs/SUVs stability and does not alter the secondary structure of the E1 protein in solution (Figure 1.8). In addition, we noticed that addition of this mild amount of detergent (0.01% SDS) in our E1 sample buffer actually increased the stability of the protein at RT.

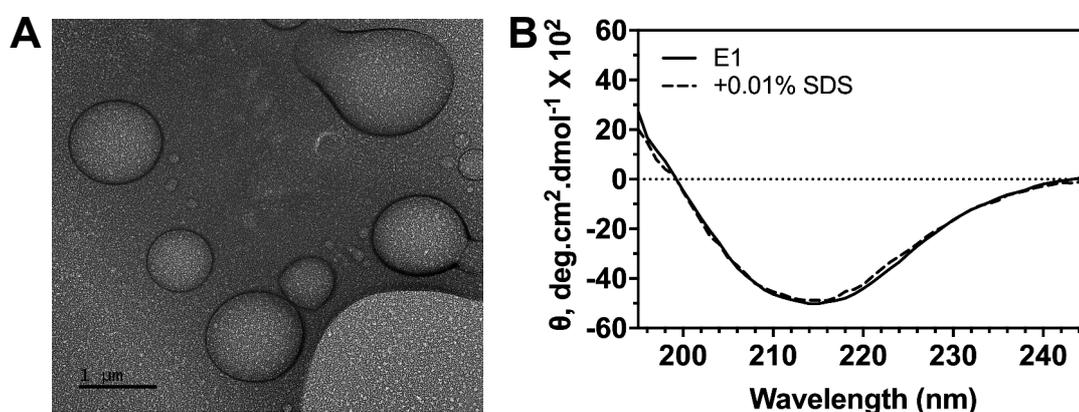


Figure 1.8 Analysis of liposome integrity and E1 secondary structure in presence of SDS. (A) TEM analyses on negatively stained liposome samples in TN plus 0.01% SDS buffer. (B) Far UV-CD spectra of E1 WT protein in presence and absence of 0.01% SDS in TNE buffer is shown.

In any case, from the gel filtration profile of E1 ran at neutral pH in presence of 0.01% SDS sample we calculated molecular masses of different peaks (Figure 1.7a). Comparison of molecular masses of E1 from each of these peaks with a standard curve prepared from known molecular weights of gel filtration standard proteins (Figure 1.7b) ran under similar experimental conditions indicated these peak fractions of E1 to be closely matching to sizes of different E1 oligomeric forms (Figure 1.7c). The peaks corresponding

to monomers, dimers, trimers, tetramers, trimer of trimers and a peak corresponding to higher order oligomeric form of E1WT protein, in mild detergent containing neutral pH buffer, can be seen clearly (Figure 1.7a). Biological significance of dimer and tetramer conformations are not known. Further, predominant higher order oligomeric forms of E1 at neutral and acidic pH conditions were also observed. This property of E1 can possibly be attributed to increased interactions between E1 homotrimers through exposed fusion loops as well as through lateral interactions. Similar oligomeric conformations (144, 145) and interactions among homotrimeric conformations with SFV viral surface E1 protein were reported previously (28, 88).

In presence of 0.01% SDS, the equilibrium of E1 oligomeric conformations shifted more towards smaller oligomeric populations (for example monomers, dimers, and trimers) at neutral pH. We expected this pattern to be more effective at acidic pH since acidic pH triggers formations of E1 homotrimer structures (88), and these structures are reported to interact among themselves to arrange into larger oligomers at acidic pH (28). Hence, we checked the gel filtration profile at acidic pH in presence of 0.01% SDS. However, though we saw a peak corresponding to larger oligomeric populations of E1 (as expected), we did not see more of the smaller oligomeric populations. We observed that while similar amount of proteins (equal to protein amount loaded in presence of 0.01% SDS at neutral pH) were loaded on to the gel-filtration column, lesser amount of protein passed through the column. The remaining proteins (likely larger oligomers) that were retained in the column could be recovered when running buffer was allowed to back flow in the column.

In addition to this, we took an alternative approach to see if lower oligomeric populations can be observed at a higher proportion at acidic pH condition in comparison to

neutral pH as we hypothesized. We did acidic pH triggering followed by glutaraldehyde crosslinking (within a short time frame of 5-15 minutes) at room temperature to test our hypothesis (Figure 1.9).

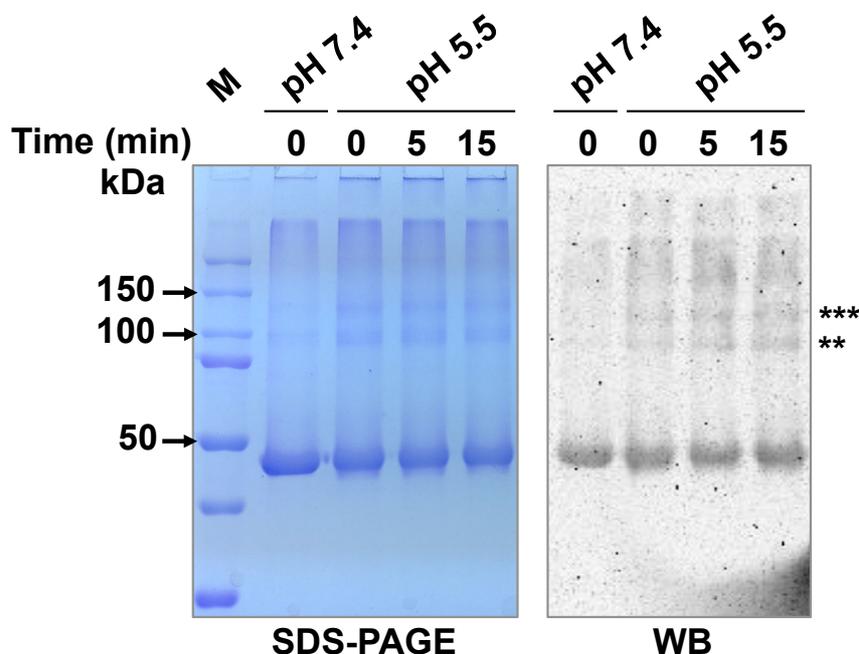


Figure 1.9 Crosslinking of *E. coli* purified CHIKV E1 WT ectodomain protein. Left – SDS-PAGE analysis on glutaraldehyde cross-linked E1 WT protein sample. Right - WB analysis on glutaraldehyde cross-linked E1 WT protein sample. ** and *** represent dimer and trimer populations of E1 in SDS-PAGE gel and in WB image.

For this, we checked formation of E1 oligomer conformations by acid pH (pH 5.5) triggering of the recombinant protein, and then running it on SDS-PAGE (after cross-linking) (Figure 1.9). In fact, like seen with other alphavirus surface proteins, the recombinant E1 protein formed dimer and trimer forms as well. Such conformations formed by several other alphavirus E1 proteins has been reported previously (144, 145).

As part of the functional characterizations, we tested both membrane interaction as well as membrane fusion properties of the purified recombinant E1 protein of CHIKV.

Results from our lipid-floatation assay (Figure 1.10a) showed, ~32-38% of total protein floated along with LUVs, when triggered with pH 5.5 buffer. This association pattern of E1 at acidic pH is ~24% more than what is observed at neutral pH (~4-8%) (Figures 1.10b, 1.10c). The neutral pH association of E1 WT protein with LUVs is probably because of accessibility of the hydrophobic fusion loops in free E1 protein. Such observations have been reported earlier with homologous fusion proteins from flaviviruses (146). However, further pH triggering might induce formation of slightly bent conformation of fusion loops dramatically increasing E1 membrane interaction properties (88).

Similar trends were also observed when we studied E1-LUV interactions using pyrene-labelled LUVs mixed with free E1 proteins in a pyrene-based fluorescence assay (Figure 1.11a). Significant increase in membrane interaction property (~15-17%) at acidic pH was observed compared to that at neutral pH (Figures 1.11b, 1.11c). Similar role of acidic pH, in triggering E1-membrane interactions has been reported with SFV viral surface cleaved E1 protein as well. BSA was used as the non-specific experimental control, and no significant membrane interaction was observed.

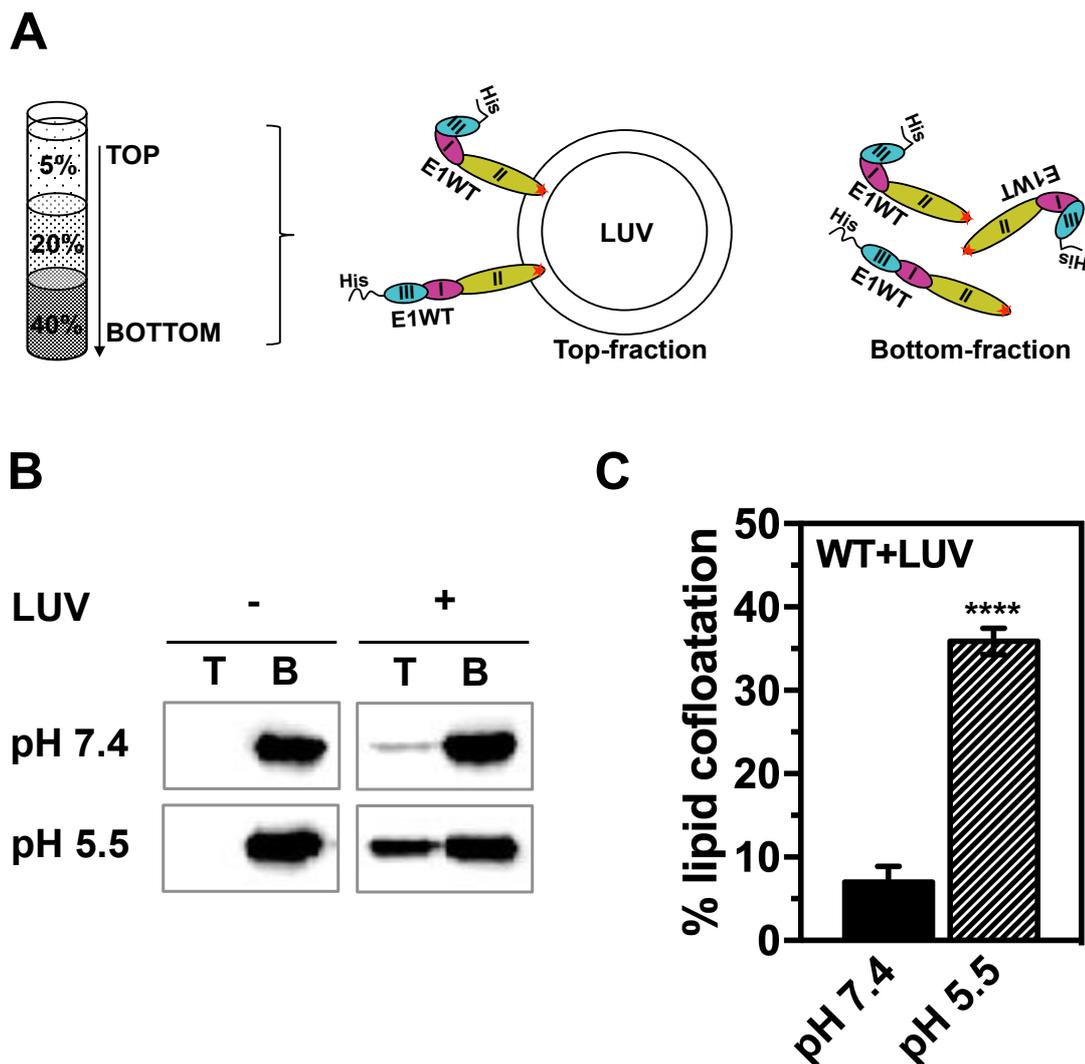


Figure 1.10 Lipid-floatation analysis of *E. coli* purified CHIKV E1 WT ectodomain protein. (A) Schematic representation of the experiment. Sucrose gradient layering is shown on the left. Schematic view of protein samples in top and bottom fractions from the sucrose gradient centrifugation step are shown. E1 WT protein domains (dIII-dI-dII) are colour coded and labelled. (B) WB analysis on top and bottom fractions from the lipid-floatation assay. T is top fraction and B is bottom fraction. (C) Quantitative analysis of E1 WT protein lipid-association pattern. Average values for each sample from three experiments with error bars are shown. One-way ANOVA was used to calculate statistical significance – **** represent $p \leq 0.0001$ level of significance.

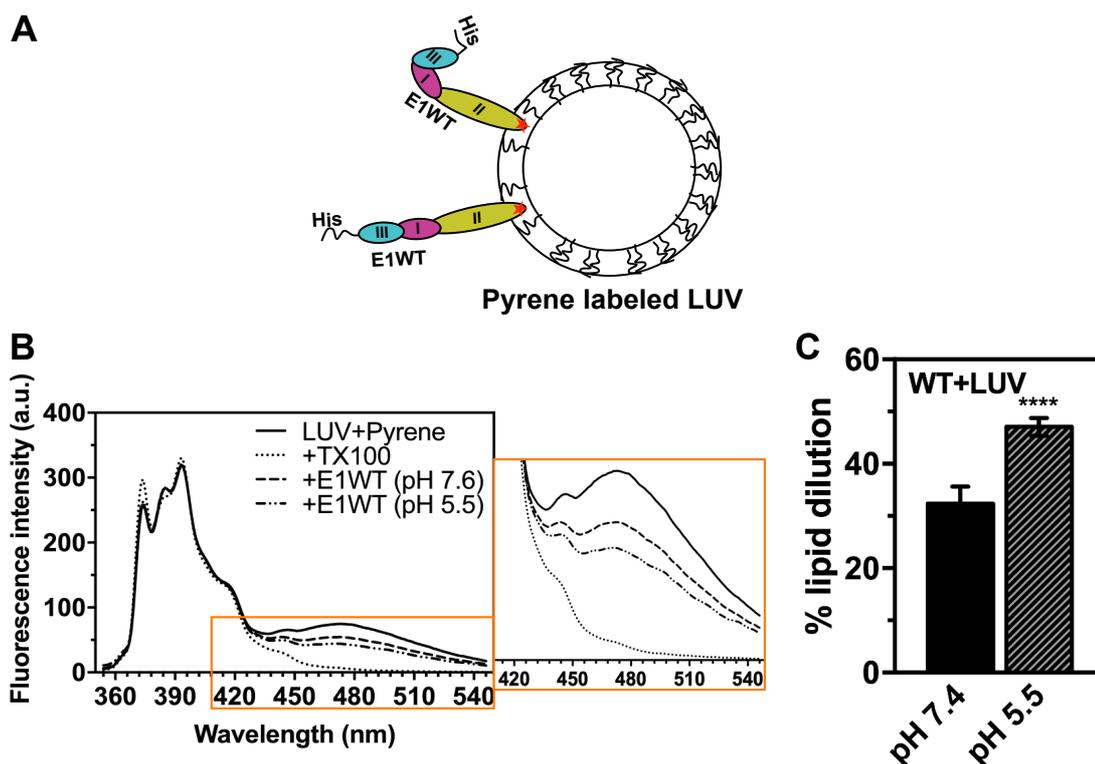


Figure 1.11 Lipid association analysis of *E. coli* purified CHIKV E1 WT ectodomain protein using Pyrene based fluorescence assay. (A) Schematic representation of the experiment. Schematic view of protein samples in membrane inserted (inserted into liposomes) forms is shown. E1 domains (dIII-dI-dII) are colour coded and labelled. Pyrene excimer and monomers in the bilayer of the liposome membrane are shown as double and single bent black lines, respectively. (B) Normalized pyrene fluorescence emission spectra of different samples with respective line format (solid and dashed lines) are shown. (C) Quantitative analysis of E1 WT protein lipid-destabilization pattern. Average values for each sample from three experiments with error bars are shown. One-way ANOVA was used to calculate statistical significance – **** represent $p \leq 0.0001$ level of significance.

In the membrane fusion assay (Figure 1.12a), the recombinant E1 protein, only when loaded onto liposomes and triggered with acidic pH (pH 5.5) showed ~43% fusogenic property as compared to null fusogenic activity observed with free E1 protein alone, tested under similar conditions (Figure 1.12b). This observation is in line with earlier reports with SFV E1 ectodomain region. The ectodomain of SFV E1 when cleaved from the viral surface and tested in different functional assays, it showed membrane association,

homotrimer formation like the full length E1 protein (134) suggesting that the E1 ectodomain could independently have a folded and functional conformation. In addition, it is also reported that due to absence of the transmembrane anchor, E1 does not carry out membrane fusion function (28, 134).

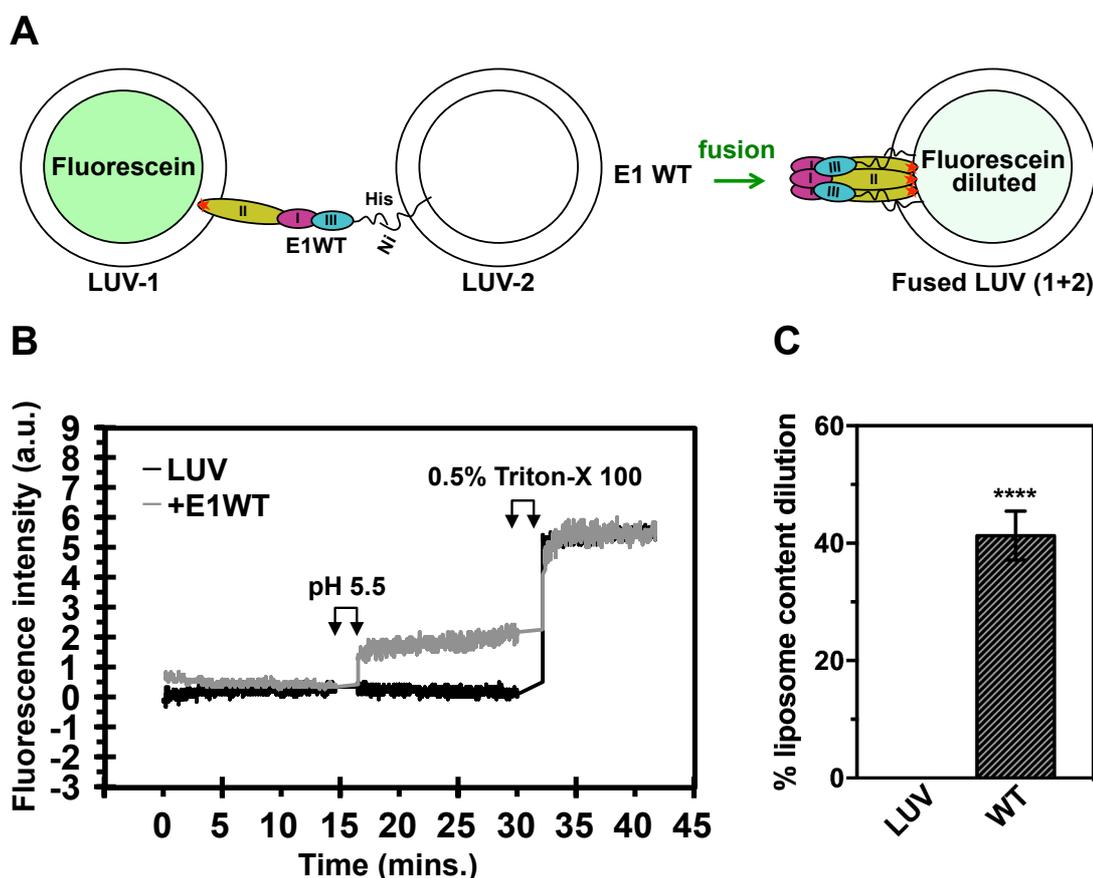


Figure 1.12 Membrane fusion analysis of *E. coli* purified CHIKV E1 WT ectodomain protein using fluorescein based liposome fusion assay. (A) Schematic representation of the experiment. LUV populations – fluorescein loaded LUV is shown with green colored center and Ni-NTA tagged LUV is shown bilayer inserted with black colored lines Ni-NTA tagged lipids with black colored lines. E1 WT protein domains (dIII-dI-dII) are colour coded and labelled. (B) Normalized fluorescein fluorescence emission spectra of different samples from the liposome-liposome fusion kinetics experiment. Down faced arrows represent point of acidic pH trigger and point of TX-100 detergent addition. (C) Quantitative analysis of E1 WT protein membrane fusion activity. Average values for each sample from three experiments with error bars are shown. One-way ANOVA was used to calculate statistical significance – **** represent $p \leq 0.0001$ level of significance.

To add to this, fusion loop interactions with the trans-membrane proximal region of the fusion protein, in the post-fusion conformation, is discussed in earlier literatures to be critical for the fusion active hairpin structure formation. When compared, our results on ~43% fusogenic property of *E. coli* purified CHIKV E1 protein (Figure 1.12c), is at a similar level of fusogenic property (~40-60%) obtained with alphaviruses fusion proteins that were reported previously (133, 147).

We also studied the recombinant CHIKV E1 protein using transmission electron microscopy on a negatively stained *E. coli* purified E1 protein sample (Figure 1.13). Our results show predominant trimeric protein populations and some pentameric and hexameric arrangements (Figure 1.13a-c). From visual observations these structures looked like homotrimers. In fact, studies on fusion between SFV E1 and liposomes have reported similar structural patterns using electron microscopy (28, 88). In these studies different starting protein concentrations led to difference in the formation of multimers of E1 homotrimers (larger multimers with higher protein concentration can be seen). Higher starting protein concentrations lead to formation of homotrimer lateral interactions forming hexameric forms of trimers, which further interact to form larger 2D lattice like arrangements. Also, through fusion loop interactions five-six homotrimers, forming rosette like structures have been observed as well.

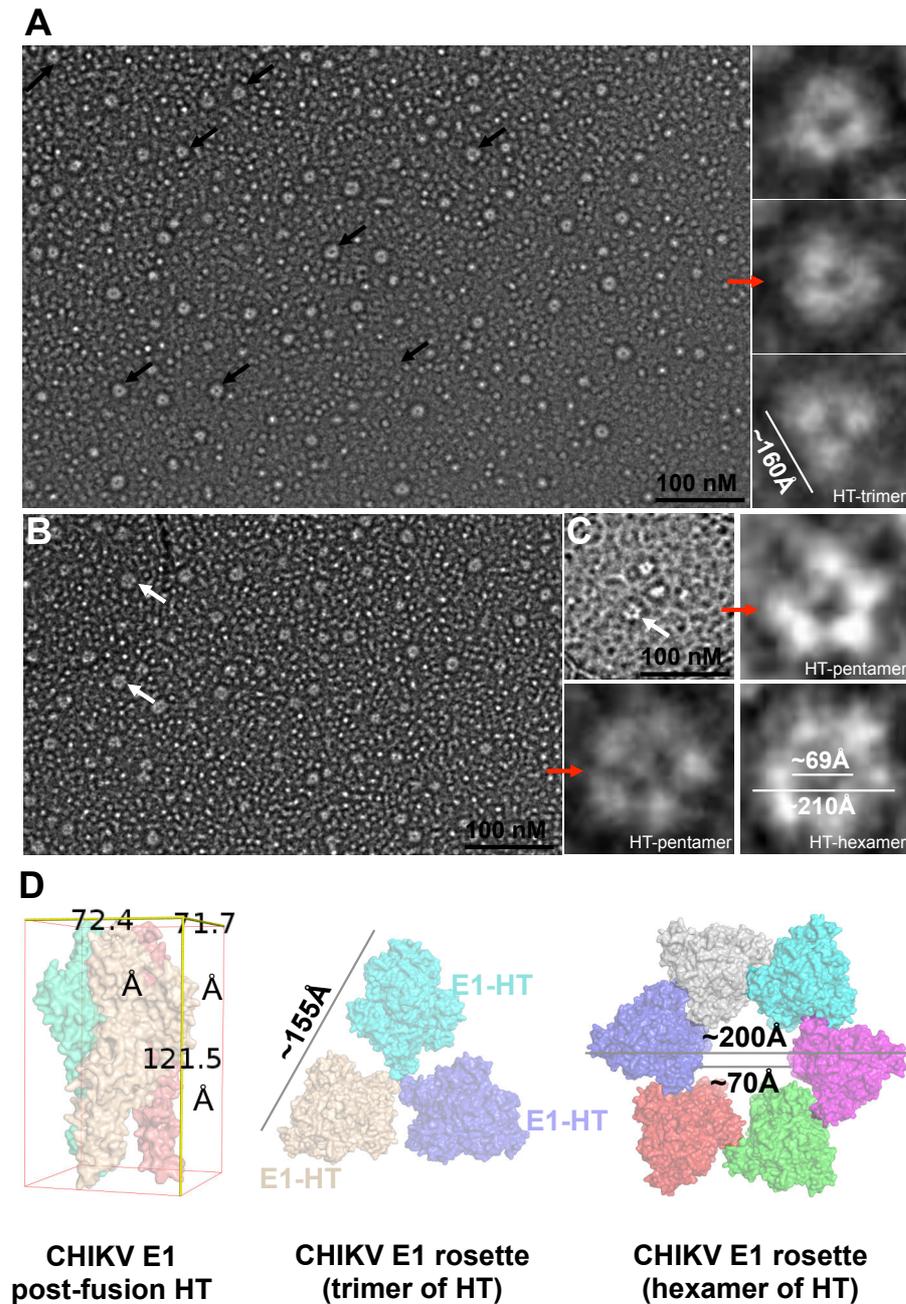


Figure 1.13 TEM analyses on negatively stained *E. coli* purified CHIKV E1 WT ectodomain protein. (A), (B), and (C) represent EM micrographs of E1 protein sample. Scale bar is shown at the bottom right. Oblique black arrows point to the trimeric homotrimer structures, and white arrows point to pentameric or hexameric homotrimer structures. Red arrows point to the zoomed in structures from respective micrographs. (D) Models of E1 trimer structure, trimer of homotrimer, and hexamer of homotrimer structures prepared in PyMol are shown with size calculations. HT is E1-homotrimer.

On the contrary, lower starting protein concentrations have lead to formation of uniform homotrimeric structures and other lower oligomeric forms (dimers) and monomers as well. Similarly, our size estimation on EM micrographs using >50 such trimeric populations reveal $\sim 160\text{\AA}$ length in size of one arm of a triangle, if that triangle covers the entire trimeric structure with three E1-HT present at three vertexes (Figure 1.13b). This led us to believe that these structures are closer to trimer of homotrimeric forms (like rosettes). We used the published EM data available on alphavirus E1 structures (28) for comparison and found that each E1-HT has a size of $\sim 5\text{nm}$ in diameter. Using the available SFV E1-HT crystal structure (PDB: 1RER) (Figure 1.13c) we build the trimer of E1-HT structure using PyMol (Figure 1.13d). Our size calculations from the trimer of homotrimer model corroborated our EM analysis indicating the oligomeric form of CHIKV homotrimer seen in our case to be a rosette of trimer of homotrimers. To further strengthen my analysis, calculations on few of the hexameric E1 homotrimers observed in our EM micrographs showed inner and outer diameters of $\sim 69\text{\AA}$ and $\sim 210\text{\AA}$, respectively. This matched significantly to our size calculations from model hexameric-HT structure (inner and outer diameters of $\sim 70\text{\AA}$ and $\sim 200\text{\AA}$, respectively) and also to earlier proposed reports of $\sim 70\text{\AA}$ inner diameters for these hexameric-HT rosettes (28).

In absence of E2, E1 readily forms the homotrimeric conformation upon triggering by acidic pH. Oligomeric forms of E1 (that includes multimers of E1 homotrimers and in our case predominantly a trimer of homotrimer) also called as rosette structures can be seen under such conditions of acidic pH trigger. These conformations that are reported to be SDS resistance (144) can be seen with free protein as well. Rosette structures are also proposed earlier with other alphavirus E1 proteins for example with SFV E1 (28, 88).

1.3.6 Testing E3-E2 and E1 heterodimer formation with *E. coli* expressed CHIKV cell entry proteins

We have also tested if the *E. coli* expressed E3-E2-ectodomain and E1-ectodomain proteins can form a heterodimeric conformation (the E3-E2-E1 conformation in which they are present on the virion surface) as seen in the crystal structure (PDB: 3N42). Though it is believed that these proteins require a complex processing and maturation step in order to attain the E3-E2-E1 heterodimeric conformation, it has also been shown that these proteins can successfully be processed into their individual forms, as in E3-E2-ectodomain as one protein and E1-ectodomain as another protein (proteolytically cleaved forms from viral surface) and still retain their function (135). We used purified E3-E2-ectodomain and E1-ectodomain proteins and tested heterodimer complex formation through crosslinking approach. However, from our SDS-PAGE analysis it is clear that these proteins upon cross-linking did not show any specific band corresponding to E3-E2-E1 protein population rather monomeric E3-E2-ectodomain and E1-ectodomain populations were observed (Figure 1.14a).

We also tried co-expressions of E3-E2-ectodomain and E1-ectodomain proteins inside *E. coli* from a dual expression vector. However, in this case also we did not observe any significant amount of protein population corresponding to molecular weight of E3-E2-E1-ectodomain region. Our SDS-PAGE analysis of co-expression trials showed expression of both E3-E2-ectodomain and E1-ectodomain proteins in insoluble fractions as opposed to soluble expression of the heterodimer if complex formation occurred inside *E. coli* expression host cells (Figure 1.14b).

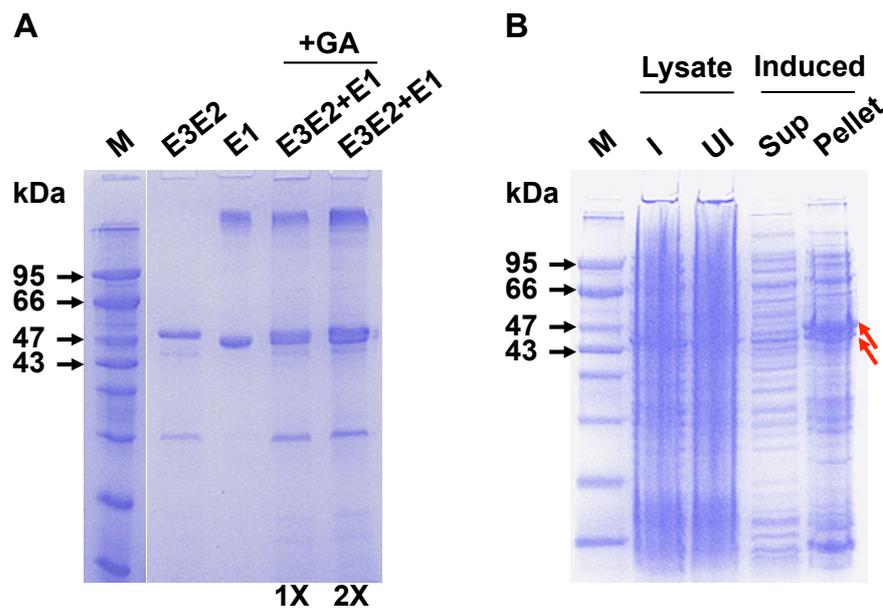


Figure 1.14 Heterodimer formation analysis on *E. coli* expressed and purified CHIKV E3-E2 and E1 ectodomain proteins. (A) Purified protein tested for complex formation by crosslinking with glutaraldehyde (GS). 1X is $\sim 2\mu\text{g}$ of total protein and 2X is $4\mu\text{g}$ of total protein. (B) Dual expression in *E. coli*. I – induced, UI - un-induced, Sup – supernatant, and M - marker.

This proved that E3-E2 and E1 once dissociated do not re-associate to form the dimeric interface either inside cell cytoplasm or in *in vitro* condition. This suggests that upon dissociation from each other these proteins might not form a heterodimer conformation as seen with viral envelope cleaved proteins as well. Further, co-expression may not be enough (may require additional post translational modification; N-glycosylation as reported for CHIKV E3 (at N12 position), E2 (N263) and E1 (at N141) (82), to form the heterodimeric conformation as seen in insect cell expression studies). The transmembrane regions of the protein though are reported to be essential for fusion function (148, 149) of the virus, these regions does not contribute to heterodimer conformation. As seen in case of both CHIKV (82) and SINV (80), both E2-E1 could be expressed and purified as

heterodimers even when the expression construct did not contain coding sequences for the transmembrane region of either of these proteins. Further supporting this, neutral pH crystal structure (PDB: 3J2W) also did not show any interactions between transmembrane regions of E2 and E1.

However, as mentioned earlier these proteins do possess full functionality in their individual forms. In fact dissociation of E3-E2 from E1 is one of the essential steps in triggering activation of the later – the key protein that performs membrane fusion during entry (upon dissociation from E3-E2 followed by acidic pH triggered conformational changes in its structure). Taken together, our results suggest that the *E. coli* purified CHIKV E1 ectodomain protein behaves more or less likely as its native counterpart and can definitely be used for various structure-function studies.

1.4 DISCUSSION:

CHIKV enters into host cells through the endocytic route. Its envelope proteins E3-E2 and E1 mediate cell entry. E3-E2 binding to receptor triggers entry through endocytosis and dissociation from E1, then E1 performing membrane fusion (62). Conformational dynamics in these proteins remain poorly understood because of unavailability of more amount of each of these proteins in their purified and functional forms. We, using *E. coli* expression-purification system, purified these proteins and characterized these for their structure and function. Here, we discuss our results in line with other available data on expression-purification as well as on structure-function studies of these proteins/of similar alpha viral envelope proteins.

Our expression-purification results suggest that these proteins can definitely be over-expressed in *E. coli* cells and these purified proteins can be refolded back to their structurally and functionally relevant forms. This is beneficial over insect cell expression approaches for isolating these proteins for studies that involve understanding their conformational dynamics. So far, the insect cell expression-purification of these proteins was only shown to be successful when performed using drosophila cell-line (80, 82). However, even in S2 cell expression studies there were variations among similar expression trial results. While SINV E3-E2-E1 was produced as trimers of E2-E1 heterodimers (80), CHIKV E3-E2-E1 was produced as E3-E2-E1 heterodimers only (82). In addition, the protein complex obtained in SINV did not dissociate into E2 and E1 individual forms in a pH range of 5.5 to 9.5 (80), even below the pH 6.4 threshold in SINV at which E2 has been shown to dissociate from E1 on virion surface (22). Moreover, such trials were limited to structural studies only and resulted as mentioned previously in functionally un-dissociable forms of these proteins. Also, similar trials in sf9 cells were limited to small-scale expression studies only indicating selective nature of expression hosts and a particular lab environment for purification of these proteins. However, use of bacterial expression system has been used (102, 139) to produce high amounts of relevant proteins (purified and refolded to their functional forms) in comparison to insect cell (sf9 cells) expressed proteins.

Our structural and functional characterizations of *E. coli* purified CHIKV envelope proteins, E3-E2 or E2 or E1 has revealed similarity both in terms of structural and functional properties to their native counterparts. All these proteins have similar secondary structural content as that of an insect cell expressed-purified and crystalized E3-E2-E1

protein complex from CHIKV (82). Our results from quaternary structure analyses on CHIKV envelope E1 protein also match to earlier proposed analyses on viral surface E1 protein from SFV (144). Our thorough analysis on the *E. coli* purified CHIKV E1 protein that was previously not done, shows very much similar structural (as seen from our quaternary structure analysis and TEM analysis) and functional (from lipid-floatation and liposome-liposome fusion assays) properties to many of the earlier studies. Especially the rosette structures that were observed with SFV E1 proteins (both in viral surface cleaved form and when used as part of the whole virion structure) match (size calculations based on EM data) to the rosette structures that we observed using the *E. coli* purified CHIKV E1 protein (28, 88). These studies have used either the whole viruses or viral surface cleaved E1 protein in similar structural and functional characterization methods and reported similar results (28, 88, 133, 134, 144, 145, 150).

Our co-expression analysis of E3-E2-ectodomain and E1-ectodomain proteins both inside and outside of the *E. coli* cell showed no heterodimer conformation formation from functionally individual forms of these proteins. Role additional post translational modification; N-glycosylation as reported for CHIKV E3 (at N12 position), E2 (N263) and E1 (at N141) (82), to form the heterodimeric conformation might be necessary, as seen in insect cell expression studies. As seen in case of both CHIKV (82) and SINV (80), both E2-E1 could be expressed and purified as heterodimers even when the expression construct did not contain coding sequences for the transmembrane region of either of these proteins. Further supporting this, neutral pH crystal structure (PDB: 3J2W) also did not show any interactions between transmembrane regions of E2 and E1. The transmembrane regions of the envelope proteins though are reported to be essential for fusion function (148, 149) of

the virus, these regions does not contribute to heterodimer conformation. Once dissociated from each other, it seems like these proteins have inherent structural features that prevents re-association between them to form the heterodimer conformation (135). However, these proteins have been shown to be functional in their individual forms (28) and it looks like the oligomeric forms are a way of regulating individual proteins function in a specific manner in *in vivo* condition. Dissociated forms of these proteins performing their respective functions are also essential part of their functioning mechanism during entry. So, this study generates useful resources for structure-function studies for understanding different steps during viral-cell entry.

CHAPTER 2

CHARACTERIZATION OF CHIKV E2 - HS INTERACTIONS TO EXPLAIN RECEPTOR BINDING REGULATION OF ENTRY

Conformational changes in Chikungunya virus E2 protein upon heparan sulfate receptor binding explain mechanism of E2-E1 dissociation during viral entry. **Sahoo B** and Chowdary TK, *Bioscience Reports*, **2019**, 39(6), BSR20191077 (**published**)

2.1 INTRODUCTION:

As it is with many other enveloped viruses, cell entry of alphaviruses also requires dissociation of receptor binding protein from membrane fusion protein. E2 and E1 interact tightly through weak interactions - majorly between domain C and β -ribbon connector of E2 and domain II of E1. E3-E2-E1 dissociation requires breaking of intra-heterodimer interactions. In the trimeric spike structure on virion surface, three E2 molecules present at the center of the spike, whereas the three E1 protomers are to the periphery (refer to figure 2.1). Inter E2 protomer interactions hold the trimeric spike. Apart from E3-E2-E1 dissociation, receptor binding may also be triggering breaking of inter E2 protomer interactions in the trimeric spike.

Receptor binding on the trimeric spike (on E2) leads to conformational changes in the E3-E2-E1 proteins (please refer to the General introduction chapter). Based on the epitope changes in E2 and E1 proteins, upon receptor binding, it was proposed that E2 and E1 dissociate from each other. However, the nature of conformational changes and the structural mechanism of E3-E2-E1 dissociation, after receptor binding, are not known.

Also, Voss et al (82) have proposed role of receptor binding in induction of E2-dB movements causing fusion loop exposure on E1 at neutral pH. In another study, Li et al (80) have speculated E2-dC rearrangements as a possible mechanism for E2-E1 dissociation. They discussed that both receptor binding and/or acidic pH could play a role in this process.

Several receptors have been proposed for chikungunya virus. MXRA8, Prohibitin, ATP synthase b-subunit, TIM1 and heparan sulfate have been shown to act as receptors/attachment factors for chikungunya virus (discussed in detail in General

introduction chapter). However, if the virus can use any of these receptors as cognate receptors for entering into multiple cell types has to be determined. Heparan sulfate (HS)/heparin is known to act as a receptor for several arboviruses (104). Multiple reports suggest role of HS in cell entry of alphaviruses. For example SINV, SFV, and EEEV can use heparan sulfate to trigger cell entry (78, 101–103, 105). Multiple reports also suggest HS as an entry receptor of CHIKV (101–103). Biochemical studies on HS binding on alphaviruses envelope proteins, and mapping mutant data that effect HS-binding on to E3-E2-E1 structure (101–103), suggest that E2-dA, E2- β -ribbon connector and E2-dB are putative receptor binding regions on E2. However, the exact type and nature of receptor binding induced conformational rearrangements in receptor binding protein that in turn governs the fusion protein structure and activity remains poorly understood. Going by the literature that many alphaviruses including CHIKV use HS as a receptor for cell entry, and that HS is present on multiple cell types; it would be pertinent to find out the exact binding site of HS on E2 and study the concomitant conformational changes in E2 structure (if any present) upon HS receptor binding.

In following sections, I discuss the results on characterization of E2 and HS receptor, and conformational changes in E2 upon receptor binding. I also discussed, based on the results, a possible mechanism of E2-E1 dissociation in alphaviruses upon receptor binding. Using *in silico* molecular docking, biochemical and molecular dynamics simulation studies, we characterized heparan sulfate (HS) receptor binding site on CHIKV E2 and concomitant conformational changes in E2 upon HS binding. I identified HS binding site on E2 at a structurally conserved positively charged pocket. Also, through published literature survey and sequence analysis we discovered a novel HS-binding motif

(HBD) with XBXXBX pattern (X can be any residue and B is a basic residue) that is present on E2 sequences of almost all alphaviruses. Further, we expressed and purified the functional ectodomain (without the transmembrane region) of CHIKV E2 protein using *E. coli* expression system and confirmed that E2 binds to HS in *in vitro* condition, through specific charge-charge interactions. I also checked the strength of HS binding to CHIKV E2 using microscale thermophoresis assay and found that the binding strength is weak ($K_d=1.8\pm 0.6\mu\text{M}$) but the strength of binding is comparable to the strength of binding by other known heparin/HS-protein interactions. Then I generated charged residue mutants in the predicted HBD (R104A, K107A and R104A+K107A) and studied their HS binding properties, which showed significant level of reduction in binding property to HS, suggesting our prediction of HS binding site on E2 to be true. Further we studied E2 conformational changes upon HS binding using MD simulation approach, and found out that upon HS binding to E2, dramatic tertiary but not secondary structural changes are observed in E2. I then validated the MD simulation predicted model of E2 conformational changes by studying E2 conformational changes upon HS binding using *in vitro* far-UV CD, intrinsic tryptophan fluorescence spectroscopy and smFRET experiments, which corroborated our MD simulation predictions. Based on these data I propose a mechanism, where HS receptor binding on E2-dA induces allosteric E2-dC movements that may push E1 away, breaking E2-E1 interface interactions and may ultimately lead to E2-E1 dissociation.

2.2 MATERIALS AND METHODS:

2.2.1 Molecular docking studies

We used AutoDock Vina (151) to perform docking of heparan sulfate disaccharide (ZINC id. 85551958) on to the trimeric spike structure E2-E1 proteins (PDB: 2XFB). We processed the PDB files prior to docking by adding polar hydrogens, removing water molecules and heteroatoms (if any present in the structure). We used Autogrid program to assign grid dimensions of 0.375 Å along XYZ axes. We then used the default docking protocol to perform a global search of HS on E2-E1 trimeric spike structure. We analysed the docking results, where resultant docking modes of HS disaccharide were listed based on their free energy of binding (least to most). We selected the HS docking pose with lowest ΔG value for further analysis. The docking process was performed using the PyMol GUI version of AutoDock Vina plugin from https://pymolwiki.org/index.php/Autodock_plugin.

We performed all the protein-protein docking as well as docking of a heparin tetra-saccharide on to the CHIKV E2-E1 trimeric spike structure using the ClusPro (152) docking server available at <https://cluspro.bu.edu/login.php>. HS/Heparin possess disaccharides as repeating units and the length of the repeating unit (for example either disaccharide or tetra-saccharide) as a functional unit of receptor binding varies. So we tested docking of both disaccharide and tetra-saccharide binding onto CHIKV E2-E1 trimeric structures. For protein-protein docking, we used the trimeric spike structure (extracted from PDB: 3J2W) as receptor and ATP synthase β -subunit (5FIL), TIM1 (PDB: 2OR8), and Prohibitin (PDB: 1LU7) as ligands. We used the default heparin tetra-saccharide (PDB: 3QMK) as ligand for performing protein-heparin docking using the advanced heparin-docking mode in the ClusPro server. We used the E2-E1 spike

conformation from PDB: 3J2W in ClusPro, which is the trimeric structure of CHIKV E2-E1 proteins obtained from a CHIKV VLP using cryo-EM data. 2XFB.pdb that we used for docking HS disaccharide in AutoDock Vina is 3D reconstruction of CHIKV E1-E2 crystal structure in to SINV cryo-EM map, where density of E3 could not resolved. Using ClusPro, we performed global search of ligand binding sites on the receptor using all default docking parameters. We used this 3J2W.pdb in this case instead of the earlier used 2XFB.pdb, as this was the recent structure available on CHIKV envelope (even though based on a VLP structure) at the time of this study. However, in both these cases the overall organization of E2-E1, as trimeric structures are highly similar. To add to this, the HS-disaccharide binding site on spike from 2XFB was similar (interacting residues from both the spikes were overlapping) to binding site of HS-tetra-saccharide docked onto spike from 3J2W. We chose the best docking pose from the top ranked cluster, sort-listed based on the balance mode of analysis. The balanced energy terms in ClusPro documentation are described with weight co-efficient values as $E = 0.40E_{rep} + -0.40E_{attr} + 600E_{elec} + 1EDARS$, in the equation $E = w_1E_{rep} + w_2E_{attr} + w_3E_{elec} + w_4EDARS$, where E_{rep} denote the repulsive and E_{attr} denote the attractive contributions to van der Waals interaction energy. E_{elec} is an electrostatic energy term and EDARS is a pairwise structure-based potential constructed by the decoys as the reference state (DARS) approach. It primarily represents contributions from desolvation i.e., change in free energy due to removal of water molecules from the interaction interface. The w_1 - w_4 coefficients define the weights of the corresponding terms.

2.2.2 Search for HBDs on E2 sequence of alphaviruses

Our final HBD prediction on alphaviruses E2 sequences was performed using the

MEME (Multiple Em for Motif Elicitation) server (153) available at <https://meme-suite.org/tools/meme>. We began our analysis by compiling all the available alphaviruses E2 sequences from UniProtKB site available at <https://www.uniprot.org/>. All together we identified 16 full length alphaviruses E2 sequences and used these for finding out the presence of any canonical HS/heparin binding sites or HBDs on ScanProsite server (<https://prosite.expasy.org/scanprosite/>). Our analysis did not show any significant/relevant matching pattern for canonical HBDs on alphaviruses E2 sequences. We reasoned that HS/heparin binding on E2 could be through a noval binding motif. Prompted by this idea, we compiled all the biochemical data where mutations in E2 sequences from different alphaviruses were shown to affect HS binding. We selected 11 amino acids sequence stretches from alphaviruses E2 sequences, where the centric residue is responsible for affecting (either increase or decrease) binding to HS. These sequences, along with linear amino acid sequence from HS docking site on E2, identified from our docking studies, is then used for finding a consensus motif in MEME server. A consensus motif with the pattern, **XXBXXBXX** (X is any residue, B is basic residue) emerged out of this motif search. However this is not a canonical HBD motif. Hence, we searched **XXBXXBXX** motif in other alphavirus E2 sequences. Altogether, we analysed 110 input-sequence stretches from 16 alphaviruses E2 sequences through the MEME server. We used the parameters such as site distribution - one occurrence per sequence, number of motifs to find = 1, motif width = 6 to 10 residues, background model = 0-order for finding the presence of a consensus motif. The final motif generated, which has 4 similar matching regions on CHIKV E2 sequence is described in results section.

2.2.3 Cloning, expression and purification of CHIKV E2, E3-E2 and E3-E2 mutant proteins

All the mutant proteins used for this part of the study (E3-E2 HBD site mutants: R104A, K107A, R104A+K107A mutant proteins as well as the S154C+S296C mutant used for FRET study) are expressed and purified exactly in the same manner as that of the WT protein (detailed in chapter 1).

2.2.4 HS/HEP binding assays

We used *in vitro* heparin affinity binding assay to characterize CHIKV E2 or E3-E2 binding to heparin/HS. Briefly, a gravitational flow column was packed with ~ 500 µl beads/ ~1000 µl of 50% slurry of heparin sepharose resin (HiTrap heparin HP, GE Healthcare Life Sciences, USA) and recombinant proteins (~ 10 µg) were then applied on to the column in TNE buffer containing 150 mM NaCl. The loaded protein was incubated for 15 minutes at RT and then the unbound fraction was collected. After an intermediate wash step with TNE buffer (5X of loading volume) the bound protein was eluted with increasing NaCl concentrations (1X loading volumes of 0.3 M, 0.6 M and 1 M NaCl elutions each). Equal volume aliquots from each of the load/input, unbound and elution fractions were analysed by SDS-PAGE. A non-specific binding control (BSA) was also included in the experiment and specific binding (through charge-charge interactions) by E2 was calculated by considering the amount of protein eluted with 0.3 M or higher NaCl concentrations. We also performed a competition-binding assay by pre-incubating E2 with soluble HS (heparan sulfate sodium salt from bovine kidney purchased from Sigma-Aldrich, USA) before loading on to the heparin column, to further prove the specificity of CHIKV E2-HS binding.

We also tested E3-E2 WT and R104A, K107A, R104A+K107A mutants as well as the S154C+S296C mutant in heparin affinity binding to estimate their extent of heparin binding property. For this purpose, we performed the heparin binding experiment as detailed above expect that elution was performed in a single fraction with 1M NaCl containing TNE buffer. The extent of protein binding to heparin (fraction bound) was calculated as percent of protein loaded (taken as/normalized to 100%), by quantifying band intensities using imageJ software. Three independent experiments were performed and Mean \pm SD values form these experiments were represented in a bar graph. The comparison between different protein groups for statistical significance of differences was calculated using one-way ANOVA analysis in GraphPad.

2.2.5 Microscale thermophoresis assay

We analysed the strength of CHIKV E3-E2 protein binding to HS using MST experiment. For the experiment, purified recombinant E3-E2 protein was labelled with NT-547, Amine reactive green fluorescent dye, which was supplied as part of the MonolithTM NT Protein GREEN-NHS labelling kit. The labelling was performed for 30 minutes at RT in dark followed by separation of free dyes from the labelled protein using a PD-10 desalting column. While performing the experiment, E3-E2 at a concentration of 0.2 mg/ml or \sim 5 μ M was kept constant and the concentration of interaction the partner (non-labelled HS) was varied from 0.03 μ M to 15 μ M. The complex mixtures were incubated for 30 minutes at RT. Then the individual sample mixture (with different E3-E2 to HS ratio) was loaded in to individual MST standard coated capillaries and reading was taken at 40% MST power using MonolithTM NT.115 instrument. The default MO. Affinity Analysis (version 2.2.7; Nano Temper) software was used for analysis of binding affinities. Final curve fitting

was performed with data points (each data point represents mean value from three independent readings) using GraphPad prism software.

The basic principle of the MST assay states that during an MST experiment, recording of sample fluorescence is performed at RT for a 3 second starting period to monitor fluorescence at steady-state followed by activation of the IR-laser for a short time period (MST-on time). Upon IR-laser activation, the change in fluorescence is calculated by taking ratio between fluorescence values obtained after MST-on time (F_1) versus fluorescence value obtained prior to IR laser activation (F_0). The following formula represents the measured fluorescence:

$$F_{norm} = \frac{F_1}{F_0}$$

F_{norm} = normalized fluorescence, F_0 and F_1 = fluorescence values before and after IR laser activation, respectively.

The final K_d (binding dissociation constant) value is calculated by plotting F_{norm} versus ligand concentration followed by a dose-response curve data fitting process. The K_d value is calculated using the following formula:

$$K_d = \frac{[A] \times [L]}{[AL]}$$

$[A]$ = Free fluorescent molecule concentration, $[L]$ = free ligand concentration, and $[AL]$ = A and L complex concentration.

2.2.6 Molecular Dynamics simulations of E2 in presence or absence of HS

We performed MD simulation experiment to study HS binding induced E2 conformational changes. The experiment was done using a workstation with 12 cores Intel

Xeon processor. We used GROMOS96 (43a1) force field as part of the GROMACS 5.1.4 9 package (154) for carrying out energy minimization, simulations and trajectory analyses on HS-docked E2 complex as well as the E2 (without HS) structure. The HS ligand parameterization was performed using the PRODRG server available at <http://davapc1.bioch.dundee.ac.uk/cgi-bin/prodrgr/submit.html>. We used explicit SPC water molecules for simulations. Briefly, using periodic boundary conditions within a cubic box solute molecules were added around the water molecules, and counter ions were incorporated for system neutralization. Energy minimization was performed using steepest descent algorithm and LINCS algorithm was applied to constrain the bonds. We set solvent density of 1 atm at 300K and used coupling time of 0.1 ps for independent coupling of solute and solvents to a modified Berendsen bath at 300K. Cut-offs of 1.4 nm for Coulomb and 1.4 nm for van der Waals was used as part of the PME method for calculations of electrostatic interactions. During MD the integration time steps were set to 2 fs, and 10 ps intervals were used for updating output coordinates. The system equilibration was performed twice prior to production MD, 100 ps under NVT and 100 ps under NPT. Final MD run was performed at 300K for a time scale of 10 ns.

PyMol was used for calculation of change in center of mass of domain structures as well as for preparing all other structure figures.

2.2.7 Far-UV CD, and intrinsic tryptophan fluorescence spectroscopy

We used far-UV CD and intrinsic tryptophan fluorescence spectroscopy (procedure detailed in chapter 1) to study HS-protein interactions and conformational changes. Both Far-UV CD and intrinsic tryptophan fluorescence spectra of E3-E2 or E2 proteins were recorded in presence or absence of HS. Prior to spectral recordings the sample mixtures

(1:50 molar ratio of protein to HS was used) were incubated for 30 minutes at RT. The concentration of protein used was 0.2 mg/ml and the experiment was performed in TNE buffer at neutral pH.

2.2.8 Fluorescence resonance energy transfer (FRET) to monitor domain movements in E3-E2 upon HS binding

For studying E3-E2 conformational changes upon HS binding the E3-E2 double cysteine mutant (S154C+S296C) was labeled with a FRET fluorophore pair. We first performed Ellman's assay and tested the availability of free cysteines before labeling with FRET fluorophores. We chose Alexa Fluor™ 488 and Alexa Fluor™ 594 C5 Maleimide (Thermo Fisher Scientific, USA) as FRET pairs for labeling the double cysteine E3-E2 mutant protein through cysteine thiol-reactive maleimide chemistry using manufacturers protocol. FRET fluorophores at a 20 molar excess concentration were used for labeling the protein by incubating the mixture in dark at RT for 2 hrs. We then purified the labeled protein by separating it from free dyes using a PD-10 desalting column (GE Healthcare Life Sciences, USA).

For FRET study we set the excitation at 493 nm (λ_{max} abs. for Alexa Fluor™ 488) and emission spectra were recorded for a wavelength range of 500-700 nm. E3-E2 protein was incubated with HS at 50 molar excess concentrations for 30 minutes at RT prior to taking spectral recordings, which were performed both in presence and absence of HS. Excitation and emission slit width settings were fixed to 5 nm.

2.3 RESULTS:

2.3.1 Heparan sulfate binds to a positively charged pocket on CHIKV E2 at the interface between domain A and arch 1 of β -ribbon connector

To predict HS binding site(s) on CHIKV E2-E1 trimeric spike surface, we performed molecular docking of heparan sulfate disaccharide on to the E2-E1 trimeric spike structure. We assessed the best mode of HS binding by considering the docking pose with lowest Gibbs free energy score ($\Delta G = -7.6$ kcal/mol) and selected this docking pose for analysing the binding site interactions of HS on E2-E1 trimer structure. Detailed crystal structure annotations of CHIKV E2-E1 heterodimer and the best-docked pose of HS on E2-E1 trimeric structure are represented in Figure 2.1a and Figure 2.1b, respectively. All the docked poses of HS are shown in figure 2.1c, which shows two major clusters of HS-docked sites on E2 (both sites are similar and on E2-dA region), and another small cluster of HS docked site at E2-E1 dimer interface, which may not be accessible in actual scenario (figure 2.1c).

HS bound at the apex of the spike on E2 domain A, towards the interior of the groove formed at the interface between dA structures from adjacent E2 protomers in the spike. The HS binding site on E2-dA is a positively charged pocket at a cleft formed between E2-dA and arch 1 of β -ribbon connector (Figure 2.1b, right top) with Lys and Arg residues lining the binding site, as expected of a HS binding site on a protein (100). HS binding at this pocket is through charge interactions between side chains of R104 and K107 with sulfates of HS (Figure 2.1b, right lower).

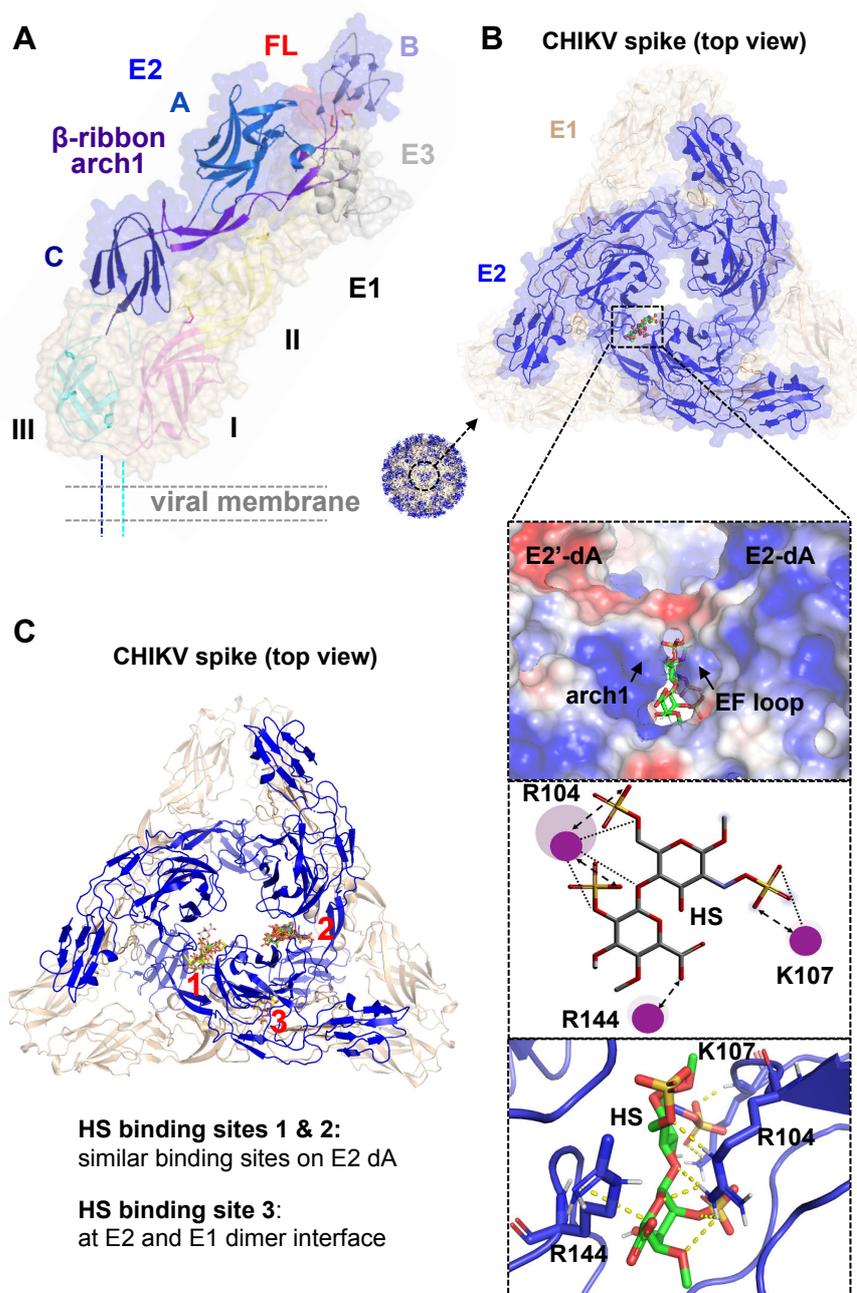


Figure 2.1 Docking of HS onto CHIKV E2-E1 trimeric spike using AutoDock Vina. (A) Domain organization of E2-E1 heterodimer structure. (B) Top: HS (shown as ball and sticks) docked trimeric spike surface. Bottom: top panel shows the electrostatic potential surface at HS binding site. Middle and bottom panels show E2-HS interactions. Hydrogen bonds as and electrostatic interactions as <---> in 2D plot generated using BIOVIA Discovery studio, and polar interactions as yellow dashed lines as generated using PyMol are shown in middle and bottom panels (both HS and HS interacting E2 residues are shown in sticks), respectively. (C) All docked poses of HS on E2, in the E2-E1 trimeric spike structure are shown. Spike is shown in cartoon representation and HS in sticks.

The two HS interacting residues, R104 and K107, are part of a linear stretch of amino acids and are part of the EF loop that connects E and F β -strands in the E2-dA β -barrel like structure. The third residue, R144, which makes backbone interactions with HS, is from the arch1 of β -ribbon connector (Figure 2.1b). While this thesis was in preparation, a new study by Chen et al (155) using cryo-EM reconstruction on HS bound eastern equine encephalitis virus (EEEV), also identified HS binding pockets that match to the HS binding site on CHIKV E2 protein that we identified in this study.

Further, we compared our HS-binding motif on CHIKV E2, which we predicted from docking analysis, with consensus HBDs reported for other proteins that possess and interact through these HBDs, with HS/heparin molecules. Interestingly, the docking site HBD sequence (AR₁₀₄CPK₁₀₇G) on CHIKV E2, did not match to any of the known HBDs (Figure 2.2a). Then based on available published literature (from CHIKV and other alphaviruses) we mapped mutations on CHIKV E2 that affect (either increase or decrease) HS binding. From our analysis, most of these reported mutations congregated to a region close to our HS binding site predictions on CHIKV E2. Based on this observation we asked a question, does alphavirus E2 have a minimal and unique HS-binding motif? To answer this, we performed motif search combining knowledge on mutations affecting HS-binding (in alphaviruses) and also from results of our docking studies (refer to material and methods) (Figure 2.2b and 2.2c). A consensus minimal HS binding sequence emerged as HBD pattern in E2 sequences of alphaviruses (Figure 2.2). Based on this we infer that XBXXBX motif sequence pattern is a novel HBD in alphaviruses (Figure 2.2c). We also observed presence of sequence patterns matching to our motif sequence pattern on almost all the alphaviruses E2 sequences that we included in our analysis (Figure 2.2b). The

conservation patterns from structural alignment of the EF-loop region that contains the docking site HBD is shown in figure 2.3.

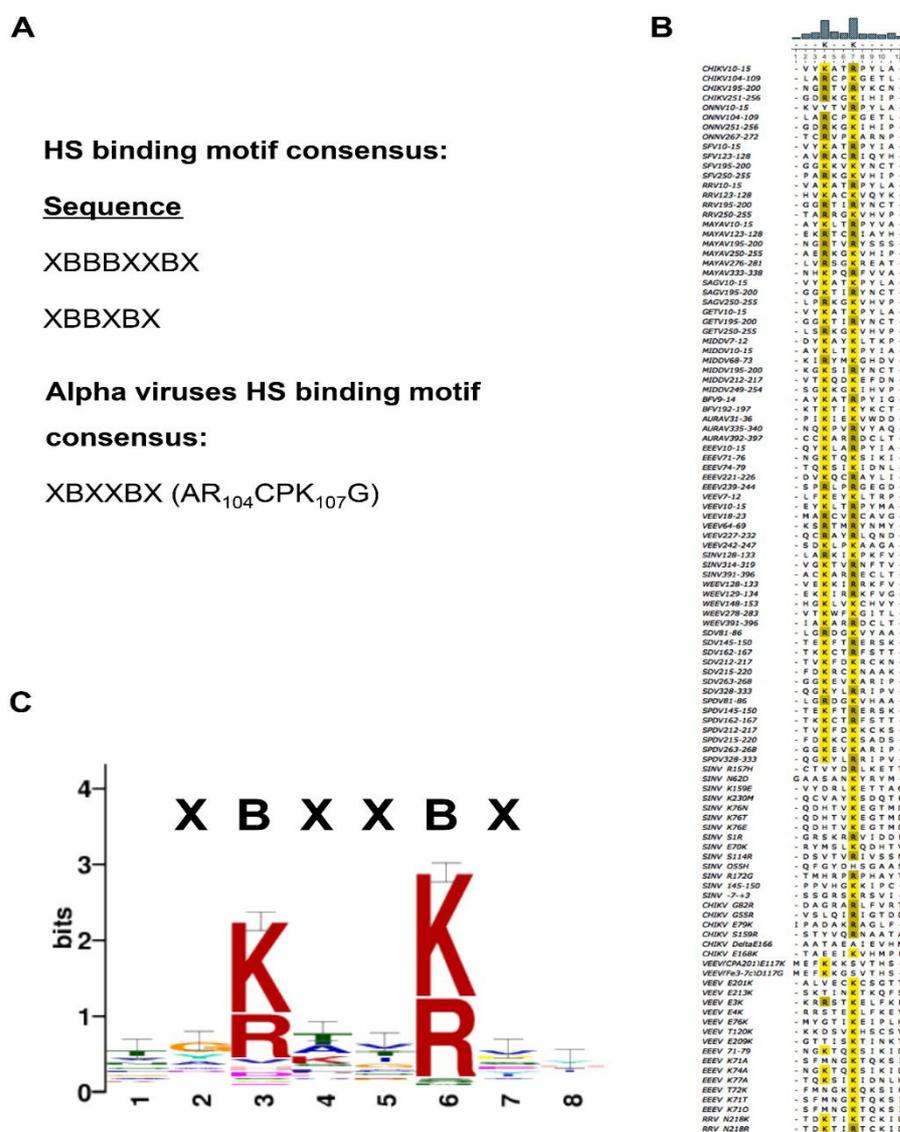


Figure 2.2 Search for Alphavirus E2-HS binding consensus motif using MEME. (A) Published HBD consensus motif on HS/HEP binding proteins. (B) Local alignment of HS binding linear sequences in E2 protein of alphaviruses (full forms of viruses are detailed in abbreviation section). (C) Consensus alphavirus E2-HS binding motif logo with XBBBXXBX pattern. Level of positional conservation is represented as bit score on y-axis.

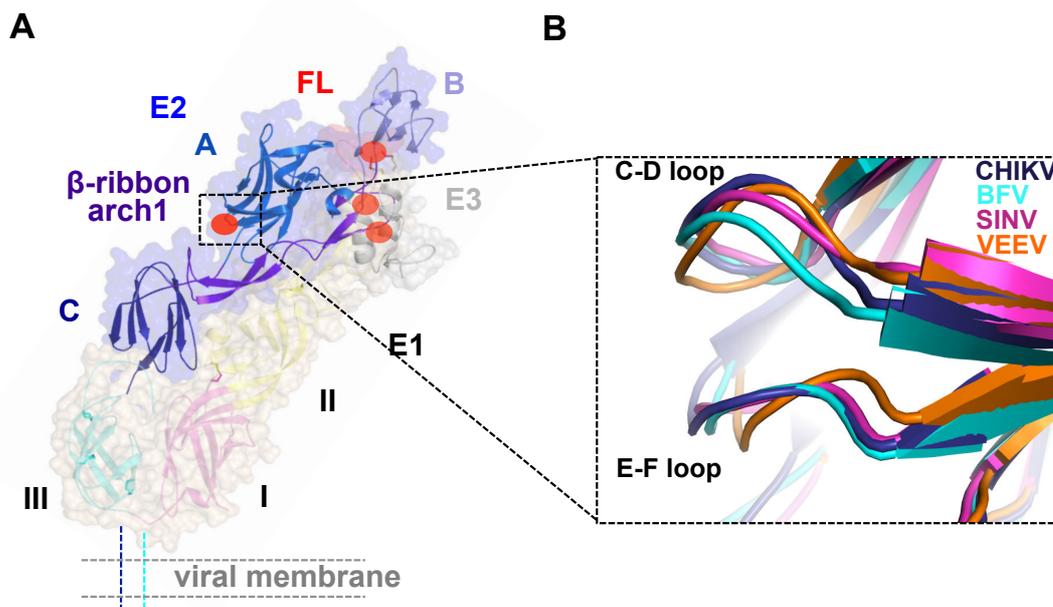


Figure 2.3. Structural analysis of HS-binding motif (HBDs) regions on E2 structure from CHIKV. (A) Four red-ovals represent four similar HBDs (with XBXXBX pattern) on E2 structure. The HBD highlighted by a black dashed box, is the docking site HBD. (B) Structural conservation of the loop regions (EF-loop) on E2 structure of alphaviruses, that makes contact (interactions through positively charged residues) with bound HS/HEP molecule.

We also performed docking and analysed the binding sites of other protein receptors as well as a heparin tetra-saccharide receptor on CHIKV E2-E1 trimeric spike structure. The binding sites of Prohibitin, TIM1, ATP synthase β -subunit as well as the heparin tetra-saccharide (R104, K107 side chains of E2 protein interact with sulfates from heparin tetra-saccharide) all mapped to the same region on E2-dA and β -ribbon connector (Figure 2.4) overlapping with predicted HS-binding site. This is suggestive of presence of a consensus receptor binding pocket on CHIKV E2.

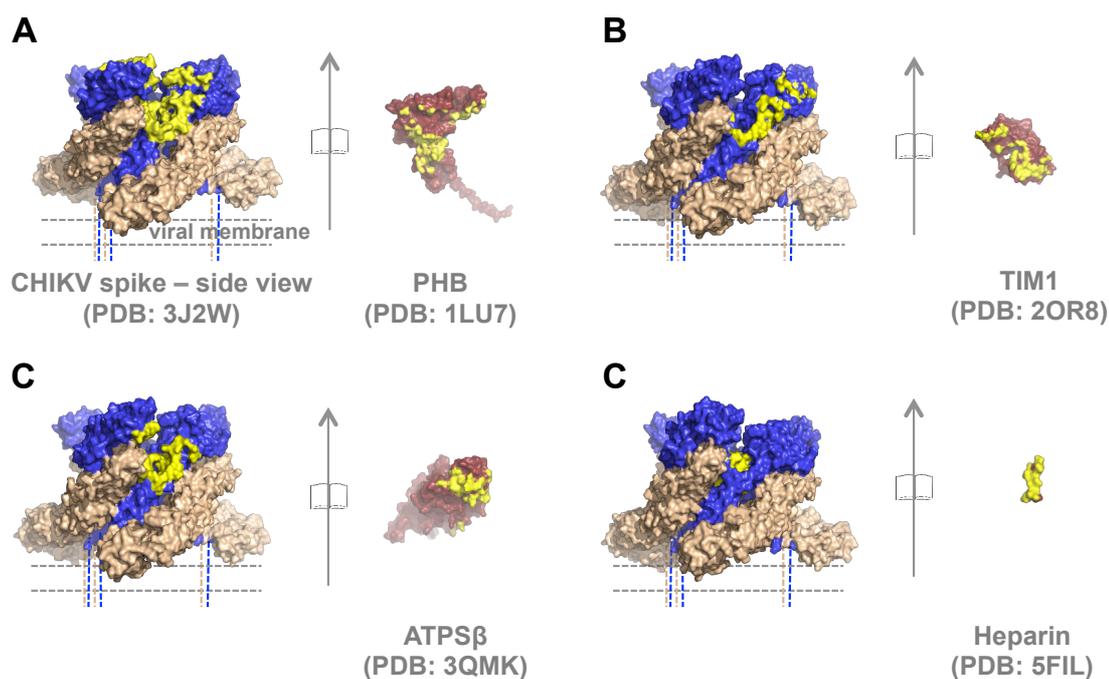


Figure 2.4 Docking of protein receptors and heparin tetra-saccharide onto CHIKV E2-E1 trimeric spike using ClusPro server. (A) Binding site of Prohibitin (PHB), (B) T-cell immunoglobulin mucin domain 1 (TIM1), (C) ATP synthase β subunit (ATPS- β), (D) Heparin (HEP) are mapped on the CHIKV E2-E1 trimeric spike structure. Each receptor structure, with E2-binding interface colored in yellow, is shown in open book orientation to the right of the spike.

2.3.2 HS binds to CHIKV E2 at predicted HBD

To experimentally test our docking predictions of HS binding site on E2, and to study E2 conformational changes upon HS binding, we cloned, expressed and purified CHIKV E2 or E3-E2 proteins (refer to chapter 1). Alphaviruses E3-E2 and E1 proteins fold together as heterodimers when produced and processed in infected cells and earlier studies expressing these proteins using insect cells have shown formation of irreversible E3-E2-E1 complexes. However, to study receptor binding induced conformational rearrangements in E2, it is desirable to have the purified E2 protein alone. We purified both CHIKV E2

and E3-E2 functional ectodomain proteins using bacterial expression system and then tested if these proteins have folded like their native E3-E2-E1 heterodimer counterparts. Our secondary and tertiary structural characterizations on E2 (refer to chapter 1) showed that purified CHIKV E2 protein is folded properly.

In the heparin-binding assay, more amount of CHIKV E2 remained bound to heparin column at 0.3 M and 0.6 M NaCl concentrations (Figure 2.5a, lanes 4-6), suggesting that the binding is through electrostatic interactions. Further, 1 M NaCl elution was required for complete release of bound E2 from heparin column. Soluble form of HS, when pre-incubated with E2, lead to reduced E2 binding to heparin column (Figure 2.5a, lane 8-10). This confirmed the specific nature of binding between E2 and heparin. To add to this, heparin bound fraction of E2 (32.6%) is significantly higher than BSA bound (7.4%) fraction (Figure 2.5b).

The HBD pattern XBXXBX, that we predicted earlier has four similar matching sites on CHIKV E2. All four of these sites are probably available on the purified E2 protein. When we mapped these sites on E2, in the CHIKV E3-E2-E1 heterodimer crystal structure, we observed that two out of these four sites are present at E3-E2 interface and would be masked on virion surface (Figure 2.3). Hence, we purified E3-E2 protein, confirmed its folding to native conformation (refer to chapter 1), and checked its binding to heparin. If the HBDs in the E3-E2 interface also contribute to HS binding, then E3-E2 protein should have lesser binding to HS than E2 alone. In our heparin-binding assays, E3-E2 binding to heparin (Figure 2.6a) is (fraction bound 32.9 %) very much similar to heparin binding by E2 alone, suggesting that the HBDs present at E3-E2 interface does not contribute to HS binding.

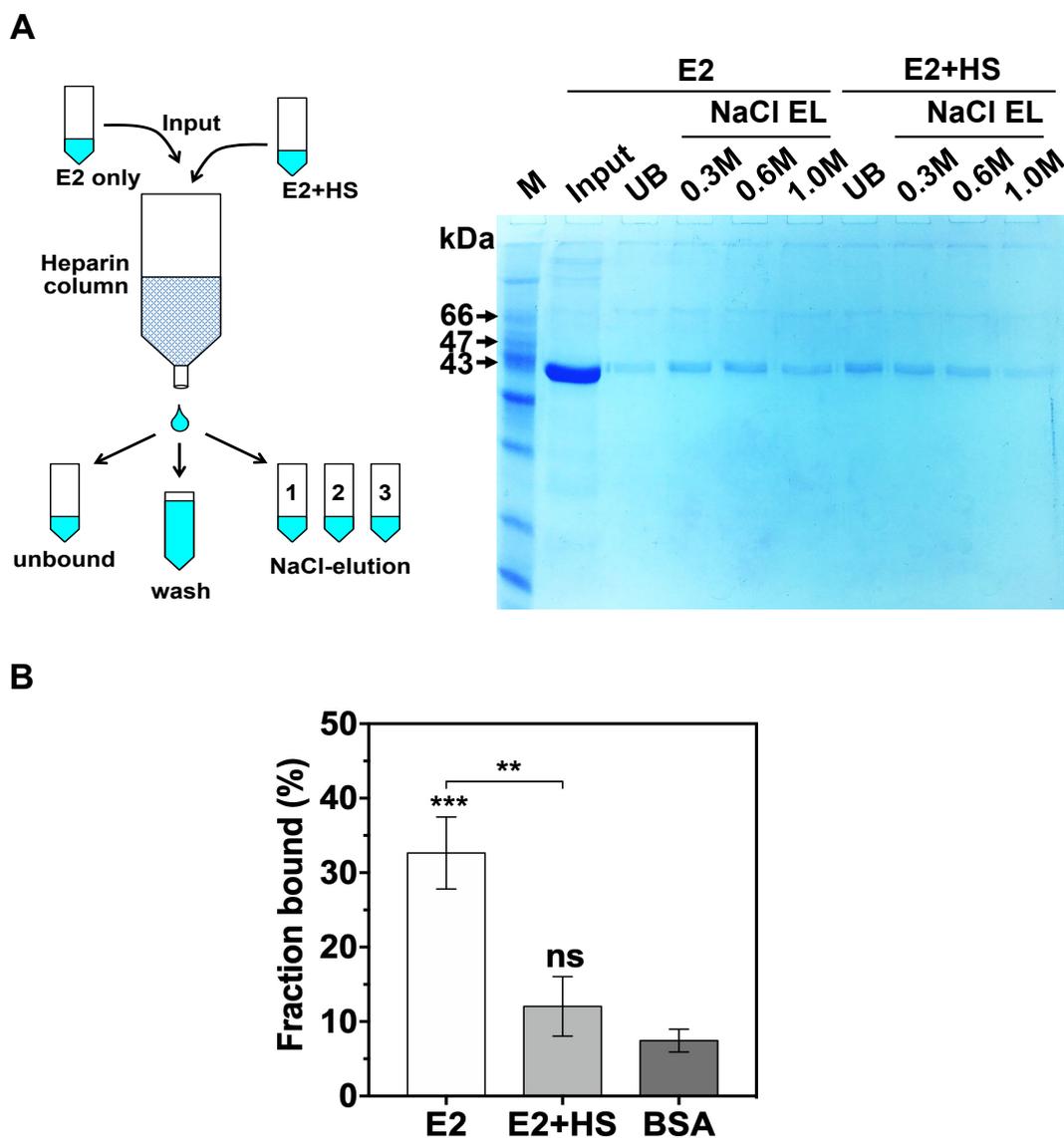


Figure 2.5 Characterization of E2-HS/HEP interactions. (A) Left: Schematic of HS/HEP binding assay. Right: Elution of E2 from heparin sepharose column. E2 (SDS gel image, lanes 4-6) and E2 pre-incubated with HS (SDS gel image, lanes 8-10). Lane labels: M - mol. wt. marker, UB - unbound, NaCl EL - bound fraction protein that eluted with NaCl. (B) Quantification of percent (%) bound protein fraction. E2+HS is E2 pre-incubated with HS in competition assay. ** and *** represent $P \leq 0.01$ and $P \leq 0.001$, respectively, 'ns' is non-significance.

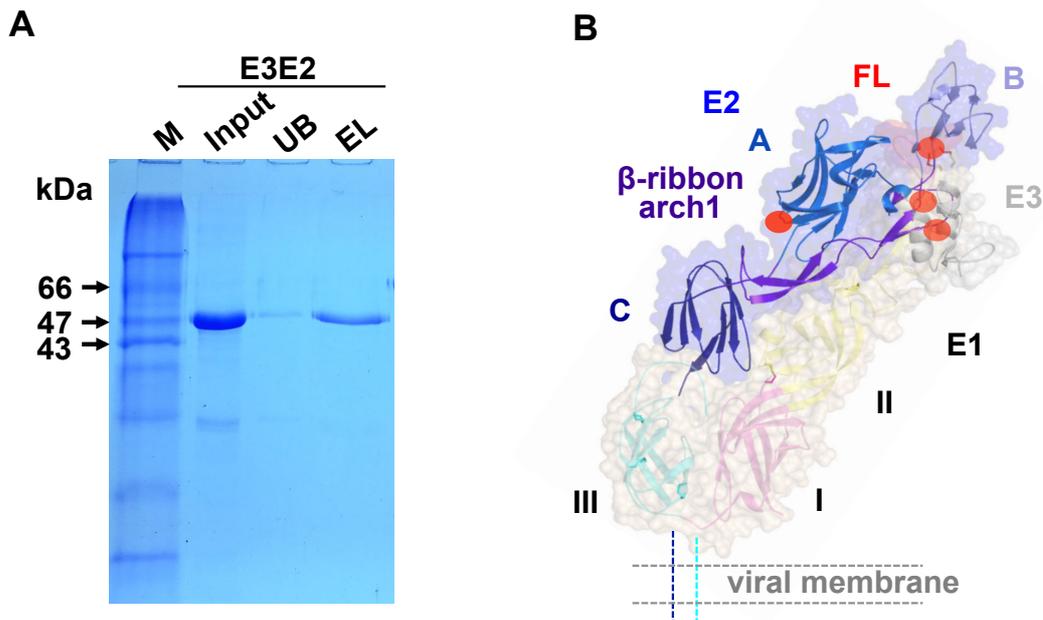


Figure 2.6 Characterization of E3-E2-HEP interactions. (A) SDS PAGE analysis on elution fractions of E3-E2 protein from heparin sepharose column. Lane EL: 1 M NaCl elution fraction. (B) All four HBDs including the two red ovals at E3-E2 interface are shown for reference.

From the predicted docking site HBD on E2, “AR₁₀₄CPK₁₀₇G”, R104 and K107 make electrostatic interactions with sulfates of HS. We generated E3-E2 R104A/ K107A/ R104A+K107A mutants, purified these mutant proteins, characterized their structure by far-UV CD and intrinsic tryptophan fluorescence, and confirmed that these mutants have folded in the same way as that of the WT protein (Figures 2.7a and 2.7b).

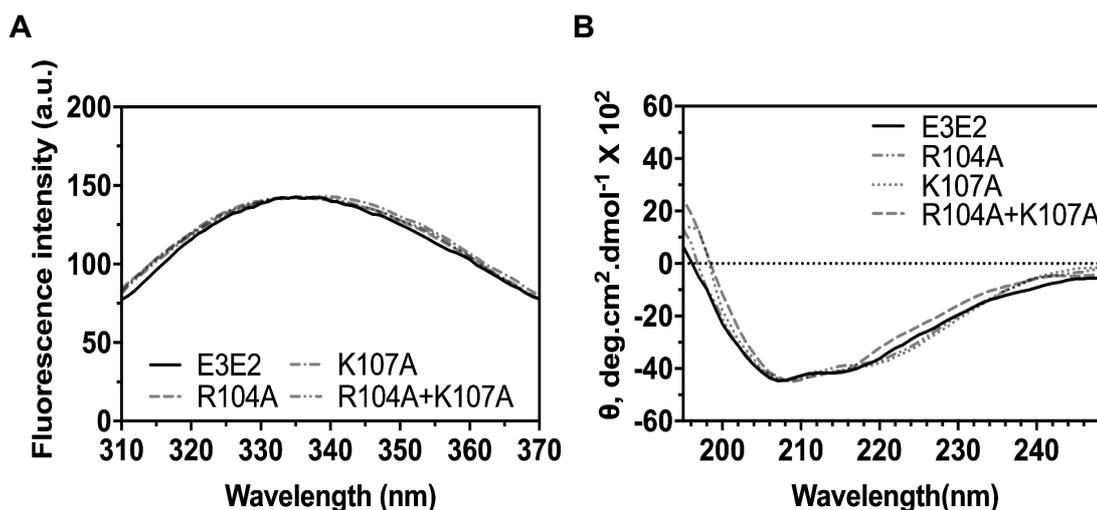


Figure 2.7 Characterization of CHIKV E3-E2 docking site-HBD mutant proteins. (A) Intrinsic tryptophan fluorescence spectra. (B) Far-UV CD spectra of E3-E2 mutant proteins, purified in a similar manner as E3-E2 wildtype protein.

Then we tested the heparin binding property of these mutants, which is significantly reduced/less compared to the WT protein (Figure 2.8). However, we did not observe complete loss of binding because of mutations in HBD of E2-dA (Figure 2.9). As mentioned earlier, there are three other HBDs in E2 structure and two of those three sites are not contributing to heparin binding by E2. However, one more predicted HBD present at the base of E2-dB would still be available for heparin binding on E2 or E3-E2 proteins. This site might be contributing to the remaining heparin binding observed with E3-E2 HBD mutants. But, this HBD that is very close to the E2-E1 interface may not be available on virion surface for HS binding (would be masked at E2-E1 interface on virion surface) (Figure 2.9). To add to this, our docking studies used a virion surface trimeric spike structure of E2-E1 and no HS docking poses were observed in E2-dB HBD.

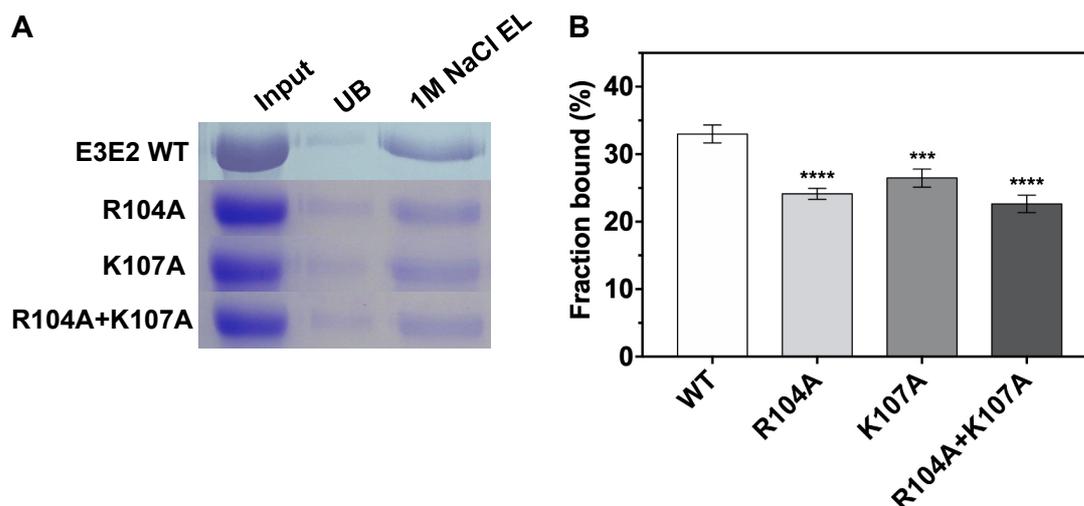


Figure 2.8 Characterization of HEP binding property of CHIKV E3-E2 mutant proteins. (A) SDS PAGE analysis on elution fractions of E3-E2 WT and mutant proteins from heparin sepharose column. Lane EL: 1 M NaCl elution fraction. (B) Comparison of E3-E2 WT and E3-E2 R104A, K107A and R104A+K107A proteins in heparin binding assay. Mean \pm SD ($n=3$) of percent (%) fraction bound to heparin is presented for each protein, with statistical differences calculated to E3-E2 WT protein. *** and **** represent $P \leq 0.001$ and $P \leq 0.0001$ level of significance, respectively.

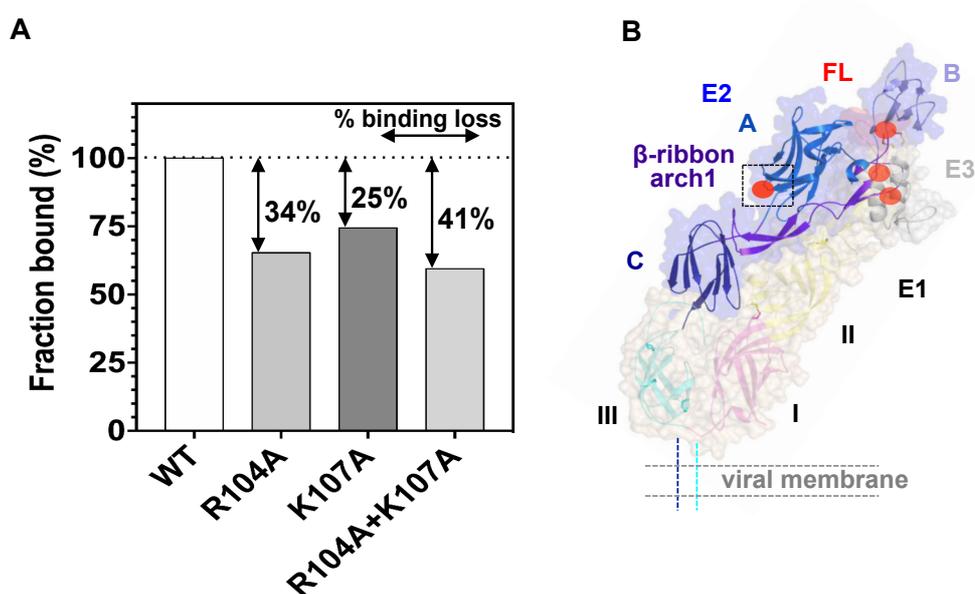


Figure 2.9 Analysis of HEP binding property of CHIKV E3-E2 mutant proteins. (A) Comparison of E3-E2 WT and E3-E2 R104A, K107A and R104A+K107A proteins in heparin binding assay. Loss of fraction bound to heparin is presented for each protein as percent (%) values. Binding of WT E3-E2 protein was normalized to 100% and binding of mutant E3-E2 proteins were calculated accordingly. (B) The red oval representing the docking site HBD is highlighted within a black colored dashed box.

Also, the extent of heparin binding observed by E3-E2 double (R104A+K107A) mutant is not additive effect of each single mutant (R104A or K107A). Earlier studies characterizing HS and other glycosaminoglycan-protein interactions also reported similar observations with HS-binding site mutants. For example in a recent study, Richard B et al (156) has shown non-additive effect of single, double or triple HBD site residue mutations on heparin binding, in anti-thrombin.

We also characterized the strength of HS binding to CHIKV E2 by Micro Scale thermophoresis assay. The K_d value observed is $1.8 \pm 0.6 \mu\text{M}$ and it indicates that the binding affinity is weak. Similar level of binding and K_d values are reported for other proteins that bind to HS (157). We also predicted presence of more than one HS binding sites on E3-E2 (Hill's co-efficient, $n=2$) (Figure 2.10). This further supports our analysis with E3-E2 R104A+K107A binding to heparin. Cryo-EM data on HS interactions with basic residues at a similar site in another alphavirus, EEEV E2 is reported recently (105).

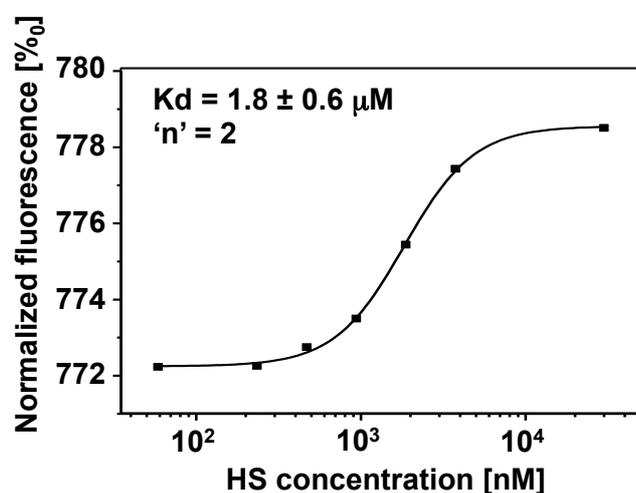


Figure 2.10 E3-E2 and HS binding affinity measurement through Microscale thermophoresis assay. Dose response curve for HS binding to CHIKV E3-E2 WT protein with K_d value and Hill's coefficient (n) value from the MST experiment.

Taken together, based on our docking predictions and comparison of heparin binding by purified E2, E3-E2 or E3-E2 R104A/ K107A/ R104A+K107A mutant proteins, we believe that the HBD between E2-dA and arch 1 of β -ribbon connector is the primary binding site of HS/heparin receptor. Also, of the four HBDs on E2 this site would be the most accessible site on trimeric spike structure at CHIKV virion surface.

2.3.3 HS binding to E2 results in domain C movement in, towards E1

Receptor binding to E2 triggering viral entry is one of the very first events in alphavirus entry. However, how receptor binding induces the breaking of E2-E1 interface interactions and dissociation is not understood. Does HS binding on E2 induce conformational changes? To understand this, we studied HS-docked E2 structure using MD simulations. Our clustering analysis of the simulation trajectory revealed two major clusters in a 6 clusters group (4 Å cut-off was used): cluster1 has 692 and cluster 2 has 284 structures (Figure 2.11a). We compared the pre-simulation HS-docked E2 structure with central structure of the major cluster1 (taken as resultant MD simulation structure in presence of HS). From overlay of both these structures it is clear that positions of E2-dB and E2-dC have changed significantly from their initial positions (Figure 2.11b). E2-dB has shifted from its completely masking position over E1 fusion loops at domain II tip to a slightly outward position. The center of mass calculations revealed that E2-dB has moved by 3.1Å from its starting position in the pre-receptor-bound conformation (Figure 2.11b, left). We did not see any such changes in E2, in a control run where E2 only MD run was conducted in absence of HS. The overall RMSD between the input E2 structure and the resultant E2 structure obtained from E2 only MD run is 1.62 Å. Radius of gyration (Rg) analysis indicated that HS binding lead to decrease in Rg of E2 and its structure become compact.

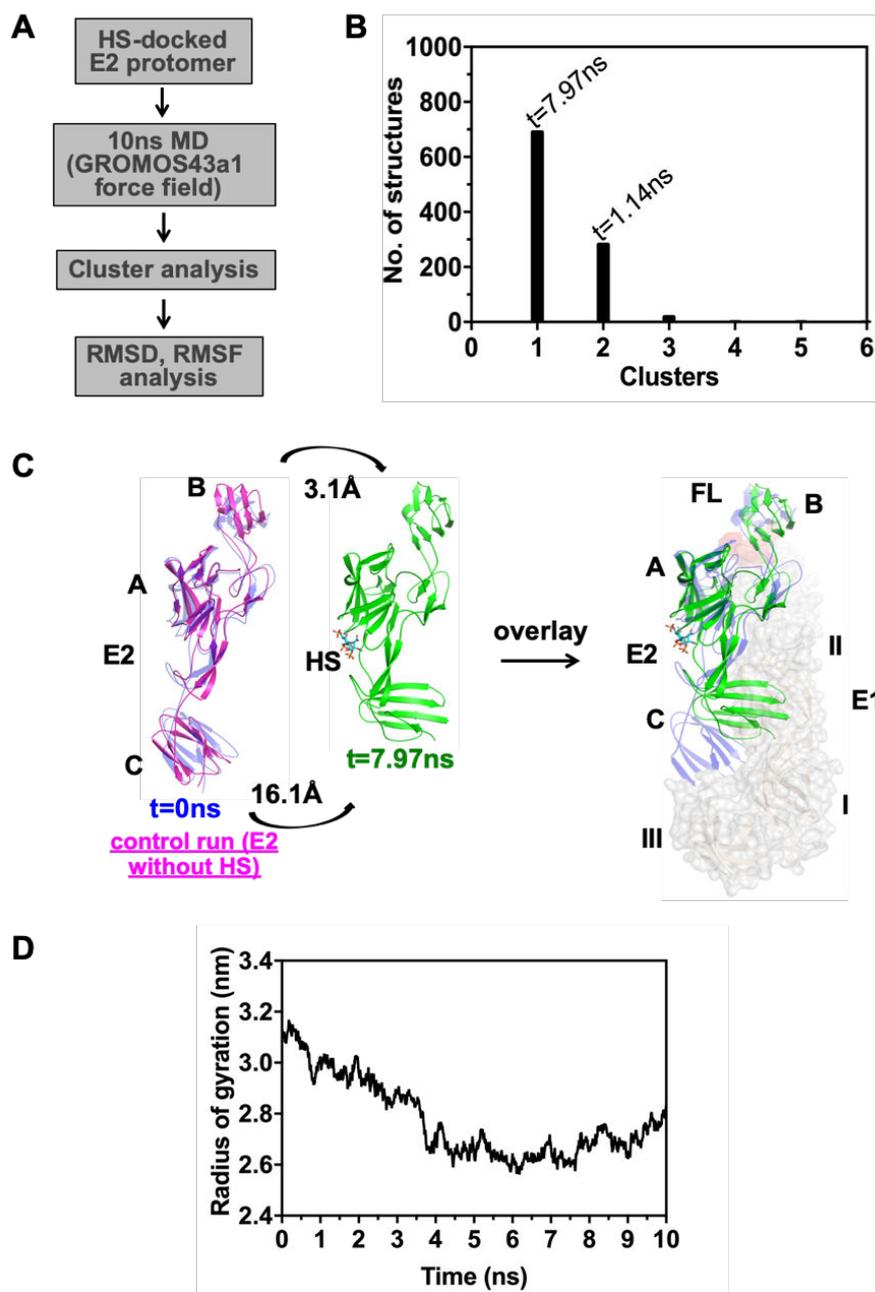


Figure 2.11 Analysis of MD simulation trajectory of E2-HS complex. (A) Schematic of the overall MD simulation analysis. (B) Clustering analysis of the MD trajectory. (C) Left: comparison of $t = 0$ ns structure with the control structure (magenta colored cartoon representation of E2, which was run for MD without HS), and $t = 7.97$ ns structure (central structure from the major cluster). Curved arrows indicate movement in domain B and domain C between $t=0$ ns and $t=7.97$ ns structures. Right: overlay of E2–E1 heterodimer structure with $t = 7.97$ ns structure. E1 (grey, surface representation) domains represented as I, II, III; and for E2 as A, B, C. E2 (blue, cartoon representation) and E2-HS, $t = 7.97$ ns structure (green) are aligned. HS is shown in stick representation. FL is E1-FL. (D) Radius of gyration analysis of the MD trajectory.

Further, backbone RMSD analysis also reflected similar results (Figure 2.12a). Major movement was seen with E2-dC - center of mass moved by 16.1 Å. (Figure 2.12a and 2.12b) and all atom RMSF analysis showed maximum deviation (Figure 2.13b). Overlay of MD resultant structure on to the spike E2-E1 heterodimer structure (Figure 2.11b) shows that domain C moved to a place where domain II of E1 would normally be positioned in the heterodimer complex. As observed, no significant secondary structural changes were seen upon HS binding.

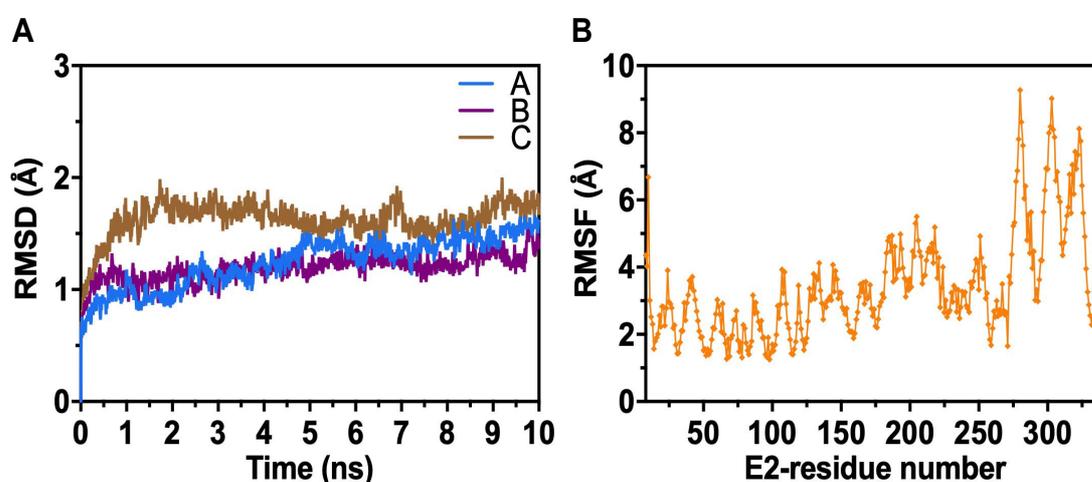


Figure 2.12 RMSD and RMSF analysis of MD simulation trajectory of E2-HS complex. (A) Domain wise RMSD and (D) RMSF analysis output from the MD trajectory of HS docked E2 complex structure.

Analysis of E2 structure shows presence of three out of five tryptophans in E2-dC. We thought, if HS binding leads to domain C movements as predicted from MD simulation studies, the environment of the tryptophan residues at least for the ones present in E2-dC region are likely to change. Tryptophan fluorescence emission spectra analysis on E2 structure in presence and absence of HS are represented as overlaid spectra in Figure 2.13.

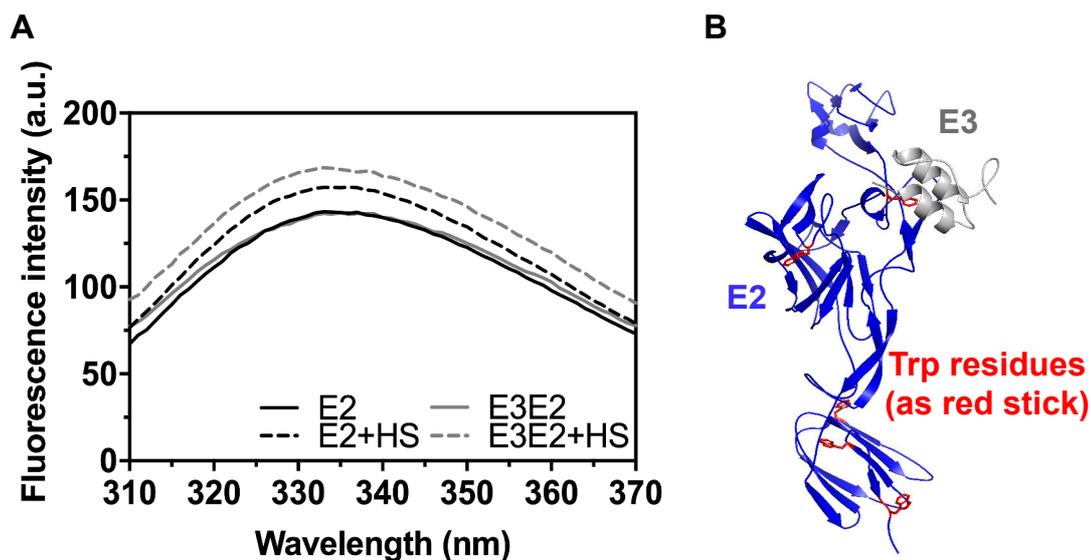
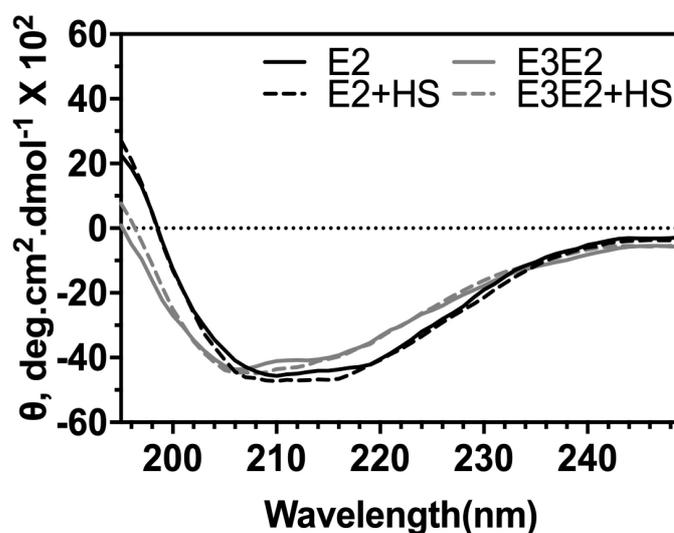


Figure 2.13 Tertiary structure analysis on HS-bound E2 and E3-E2 (A) Intrinsic tryptophan fluorescence spectra of E2 and E3-E2 in presence and absence of HS, respectively. (B) Structure of E2 (in cartoon representation) with tryptophans in E2 shown in red colored stick representation.

Intensity of intrinsic tryptophan fluorescence in presence of HS increased for both E2 and E3-E2 suggesting towards HS binding induced change in tertiary structure. On the other hand, far-UV CD spectra analysis showed no significant secondary structural change in presence or absence of HS (Figures 2.14a and 2.14b). From this, we hypothesize that tryptophan fluorescence changes observed is primarily because of E2-dC bending to bring three out of the five tryptophans close towards the β -ribbon connector.

A



B

% secondary structural content	<i>far-UV CD data</i> (K2D3 predicted values)		<i>crystal structure</i> (calculated from PDB: 3N42)		
	α -helix	β -sheet	α -helix	β -sheet	
<i>protein (-/+ HS)</i>					
E2	-HS	1.76	47.42	2.36	47.9
	+HS	2.5	47.74		
E3E2	-HS	3.09	36.41	7.07	37.70
	+HS	3.08	36.45		

Figure 2.14 Secondary structure analysis on HS bound E2 and E3-E2. (A) Far-UV CD spectra of E2 and E3-E2 in presence and absence of HS, respectively. (B) Estimated percent (%) secondary structural content in E2 and E3-E2 and their comparisons between far-UV CD data (predicted using K2D3 server) and from crystal structure data.

To validate MD simulation predictions further, we labeled a FRET fluorophore pair on E3-E2, one in domain C and another in β -ribbon connector. We reasoned that if domain

C moves towards β -ribbon connector, as predicted in our MD simulations, the FRET acceptor and donor introduced on domain C and β -ribbon connector should move closer, thus increasing the acceptor fluorescence. To choose a right FRET pair that can reflect the distance between the two E3-E2 domains, before and after HS binding, we analysed all the commercially available FRET pairs. This R_0 reported for most FRET pairs is in the range of $\sim 60\text{-}70\text{\AA}$ (158). However, the farthest distance between the domain C and β -ribbon connector is not more than 30\AA . Irrespective of where the FRET pair is placed, some amount of acceptor fluorescence is expected, even when HS is not bound to E3-E2. (refer to 2.17a). We expected that FRET to increase (increase in acceptor fluorescence and decrease in donor fluorescence) after HS binding. To introduce the FRET pair, we created a double cysteine mutant of E3-E2 (please refer to the description in Material and Methods) – a cysteine introduced in domain C, replacing S296; and another in β -ribbon connector, replacing S154 (Figure 0.15).

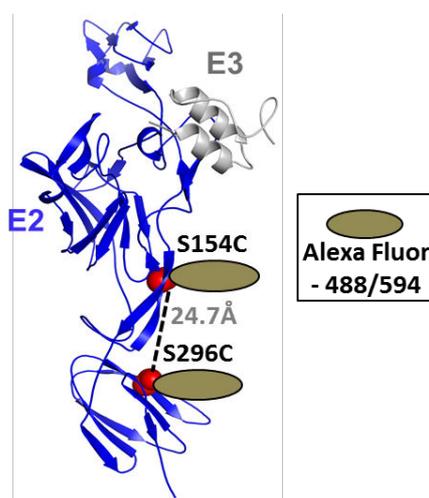


Figure 2.15 Analysis of E3-E2 structure for FRET pair introduction. Structure of E2 (in cartoon representation) with Ser154 and Ser296 (shown as red colored spheres) locations chosen for cysteine mutation are shown. $C\beta$ to $C\beta$ distance between these two residues is also shown.

The double cysteine mutant, E3-E2(S154C+S296C) is similar to WT with respect to secondary structural content and tertiary structure, as evident from our far-UV CD and intrinsic tryptophan fluorescence data (Figure 2.16a and 2.16b). We checked for the availability of free cysteines with Ellman's reagent prior to fluorophore labeling, to make sure that only introduced cysteines are available for labeling. Absorbance observed with Ellman's reagent at 412 nm for WT protein is same as buffer blank, whereas it increased for the double cysteine mutant (Table 2.1). Also, we checked heparin binding of the double cysteine mutants, labeled with FRET fluorophore pair – they bound to HS, to a similar extent as that of the wildtype protein (Figure 2.17).

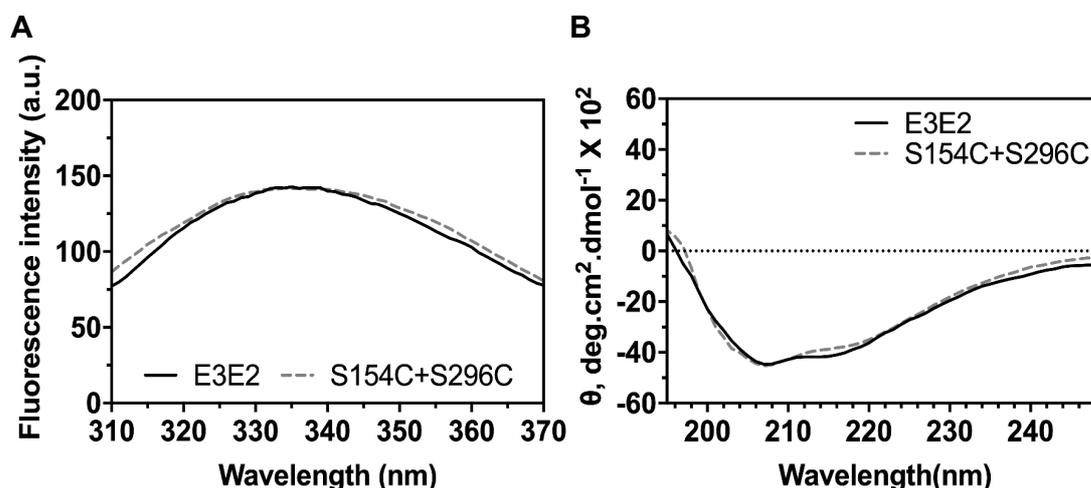


Figure 2.16 Characterization of CHIKV E3-E2 double cysteine mutant protein. (A) Intrinsic tryptophan fluorescence spectra and (B) Far-UV CD spectra of E3-E2 S154C+S296C mutant protein, purified in a similar manner as E3-E2 wildtype protein.

Table 2.1: 412nm absorbance values of BSA, E3-E2 WT and S154C+S296C mutant proteins and the buffer control from the Ellman's assay, which tests for free sulfhydryl groups in proteins. All proteins were tested in TNE pH 7.6 buffer for the experiment.

	TNE buffer	E3-E2-WT	Unlabeled E3-E2-S154C+S296C	Labeled E3-E2-S154C+S296C
Absorbance@412nm	0.14	0.14	0.24	0.14

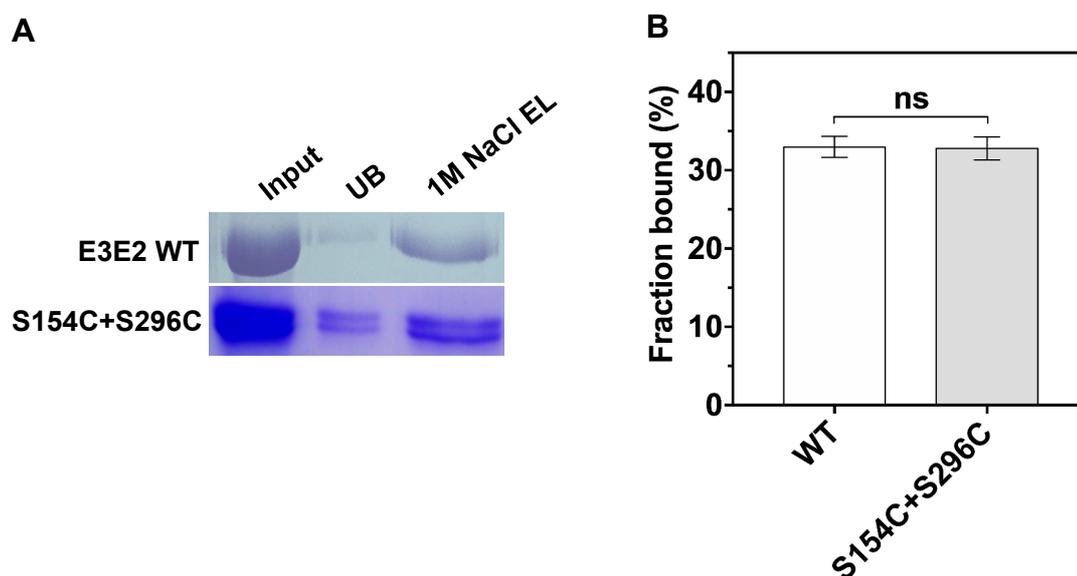


Figure 2.17 Characterization of HEP binding property of CHIKV E3-E2 double cysteine mutant protein. (A) SDS PAGE analysis on elution fractions of E3-E2 WT and double cysteine mutant protein from heparin sepharose column. Lane EL: 1 M NaCl elution fraction. The double cysteine mutant seen running as doublet in SDS-PAGE could be because of protein degradation. (B) Comparison of E3-E2 WT and E3-E2 S154C+S296C proteins in heparin binding assay. Mean \pm SD ($n=3$) of percent (%) fraction bound to heparin is presented for each protein, with statistical differences calculated to E3-E2 WT protein. 'ns' is non-significance.

We did FRET assay (Figure 2.18) with labeled E2 protein alone first. As expected, FRET was observed between donor and acceptor in labeled protein alone (Figure 2.18). This further suggests that our purified E3-E2 protein has domain organization similar to what is seen in the crystal structure. However, we did not observe significant change in acceptor or donor fluorescence emission in presence of HS (Figure 2.18).

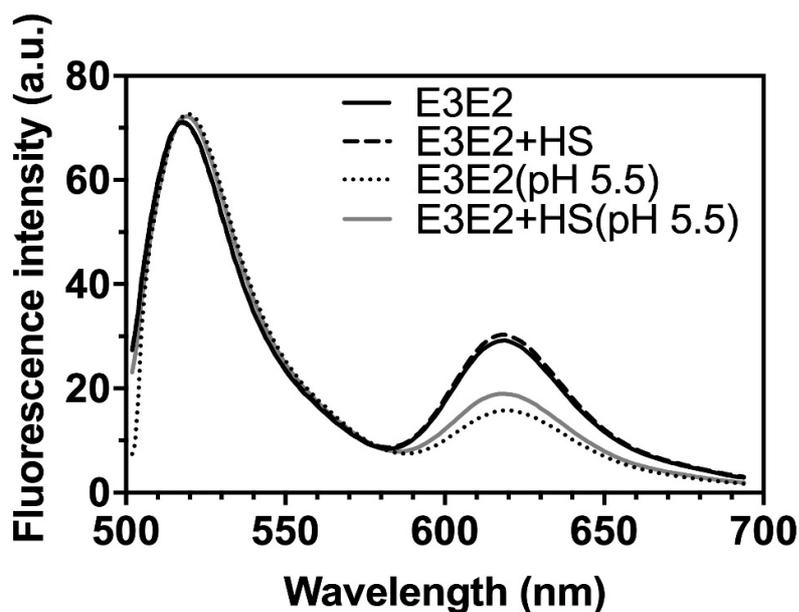


Figure 2.18 FRET analysis on HS bound E2 and E3-E2 structure. FRET data represented as fluorescence emission spectra of Alexa Fluor labelled E3-E2 (either at pH 7.4 or pH 5.5) WT protein, in presence or absence of HS.

MD simulation predictions suggested that HS binding to E2 may cause domain C to move by ~ 16 Å, close towards β -ribbon connector and if this is right we should expect more FRET efficiency in presence of HS compared to absence of HS with labeled E3-E2 protein. This could be because of donor and acceptor fluorophores being much closer (~ 24.7 Å) on the protein than the R0 (~ 60 Å) distance, and hence already exhibiting maximum FRET efficiency (close to 100%, to be specific 99.52%). In this scenario, even in the presence of HS, domain movements upon HS binding would not result in perceivable

change in FRET. If this assumption is right, then at a mildly acidic pH (5.5) condition, where E2 β -ribbon connector structure has been reported to be disordered that results in dissociation from domain A contacts (80), should result in reduced FRET because of increase in distance between FRET donor and acceptor pairs. In fact, FRET observed with E3-E2 at pH 5.5 is significantly reduced (Figure 2.18) (higher donor emission and lower acceptor emission) compared to FRET at pH 7.4, suggesting our assumption about distance measurements between FRET pair on E3-E2 at pH 7.4 to be right. Presence of HS with E3-E2 also did not alter the lowered FRET observed at pH 5.5 (Figure 2.18).

2.4 DISCUSSION:

Like other alphaviruses, CHIKV enters into cells through receptor binding triggered endocytic entry pathway. Its surface envelope protein E2 is the receptor binding protein. Despite years of study by many different groups, our knowledge on E2 - receptor interactions is very limited. Also, the mechanistic details of E2-E1 conformational rearrangements upon receptor binding are poorly understood. In current study, using *in silico* docking, MD simulations and biochemical approaches, I have characterized heparan sulfate and heparin receptor interactions with CHIKV E2 protein. We have also studied HS binding induced E2 conformational changes and we believe our results could possibly explain the E2-E1 dissociation mechanism - an essential step before membrane fusion. Here, we discuss our results in line with available literature on alphavirus E2-E1 dissociation mechanism during cell entry.

Many viruses make use of cell surface HS as attachment receptors. Alphaviruses such as CHIKV (78, 101, 103), SFV (159), SINV (159, 160), Ross River virus (161), EEEV

(162) can utilize cell-surface HS for triggering viral cell entry. Our docking results suggested that HS binds on to CHIKV E2-E1 trimeric spike structure at a positively charged pocket. This pocket is formed between E2-dA and arch 1 of β -ribbon connector. From the available structural data (based on structure alignments), analogous positively charged pockets can be seen in other alphaviruses (SINV (22) and VEEV (91)) trimeric spike structures as well. Our predicted HBD on CHIKV E2-dA is part of the highly conserved (both sequence wise and structurally) EF loop of alphaviruses (163).

Consensus HBD motifs are present on other viral envelope proteins such as HIV gp120 (164) and RSV G (165), DENV E (166) and HSV gD (167), which bind to HS through these HBDs. No such consensus HBDs are observed on alphaviruses E2 sequences. We identified a novel consensus HBD motif pattern, XBXXBX, in alphaviruses that is present most available alphaviruses E2 sequences and mostly map to domain A. In SINV, the HBD is present in CD loop of domain A with Lys and Arg residues side chain projections forming the HS binding pocket, in similar structural context as seen with CHIKV E2. In an earlier study, Voss JE et al (82) mapped mutations that effect HS binding (data published on alphaviruses) onto CHIKV E3-E2-E1 crystal structure and inferred E2-dB, E2-dA and β -ribbon connector to be putative receptor binding regions. Published reports show alphavirus neutralizing antibody escape mutations also map to same regions on E2. Kam Y et al (168) mapped CHIKV neutralizing antibody epitope on to β -ribbon connector at a region close to the HBD proposed in this work. Interestingly, predictions of binding sites for other protein and heparin receptors map to E2-dA or β -ribbon connector and E2-dB as well, overlapping with HS binding site on E2 (Figure 2.4). Zhang R et al (96) recently mapped antibody epitopes overlapping with MXRA8 receptor binding site on E2

to domain A and domain B. In fact, two other publications that have come up in parallel to this work and published in June 2019 by Basore et al (90) and Song et al (97) showed the first co-complex structure of a receptor (MXRA8) bound CHIKV E2-E1 trimeric spike structure and it is the 1st structural evidence of a receptor bound state of any alphavirus envelope protein structure. This study elucidated the MXRA8 receptor binding site, which is very much overlapping with our predictions of the receptor binding site on domain A and domain B regions of CHIKV E2 protein (Figure 2.18). In addition, in a very recent cryo-EM study by Chen et al (155) the authors have identified the HS binding site on another alphavirus, eastern equine encephalitis virus (EEEV) to be in a region on the homologous E2 protein that is highly comparable to the HS binding pocket that we proposed from this study on CHIKV E2 protein (Figure 2.19).

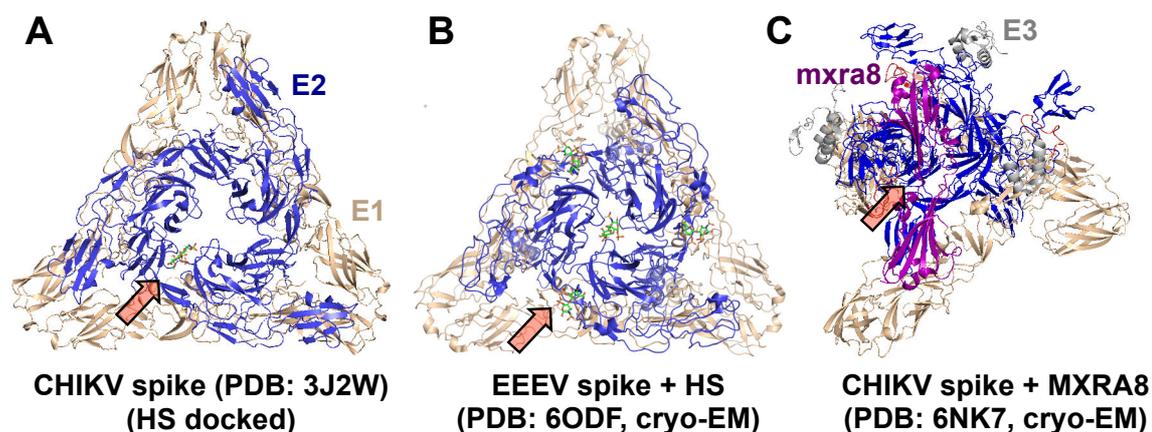


Figure 2.19 Comparison of reported MXRA8 and HS binding sites on alphavirus spike to HS docked site on CHIKV E2 from this study. Trimeric spike in CHIKV and EEEV are shown as cartoon representation. E2 is shown in blue and E1 in wheat color. E3 if present is shown in gray color. Only top and angled top views of the spike are shown for better visualization of receptor binding sites. (A) HS (shown as sticks) docked pose on CHIKV spike. (B) Cryo-EM determined sites of HS (shown as sticks) on EEEV spike. (C) Cryo-EM determined site of MXRA8 (shown as cartoon, purple color) on CHIKV spike.

Further, we validated our *in silico* predictions using recombinant CHIKV E2 or E3-E2 proteins, both of which bound to HS and heparin specifically, through weak electrostatic interactions ($K_d = 1.8 \pm 0.6 \mu\text{M}$). Mutation of HBD site charged residues, either R104 or K107 or both lead to significant reduction in HS binding. Altogether, our *in silico* and biochemical data support that the HS binding site at E2-dA and β -ribbon connector is the primary binding site of HS receptor on CHIKV E2.

Trimeric spike structures from CHIKV (82, 90, 141) SINV (22), VEEV (91) show three E2 molecules at the center of the spike, where E2-dA and β -ribbon connector region make most of the intra spike contacts. Also, interaction between E2-dC and β -ribbon connector with E1-dII hold E1 in place at the periphery of the spike. However, the post-fusion structure of E1 is a homotrimer. Role of both receptor binding (35, 82, 106) and endosomal acidic pH (22, 80, 169) in E2-E1 dissociation have been proposed. It is also likely that receptor binding to E2 may initiate heterodimer dissociation, which can then be completed during the endosome acidification step. In either case, E2-E1 dissociation would involve breaking of major inter-protein contacts (mostly between domain C of E2 and domain II of E1). E2-dA-dB ‘clamp’ opening around E1-FLs may happen subsequently or simultaneously to E2-dC and E1-dII dissociation. To evaluate how receptor binding affects/triggers these events, we did MD simulations on HS-bound E2 structure. Our prediction from simulations suggests movement of E2-dC, and E2-dB regions, upon HS binding. The HS binding region, the E2-dA however, was relatively stable and did not show significant change in conformation. Intrinsic tryptophan fluorescence, and FRET studies with CHIKV E2 and E3-E2 in the presence of HS indicated movement of domain C region. As imaginable, this movement of E2-dC would sterically clash with E1-dII in the E2-E1

heterodimer structure and probably would push E1 away from the hypothetical central axis of the trimeric spike structure. The marginal E2-dB displacement that is observed would open up the domain A and B clamp that locks E1-FLs and would probably break a critical hydrogen bond between H226 on E2-dB and A92 on E1-FL. Li et al (80) using SINV E2-E1 trimer crystal structure earlier reported that E2-dA-dB ‘clamp’ opens up, exposing the fusion loops on E1-dII tip, as an intermediary conformational change step during entry. Also, the authors have discussed role of E2-dC rearrangements as a possible mechanism for E2-E1 dissociation.

Based on experimental observations and interpretations of data available on other alphaviruses E2-E1 proteins, we propose a model-explaining role of receptor binding trigger in alphavirus entry (Figure 2.20). The positively charged groove between domain A and arch 1 of β -ribbon connector is a conserved heparan sulfate receptor binding region on E2 in alphaviruses. Receptor binding to E2-dA allosterically triggers movement of E2-dC in, towards E1, pushing E1 away from trimeric spike central axis. Subsequently or simultaneously, E2-dA-dB clamp opens up exposing E1 fusion loops. Envelope protein domain movements upon receptor binding triggering fusion protein activation have been reported for other viruses as well. In HIV-1, receptor and co-receptor binding to receptor binding protein induces allosteric conformational changes in receptor binding protein that induces conformational changes and triggers fusion protein activity (32). In measles virus, binding of receptor to receptor binding protein results in head domain movements, triggering fusion activity of fusion protein (36). Similarly, in Ebola virus, cell surface receptor binding followed by endosomal receptor binding to receptor binding protein induces allosteric conformational changes and triggers fusion protein activity (38).

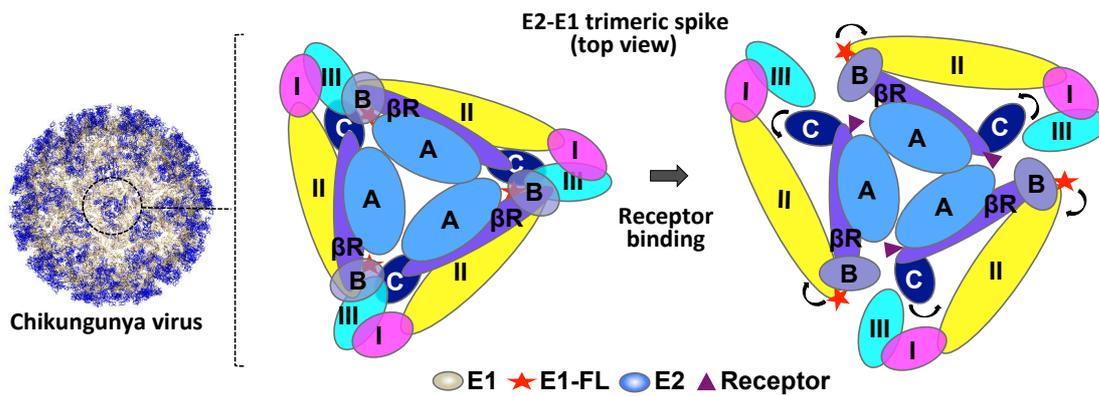


Figure 2.20 Mechanism of alphavirus E2-E1 dissociation upon receptor binding. Top view of E2-E1 trimeric spike (zoomed in view from the viral surface) before (left), and after receptor binding (right). E1 domains are represented as I, II and III (magenta, yellow and cyan colored ovals, respectively), and E2 domains are represented as A, B and C (marine blue, light blue and deep blue colored ovals, respectively) and the β -ribbon connector (shown as purple colored oval). Receptor (purple triangle) binding to E2 on domain A or β -ribbon connector allosterically induces domain movements in E2: E2-dC movement pushes E1-dII, resulting in E2-E1 dissociation. E2-dA-dB clamp opening exposes E1-FL (red star) at the tip of domain II.

Despite several studies on alphavirus receptors, the mechanism of receptor binding and triggering of cell entry remained elusive. Our study provides first clues for understanding receptor binding as a regulation mechanism in alphavirus cell entry. Interestingly, our is corroborated by a recent study by Chen et al (155), through cryo-EM study the authors have identified HS binding site(s) on another alphavirus, eastern equine encephalitis virus (EEEV) to be in a region on the homologous E2 protein that is highly comparable to the HS binding pocket that we propose in this study on CHIKV E2 protein (Figure 2.19). The potential use of HS sulfate as an antiviral agent in alphavirus entry is proposed by Chen et al (155) and Weber et al (102). Such approach is an extrapolation of the knowledge that we have, based on receptor binding site blockers developed against influenza (51) and HIV viruses (52).

CHAPTER 3

CHARACTERIZATION OF CHIKV E1 DI-DIII INTERFACE MUTANTS TO EXPLAIN ACIDIC PH TRIGGERED CONFORMATIONS DURING ENTRY

Acidic pH triggered conformational changes in Chikungunya virus fusion protein, E1 – a spring twisted region in domain I-III linker acts as a hinge point for swivelling motion of domains. **Sahoo B**, Gudigamolla NK, and Chowdary TK (**Under review in Journal of virology**)

3.1 INTRODUCTION:

CHIKV membrane fusion protein, E1, has structural features of class II viral fusion proteins: has three β -sheet rich domains (domains III, I and II, from viral membrane proximal to distal order); has internal fusion loops at the tip of domain II; form a heterodimer with receptor binding protein, E2, in pre-fusion state, but form homotrimers post fusion (refer to figure 3.1 and refer to General introduction for detailed structure description). Owing to the structural similarity with other viral membrane fusion proteins (FP), it is proposed that CHIKV E1 follows the universal structural mechanism explained for viral fusion proteins (170, 171). Viral fusion proteins undergo pre- to post-fusion conformational changes, through an extended intermediate conformation to facilitate membrane fusion (please refer to General introduction for detailed description of the universal mechanism). In class II viral fusion proteins, the three domains of the fusion protein move and rearrange into a post-fusion hairpin conformation by the end of the membrane fusion (172–175). Domain rearrangements in pre- to post-fusion conformation change in alphavirus E1 are triggered by acidic pH (172–175).

E3-E2-E1 protein organization on virion surface follows that of other alphaviruses. Crystal structure of CHIKV E3-E2-E1 heterodimer in its pre-fusion conformation (PDB: 3N42) (82) and cryo EM reconstruction of the trimeric E2-E1 on virion spike (176) revealed that E2 is at the center of the spike, whereas E1 is at the periphery. The spike structure is held by the inter-protomer interactions of E2. The three E1 protomers do not interact with each other (refer to figure 3.1).

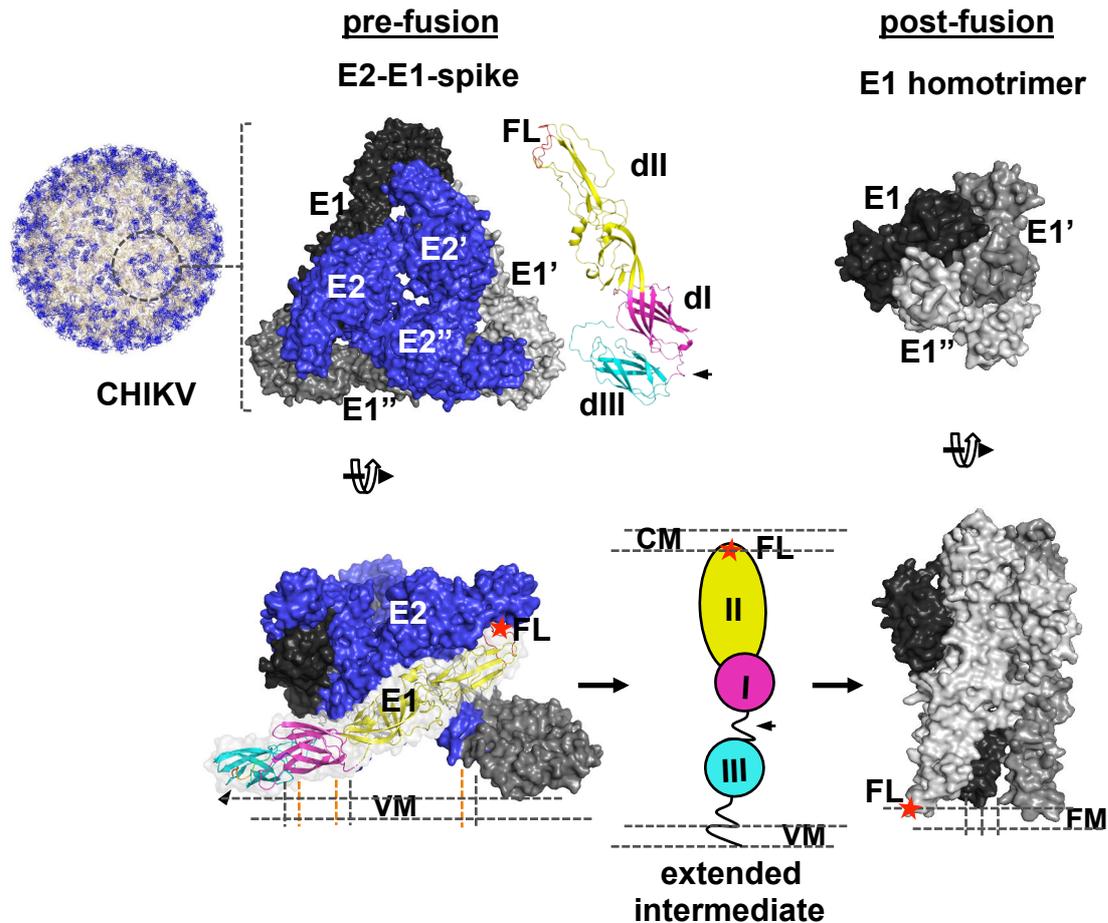


Figure 3.1 CHIKV envelope proteins pre- to post-fusion conformation switch. Top (left to right) - Image from 3D reconstruction of CHIKV E2-E1 into CHIKV VLPs (EMD-5577), surface view (labelled CHIKV), is shown. One trimeric spike structure is shown in circle. Zoomed in image of one E2-E1 trimeric spike is shown to the left (PDB: 3J2W). Top view of E1 (different shades of gray) and E2 (blue) are shown in surface representation. Three protomers of E1 and E2 are labelled, E1, E' and E'', and E2, E2' and E2'', respectively. A single monomer of E1 (from PDB: 3N42), in cartoon representation, is shown to explain the pre-fusion hairpin conformation, with the three domains (dI, dII and dIII) and fusion loops (FL) labelled. E1 dI-III linker is marked with a left black arrow. Top view of the post-fusion E1 homo-trimer (PDB: 1RER) is shown to the right. Bottom: shows side view of the trimeric spike in the left, schematic of a possible extended intermediate in the middle (dIII, dI, dII and fusion loops depicted in cyan, magenta and yellow ovals and a red star, respectively), and post-fusion E1 homo-trimer conformation in the right. VM, CM and FM refer to viral membrane, cell membrane and fused membrane, respectively. Dashed lines connecting protein structure figures with VM and FM, represent stem and transmembrane region that anchor the proteins into VM and FM.

Post-fusion structure of CHIKV E1 is not available. However, crystal structure of E1, which is a homotrimer of E1, from Semliki Forest virus (SFV), determined at acidic pH (pH 4.0) is published (PDB: 1RER) (172). In this conformation, dI-II regions from three E1 protomers form the central core of the homotrimer, surrounding the vertical axis of the trimer. E2 is not present in the post-fusion conformation. So, during pre- to post-fusion conformation switch, dI-II region from three E1 protomers should come together and establish inter-protomer interactions. Domain III from each protomer is arranged, folded back onto dI-II central core, like an outer whorl of the structure (refer to figure 3.1, 'post fusion'). Owing to the similarity of CHIKV E1 pre-fusion structure with that of SFV and SINV pre-fusion structure, it is reasonable to assume that CHIKV E1 also forms a post-fusion homo-trimer, very similar to SFV E1 post-fusion homo-trimer.

An evidence for an extended conformation of E1, though an indirect one, came from cryo-EM studies of Cao et al (177). They exposed SINV viruses to pH 6.4 in presence of liposomes. A weak electron density bridging virus and liposomes at the periphery of the trimeric spike they interpreted as extended conformation of E1 (refer to figure 0.17 in general introduction chapter). Several other biochemical studies also indicated that domain III and dI separate and form an extended intermediate during early steps of membrane fusion (174, 178–180).

The pre- to post-fusion conformational switch, from E3-E2-E1 trimeric spike to E1 homotrimer, is triggered by two sequential events: 1. Receptor binding on E3-E2-E1 spike and 2. Acidic pH of the endosome. Earlier, we have shown that receptor binding leads conformational changes in E2 that explain triggering of E2 dissociation from E1. We hypothesized that receptor binding may initiate dissociation of E2 and E1, however

complete dissociation may require acidic pH of the endosome(181) (refer to chapter 2, Discussion). Several other studies (172, 173, 177, 182, 183) report dissociation of E2 and E1 in acidic pH of the endosome. As part of the pre- to post-fusion conformation switch, once E2 dissociates from E1, the dI-II region from the three E1 protomers have to swivel to the center of the spike, forming the homotrimer. Li et al (173) proposed that, upon acidic pH trigger, dI-II region would move by 180° to the center of the spike, displacing E2. Since, the dIII is anchored into the transmembrane region, one possible way that the dI-II region of each protomer moving to the center of the spike is by a swivelling motion over the dI-III linker. By time-resolved cryo-electron microscopy on SFV, Fuller et al (184) also proposed that at acidic pH, E1 shows a centripetal movement by a swivelling motion towards center of the spike to form a trimer. A similar proposal was done by Mukhopadhyay et al (185).

However, mechanistic details on how and where in the protein, acidic pH triggers swivelling motion of domains is not known. During pre- to post-fusion conformation change, domains of E1 are likely to move as rigid bodies — no major change in the structure of individual domains is seen. The domain I-III linker (the hairpin bend) is the only flexible region in the protein (please refer to General introduction for description of the dI-III linker structure). We wondered if the swivelling motion of domains is brought about by conformational changes in the dI-III linker in acidic pH.

In parallel or subsequently, the hairpin structure may extend linearly to arrange all three domains of E1 into a linear axis. Before linear extension of the molecule, the interactions between dIII (one arm of the pre-fusion hairpin) and dI (the other arm of the hairpin) have to be broken. Dissociation of domain I from III is also triggered by acidic pH.

Conserved histidines and other charged residues in the dI-dIII interface are proposed to be critical for acidic pH sensing. In the pre- to post-fusion conformation switch, a conserved histidine, H331, shows altered interactions other charged residues in the dI-dIII interface. H331 makes electrostatic interactions with K16 in pre-fusion and with N149 in post-fusion with an adjacent protomer of E1 (refer to General introduction chapter for details). To answer how acidic pH triggers dissociation of domain I from III, we studied the role of a hydrogen bond between a conserved histidine in domain III (His₃₃₁ in CHIKV E1) and Lys₁₆ in domain I.

This chapter narrates the results of the study and discusses structural changes in E1 dI-dIII linker region and dI-dIII interface triggered by acidic pH that explain swivelling of E1 dI-dII over dIII.

3.2 MATERIALS AND METHODS:

3.2.1 Sequence and structure analysis of CHIKV E1

We used sequence corresponding to the dI-III linker region (residues 283-294) of CHIKV E1 (protein_id="ABD95938.1), to align with corresponding sequence region from other alphavirus E1 sequences. Altogether, 16 sequences were compared in multiple sequence alignment (MSA). We used Clustal Omega program with default parameters for MSA analysis.

Crystal structure of CHIKV E1-E2-E3 soluble ectodomain complex, PDB: 3N42 (186), is used for analysis of pre-fusion conformation of E1. Post-fusion homo-trimer structure of CHIKV E1 was modeled using CHIKV S27 isolate E1 sequence (protein_id="NP_690589.2") as input and SFV post-fusion E1 structure (PDB: 1RER) as template. We used the SWISS-MODEL web server for homology modeling. We did Propka analysis on E1 protein in pre-fusion (CHIKV E1, PDB: 3N42), and in post-fusion homo-trimer conformation) using the PDB2PQR server available at http://nbc-222.ucsd.edu/pdb2pqr_2.0.0/ with default parameters. We used PyMol for studying pre-fusion and post-fusion conformations of domain I, and III interactions involving conserved H₃₃₁ and preparing structure figures.

3.2.2 Constant pH molecular dynamics (MD) simulation on E1 at pH 5.5

We performed constant pH molecular dynamics (MD) simulation on E1 at pH 5.5. We used the AMBER99SB force field as part of GROMACS 5.1.4 package for carrying out energy minimization, simulations and trajectory analyses on CHIKV E1 structure. CHIKV E1 structure extracted from PDB: 3N42 was prepared for MD simulation by removing water molecules and HET atoms first. The 'cleaned' pdb was then used as an

input at PlayMolecule server available at <https://playmolecule.org/> for assignment of protonation states at a given pH value. In Gromacs, all other steps till final MD run were performed as described in chapter 2 for E2-HS complex MD simulation. Production MD run was performed at 300K for 50ns time scale.

3.2.3 Cloning, expression and purification of CHIKV E1 WT and mutant proteins

All the mutant proteins used for this part of the study (E1 dI-dIII interface mutants: H331D, H331E, and the H331C+L18C proteins) are expressed and purified exactly in the same manner as that of the WT protein (detailed in chapter 1).

3.2.4 Far-UV circular dichroism and intrinsic tryptophan fluorescence spectroscopy

To ascertain that the mutant E1 proteins are structurally similar to the WT protein, we recorded far-UV CD and intrinsic tryptophan fluorescence spectra on E1 dI-dIII interface mutants, E1 H331D, H331E, and the H331C+L18C. Experimental procedure for recording CD spectra and fluorescence spectra were the same as used for the WT protein (described in chapter 1).

3.2.5 Size exclusion chromatography

We analysed the quaternary structures of E1 dI-dIII interface mutants: H331D, H331E, and the H331C+L18C proteins using size exclusion chromatography in the same way as the WT protein (procedure detailed in chapter 1).

3.2.6 Lipid co-floatation assay

We used this assay to study the membrane interaction properties of E1 H331C+L18C and H331D/E mutant proteins in the same manner as the WT protein (procedure detailed in chapter 1).

3.2.7 Pyrene excimer fluorescence assay for studying E1-membrane association

We used this assay to study the membrane destabilization properties of E1 H331C+L18C and H331D/E mutant proteins in the same manner as the WT protein (procedure detailed in chapter 1).

3.2.8 Liposome content mixing fusion assay

We used this assay to study the membrane fusion properties of E1 H331C+L18C and H331D/E mutant proteins in the same manner as the WT protein (procedure detailed in chapter 1). All readings were taken as described previously except for the E1 H331C+L18C + β -ME sample, in which case 5mM β -ME was added to the reaction mixture before acidic pH triggering step.

3.3 RESULTS:

3.3.1 A conserved arginine-aspartic acid/glutamic acid pair in alphavirus E1 dI-III linker

In order to understand the role of dI-III linker region in E1 domain swivelling movements during membrane fusion, we started by aligning and comparing the linker region sequences from different alphaviruses. Sequence alignment is shown in figure 3.2A. There are two conserved prolines, one on either side, as bounds of the linker in most of the sequences. Overall, no sequence conservation is noted, except for an arginine and a phenylalanine in the middle of the sequence. As per CHIKV E1 sequence numbering, the conserved arginine is at 289 position. We noted that an aspartic acid or a glutamic acid residue, 2 or 3 residues in the C-terminal side of the Arg₂₈₉, is seen in 14 out of 16 sequences that we compared. Zheng et al (179) studied importance of the dI-III linker in SFV E1 by creating point mutations and deletion mutants. Their observations gave insights into importance of the linker in acid pH-induced homo-trimerization of E1. They inferred that an acid-sensing conserved histidine residue in domain I, H3, forms a π -cation interaction with the conserved arginine and phenylalanine of dI-III linker, in post-fusion homotrimer. This suggests that the conserved arginine and surrounding residues in the dI-III linker may have a critical role in acid pH-induced conformation changes in E1.

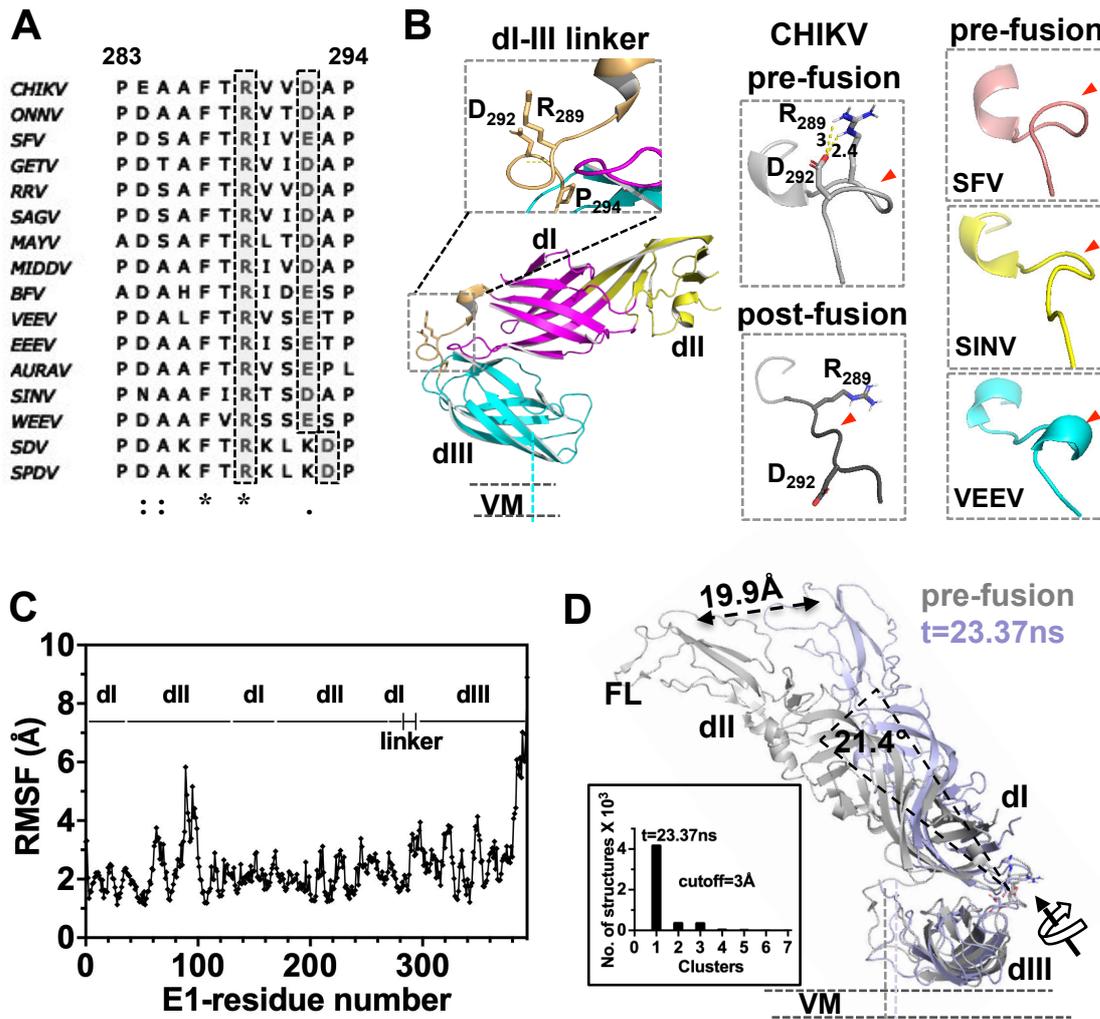


Figure 3.2 Analysis of alphavirus E1 di-III linker region in pre- and post-fusion state. (A) Multiple sequence alignment of alphavirus E1 di-III linker: conserved arginine – aspartate/glutamate pair is highlighted in gray. Sequence numbering is as per CHIKV E1 sequence. (B) Left column - structure of di-III linker from CHIKV E1 (from PDB:3N42) in pre-fusion conformation (cartoon representation). Conserved arginine (R289) and aspartate (D292) pair is shown in sticks. Zoomed in image of the boxed region of di-III linker is shown in inset above. Main chain hydrogen bond holding the spring-twist region is shown by yellow dotted line. Middle column – CHIKV E1 di-III linker region in pre-fusion and post-fusion conformation. Distance between R289 and D292 side chain atoms is shown and shown by yellow dotted line. Right lane – di-III linker regions from pre-fusion structures of SFV, SINV and VEEV are shown. In middle and right lanes di-III linker is shown in a different orientation compared to that in left lane, and red arrow heads point to the spring-twisted region of the linker. A different orientation of the linker (compared to left lane) is shown (C) Graphical presentation of root mean square fluctuation (RMSF) analysis of the MD simulation trajectory. Sequence stretch is mapped to the

corresponding domain (dI, dII, dIII or dI-III linker). (D) Structural overlay of $t=23.37\text{ns}$ cluster central structure from MD simulations (slate) over pre-fusion crystal structure of E1 (from PDB 3N42). Cluster analysis of the MD simulation trajectory is graphically presented in the inset. The structures were aligned over domain III. Centre of mass of dI-II regions were estimated using `center_of_mass.py` script from PyMol script repository. A double-headed arrow depicts displacement of center of mass of dI-II regions in MD simulation. Rotation of dI-II virtual axis over the dI-III linker is shown.

3.3.2 A conserved spring-twisted structure in the dI-III linker is untwisted in pre- to post-fusion conformation change

We then analysed structure of the dI-III linker from different alphaviruses, where a pre-fusion structure of E1 is available. The linker region does not have any regular secondary structure, except for a 3_{10} helix at the N-terminus (figure 3.2B). However, two features of the dI-III linker structure caught our attention. In CHIKV E1, side chains of the conserved arginine (R289) and the conserved aspartic acid (D292) face in the same direction, away from the protein interior (figure 3.2B, left column and inset). The distance between the Arg₂₈₉ and the Asp₂₉₂ side chains (3.0 \AA between NH₂ on R289 and OD₂ on D292; PDB: 3N42) would allow a salt bridge formation (figure 3.2B, middle column). Similarly, in SFV and VEEV as well, the Arg₂₈₉ and Glu₂₉₂ side chains are at a distance and orientation to form a salt bridge. Thus, salt bridge between the conserved Arg and Asp/Glu pair in the pre-fusion conformation seems to be a conserved structural feature in alphavirus E1 dI-III linker region.

Another interesting structural feature that we noted is the backbone conformation of the region between the conserved Arg₂₈₉ and Asp₂₉₂. Backbone of this region is twisted like a spring. And the spring twist of the backbone, appears like, is stabilized by the Arg-Asp/Glu salt bridge, apart from a main chain hydrogen bond (figure 3.2B, inset in left

panel). Except for VEEV E1, where this region is modeled as a 3_{10} helix, in CHIKV, SFV, SINV E1 this region is seen as a spring-twisted structure (figure 3.2B, panels in right column). Considering that the dI-III linker is critical for viral entry (179), and the conserved Arg-Asp/Glu salt bridge of the linker may break at the acidic pH, we wondered if the spring-twisted region of the dI-III linker acts as a ‘hinge’ for dI-II swivelling (over dIII) during membrane fusion. If this were to be the case, then in the extended intermediate and in post-fusion structure of E1 we expect to see the spring-twisted region to be untwisted, and the Arg-Asp/Glu salt bridge to be broken. Post-fusion structure of CHIKV E1 is not available. Hence, we modeled post-fusion structure of CHIKV E1, using SFV E1 post-fusion homo-trimer structure (PDB: 1RER). Indeed, in the post-fusion structure of SFV, and in post-fusion model of CHIKV E1 (figure 3.2B, middle column, panel labelled ‘post-fusion’), the conserved arginine side chain is on the opposite side to that of aspartic acid (glutamic acid in SFV). More importantly, backbone of the region does not show the spring twist seen in the pre-fusion conformation.

Thus, our sequence and structure analysis suggest that the spring-twisted region of the E1 dI-III linker acts as a ‘hinge’ for the domain swivelling, upon acidic pH trigger.

3.3.3 MD simulations of E1 at acidic pH show dI-II rotation on dIII over the linker spring-twisted hinge

Histidine residues in the dI-III interface have been proposed as key acid-sensing residues (187). One of the proposed ‘acid-sensing’ residue is histidine 331 in domain III (188). In pre-fusion conformation of CHIKV E1, histidine 331 is in a hydrogen bond interaction with lysine 16 in domain I. This hydrogen bond staples both domains from

moving apart in the pre-fusion conformation. We reasoned that, in acidic pH, histidine 331 gets bi-protonated and the domain I-III stapling His₃₃₁-Lys₁₆ interaction breaks. Only then, domain I-II region can swivel towards the center of the spike. The swivelling-in motion, as such, is brought about by parallel untwisting of the dI-III linker spring-twisted region. In order to test this idea, we performed constant pH molecular dynamic simulations on E1 structure. The optimum pH for triggering membrane fusion for CHIKV is 5.5 (189). CHIKV E1 structure (corresponding to ectodomain of a monomer) was ‘titrated’ to pH 5.5. We, using PROPKA on PDB2PQR server, assigned protonation states on all titratable residues and estimated the pKa values. We compared pKa values of ‘acid sensing’ residues at pH 7.4 and pH 5.5. We expected that protonation of the His₃₃₁ at pH 5.5 would be different from that seen in neutral pH. A similar change in protonation status of the acid sensing histidine residue, His₃₂₃, has been proposed as a trigger for conformational change in E protein of tick-borne encephalitis virus, at fusion-triggering acidic pH (190).

Contrary to our expectation, His₃₃₁ did not show any change in protonation when pre-fusion structure was titrated to pH 5.5 (pKa of His₃₃₁ at pH 7.4 and at pH 5.5 is 4.89). Only change noted for His₃₃₁ side chain is that it changed to ϵ -tautomer (from δ -tautomer at pH 7.4). We then used this pH 5.5-titrated E1 structure for MD simulation to predict conformational changes in the protein. RMSF deviation per residue for the trajectory is shown in figure 3.2C. As expected, there is no major change seen in (less than 3 Å RMSF deviation) structure of domains I, II and III, except for fusion loops region of domain II (RMSF deviation of ~ 6 Å). Another region that showed RMSF change above 3 Å is dI-III linker (figure 3.2C, marked as ‘linker’). Cluster analysis on the conformations of MD trajectory revealed a major cluster at $t=23.37$ ns (figure 3.2D inset). Interestingly, His₃₃₁ -

Lys₁₆ interaction did not change in $t=23.37$ ns structure. This implies that breaking of His₃₃₁-Lys₁₆ interaction is not an acid pH-triggered structural change. Important to note that, in post-fusion conformation this interaction is broken, and His₃₃₁ makes a new interaction with Asn₁₄₉ from another protomer in the homo-trimer structure (refer to figure 3.3A and B).

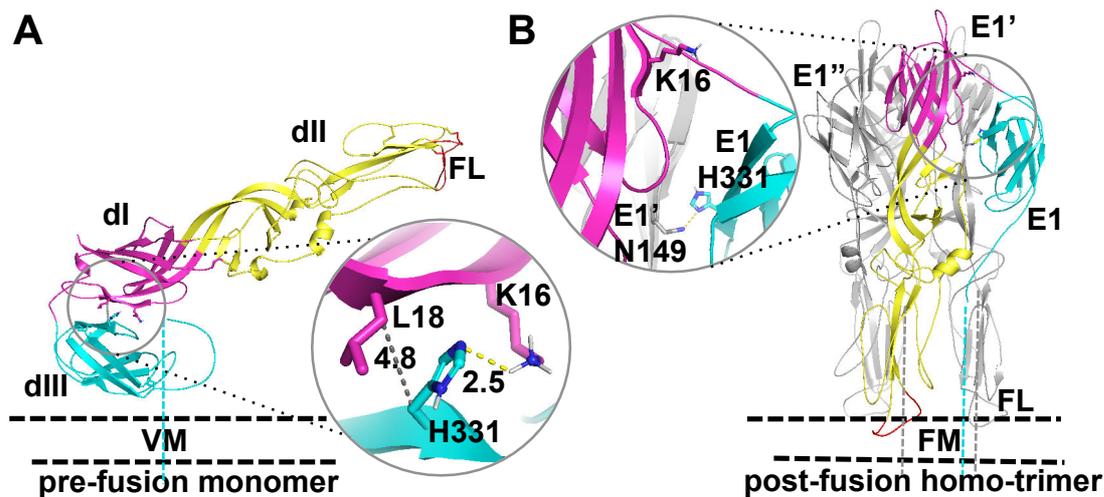


Figure 3.3 Structure analysis of pre-fusion (PDB: 3N42). (A), and post-fusion (homology model) (B) domain I-III stapling interactions by His331 in CHIKV E1. In (A), hydrogen bond between His331 and Lys16 in pre-fusion conformation is shown with yellow dotted line. C_b-C_b distance between His331 and Leu18 is shown by grey dotted line. Distance units are in Å.

We then compared the dI-III linker region of the $t=23.37$ ns cluster central structure to that of the pre-fusion conformation. We noted a rotational movement of the dI-III linker compared to pre-fusion structure. Though there is no major change in the spring-twisted region of the linker, the spring twist is slightly opened up compared to the pre-fusion conformation. This resulted in movement of the Arg₂₈₉ side chain slightly away from Asp₂₉₂. In this orientation the Arg₂₈₉ cannot form a salt bridge with Asp₂₉₂. We also noted

that, when we align the $t=23.37$ ns structure over pre-fusion structure, the domain I-II appeared slightly rotated over the dI-III linker. Since, domain III is anchored into the membrane through stem and TM region, we assume that the dIII does not move during the conformational change. In the domain III aligned structures (figure 3.2D), it is apparent that the dI-II region has swivelled out by $\sim 22^\circ$. The center of mass of domain I-II region is displaced by ~ 20 Å in $t=23.37$ ns structure compared to pre-fusion structure (figure 3.2D). Change in ϕ , ψ , and ω angles between R₂₈₉VVD₂₉₂ residues of the linker region, between pre-fusion and $t=23.37$ ns cluster central structure (comparative table shown in table 3.1) also indicated untwisting of the spring-twisted region in acidic pH.

Table 3.1. Comparison of dihedral angles of “RVVD” residues from the dI-dIII linker region between $t=0$ ns and $t=23.37$ ns structures from MD simulation studies.

$\text{N} \xrightarrow{\phi} \text{C}_\alpha \xrightarrow{\psi} \text{C} \xrightarrow{\omega} \text{N}$				
Dihedrals	Structure	Phi (ϕ)	Psi (ψ)	Omega (ω)
R289	$t=0$ ns	-64.7	150.5	-179.9
	$t=23.37$ ns	-53.0	139.2	-167.0
V290	$t=0$ ns	-55.9	-38.7	178.7
	$t=23.37$ ns	-70.2	-25	166.5
V291	$t=0$ ns	-70.2	-40.2	179
	$t=23.37$ ns	-79.6	-5.3	-162.8
D292	$t=0$ ns	-72.6	-18.9	179
	$t=23.37$ ns	-120.4	-40.1	164.1

Based on our MD analysis, we propose that untwisting of the spring-twisted region in the dI-III linker is the first step in acidic pH-induced conformational changes of alphavirus E1. This would initiate the swivelling motion of domain I-II to the center of the spike. Breaking of the domain I-III stapling interaction, between His₃₃₁ and Lys₁₆, may be a later downstream step in the mechanism.

3.3.4 Purified mutant ectodomain proteins are functionally active in membrane association and membrane fusion assays

To validate our MD simulation predictions on E1 conformational changes at fusion-triggering acidic pH, we purified the soluble ectodomain of E1 H331D/E mutants following the same protocol used to purify the WT protein (refer to chapter 1). The purity and the far-UV CD /intrinsic tryptophan fluorescence spectra analysis of both these mutants were comparable to the WT protein (Figure 3.4A-C).

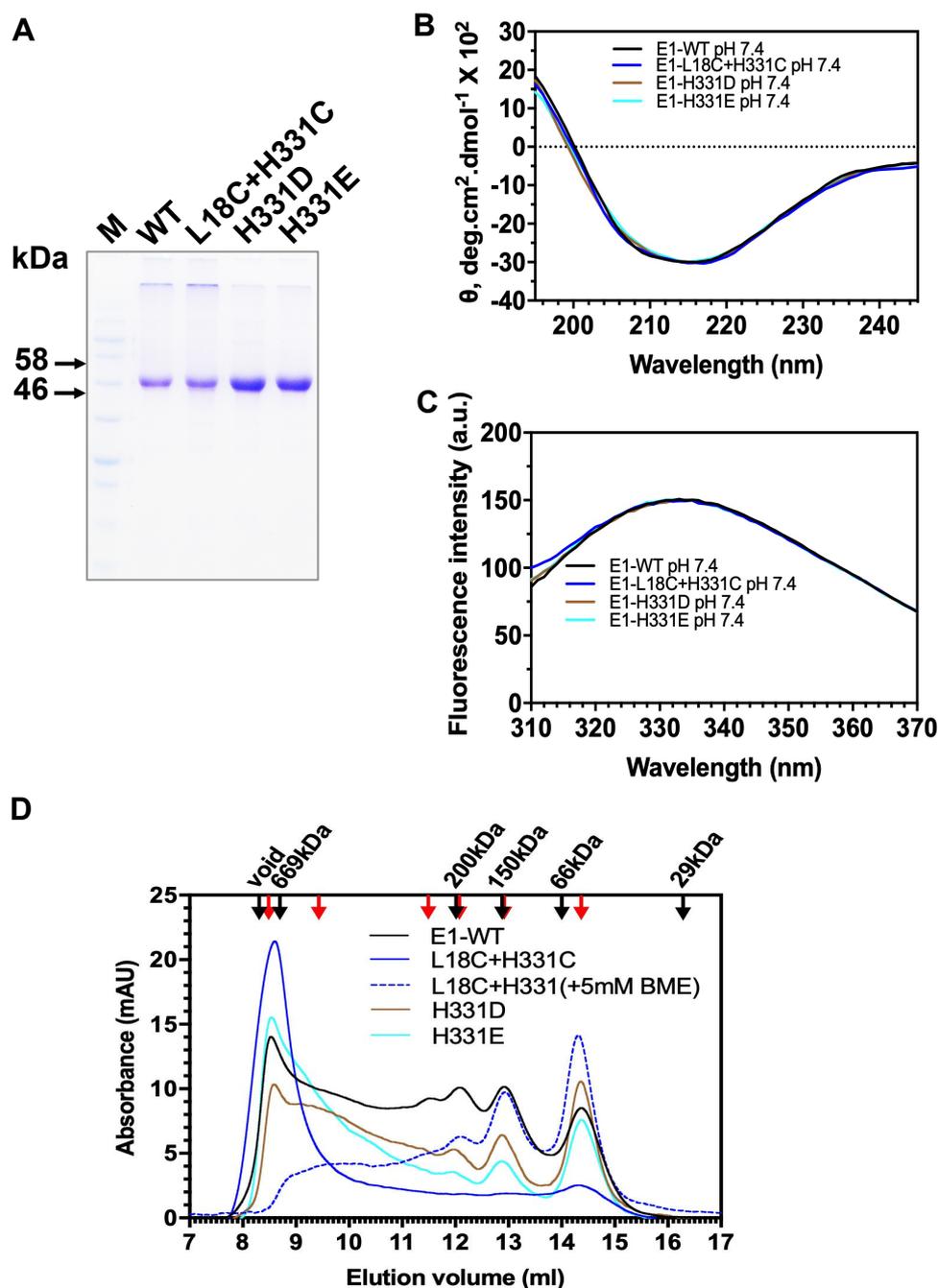


Figure 3.4 Purification and characterization of E1 dI-dIII interface mutants. (A) SDS-PAGE gel image of purified WT and mutant protein samples. (B) Far-UV CD spectra analysis on purified E1 WT and mutant protein samples. (C) Intrinsic tryptophan fluorescence spectra analysis on purified E1 WT and mutant protein samples. (D) Quaternary structure analyses of CHIKV E1 WT and mutant proteins. Gel filtration profiles of all proteins are shown with color-coded lines for respective protein samples. Positions of the gel filtration standards are marked with down faced arrows above the chromatogram.

We set out to test these mutants in a liposome fusion assay. Our reasoning is that H331D/E mutant E1 should show a higher threshold of acidic pH (lower than pH 5.5), as salt bridge between the introduced Asp (in place of H₃₃₁) and Lys16 should not be broken at pH 5.5. We also prepared a double mutant of E1 protein where His₃₃₁ and Leu₁₈ are changed to cysteines. The rationale of creating the double cysteine mutant is described in a later section. We introduced a pair of cysteines: one replacing His₃₃₁ and another in domain I, replacing Leu₁₈. We chose Leu₁₈ for introducing cysteine in domain I, as C β -C β distance between His₃₃₁ and Leu₁₈ is 4.8 Å (refer to figure 3.3A). This C β -C β distance is taken as optimal for introducing cysteine disulphide bond (191). We refer to this double cysteine mutant as cystine staple mutant from here onwards. In the cystine staple E1 scenario, acidic pH should be able to trigger membrane fusion between the liposomes (as His₃₃₁ is not an acid sensing residue in this scenario), but only when cystine bond is reduced by a reducing agent.

We purified the double cysteine mutant (Figure 3.4A) in the same way as the WT protein (detailed in chapter 1). To ensure that this purified mutant protein is in its native conformation and functionally active we compared its purity and secondary and tertiary structural properties to that of the WT protein. Briefly, the far-UV CD spectrum of cystine staple mutant, overlapped with that of E1 (figure 3.4B). Intrinsic tryptophan fluorescence emission spectrum of E1 WT and mutants (figure 3.3C) showed a peak at 333 nm, which implied a properly folded conformation of purified E1. We also confirmed that the introduced cysteines in E1 formed a disulphide bond, as expected. Ellman's reagent reactivity, to estimate free sulfhydryl groups in the protein, was comparable between wildtype E1 and the cystine staple mutant (table 3.2).

Table 3.2 Ellman's assay results with purified CHIKV E1WT and cysteine staple mutant proteins. The 412nm absorbance values of E1 WT and H331C+L18C mutant proteins and the buffer (TNE pH 7.4) control from the Ellman's assay, which tests for free sulfhydryl groups in proteins, are listed.

	TNE buffer	E1-WT	Purified and refolded E1 H331C+L18 C
Absorbance@412nm	0.14	0.14	0.14

We also analysed the quaternary structure of all the purified E1 mutants and compared those to the WT protein (Figure 3.4D). The gel-filtration profiles of H33D/E mutants were comparable to the gel-filtration profile of the WT protein. However, the double cysteine mutant mostly formed higher order oligomer (most of the protein is in the fractions corresponding to the first peak of the chromatograph) (Figure 3.4D). However, when 5mM β -ME was added to the elution buffer, the double cysteine mutant elution profile closely matched with that of the WT protein.

We then used the purified E1 and the double cysteine mutant, in a liposome co-floatation assay (Figure 3.5A and B). Earlier studies by others (174, 180, 192) have shown that E1 protein, either purified in insect cells or mammalian cells or reconstituted into lipid vesicles from viral membranes, associates with lipid membranes. This is taken as a measure of the protein's fusion function. The fusion loops of E1 would be exposed when E1 is expressed alone. These fusion loops can insert into membranes. Nearly 4-8% of total E1 protein (both WT and mutants) floated along with LUVs at neutral pH (pH 7.4).

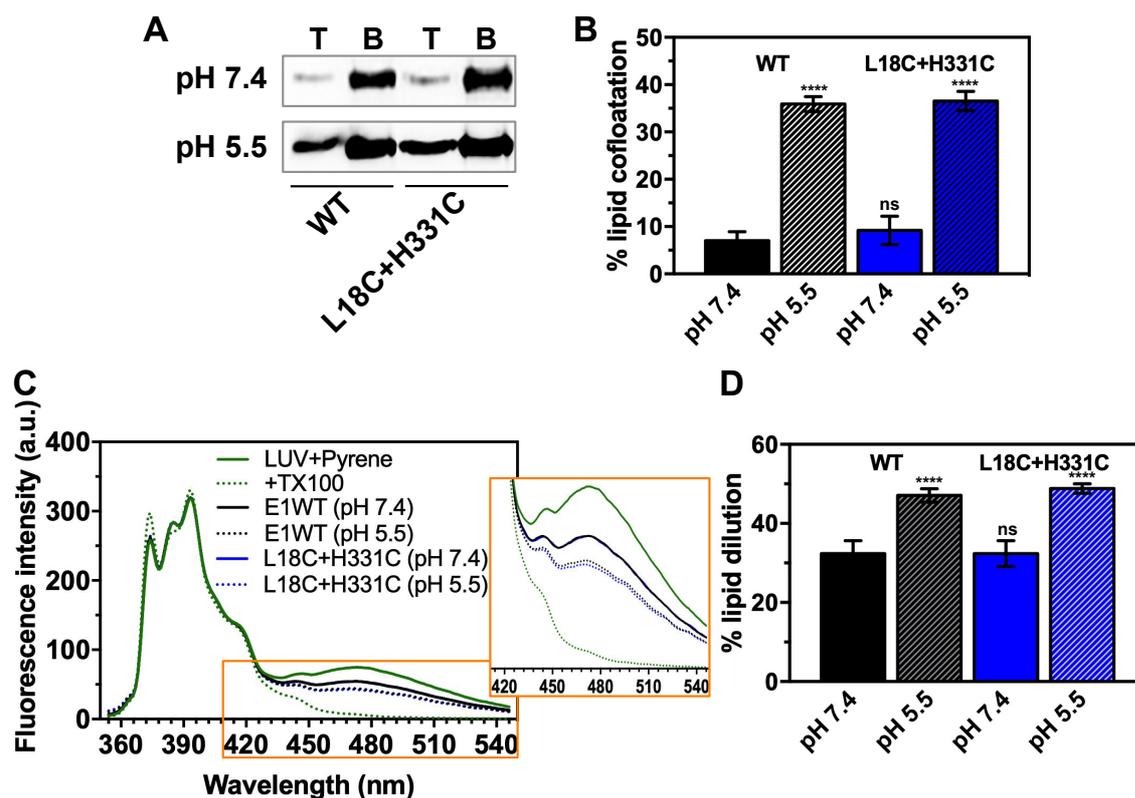


Figure 3.5 Functional characterization of E1 WT and cystine staple mutant in liposome co-floation and pyrene excimer fluorescence assays. (A) Image of Western blots for E1 WT and cystine staple mutant protein in top (labelled ‘T’) and bottom (labelled ‘B’) fractions from the lipid-floation assay. (B) Fraction of E1 co-floating with LUVs (average of three experiments) is presented as a bar graph. (C) Pyrene emission spectrum in presence of E1 WT or cystine staple mutant at pH 7.4 and pH 5.5. LUV+pyrene (green solid line) acted as control (zero value of lipid dilution because of E1 FL insertion). Spectrum with pyrene incorporated LUVs + TritonX100 (TX100) (green dotted line) is used for calculating 100% value of lipid dilution. E1 WT (pH 7.4) (solid black line) and E1 WT (pH 5.5) (dotted black line) is one comparison set, while spectra with E1 cystine staple mutant (L18C+H331C) at pH 7.4 (solid blue line) and pH 5.5 (dotted blue line) is another comparison set. Pyrene fluorescence emission spectrum was recorded in 350-550nm wavelength range. Excitation wavelength was set to 340nm and excitation and emission slit widths were set to 5nm. Each spectrum was an average of three readings and blank corrected. Emission spectra were normalized by setting constant the pyrene emission monomer peak at 397nm. Inset shows zoomed in picture of the spectra region between 420 – 540 nm. (D). Effect on lipid membrane due to E1 association, as percent dilution of lipids, was estimated by measuring drop in pyrene excimer peak values at ~475nm, in presence of WT or mutant. Average values for each sample from three experiments is plotted in the bar graph along with error bars. One-way ANOVA was used to calculate statistical significance – **** represent $p \leq 0.0001$ level of significance and ‘ns’ is non-significance.

However, when the protein and LUV mixture was incubated in a pH 5.5 buffer, membrane association increased to ~32-37%. A similar observation was done with SFV or SINV E1 association with LUVs. In co-floatation experiment, more E1 co-floated with LUVs when the protein and LUV mixture was incubated at pH 5.5 (180, 193). Ectodomains of Tick borne encephalitis virus E protein trimerizes on membrane surface and co-floats with LUVs (194).

In order to rule out the possibility that the observed LUV co-floatation is because of some exposed hydrophobic patch on the E1 protein, we performed a lipid membrane destabilization assay. Pyrene excimer fluorescence peak intensity decreased in presence of E1 protein (figure 3.5C and inset). And, this effect is more pronounced when the protein is pre-exposed to fusion-triggering pH (pH 5.5) (figure 3.5D). A similar result was noted when the cystine staple mutant of E1 was used in the experiment (figure 3.5D). This suggests that LUV association of E1 is through insertion of its fusion loops into LUV membrane. Similar observations were made with other alphavirus fusion proteins and LUV association (193, 195). We observed a similar behaviour for H331D and H331E mutant of E1 in LUV co-floatation assay and pyrene excimer fluorescence assays (figure 3.6 A-D).

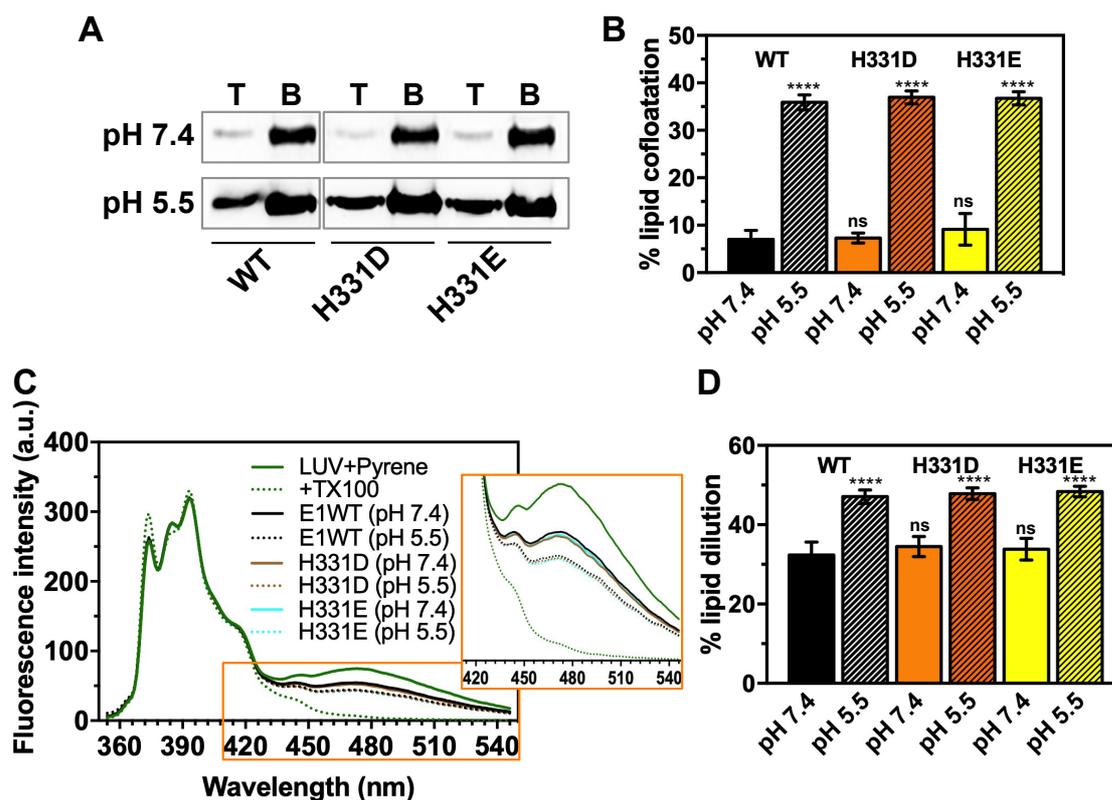


Figure 3.6 Analyses of membrane interaction properties of CHIKV E1 WT and H331D/E mutant proteins in lipid-floatation, and pyrene excimer fluorescence assay. (A) Images of western blot (WB) analysis on E1 WT and H331D/E mutant protein samples from lipid-floatation assay at pH 7.4 (top lane) and pH 5.5 (bottom lane). (B) Quantitative analysis of WB results from lipid-floatation assay at pH 7.4 (solid bars) and pH 5.5 (dashed bars). Average values from three experiments with error bars are shown for each sample. (C) Pyrene fluorescence spectra analysis on E1 WT and H331D/E mutant protein samples from lipid-association assay at pH 7.4 (solid lines) and pH 5.5 (dashed lines). Zoomed in spectra shows changes in the pyrene excimer peak region. (D) Quantitative analysis on pyrene fluorescence from the lipid-association assay at pH 7.4 (solid bars) and pH 5.5 (dashed bars). Average values for each sample from three experiments with error bars are shown. One-way ANOVA was used to calculate statistical significance – **** represent $p \leq 0.0001$ level of significance and ‘ns’ is non-significance.

We confirmed that the E1 and the cysteine staple mutant proteins are not unfolded at the fusion triggering pH by recording far-UV CD spectra in pH 5.5 buffer. The spectra at pH 5.5 almost overlapped with those recorded in pH 7.4 buffers (figure 3.7).

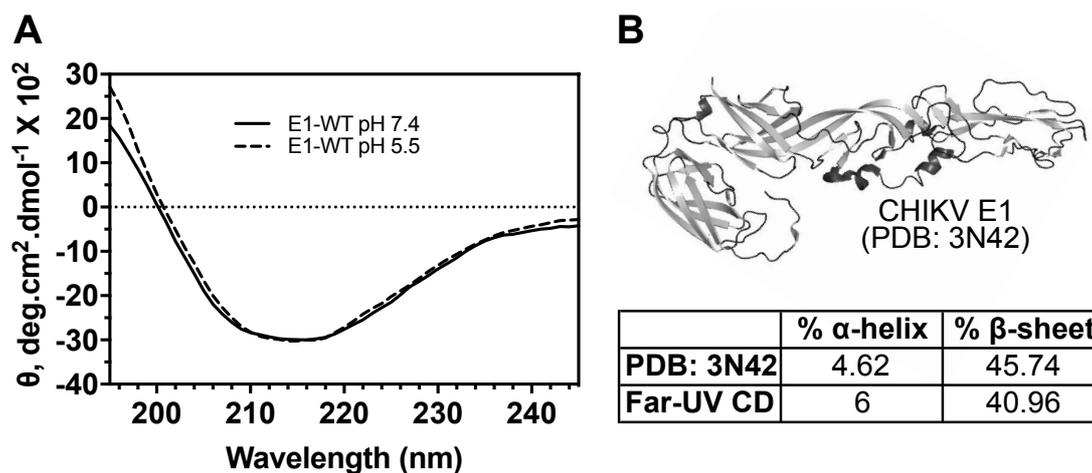


Figure 3.7 Comparison of CHIKV E1 percent secondary structural contents predicted from far-UV CD data and calculated from crystal structure. (A) Far-UV CD spectra for *E. coli* purified CHIKV E1 WT at pH 7.4 and pH 5.5. (B) Top - Crystal structure of CHIKV E1 (PDB: 3N42), β -sheets are shown in gray and α -helices in deep-gray colors. Bottom - Comparison of percent α -helical and β -sheet contents from the crystal structure and from the far-UV CD data of E1 taken at pH 7.4.

Taken together, our structure analysis and membrane association studies establish that the CHIKV E1 protein and the mutant proteins that we expressed in *E. coli* are structurally and functionally similar to insect cell-expressed and viral surface-extracted E1.

3.3.5 Domain I and III separation at acidic pH is essential for fusion activity of E1, and domain I-III interface breaking is not triggered at the conserved His331

Our MD simulations did not show any major change in the conformation surrounding the His331, at least within the time scale of the simulation study (50ns). This suggests that at pH 5.5, His₃₃₁-Lys₁₆ interaction in the dI-III interface is retained. The domain I-III separation in acidic pH might start at a place other than His₃₃₁-Lys₁₆ interaction in the dI-III interface. In this scenario, replacing His₃₃₁ with another residue, should not affect acidic pH triggering of E1 for fusion function. However, in the cystine staple mutant,

fusion function should be seen only when the disulphide bond between domain I and III is broken by a reducing agent.

To test this idea, we used WT and cystine staple mutant of E1 in a liposome fusion assay. Our fusion assay is designed such that when E1 on the coated LUVs performs fusion with the fluorescein-loaded LUVs, fluorescein fluorescence is de-quenched due to dilution effect (refer to schematic in figure 3.8A). In the fusion assay, WT E1 showed 41% fusion activity. Whereas when the domain I and III-stapled (cystine staple mutant) E1 is used, only 15% fusion activity was seen (figure 3.8B, green line and green bar in figure 3.8C).

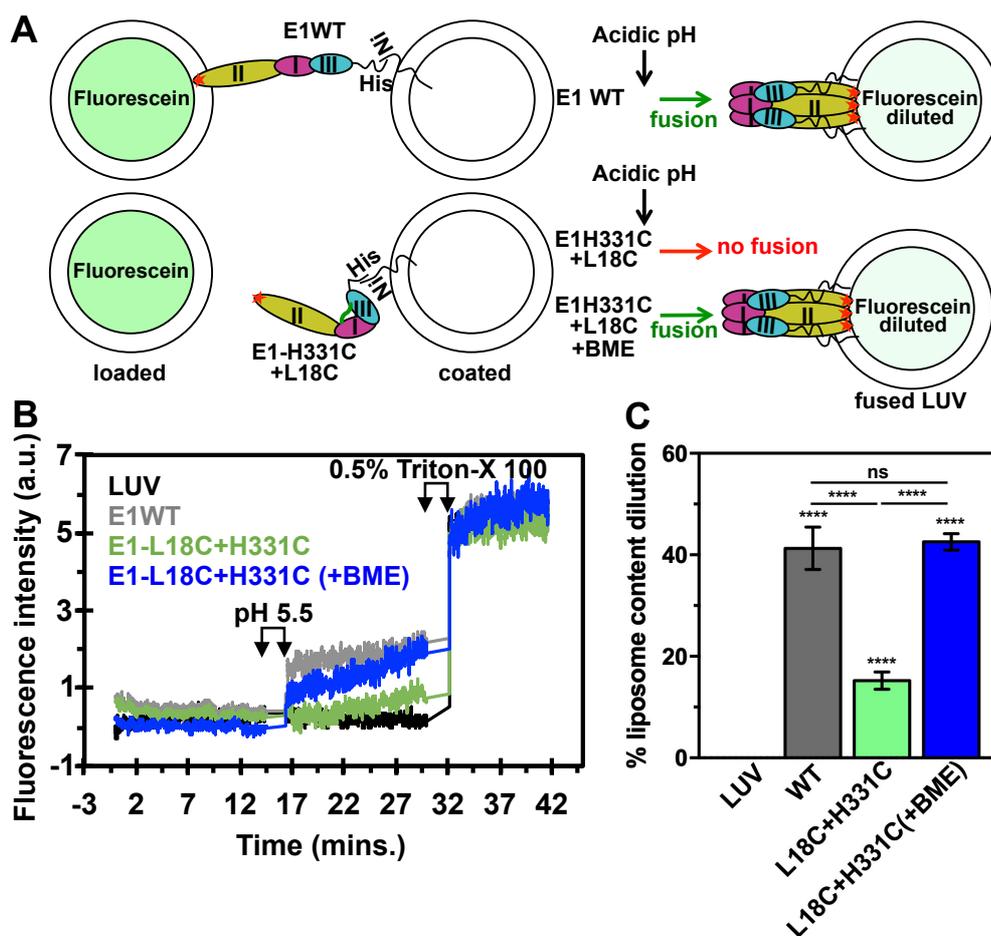


Figure 3.8 Liposome fusion assay for E1. (A) Schematic explaining fusion assay used in this study. Top panel: fluorescein-loaded and E1-coated LUVs are mixed in pH 7.4 buffer and fluorescence intensity at 520 nm (emission maximum for fluorescein) is recorded as a time-scan. Excitation wavelength was set to 490 nm. The excitation and emission slit widths were set to 5nm. After first 15 minutes of recording the fluorescence intensity, calculated amount of 1M sodium acetate (say, for 1 ml assay mixture 15 mL of the sodium acetate was added to bring the pH down to 5.5) is added to the assay mixture. After that, fluorescence intensity was recorded for another 15 minutes, before adding TritonX-100 to a concentration of 0.5% v/v. After adding TritonX-100, fluorescence intensity was recorded for another 10 minutes. Bottom panel: the assay was performed exactly the same as above, except that either b-mercaptoethanol was added before the acidification step or not. (B) Fluorescence time-scan in presence of LUV alone (black), E1 WT (gray), E1 cystine staple mutant (E1-L18C+H331C) (green) and E1 cystine staple mutant (E1-L18C+H331C) with b-mercaptoethanol (plus β -ME) added (blue). (C) percent liposome content dilution was estimated as described in Material and Methods and average values from three experiments is plotted in the bar graph along with error bars. One-way ANOVA was used to calculate statistical significance – **** represent $p \leq 0.0001$ level of significance and ‘ns’ is non-significance.

In another assay with cystine staple mutant, β -ME was added, before acidification of the assay mixture. The fusion activity recorded in this case (figure 3.8B, blue line and blue bar in figure 3.8C) is comparable to the WT protein. We confirmed that lack of fusion in cystine staple mutant (without β -ME) is not because of inability to insert its fusion loops into target membrane. Cystine staple E1 mutant associated with membrane vesicles and performed equal to the WT protein in membrane destabilization assay (figure 3.5B and D). These results indicate that domain I-III separation is triggered at pH 5.5, which is a required conformational change step for fusion function of E1. And, this separation step does not get triggered by acid-sensing of the conserved His₃₃₁.

Further corroborating this idea, H₃₃₁D/E mutants of E1 fusion activity is comparable to the wildtype E1, in the fusion assay (figure 3.9). If His₃₃₁ were to be an acid-sensing residue, replacing it with an acidic residue (D or E) should have increased the acidic pH threshold for fusion triggering – would have required more acidic pH (lower than 5.5) for fusion function.

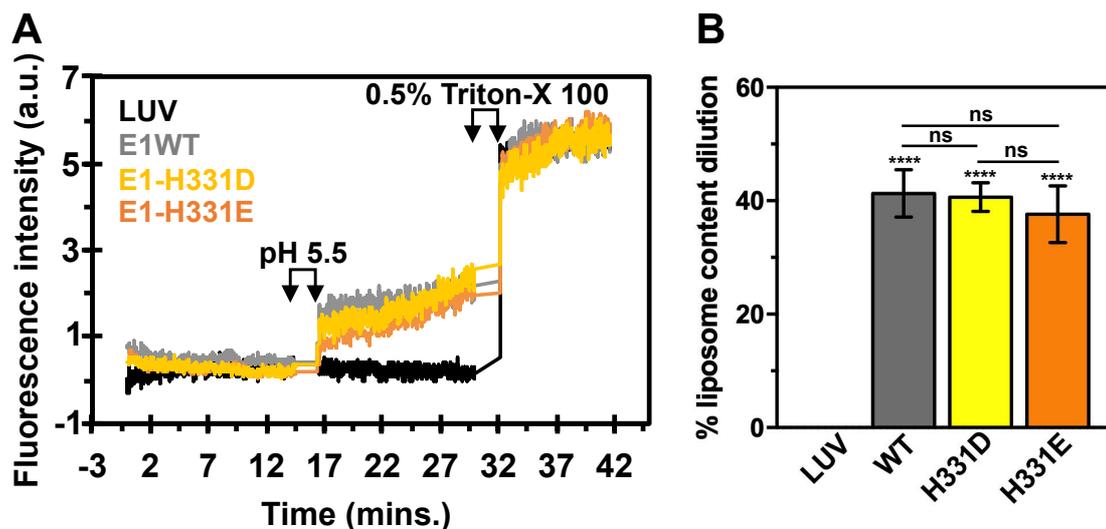


Figure 3.9 Analyses of membrane fusion by CHIKV E1 WT and H331D/E mutant proteins in lipid-mixing assay. (A) Results of the fusion properties of E1 WT and H331D/E mutant proteins from the fluorescence based kinetic assay. (B) Quantitative analyses of the extent of fusion by E1 WT and H331D/E mutant proteins. Average values for each sample from three experiments with error bars are shown. One-way ANOVA was used to calculate statistical significance – **** represent $p \leq 0.0001$ level of significance and ‘ns’ is non-significance.

DISCUSSION:

Viral fusion protein undergoes a pre-fusion hairpin conformation to a post-fusion conformation change as it performs viral and cell membrane fusion. An extended conformation intermediate is formed during the pre- to post-fusion conformation change. In alphaviruses and other viruses that take endocytic pathway for cell entry, the pre- to post-fusion conformation change is triggered by acidic pH of the maturing endosome. In pre- to post-fusion conformation change two major changes are seen: 1. E1 moves from the periphery, to the centre of the E1-E2 heterodimer trimeric spike and forms a homo-trimer. 2. Inter-domain interactions in the interface of dI and III are broken and new interactions

are formed, post fusion. How these conformational changes are brought about in E1, after exposure to fusion-triggering acidic pH are addressed in this study.

In the pre-fusion conformation, each E1 protomer in the trimeric spike structure is complexed with the receptor binding protein, E2. Post-fusion, E1 forms a homo-trimer with dI-dII regions at the center of E1 homo-trimer vertical axis. So, in the pre- to post-fusion conformation change, E1 dI-II region of each protomer must have moved in towards the center of the trimeric spike, replacing E2. A swivelling motion of domains for E1 in pre- to post-fusion conformation change was proposed (173, 184–186). The only region of E1 that is relatively flexible, and thus can act as a hinge for swivelling of domains, is the linker loop connecting the domain I, and III. Our analysis showed that a highly conserved R-(X)₂₋₃-D/E stretch in the linker sequence might have a role. Interestingly, in the pre-fusion structure of E1, backbone of the region spanning the R-(X)₂₋₃-D/E is twisted like a spring, stabilized by a salt bridge between R and D/E of the sequence. We reasoned that untwisting of the spring-twist region in the dI-III linker would swivel the dI-II, like a hinge, on dIII. Findings from our MD simulation studies on CHIKV E1 structure, at pH 5.5, corroborated our hypothesis. Zheng et al (179) in their studies, to understand acid pH-induced conformational changes in SFV E1, also purported the idea that Arg₂₈₉ in dI-III linker is a key residue for homo-trimer formation of E1.

Parallel to the untwisting of spring-twisted region, the domain I-III stapling interactions should also break to allow swivelling motion of the domains. Several charge-charge interactions between domains I, and III hold both the domains in the hairpin conformation. In TBE E, a hydrogen bond between a conserved histidine (structurally in the same context as His₃₃₁ of CHIKV) and a lysine in domain I is a domain stapling

interaction and would break in the acidic pH (190). We expected H331-K16 hydrogen bond in CHIKV (188) E1 would play a similar role in pre- to post-fusion conformation switch. Protonation changes in acidic pH on acid-sensing histidine residues has been proposed as a trigger for pre- to post-fusion conformational changes for several other viral fusion proteins (196). We expected that the His₃₃₁ protonation would change when E1 structure was titrated to pH 5.5, and the H331-K16 hydrogen bond would break. Contrary to our expectation, our MD simulation study hinted the His₃₃₁, unlike in flavivirus E, is not an acid-sensing residue, at least in the time scale of our MD simulation. Our observations are corroborated by those of Qin et al (197). In a similar MD simulation study on CHIKV E1, Zeng et al (188) did estimate change in pKa of His₃₃₁ from monomeric (pre-fusion) state to, fusion intermediate, to post-fusion conformation. They also noted that, in pre-fusion state and extended intermediate (acid-pH triggered, intermediate conformation) His₃₃₁ has a pKa of 4.8 – a value very close to the pKa that we estimated for His₃₃₁ in pre-fusion and acid-pH titrated state. However, in post-fusion conformation, pKa of His₃₃₁ increased to 8.6. This implied that His₃₃₁ environment change happens during extended intermediate to post-fusion conformation change.

Results of our fusion functional assays using purified E1 protein, H331D/E mutants and cystine staple mutants validated our MD simulation predictions on His₃₃₁. Our observations with cystine staple mutant also prove that domain I-III dissociation is an essential step in the acid-pH induced conformational changes required for membrane fusion.

Based on interpretation of our results from this study, in the context of the universal mechanism proposed for viral fusion proteins, we present a mechanistic model to explain

acidic pH-triggered conformational changes in alphavirus membrane fusion protein, E1 (figure 3.10). E2-E1 dissociation is initiated after receptor binding on to E2 (181). The first conformational change that is triggered in E1 (after the endosome gets acidified to pH 5.5) is in the dI-III linker. Acidic pH triggers breaking of salt bridge between the conserved R-(X)₂₋₃-D/E pair in the linker region. This results in untwisting of the spring-twist structure. Because of the untwisting, dI-II region swivels outwards from the trimeric spike. Thus, spring-twisted structure is the hinge point for the domain swivelling motion.

In parallel, acidic pH also triggers domain I-III separation. And, the domain I-III separation does not start with the conserved His₃₃₁-Lys₁₆ hydrogen bond breaking. Our observation is in agreement with that of Qin et al (197). In a virus mediated fusion assay, they have shown that E1 H331A mutant of SFV is not defective in growth properties compared to WT E1 containing virus. It is possible that, despite the structural similarity, alphavirus fusion protein, E1, has a different mechanism of acid pH-induced pre- to post-fusion conformation switch from that of flavivirus E protein. In alphaviruses, dI-III separation may be triggered at another location in the interface. One possibility is, another highly conserved histidine residue near to the interface, His₃ from domain I. Interestingly, H3A mutation in E1 of SFV resulted in reduced viral infection and change in pH threshold (197). Zeng et al (188) also suggested His₃ (along with His 331) as an acid-sensing residue in CHIKV E1. It would be interesting to see how His₃ interactions change at acidic pH, and how that would trigger dI-III dissociation.

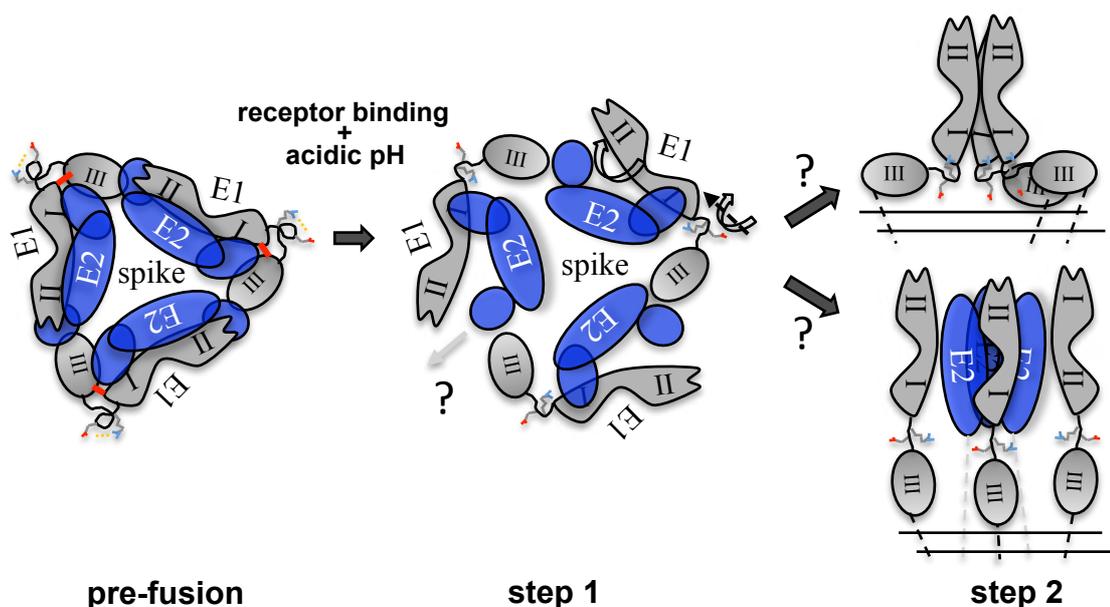


Figure 3.10 Mechanism of acid pH-induced conformational change in E1 during viral entry. Alphavirus fusion protein, E1, undergoes pre- to post-fusion conformational change, through an extended intermediate, during membrane fusion step of viral entry. Three E2-E1 hetero-dimers, curl around a central vertical axis to form a spike structure on virion surface (pre-fusion). **Step1:** E2-E1 dissociation and E1 domain swiveling. Receptor binding and acidic pH trigger this step. Acidic pH breaks a conserved arg-asp/glu salt bridge in the dI-III linker of E1 and triggers untwisting of a spring-twisted region of the linker. This brings about swiveling motion of dI-II region over dIII. As a result, E1 moves away from the spike axis. The dI-III stapling interactions also break in this step. E2 may escape out from the spike center at this stage. **Step2:** possibility 1 (top) - extended intermediate and E1 homo-trimer formation. Swiveling-in motion of dI-II regions may bring these regions from three protomers to form the central core of the post-fusion homo-trimer. E1 domains may extend linearly following this. Alternate possibility (below) – swiveling motion will also extend the dI-III linker, to form three monomeric extended intermediate E1 protomers first. Trimerization, through dI-II region may follow at a later step. E2 may escape from the center at that stage. In pre-fusion and step 1, top view of the spike is shown in the schematic. In step 2, side view of the spike schematic is shown for clarity. E2 is shown with brown ellipses and E1 in gray shapes. The dI-III linker is shown in hand-drawn curve shape. The spring-twisted region is depicted with a spring circle shape. Yellow-dotted lines show salt bridge between the conserved Arg-Asp/Glu pair in the dI-III linker. In pre-fusion structure, a thick red line shows dI-III stapling charge interactions. In step 2 schematic, double parallel lines below the spike shape is viral membrane.

What follows after the domain swivelling motion is triggered on the spring-twist region and dI-III interactions are broken? There are two possible scenarios: 1. By swivelling-in, domain I-II region of the three E1 protomers in the trimeric spike come to the centre of the spike, establish inter-protomer interactions and form the E1 homo-trimer central core. 2. Alternatively, the dI-III linker region becomes linear (as the dI-III hairpin-stapling interactions are broken) extending all three domains on to a linear axis. Homo-trimerization may follow this step. Earlier studies propose both the possibilities. Claudia et al (180) showed that when only dI-II region of SFV E1 is expressed, replacing dIII and the dI-III linker with a short flexible region, acidic pH triggered formation of the dI-II region trimers. This implies that homo-trimerization precedes extended intermediate formation (top, in step II, figure 3.10). In contrast to this, Cao et al (177) proposed a mechanism where E1 forms an extended intermediate as a monomer, implying that homo-trimerization is a later event in membrane fusion (bottom, in step II, figure 3.10).

Taken together, our study gives first mechanistic insight on acid-pH conformational changes in E1. We explained role of a conserved arginine – aspartate/glutamate pair in the dI-III linker, in maintaining a spring-twisted backbone conformation in pre-fusion conformation of E1. We also showed that, in acidic-pH, salt bridge between the conserved Arg-Asp/Glu pair breaks and untwists the spring-twisted region. This results in domain I-II swivelling motion. Explaining different steps in the mechanism of E1-mediated membrane fusion will help understand structure-function relationship of class II fusion proteins, in general. And, also will help in devising strategies to prevent viral entry.

CHAPTER 4

***CHARACTERIZATION OF CHIKV E3-MEMBRANE
INTERACTIONS TO EXPLAIN ROLE OF E3 IN
REGULATION OF ENTRY***

Acknowledgement: I thank Mr. Satyamurthy Kundarapu for his help in cloning and expression of the E3 F15W mutant protein.

4.1 INTRODUCTION:

E3 is the smallest of the three alphavirus envelope proteins. E3 is the resultant smaller product of proteolytic processing of p62 protein. E3 has a chaperoning function proper folding and maturation of E2 and E1 into heterodimer conformation (79, 83, 198). During maturation of these proteins, E3 guides E2-E1 folding by likely anchoring into ER membrane and then later retracting itself as a result of both capsid autocleavage and E3 glycosylation at N12 site. After retrieval from the ER membrane, E3 becomes the new N-terminus for the p62-E1 polypeptide. After maturation, in trans-golgi network, E3 interacts with E2 acid sensitive region (ASR), and held at the apex of the trimeric spike.

As discussed in introduction chapter, E3 is retained on virion surface in some alphaviruses (82, 90, 91, 199) and not in others (86, 87). In chikungunya virus E3 is retained. E3 is held on virion surface (where it is retained) through its interactions with E2. On the trimeric spike E3 is at the most apical site (detailed organization of trimeric spike presented in General introduction). Crystal structure of E3-E2-E1 (PDB: 3N42) shows that E3 is positioned like a brace at the apex of the structure, covering domain A and B of E2. Core of E3 structure is like a horseshoe of three α helices, and all three helices of the core make several contacts with E2 at one face of acid sensitive region (ASR) - E1 interacts with E2 ASR on the opposite side.

Role of E3 in biogenesis of alphavirus envelope proteins and its role in ‘protecting’ E1 from pre-mature fusion function activation by the acidic pH of the trans-Golgi network are thoroughly studied. However, E3 role in viral entry is not known. It is possible that E3’s function in viral entry, if any, may be virus species specific. Studies where one alphavirus E3 sequence was replaced with E3 sequence from other alphaviruses resulted in formation

of reduced infectious particles and defective spike arrangements, suggesting additional role of E3 than the presumed signal peptide function (200). Also, in alphaviruses where E3 is not retained on mature virion surface, mutation in E3 furin cleavage site, thus retaining E3 on the mutant virus surface, affected fusion function (required more acidic pH than the optimum pH threshold for fusion) (85, 87, 199). These observations suggest that though E3's association with E2 in secretory pathway prevents premature fusion activation by E1, once exported to outside of the cells its release from E2-E1 complex is necessary for priming the virus fusion competent.

E3 'brace' release from E2, to unlock domain A-B 'clamp' surrounding the E1 fusion loops is suggested as an important regulatory mechanism by Voss et al (82), from analysis of the crystal structure of E3-E2-E1. Unless E3 'brace' is removed, the fusion loop unmasking is not possible. Thus, biochemical and mutant studied discussed above, and crystal structure analysis, strongly suggest that E3 has a regulatory role in viral entry (at least in those alphaviruses where E3 is retained on surface), other than the chaperoning function during biogenesis of the proteins. Dissociation of E3 from E2, as it appears, is an important step for fusion function of E1.

Now, the obvious question in the context of viral entry regulation is what triggers E3-E2 dissociation during Chikungunya virus entry? There are two possibilities: i. N-terminal region of E3 interacts with membrane, resulting in conformational change in E3, and eventual E3-E2 interaction loss. ii. Acidic pH of the endosome (acidic pH triggering of E3-E2 interface as a result of transition from extracellular neutral to endosomal fusion triggering acidic pH), breaks interactions between E3 and E2, resulting in E3 release. My rationale for the above stated hypothesis is as following: The N-terminal few residues of

E3 (that might have inserted into ER membrane during biogenesis) may interact with cell membrane (or endosome membrane) during viral entry because of its apical positioning in spike (would be close to target membrane). And, this might result in change in conformation of E3, eventually leading to E3-E2 interaction loss. Rationale for the other possibility (acidic pH triggering) is that E3 interacts with E2 ASR, which would get disordered during acidic pH trigger, several charge-charge contacts between E3 and E2 may change upon acidification. For example, in the neutral pH structure of CHIKV E3-E2-E1 complex, E3 Y47 forms a hydrogen bond with E2/K254; E3 Y48 forms a hydrogen bond with E2 E166 and a π - π interaction with E2 H256. In addition to these, E3 E39 and D40 form salt bridges with E2 R251. All these neutral pH interactions of E3-E2 interface could possibly be disrupted by endosomal acidic pH during cell entry. Because, these interactions involve a highly conserved acid-sensing residue (E2 H256) and other acid sensing charged residues (E3 - E39 and D40, and E2 - E166 and K254) from both E3 and E2, these residues may undergo protonation changes (and hence may trigger breaking of interactions involved) at acidic pH. So neutral to acidic pH trigger but not acidic to neutral pH, may possibly trigger breaking of E3-E2 interface and E3 release in alphaviruses (at least in those, which contain E3 as part of the mature virus particle structure). Results of from this study are presented in subsequent sections.

4.2 MATERIALS AND METHODS:

4.2.1 E3 sequence and structure analysis

We performed E3 sequence analysis using sequence information from LR2006_OPY1 strain of CHIKV and the details of the protein sequences (protein_id="ABD95938.1") used in this study can be found at <https://www.ncbi.nlm.nih.gov/protein/90654094/>. We used E3 sequence for predictions of hydrophobic amino acid residues by using ProtScale module in ExPASy server available at <https://web.expasy.org/protscale/>. We used different prediction algorithms by Kyte and Doolittle and by Hopp and Woods, with their default parameters, for predicting hydrophobicity of E3 amino acid residues. Amino acid hydrophobicity scales used by these two studies are published and summarized in following references (201, 202). We also studied membrane spanning region predictions on E3 sequence using default parameters in Tmpred server available at https://embnet.vital-it.ch/software/TMPRED_form.html. We also, analysed E3 structure (PDB: 3N41 alongwith E2 and E1) and E3-E2 interface interactions using this structure in PDBePISA server available at https://www.ebi.ac.uk/msd-srv/prot_int/cgi-bin/piserver.

4.2.2 Cloning, expression and purification of CHIKV E3 F15W mutant protein

The E3 F15W mutant used for this part of the study was expressed and purified exactly in the same manner as that of the E3 WT protein. The F15W mutant was generated by site-directed mutagenesis approach.

4.2.3 SDS-PAGE, chloroform staining, and Western blotting

15% SDS-PAGE gels were always used for analysing E3 WT and mutant proteins.

Coomassie brilliant blue staining was used for assessing purity of recombinant E3 proteins. The F15W mutant was verified using chloroform staining on SDS-PAGE gels ran with F15W protein samples and then detected under UV lights in a gel-doc system. We used this method to check the presence of the tryptophan residue introduced in place of phenylalanine in the E3 F15W mutant protein (we sequence verified the F15W construct also to verify the mutant clone). In presence of chloroform the protein bands under UV light illuminate as fluorescent bands because of addition of a formyl group to the indole ring of tryptophan (203). Since, E3 does not possess any tryptophan residues in its WT form, the F15W can be easily verified on SDS-PAGE gel by chloroform staining. Western immunoblotting with penta-HIS HRP conjugated antibody was used to confirm E3 presence in fractions from lipid co-floatation assays. Western blotting was performed using a standard protocol and enhanced chemiluminescence (ECL) substrate (purchased from Thermo Scientific™, MA, USA).

4.2.4 Far-UV CD and intrinsic tryptophan fluorescence analysis of E3 structural changes upon interactions with liposomes

In order to probe structural changes in E3 upon interaction with lipid LUVs and SUVs (as membrane mimetic models) we recorded far-UV CD and intrinsic tryptophan fluorescence spectra of the protein in presence and absence of liposomes. Experimental conditions and data processing and analysis methods for far-UV CD and fluorescence spectroscopy are the same as described for E1, E2 or E3-E2 proteins (refer to chapter 1). The samples for the spectroscopy were prepared in the following manner: concentration of E3 was kept constant at 0.2 mg/ml, while SUVs/LUVs concentrations were varied (molar excess of 20 and 50 were used in addition to 1:1 protein to liposome ratio). LUVs/SUVs

were prepared using individual lipids at PC: PE: PA: Cholesterol = 1: 1: 0.3: 2 molar ratios.

4.2.5 Size exclusion chromatography for studying E3-liposome interactions

We did gel filtration chromatography analysis on E3-liposome mixtures to check if E3 can form a stable interaction with the membranes. Approximately, 200 µg of E3 in TNE buffer either in presence or absence of liposomes was loaded on to Superdex G75 10/300 GL for the experiment. We used SUVs and LUVs at 50 molar excess to E3 protein for this experiment. The peak fractions for the purified E3 proteins and liposomes (in the void volume fractions) were then verified by WB analysis and by pyrene excimer fluorescence assay. Pyrene excimer fluorescence assay was described in material and methods section of chapter 1.

4.2.6 Lipid co-floatation assay for studying E3-liposome interactions

We used this assay to study E3-liposome interactions. We followed the same protocol described for studying E1-liposome co-floatation in chapter 1.

4.2.7 Liposome content release assay

We also studied E3-membrane insertion induced content release from a fluorophore loaded liposome population using fluorescence spectroscopy. Bacterially purified recombinant E3 proteins (at 0.2mg/ml concentration) in TNE pH 7.4 buffers were used for the assay in presence of 0.5mM fluorescein loaded LUVs. Finally, 0.5% Triton-X 100 detergent was added for liposome solubilization and complete dilution of fluorescein molecule. Excitation maximum for the experiment was set to 490nm and emission spectra was recorded from 500 to 550 nm wavelengths. All spectra represented in figures are accumulations of three spectral readings each and blank corrected.

4.3 RESULTS:

4.3.1 E3 sequence and structure analysis reveal key interactions at E3-E2 interface

To find out the possible mechanisms how E3 could be released from E2, we analysed both its sequence and structure. Multiple sequence alignment of E3 sequence from different alphaviruses, revealed a conserved ‘GY₄₇Y₄₈’ motif (Figure 4.1a). The conserved aromatic residues, Y47 and Y48, from this motif make critical π -cation or π - π interactions with E2 at the E3-E2 interface, in a pH dependent manner (83). PISA analysis of E3-E2 interface shows several charge-charge interactions between E3 and E2, as noted by others as well (82). Notable interactions are, E3 Y47 - E2 K254, E3 Y48 – E2 E166 and E2 H256, E3 E39 and D40 with E2 R251. Specifically, E3 Y47 forming hydrogen bond with E2 K254 and E3 Y48 forming a hydrogen bond with E2 E166 and a pi-pi interaction with E2 H256. In addition, a salt bridge between E3 D40 and E2 R251 also stabilizes E3-E2 interaction (this analysis was done using PDB: 3N42, CHIKV E3-E2-E1 crystallized at pH 7.0). The important interactions involving Y47, Y48 and E39, D40 from E3 side at the E3-E2 interface are shown in Figure 4.1b and listed in Table 4.1.

Tables 4.1. Key interactions at E3-E2 interface involve charged residues.

Interactions	Type	Distance
E3:Y47:OH-E2:K254:O	Hydrogen bond	2.99 Å
E3:Y48:OH-E2:E166:OE1	Hydrogen bond	2.72 Å
E3:Y48-E2:H256	π - π stacked	3.89 Å
E3:E39:OE1-R251:NH2	Electrostatic	3.86 Å

E3:E39:OE1-R251:NE	Hydrogen bond	3.32 Å
E3:D40:OD1-R251:NH2	Hydrogen bond; Electrostatic	2.76 Å
E3:D40:OD2-R251:NH1	Electrostatic	4.86 Å

4.3.2 Interaction of E3 with membranes

Rationale of studying interactions of E3 with LUVs and SUVs are: 1. E3 is present at the most apical part on the viral surface spike, hence is likely to contact the host membrane first during cell entry, 2. The N-terminus of E3 has an ER signal sequence and is predicted to insert into ER membrane during transport of the E3-E2-E1 complex into ER lumen during biogenesis. However, which sequence from at the N-terminus of E3 inserts into the ER membrane is not known. To test our hypothesis that E3-membrane interactions contribute to E3 release from E2, we analysed the sequence of E3 protein for presence of hydrophobic amino acid patches using different prediction algorithms to find putative membrane interacting stretches.

Our results from ProtScale analysis using multiple algorithms predicted a patch of ~15 amino acids at N-terminus of E3 (Figures 4.2a and 4.2b). Hopp & Woods algorithm evaluated E3 sequence based on hydrophilic and hydrophobicity properties, and predicted E3 N-terminal residues to be hydrophobic (Figure 4.3a). Kyte & Doolittle algorithm however based on a positive scoring pattern assigned the same E3 N-terminal residues to be hydrophobic (Figure 4.3b).

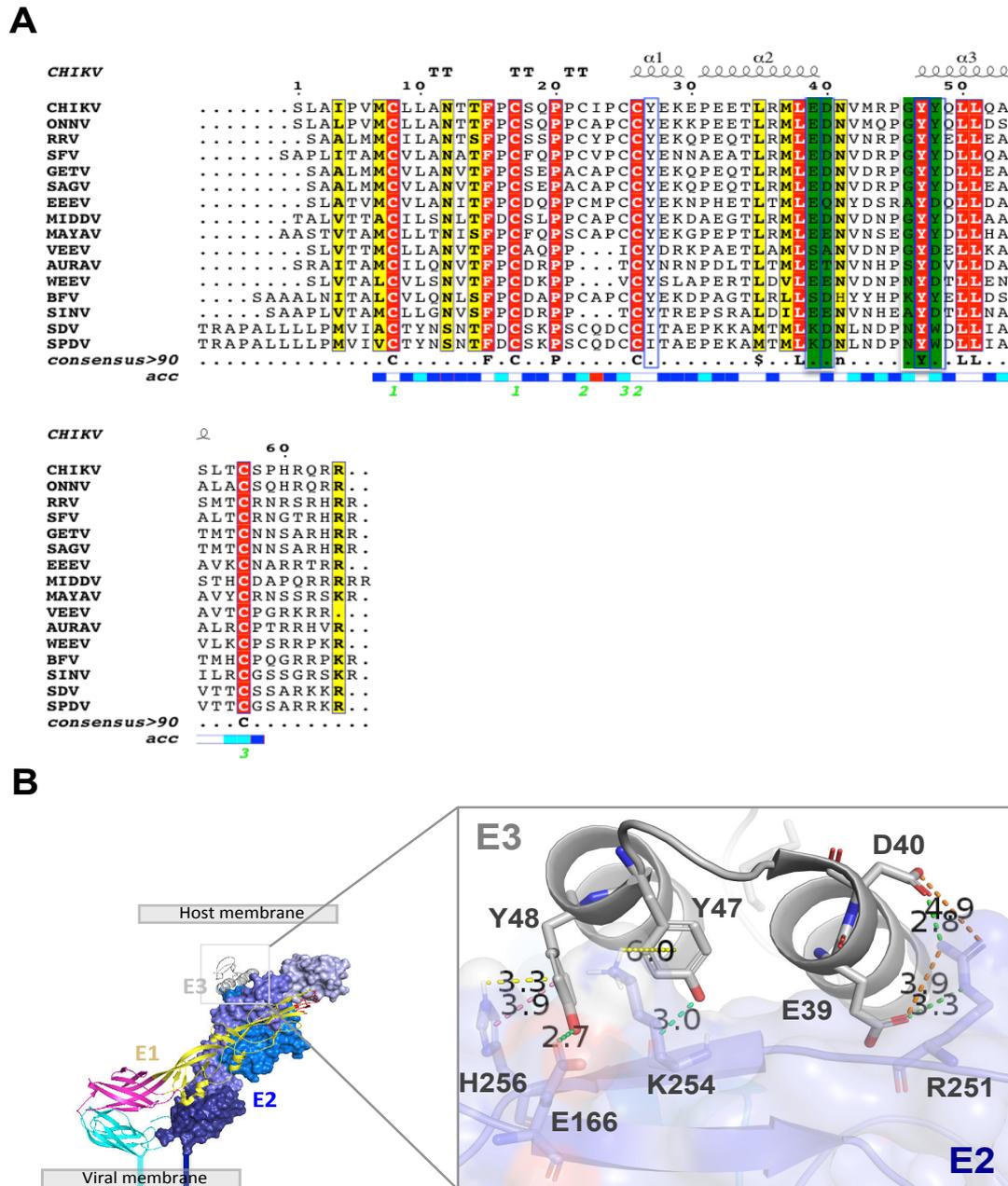


Figure 4.1 CHIKV E3 sequence and structure analysis. (A) MSA of CHIKV E3 sequence among alphaviruses. Key residues that make E3-E2 interface interactions are shown with background color of green. (B) Specific interactions at E3-E2 interface involving highlighted residues from E3.

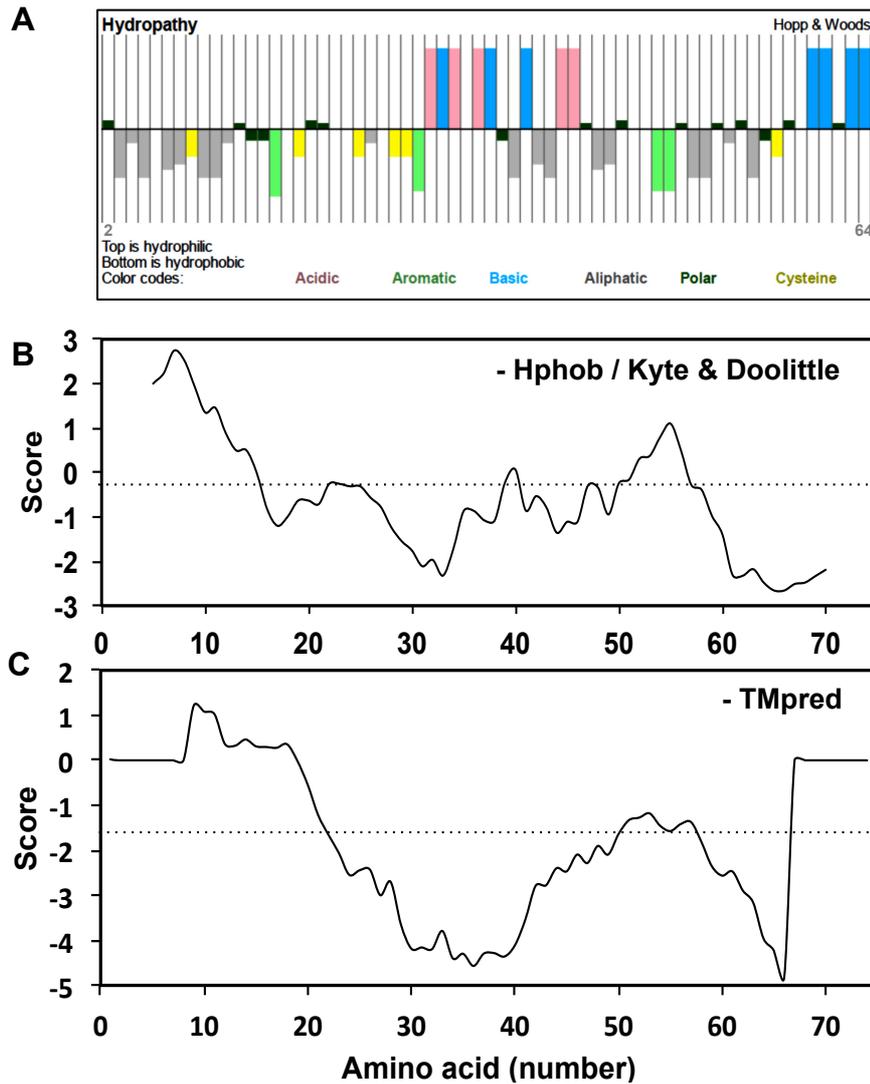
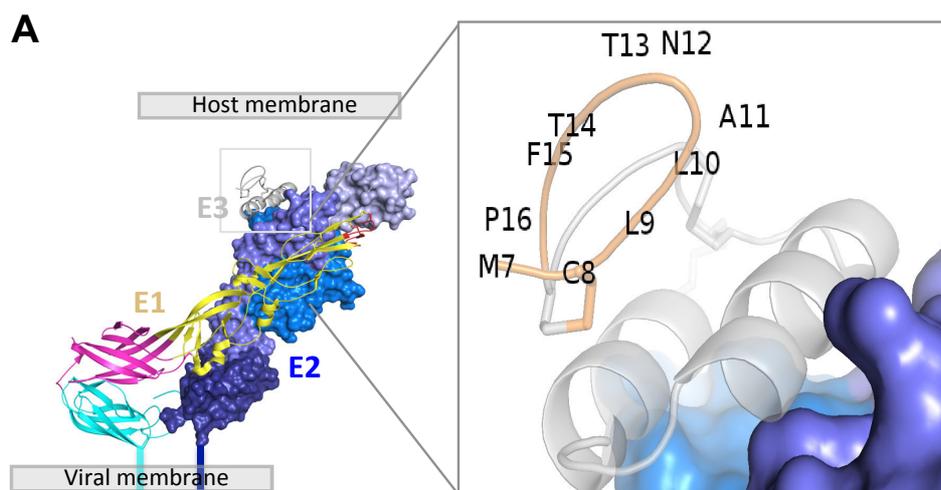


Figure 4.2 CHIKV E3 sequence analysis for membrane interactive region(s). (A) Hydropathy plot analysis using Hopp and Woods method. (B) Hydropathy plot analysis using Kyte and Doolittle method. (C) Tmpred server analysis for presence of putative trans-membrane region.

This patch “L₃APVMCLLANTTF₁₅PC₁₇” has multiple hydrophobic amino acids, an indication of a membrane interacting region. We used the Tmpred server to analyse E3 sequence. Tmpred server predicts for membrane spanning regions and their orientations in membrane based on statistical analysis on TMbase that contains data on many naturally occurring membrane proteins. Positive values from the scoring matrices on a particular

sequence suggest for a putative membrane-spanning region. Our analysis of trans-membrane spanning region using Tmpred server on E3 sequence also suggested the N-terminal amino acids to have a positive Tmpred score. These N-terminal amino acids mostly include residues at E3's N-terminal hydrophobic patch and hence, are likely to be part of a potential membrane-spanning region (Figure 4.2c).



B

E3WT:

MSLAIPVMCLLANTTFPCSQPPCTPCCYEKEPEETLRMLE
NVMRPGYYQLLQASLTCSPHRQRRSLEHHHHHH

E3F15W:

MSLAIPVMCLLANTTWPCSQPPCTPCCYEKEPEETLRMLE
DNVMRPGYYQLLQASLTCSPHRQRRSLEHHHHHH

Figure 4.3 CHIKV E3 structure analysis. (A) Zoomed in view of E3 on E2. E3 N-terminal region is highlighted in yellow with some missing residues from extreme N-terminus. (B) Sequence of E3-WT and F15W mutant proteins with F to W mutation region in the predicted N-terminal patch (yellow background) highlighted in red.

Interestingly, a study by Wu et al (204) using a E3-E2 furin cleavage site mutant SFV (SFV^{SQL}) in cryo-EM imaging, showed that E3 N-terminus protrudes downwards from top of the spike to a region in E1-dIII where it contacts with hydrophobic amino acids from E1-dIII forming a gripper around the E2-E1 heterodimer complex. In CHIKV E3-E2-E1 crystal structure, this part of the E3 N-terminus was missing (less by 5 residues), probably because of high flexibility (or disordered ness).

Taken together, our *in silico* sequence predictions and structure analysis suggested that E3 may interact with membrane through the N-terminal 15 residues hydrophobic amino acids patch. This may result in change in conformation of E3. Altered conformation of E3 may break the interactions between E3 and E2. Prompted by this idea, we set out to test interaction of E3 with membrane using SUVs/LUVs as membrane mimetic models.

4.3.3 E3 does not interact with liposome membrane

To study E3 interaction with liposomes, using fluorescence spectroscopy, a tryptophan residue is introduced in E3, replacing a phenylalanine (15th residue in the sequence) in the predicted trans membrane region (Figure 4.3b). E3 does not have tryptophans in its sequence. Both the WT E3 protein and E3 F15W mutant were purified and used in liposome interaction assays. A representative gel image of SDS-PAGE analysis of the purified proteins is shown in figure 4.4a. Chloroform staining on purified F15W mutant confirmed presence of introduced tryptophan molecule (Figure 4.4b).

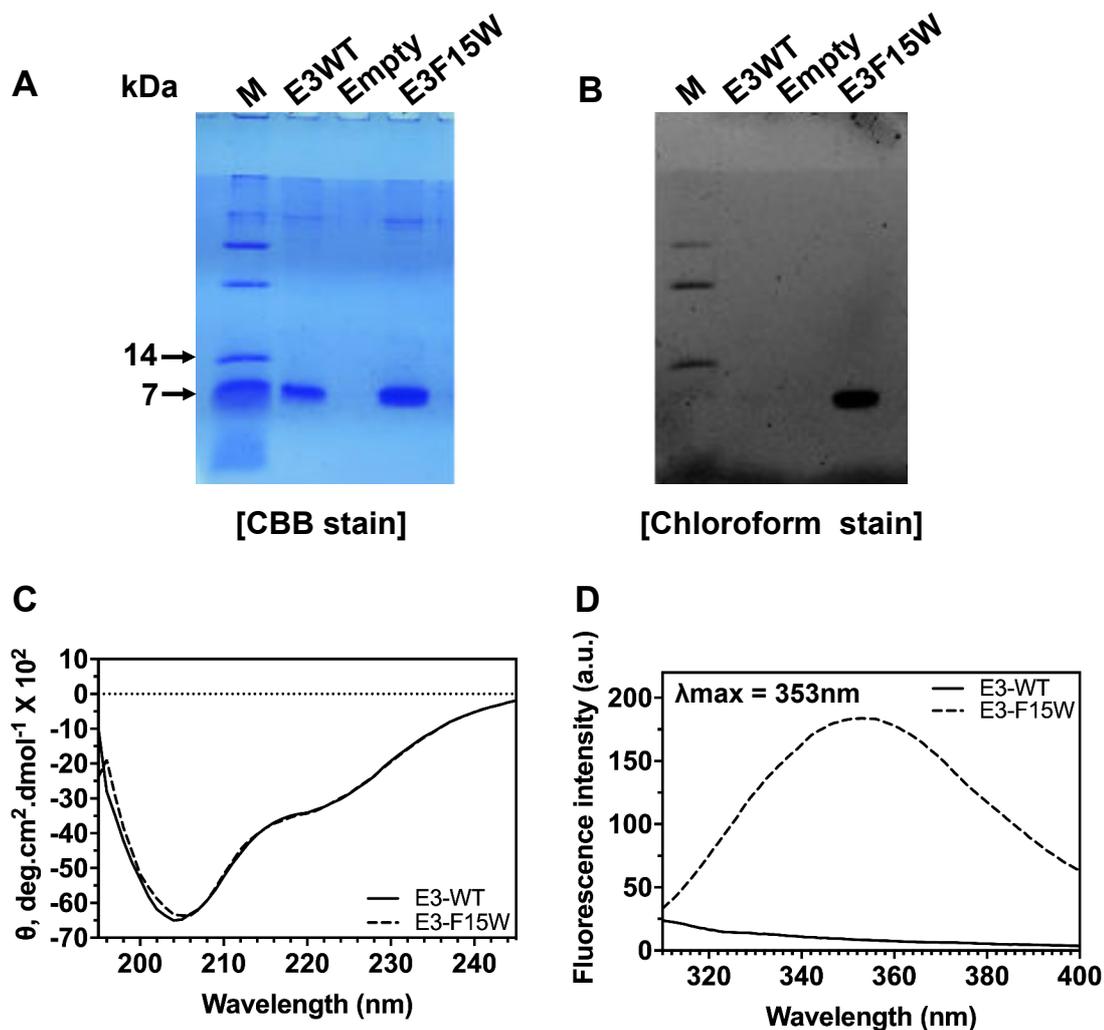


Figure 4.4 Purification and characterization of CHIKV E3F15W mutant. (A) CBB stained and (B) Chloroform stained SDS-PAGE images. (C) Far-UV CD and (D) Intrinsic tryptophan fluorescence spectra analysis on purified F15W mutant protein. E3WT protein is used for comparison.

Further, we characterized secondary structural content of both WT and mutant E3 protein, which showed presence of predominant α -helical content. Comparison of E3 secondary structural content estimated from our far-UV CD data (α -helix = 38%, β -sheet = 13.9%) showed similarity to secondary structural content (α -helix = 40%, β -sheet = 13.3%) calculated from the crystal structure, indicating protein folding into α -helix rich native like

structure (Figure 4.4c). Intrinsic tryptophan fluorescence spectroscopy analysis on E3-F15W mutant (emission max. at 353 nm) indicated that the introduced tryptophan is present in an open and solvent accessible region (the N-terminal region is flexible and not buried) (Figure 4.4d). Secondary structure and tertiary structure analysis (using intrinsic fluorescence spectroscopy) suggest that the E3 WT and F15W are folded into a native conformation.

We studied E3-membrane interactions through the intrinsic tryptophan fluorescence approach by analysing introduced tryptophan residue microenvironment changes in presence of liposomes (Figures 4.5a and 4.5b).

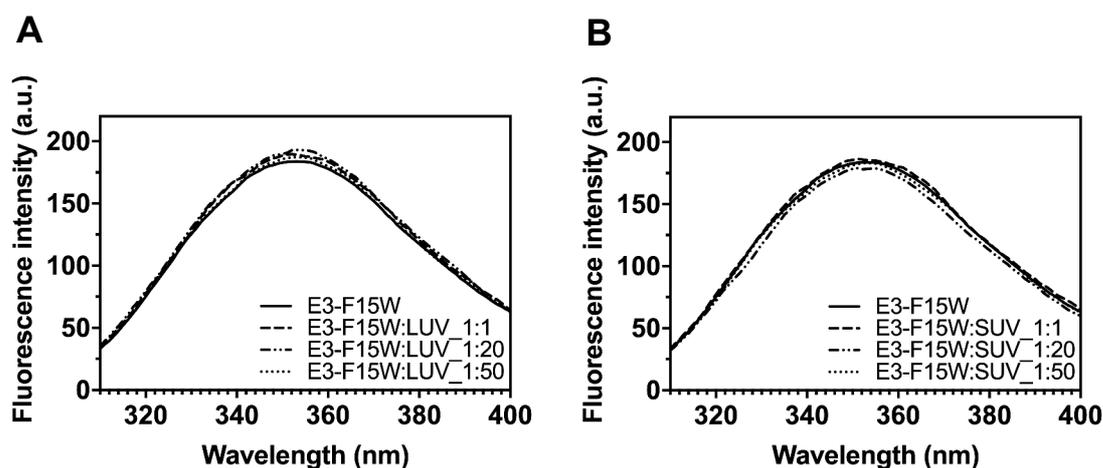


Figure 4.5 Intrinsic tryptophan fluorescence spectra analysis of E3-liposome interactions. (A) Titration of F15W mutant in increasing molar ratios of LUVs. (B) Titration of F15W mutant in increasing molar ratios of SUVs.

The fluorescence intensity and emission maximum of tryptophan (indole group) is altered when the solvent microenvironment or dielectric is changed – in hydrophobic environment tryptophan fluorescence intensity is increased and emission maximum is blue shifted. We used this fluorescence property of tryptophan to probe membrane interaction

of E3. Thus in presence of liposomes, and if E3 interacts with liposomes, then tryptophan fluorescence intensity of E3-F15W mutant should increase and emission maximum blue shifted. However, we did not see any changes in intrinsic tryptophan fluorescence spectra of E3-F15W protein, even in presence of excess molar ratios of liposomes.

Further, we studied E3-F15W membrane interactions using size-exclusion chromatography (Figure 4.6a). Liposomes elute in the void volume fraction of the gel filtration column, and E3 is a smaller protein (~8.6 kDa in size), which elutes farther away in the gel filtration profile (~ at 15ml retention volume) corresponding to its monomeric size. Our reasoning was, if E3 interacts with liposomes, some amount of protein should also be seen in void fraction, since interacting liposomes would elute in this fraction - F7-F8, (confirmed by pyrene excimer fluorescence analysis on all the fractions from this gel-filtration study). However, our WB analysis on fractions corresponding to void elution fraction to E3 monomeric protein elution fractions, clearly showed no E3 protein in void fraction (F7), and showed all E3 protein around monomeric elution fraction (F14-F16) when E3 protein mixed with liposome samples were tested (Figure 4.6b).

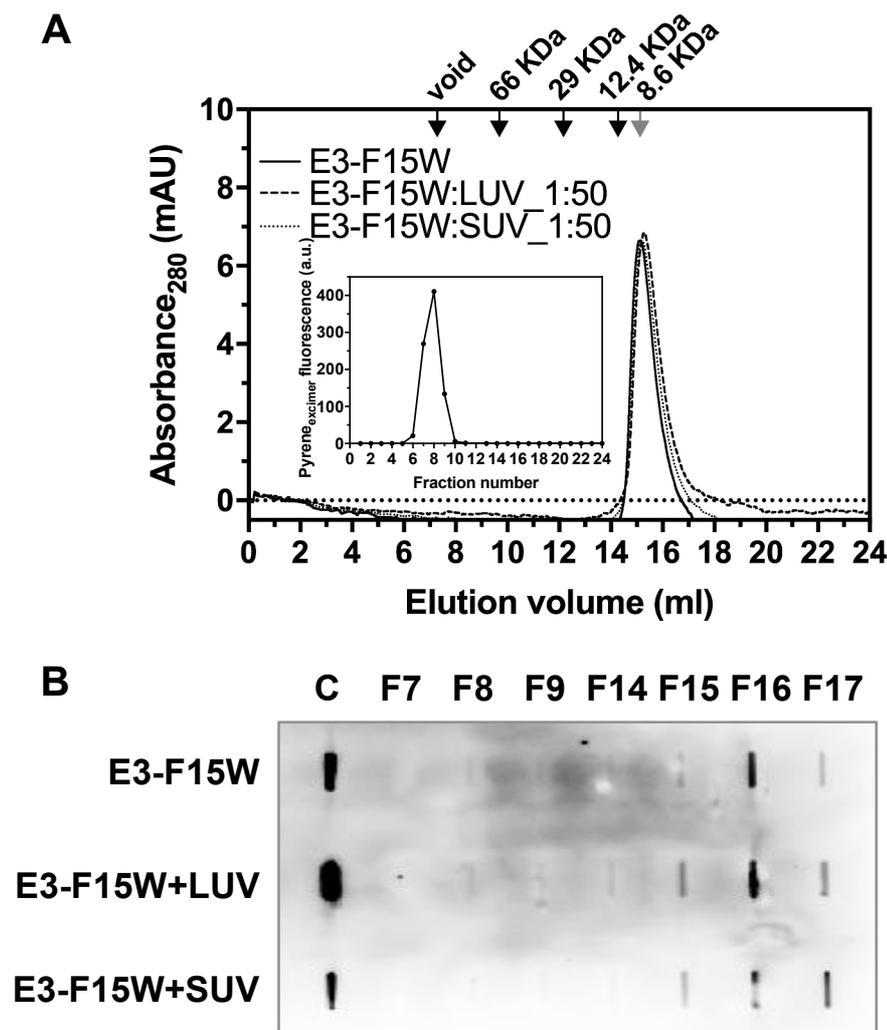


Figure 4.6 Size-exclusion chromatography analysis of E3-liposome interactions. (A) Gel filtration profile of F15W mutant in presence of molar excess of LUVs/SUVs. The fractions 7 and 8, where most of the liposomes eluted were confirmed by pyrene excimer fluorescence assay performed on all the fractions and the graph representing this is shown as an inset. (B) Slot blot analysis on fractions (F7-F17) from the above shown gel filtration experiment.

From this experiment, the gel filtration profile with E3+liposomes sample was similar to only E3 sample and thus, our observation from this experiment was also negative for E3-membrane interactions. We then studied E3 WT protein interactions with liposomes using the lipid-floatation assay (Figure 4.7a).

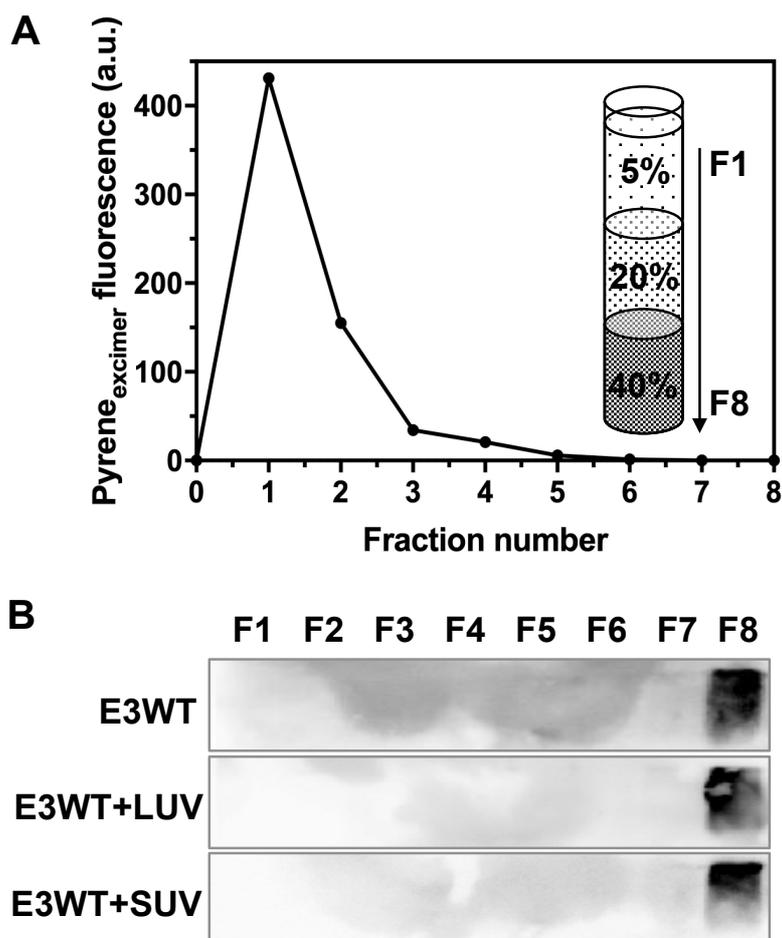


Figure 4.7 Lipid-floatation analysis of E3-liposome interactions. (A) Pyrene excimer fluorescence analysis on fractions (F1-F8) from the lipid-floatation assay. (B) Slot blot analysis on fractions (F1-F8) from the above shown lipid-floatation experiment.

However, in this case also, we observed free liposomes in top fractions of the sucrose gradient without presence of any interacting E3 protein, and all non-interacting E3 proteins were settled at the bottom fractions of the sucrose gradient, as analysed by WB (Figure 4.7b). Taken together, our experimental results suggest that despite presence of a hydrophobic patch of amino acids at its N-terminus, CHIKV E3 does not show any membrane interaction property through it's N-terminus or otherwise, when tested in *in vitro* membrane interaction studying assays at least in the conditions we tested. Further

characterization of the E3 interaction with membrane using liposomes of different composition and curvature are necessary to draw conclusions.

4.3.4 E3 N-terminal hydrophobic patch neither inserts into lipid bilayer nor does it undergo structural transition in presence of LUVs

Further we asked could E3 insert into a lipid bilayer? This is still possible as a result of transient unstable interaction between E3 and a lipid membrane. And also, can this interaction through membrane insertion of E3's N-terminus induce structural transitioning of this flexible region to a more stable α -helical structure? To address all these points, first we studied leaking of contents from a liposome population that has fluorescent molecules trapped inside its core at quenching concentrations, in presence of E3 WT protein (Figure 4.8a). Complete dissolution of liposomes and dilution of the trapped fluorescence molecule with concomitant fluorescence increase was checked successfully in presence 0.5% TX-100. However, when E3 was used in the reaction buffer, no significant change in intensity was observed. In fact, the spectra of the labelled liposomes in presence or absence of E3 WT protein were almost overlapping in nature. This indicated that in *in vitro* conditions E3 does not pierce through the liposome membrane and does not cause content leakage from fluorescently loaded liposomes through insertion of its N-terminus into liposome membrane.

We also analysed E3 structural changes during this process, by studying E3 secondary structure in presence and absence of liposomes using far-UV CD spectroscopy (Figure 4.8b). Our results indicate that, both in presence and absence of liposomes the far-UV CD spectra of E3 WT protein overlapped with each other. So, no significant change in secondary structural content opposite to our expectations of α -helical content increase as a

result of E3-N-terminus membrane insertions was observed. This rules out the possibility that E3, through its N-terminus, may transiently interact or insert into lipid membrane and that, this interaction pattern may lead to α -helical transition of its flexible N-terminal region.

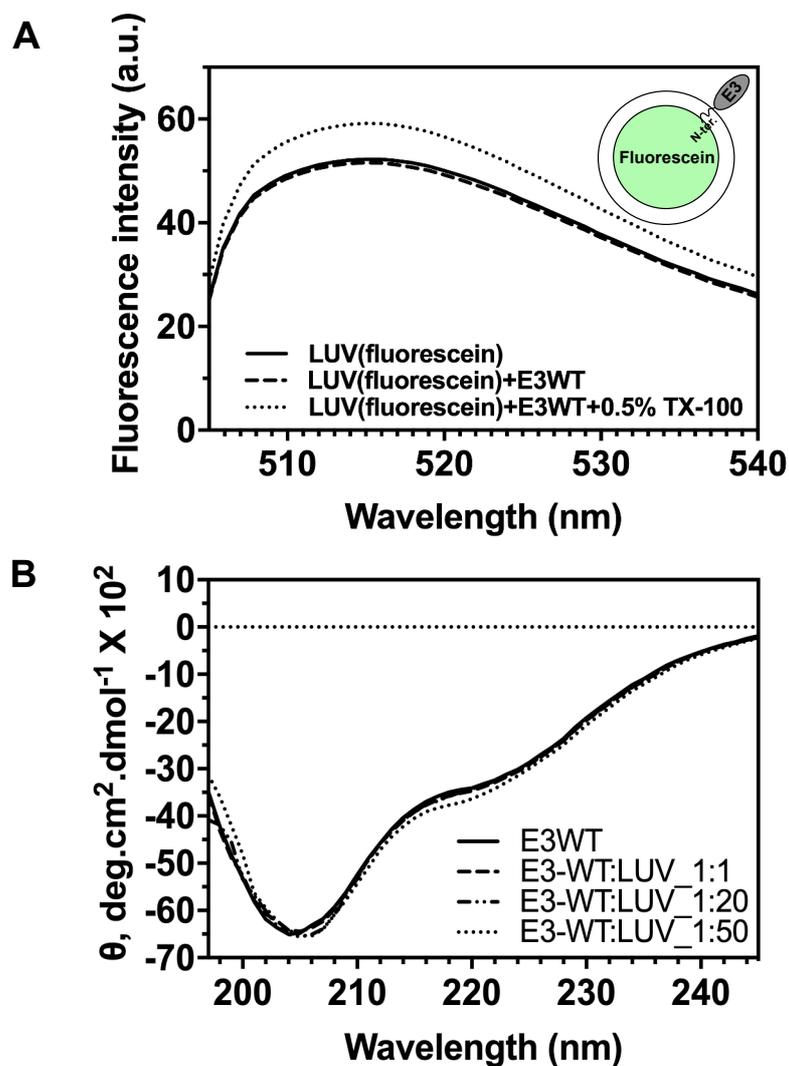


Figure 4.8 Characterization of E3-N-terminus insertion into liposomes. (A) Analysis of fluorescein content release from liposomes in presence of E3-WT protein. (B) Far-UV CD analysis on E3-WT protein in presence of increasing molar ratios of LUVs.

Taken together, while our *in silico* predictions indicated towards a potential membrane interacting region in E3 at its N-terminus, our experimental validations ruled out this possibility. Hence, our observations and experimental results suggest that E3 does not interact with lipid membranes, and thus may have an alternative mechanism of regulation of E2-E1 structure and activity. Alternatively, E3-E2 interactions involve charged residues that are believed to undergo protonation and rotameric changes at acidic pH. Such changes are likely to affect these interactions at acidic pH and may break open the E3-E2 interface at acidic pH. This neutral to acidic pH transition could be a possible trigger for E3 release from E2. We have generated Y48A, Y47A+Y48A, and H256A mutants of E3-E2 recombinant protein and have purified and characterized these mutants for secondary and tertiary structural properties. Using these, further experimentation is currently going on in the lab for proving this hypothesis.

4.4 DISCUSSIONS:

Role of the E3 protein in alphavirus cell entry is not studied. As understood from the crystal structure of E3-E2-E1, E3 is positioned in a way in the structure that unless E3 dissociates from the E2, E1 fusion loops cannot be unmasked. Thus, E3 may be acting as a regulator in the cell entry. Supporting this notion, Basore et al (90), in a cryo-EM study, reported structure of MXRA8 bound with E3-E2-E1 trimeric spike of CHIKV, where E3 is present on the trimeric spike after MXRA8 binding.

In this study, we explored possible mechanisms by which E3 dissociation from E2 may be initiated. There are two possible scenarios that we explored: acidic-pH triggering E3-E2 dissociation and E3 membrane interaction leading to conformational change in E3, possibly leading to E3-E2 dissociation.

There are a number of key interactions at E3-E2 interface (11 weak interactions, either hydrogen bond or salt bridges or cation-pi or pi-pi interactions) that can be affected by change in pH. Change in conformation in E3 protein may lead to breaking of these interactions and release of E3 from E2. Our sequence comparisons (E3 sequences from different alphaviruses), identified two highly conserved motifs in E3. Of the two motifs the 'GY₄₇Y₄₈' motif is particularly interesting. The conserved Y₄₈ in the 'GY₄₇Y₄₈' motif, is involved in a π - π interaction with a acid-sensing histidine (H256 in the ASR region) residue from E2. It is possible that endosome acidification may trigger breaking of this π - π interaction, triggering E3-E2 dissociation. This notion is supported by the observation that E3 Y47 residue interaction with a charged residue (in a π -cation interaction) of E2 in SFV is pH sensitive (83). Further studies to understand the role of GYY motif and other conserved charged residues in E3, in acidic pH-mediated dissociation of E3-E2 are being

continued in the lab.

Alternative (to acidic pH) possible mechanism of E3-E2 dissociation, is through E3 interaction with the membrane. We tested E3-membrane interactions using SUV and LUV liposomes as membrane models. However, our studies are not conclusive. We did not see E3 interaction with SUV or LUV liposomes, at least in the *in vitro* conditions that we have tested. Further characterization using different lipid composition, or using E3 protein in complex with E2, may be required.

CHAPTER 5

DEVELOPMENT OF A BACULOVIRUS PSEUDOVIRION SYSTEM FOR USE IN FUNCTIONAL ENTRY ASSAYS

Acknowledgement: I thank Mr. Debashis Panda who initiated this work in the lab. I acknowledge his help in preparation of the homologous recombination construct.

5.1 INTRODUCTION:

To study viral biology in a lab setup, especially the early events in a virus life cycle (for example viral entry), several different assay systems or approaches are available. Each of these assay systems make use of either the infectious virion particle or use other ways to have a mimic of the infectious virion particle (lacks one or more components) to study its function/function of its components (205–207). These methods are referred to as functional assays. Viral cell entry is mainly dependent on the function of the viral envelope proteins. Hence, different approaches other than using the infectious virion particles are employed for structure-function studies on viral envelope proteins. Among these other methods, use of reverse genetics approach where viral RNA transcription from a cDNA clone is performed and then these RNAs are used for transfecting cells in a cell-culture based assay for generation of infectious virion particles. Mutations can be created in the cDNA clone and the mutant viruses can be generated by transfecting with the transcribed mutant RNA samples (208–211). Further, in addition to using the infectious virion particles, this approach that involves generation of infectious viruses requires certified class III BSL facility for research work and hence, limits the availability of these methods to specific labs equipped with such facilities. However, there are other methods that does not require such level of facilities and use of infectious virus particles. These systems can be used in a common lab setup, which involves use of either VLPs (205) or pseudoviruses (206, 207). Both these methods have their advantages and disadvantages. VLPs are specific to respective viruses (includes capsid+envelope+envelope proteins only) and can generate the closest mimicking particles (structure wise) to the infectious clones (205). However they lack the entire replication machinery and hence the system is limited to use in structural

and in certain functional studies. For structural studies, mostly EM based analysis on isolated VLPs are performed. For functional studies, fluorescence dye labelled VLPs are used for fusion with a lipid bilayer or liposome membrane in *in vitro* conditions. One of the limitations of using the VLP approach is the trial and error approach of VLP production with minimum success rate (205). Since for each virus, the virus specific components are used, the protocol needs to be optimized each time for making VLP for a specific virus. Though a single plasmid carrying the entire structural polyprotein cassette is enough for generating VLPs, creating mutants of VLPs could be problematic especially when the mutations affect the structural stability of the VLPs. On the other hand, pseudo viruses are non-infectious forms of actual virions and in general carry few viral components from the actual virus such as envelope entry proteins (206, 207). All other components required for making a virion particle is supplied from a number of plasmids that code for other necessary viral components of a different virus (other than the one from which the envelope protein coding sequence is used). In case of pseudo viruses, a simpler genome incorporated reporter protein coding sequence (for example GFP, expression of which can be studied through fluorescence microscopy) can be used to study viral cell entry using cell culture based assays, in a more realistic way to understand the functionality of envelope proteins during viral cell entry. Pseudo viruses has the features of a VLP as well and can be used for structural analysis using EM based approaches. There are two well studied pseudo typing systems available – the lentiviral packaging system (207) and the VSV G-VSV system. For lentiviral packaging either a 3rd generation system is required, which uses three plasmids or a 4th generation system is required that uses four different plasmids. These plasmids generally are classified as packaging plasmid (one or two plasmids code for structural

components for pseudovirion making), transfer plasmid (contains a reporter or transfer sequence and a packaging signal for packaging into pseudo viruses) and an envelope plasmid (codes for foreign viral envelope proteins). For VSV G-VSV pseudovirion system, a VSV helper virus is used along with an envelope protein expressing plasmid or a stable cell line, for generation of pseudo viruses (206). Thorough protocols on how to use these systems are available for gene delivery or siRNA delivery methods (212, 213). However, for studying viral cell entry, these systems are time consuming, tedious and require simultaneous transfections with multiple plasmids (or a helper virus + plasmids in case of VSV G-VSV pseudo type system) (206, 207). We took a similar approach and tried pseudo typing of CHIKV envelope proteins using the 3rd generation lentiviral packaging system. However, our trials were not successful. The lentiviral/VSV G-VSV systems are made using respective viral components that involves use of a single or at max two foreign envelope proteins. So, it is easier to pseudo type a viral envelope where only a single envelope protein is present. In other cases for example in case of CHIKV, in addition to the envelope fusion protein a companion receptor binding protein and two other small proteins are also present. Packaging with such complex envelope protein systems though tried previously are not thoroughly studied. More data related to these are necessary to understand how such envelope protein complexes can also be successfully used for pseudo typing.

Baculoviruses on the other hand are genetically simple and easier to manipulate and hence, are also used for development of pseudovirion systems (214). These are a group of viruses from baculoviridae family that contains ~76 species divided among four different genera – alpha, beta, gamma, and delta-baculoviruses (215–217). These viruses infect a

wide range of invertebrate hosts; larval forms of the moth species are most common. Baculoviruses also infect flies, shrimps and mosquitoes. Though these viruses can enter into mammalian cells in a cell culture based system, these viruses fail to replicate inside mammalian cells or any other vertebrate animal cells. They carry a double stranded DNA genome of ~80 to 180 kb in size. The most studied of the baculoviruses is the alpha-baculovirus *Autographa californica multicapsid nucleopolyhedrovirus* (AcMNPV), which has a DNA genome of ~133 kb. This virus has two forms; one the occlusion derived virus (ODV) and the other is the budded virus (BV) form. These viruses are smaller particles (<1/1000 of a millimetre across) and are protected by a protein coat called polyhedron in the ODV form. These forms are eaten by insects from an infected plant, from where the ODV form of the virus is transferred to insect gut and dissolves there in the alkaline environment for genome release and infection. The BV however performs the cell-cell transmission for systemic infection. The viral envelope protein of AcMNPV is gp64, which not only helps in formation of the virion but also helps in attachment to host cells (218), acidic pH dependent virus-host cell membrane fusion, entry from within the endocytic vesicles (219–221), and budding, where BV acquires cell envelope with surface expressed glycoproteins from an infected cell etc. (222). Deletion of gp64 can be lethal unless complemented with other viral envelope proteins for virus production (223–225). Baculovirus infection can be seen as early (0-6 h.p.i), late (6-24 h.p.i), or very late (72 h.p.i). BV is produced towards end of the late phase and ODV is produced towards end of the very late phase.

With the available methods and commercial strategies (214, 226–228), it is now easier to manipulate the single DNA genome of the baculovirus for different purposes.

Moreover over all procedure for these methods requires maintenance of low cost derived insect cell culture system and a basic lab set up with molecular biology facilities. As mentioned earlier these viruses, though insect infecting viruses, can also enter into mammalian cells. Taking into consideration all these parameters, several trials have been made to manipulate their genomes to use these viruses for development of pseudo typing systems. One widely used approach is the pRED/ET based recombination approach (226) to remove the sole viral envelope or entry protein, gp64, and replacing it with any other envelope protein of interest. However, the earlier strategies included a cumbersome approach to achieve this process with minimum efficiency of recombination and hence, less success. Example of one such approach is the use of a transfer vector with gene of interest that needs to be recombined with the co-transfected baculovirus genome inside insect cells in a cell culture based system. Post co-transfection the progeny viruses were titrated and selected for recombinants based on the following strategy. Selection methods in certain cases depended on the blue white selection by transfer of a lacZ locus, which was kept upstream to the recombination cassette in the transfer vector. Such methods required screening multiple plaques of viruses generated inside insect cells for recombinants, and the efficiency of recombination observed remained below 1% (214).

We took advantage of the available gene manipulation methods that depends on engineering through recombination and transposition approaches (214, 226–228). Presence of transposition sites in the single baculovirus genome (as per Bac-BacTM system) for ease of incorporation of reporter as well as different viral envelope protein-coding sequences makes it easy to manipulate the bacmid genome once the envelope protein, gp64, coding sequence is replaced with a selection marker or a reporter or both through recombination

approach for making pseudo viruses. Recombination can be performed by use of pRED/ET system inside the same *E. coli* host that carries the transposition machinery for recombination, rather than depending on insect cells for recombination (214, 226). Also, a fluorescence reporter-based screening can be performed in *E. coli* itself and is simpler compared to screening of multiple plaques post-insect cell transfection and recombinant virus production. To add to this, manipulation of a single bacmid genome is beneficial over manipulation and use of multiple plasmid systems used for other pseudovirion systems. Also, the use of an insect virus significantly reduces the chances of human infection/for that matter reactivation of a pathogenic virus inside a mammalian host cell.

We planned for developing a baculovirus pseudovirion system for use in functional entry assays since our trials with Lentiviral packaging systems involving pseudo-typing of CHIKV envelope proteins were not successful. Though Lentiviral and VSV G-VSV pseudo-typing systems are used extensively for gene transfer methods and generation of pseudo-envelops with envelope proteins from many different viruses, pseudo-typing with CHIKV envelope proteins were not tested thoroughly (206, 207). Our goal is to knock out the baculovirus envelope protein, gp64 and then knock-in ‘an antibiotic plus a fluorescence reporter (Cm-GFP)’ cassette in place of gp64 coding sequence in frame with gp64 promoter. Our reasoning for this strategy is, while presence of Cm adds an extra level of selection marker during screening inside bacterial cells, presence of GFP in frame with Cm promoter makes GFP active inside bacteria as well, also contributing to ease of selection during screening for the right recombinants. In addition, GFP can also be driven from the gp64 promoter and hence once the right recombinant is transfected into sf9 insect cells, transfection verification and efficiency can be checked easily by utilizing GFP reporter

fluorescence. Once the pseudovirion particles assemble into functional infectious agents (at least in sf9 cells), GFP reporter fluorescence can be utilized to study both syncytia as well as functional cell-transduction in insect cells (Figure 5.1). In this way, the functionality of the viral envelope proteins can be verified (and compared between WT and mutant proteins) with ease in cell entry assays.

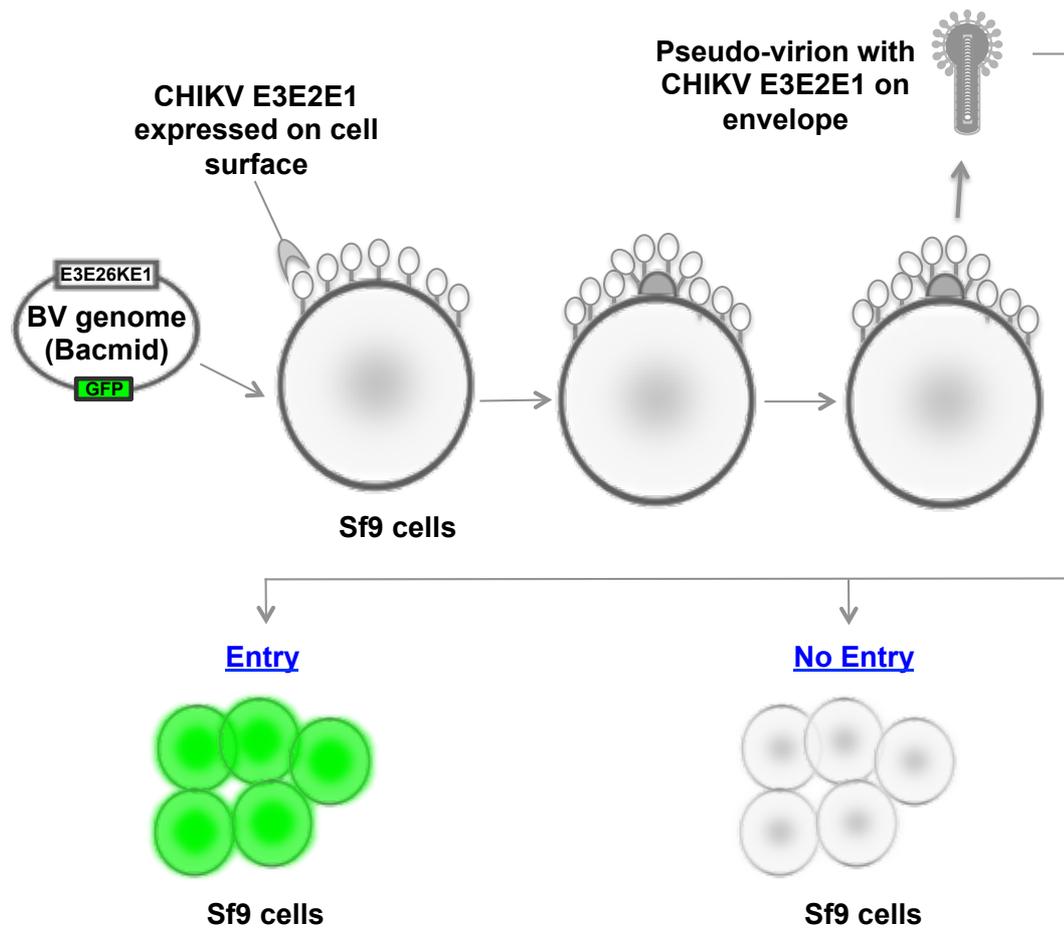


Figure 5.1. Scheme showing overview of baculovirus (BV) pseudovirion development plan.

5.2 MATERIALS AND METHODS:

5.2.1 Lentiviral packaging of CHIKV envelope proteins

We used the pcDNA3.1 based pBS002 construct (pcDNA3.1 + E3-E2-fl-6K-E1-fl, construct making is detailed in the materials and methods section of chapter 1) for cell-surface expression of full-length structural envelope proteins of CHIKV in HEK293T, and Hela cells. Details on the expression procedure are discussed in chapter 1. For lentiviral packaging, we used a 3rd generation based lentiviral packaging system, where psPAX2 (the packaging plasmid), pLL3.7 (the transfer plasmid with GFP reporter), and pMD2. G (envelope plasmid with VSV G envelope protein) or pBS002 (envelope plasmid with CHIKV envelope proteins) plasmids were used at 1:1:1 molar ratios (total of 6µg of plasmids per well of a 6 well plate was used) for transfections. All items used are purchased from Invitrogen™. We collected the viral stocks from these transfected samples at 5 d.p.t, and used these stocks (after removal of cell debris by centrifugation and syringe filtration) for infection of fresh batch of cells at ~80-90% confluency with >95% viability. We used polybrene as a transduction enhancer at 6-8µg/ml of media. 48-72 hrs post transduction, treated cells were checked for successful infection by recording fluorescence images for GFP reporter fluorescence using a fluorescence microscope.

5.2.2 Construction of ‘gp64-start+Cm-GFP+gp64-end’ cassette for recombination

We prepared the gp64-start+Cm-GFP+gp64-end cassette in sequential cloning steps. Small regions from the beginning and end of the gp64 coding sequence were PCR amplified from the baculovirus genome. Similarly the coding sequences for Cm and GFP were PCR amplified from pRIG and pIZT-V5-HIS plasmids respectively. The gp64

beginning and end regions (termed as gp64-start and gp64-end) were cloned into pUC19 vector in two sequential cloning's, one generating the gp64-start-Cm and the other generating the GFP-gp64-end fragments. The final construct was prepared by cut-pasting GFP-gp64-end region next to gp64-start-Cm region in pUC19 vector (Figure 5.2). The primers used for making the cassette are listed in Appendices, Table a1.4. The entire cassette “gp64-start+Cm-GFP+gp64-end” was PCR amplified and transformed into a recombination host carrying the baculovirus genome with WT gp64 coding sequence for recombination.

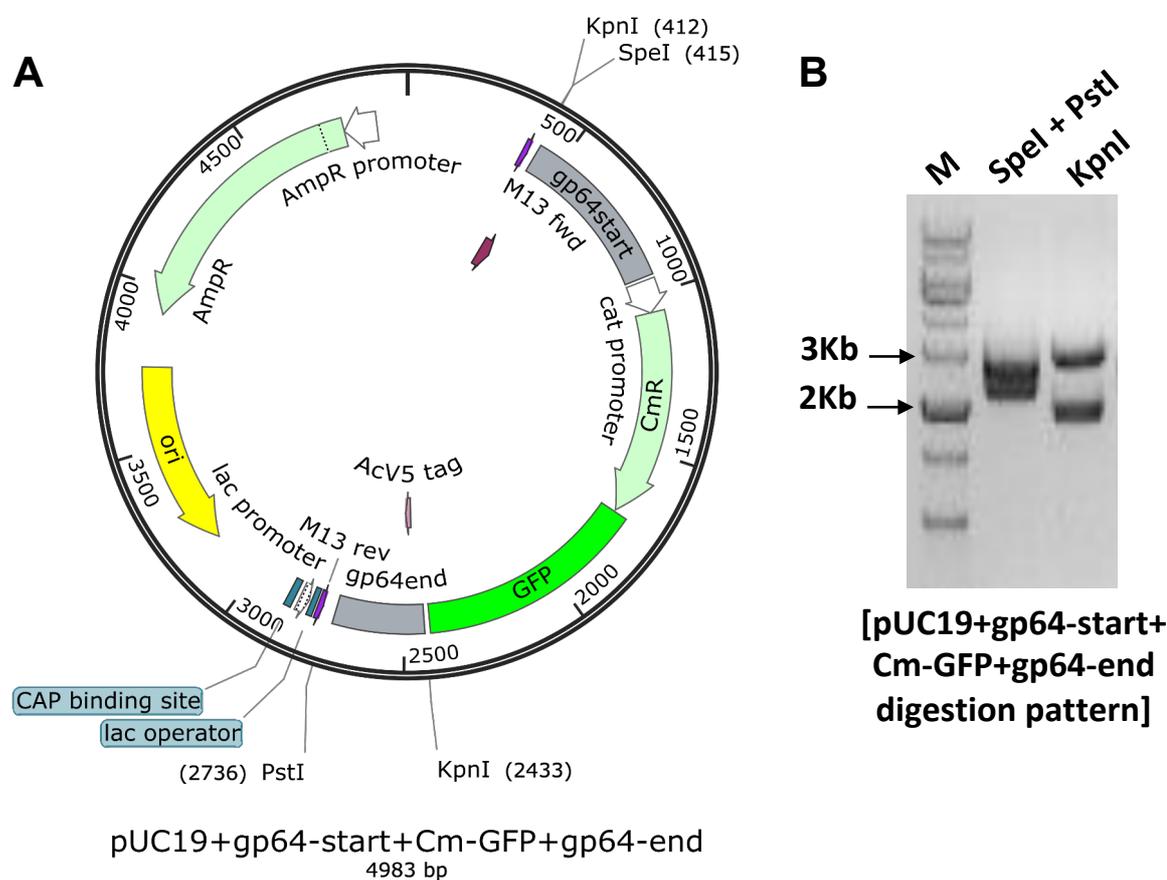


Figure 5.2 Homologous recombination cassette. (A) Map of homologous recombination cassette containing pUC19 clone. All features are labelled and base pair numbering is also shown. (B) Integrity check of the clone shown in (A) that carries the recombination cassette through restriction digestion. Name of restriction enzymes used are labelled.

5.2.3 Maintenance of pRED/ET recombination plasmid inside *E. coli* DH10Bac cells

We chose the DH10Bac strain of *E. coli* as a recombination host because of following reasons. These *E. coli* strains contain a baculovirus genome called bacmid as a shuttle vector (bMON14272) and a plasmid called as helper plasmid (pMON7142) that codes for an enzyme called transposase inside the *E. coli* cell for site specific transposition of a gene of interest, cloned into the insect cell recombination vector pFastBac, to the bacmid genome. bMON14272 code for kanamycin resistance, pMON7142 codes for Tetracyclin resistance, and upon transposition Gentamycin resistance is transferred from pFastBac to bMON14272. All these antibiotic selection markers in addition to presence of the blue white screening locus at the Tn7L and Tn7R transposition site makes this *E. coli* strain a good host for baculovirus genome manipulation. The exact genotype of the *E. coli* DH10Bac host is $F^{-mcrA} \Delta(mrr\text{-}hsdRMS\text{-}mcrBC) \Phi80lacZ\Delta M15 \Delta lacX74 recA1 endA1 araD139 \Delta(ara, leu)7697 galU galK \lambda rpsL nupG/pMON14272/pMON7124$.

We used the Gene Bridges' RED/ET homologous recombination method that makes use of a set of site specific recombination promoting enzymes by expressing them in a controlled manner inside a recombination host with all other necessary resources required for recombination (for example the DNA to be inserted with homology arms that match to the homologous locus of a recombination target) present inside it. This approach is beneficial over other similar methods as it is less time consuming (can be done in few days), is independent of use of restriction sites, and there is no size limit as such for the recombination cassette for homologous recombination. The basic requirement for recombination using the RED/ET system is a ~50 bp homology arm that can be chosen

freely for recombination at any selected site. *E. coli* genome, BACs, and even the bacmids were successfully used as recombination targets using this system previously. The process is simplified by use of a single plasmid (pRED/ET) to carry the lambda phage derived recombination protein coding sequence along with other necessary components. We used this pRED/ET coded set of enzymes (gam, beta, exo and recA) kept under the control of an inducible arabinose promoter for recombination. Once this plasmid is transferred into a recombination host where both the recombination cassette and the recombination target sequences are present, Gam protects the introduced DNA cassette from degradation by RecBCD and SbcCD systems. Exo is a 5'→3' exo-nuclease that creates 3' single stranded DNA overhangs for recombination, and Beta binds to these overhangs, protects these overhangs and promotes annealing of these regions to a homologous region on the recombination target. The recA performs the site-specific recombination. This pRED/ET plasmid has pSC101ori and the growth of *E. coli* hosts maintaining this plasmid are sensitive to temperatures above 30°C. The plasmid has Amp^R and post-transformation into *E. coli* DH10Bac cells, the cells were grown on LB agar plates containing 100µg/ml ampicillin at 30°C inside a temperature incubator till the colonies appeared (after ~20 hrs). The transformation protocol used in the previous step is detailed further. ~50ng of pRED/ET (size of 9270bp) plasmid was transformed into *E. coli* DH10Bac host recombination strain by giving heat shock at 42°C for 1 minute. For recovery, transformed cells were grown at 30°C, 250rpm for 1.30 hrs in 1ml of fresh LB broth without any antibiotics. Post incubation, cells were spun down at 4000 rpm for 2 minutes, then ~900µl of media from the top was poured off and cell pellet was re-suspended in remaining ~100µl of media followed by plating on LB-agar-Amp plus other antibiotics of interest containing

plate for growth. Cells were incubated at 30°C till single colonies appeared.

5.2.4 Preparation of electro-competent pRED/ET bearing *E. coli* DH10Bac cells

For preparation of electro-competent cells for homologous recombination single colonies of pRED/ET bearing *E. coli* DH10Bac cells were picked and used for inoculation of 2ml of primary culture. From the O/N grown primary culture (at 30°C, 250rpm), 2% inoculum was used for 10ml of secondary culture (at 30°C, 250rpm). It takes ~2hrs to reach OD₆₀₀ of 0.3-0.4. At this point L-arabinose was added to a final concentration of 0.4% (from a fresh 10% stock) for induction. Incubation at 37°C, 250rpm for another 1hr post induction was allowed (during this 1hr induction at 37°C, all proteins necessary for subsequent recombination are expressed. By the end of this 1hr induction step, cells will still have ~2-3 copies of pRED/ET. The plasmid is actually lost after electroporation and recombination of gene of interest when cells are incubated at 37°C for O/N). Now, centrifugation was done with ~1.5ml of secondary culture in pre-chilled 1.5ml tubes at 11000rpm for 30secs at 2°C. The supernatant was removed and cell pellet was re-suspended in 1ml of ice-cold 10% glycerol under sterile conditions. This step was repeated twice and finally the electro-competent cells were re-suspended in ~100µl of leftover 10% glycerol from the last wash step. The tubes were kept on ice for addition of ‘recombination cassette DNA PCR product’ for electroporation and subsequently recombination.

5.2.5 Homologous recombination of “gp64-start+Cm-GFP+gp64-end” cassette with baculovirus genome

For homologous recombination, ~1000ng of linear DNA cassette with homology arms (‘gp64-start+Cm-GFP+gp64-end’ PCR product) were added to chilled electro-competent *E. coli* DH10Bac cells and then pipetted slowly for mixing, inside the pre-chilled

electroporation cuvettes. Electroporation was performed at 1350V for a brief 5ms pulse. Immediately 1ml of fresh LB media without any antibiotics were added to the cuvette and pipetted up and down carefully (at least once) to mix the cells, and then kept back on ice. The mixture was then transferred into a fresh 1.5ml tube followed by incubation at 37°C, 250rpm for 3hrs. Cells were spun down at 4000rpm for 2 minutes and then ~900µl supernatant media were poured off. The remaining cell pellet was re-suspended in ~100µl of leftover media and plated on LB agar + Cm (used at 10mg/ml) + other antibiotics of interest (excluding Amp) containing plate. Then this plate was incubated at 37°C till single colonies appeared. It takes ~2-3 days for single colonies to appear on plate. The colonies were re-streaked onto fresh LB agar + antibiotic plates (now with Cm at 30mg/ml concentration) and used for verification of recombination by isolating recombinant bacmids from these cells and checking for deletion of gp64 and insertion of 'gp64-start+Cm-GFP+gp64-end' through PCR using these gp64 forward and reverse primers (Appendices, Table a1.4).

5.2.6 Preparation of viral envelope protein coding constructs and their transposition into Δ -gp64-Cm-GFP bacmid genome

Constructs for viral envelope protein expression were prepared by cloning the coding sequences of VSVG (amplified from pLP-VSVG plasmid), CHIKV E3-E2-6K-E1 (details about other coding regions of E3, E2, 6K and E1 are explained in chapter 1, materials and methods section), and Baculovirus gp64 (amplified from WT bacmid genome isolated from *E. coli* DH10Bac cells) into the MCS of pFB1 vector. Primers used for making each of the above mentioned constructs are detailed in Appendices, Table a1.4. The constructs were transformed into the Δ gp64-Cm-GFP bacmid carrying *E. coli* DH10Bac

cells using a standard protocol. The transposed recombinant bacmid carrying colonies were selected through blue-white screening (white colonies are recombinants, +ve for transposition). Recombinant bacmids were isolated from re-streaked fresh white colonies using a modified mini prep method (where except mini-prep column, isopropanol precipitation approach was used for DNA purification). After brief washing steps with 70% ethanol, the precipitated DNA pellet (white colored pellet) was air dried and then re-suspended in elution buffer for transfections into sf9 cells. PCR verifications for positive transposition in these bacmids were done using M13F/M13R primers.

5.2.7 Transfection of purified recombinant bacmids into sf9 cells for pseudo-virus production

Recombinant bacmids purified were always used freshly for transfection experiments. Sf9 cells growing in their log phases (at a density of $\sim 2-3 \times 10^6$ cells/ml) were seeded to $\sim 80\%$ confluency, 12 hrs prior to transfection, in 6-well dishes (in 1X Sf-900II SFM media with 1X PS). Just before transfection, media was changed to without antibiotic containing 1X Sf-900II SFM (~ 1 ml/well). Transfection was performed on to adherent sf9 cells in a 6-well dish following the manufacturers protocol except that $\sim 5 \mu\text{g}$ of recombinant bacmid DNA was used per well. Post transfection, ~ 5 hrs of incubation step was followed in the transfection media. Post this incubation step media was changed to 1X Sf-900II SFM with 5% FBS and 1X PS, and cells were incubated further (for 2-6 days) depending on the experimental requirement.

5.2.8 Syncytia assay

For syncytia assay, recombinant bacmids lacking gp64 and with complementing viral fusion proteins from Baculovirus gp64, VSVG, and CHIKV E3-E2-6K-E1 were

transfected into sf9 cells and incubated for post transfection period of 48 hrs. Then, sf9 cells were washed with PBS buffer and then treated with acidic pH buffer (PBS buffer adjusted to pH 5.5) for 2 minutes. Acidic pH buffer was removed and after a brief washing step with PBS buffer, cells were kept in 1X Sf-900II SFM media with 5% FBS and 1X PS for another 4 hrs. Syncytia formation was observed by looking for multinucleated large and fused cells under phase contrast microscope. GFP reporter fluorescence from large fused cells was also observed under fluorescence microscope for conformation of syncytia formation.

5.2.9 Pseudovirion transduction/cell entry assay

For pseudovirion based entry assay, recombinant bacmids lacking gp64 and with complementing viral fusion proteins from Baculovirus gp64, VSVG, and CHIKV E3-E2-6K-E1 were transfected into sf9 cells and incubated for post transfection period of 5-6 days. Used media from these incubated cells were collected and centrifuged at 1000 rpm for 5 minutes at RT. Supernatant from this step were collected and filtered through 0.45 μ m syringe filters. The filtered supernatants containing viral particles were used freshly at different dilutions (1-1000X dilution) for transduction into fresh batch of sf9 cells. Cells for transduction were prepared the same way as for transfection except confluency was kept to ~90% at the time of transduction (as a monolayer of cells). Cells were incubated for another 3-5 days to observe ~100% transduction through visual observations using phase contrast microscope or through observation of GFP fluorescence using fluorescence microscope. Number of cells with GFP fluorescence over total number of cells was used for calculation of percent-transfected cells at different time points (at 48 hrs representing 1st round of transduction and at 120 hrs representing cell-cell infection).

5.2.10 Immunofluorescence assay to check surface expression of CHIKV cell entry proteins in sf9 cells using CHIKV E1-pAb

For checking surface expression of CHIKV envelope protein through immunofluorescence assay, recombinant bacmids lacking gp64 and with CHIKV E3-E2-6K-E1 were transfected into sf9 cells (grown on cover slips in a 6 well dish) and incubated for post transfection period of 48 hrs. Used media from these incubated cells were removed and thoroughly washed with PBS. Washed cells were blocked in blocking buffer (3% BSA containing PBS with 0.01% Triton-X 100) followed by incubation with E1-pAb (used at 1:25000 dilution) containing buffer for 2 hrs. After another brief washing step, 1hr incubation with secondary antibody (with Alexa Fluor™ 594 tag) at 1:5000 dilution was performed. Then following thorough wash (3times 5minutes wash with PBS), fluorescence images were recorded using fluorescence microscope and analysed for presence of GFP fluorescence (especially at cell surface).

5.3 RESULTS:

5.3.1 Production of CHIKV pseudo viruses using lentiviral packaging system

We planned and tested expression of full-length forms of CHIKV envelope proteins for cell surface expression in mammalian cells. Our purpose for mammalian expression was to test packaging of CHIKV envelope proteins into lentivirus based pseudo viruses. Cell surface expression of CHIKV envelope proteins would allow newly budded pseudo-lentiviruses to take these proteins along with the host membrane on their envelope while budding and hence can be used to study CHIKV envelope proteins in cell entry assays. Our construct making included the trans-membrane regions of E2 and E1 proteins and also included the native 6K sequence that joined E2 end and E1 beginning coding sequences. This entire ORF was kept in frame with the cytomegalovirus (CMV) promoter for expression in mammalian cells. We used this construct for mammalian cell expression for lentiviral pseudo-typing of CHIKV envelope proteins using a 3rd generation based lentiviral packaging system. However, we did not observe any functional pseudo viruses with CHIKV envelope from our transduction experiment (Figure 5.3). On the contrary, the VSV G protein (used as positive control for this experiment) was successfully pseudo-typed into the pseudo-lentiviruses and was able to successfully enter into HEK293T cells upon transduction. This was evident from the GFP reporter fluorescence obtained only in case of VSV G pseudo-typed lentiviruses (Figure 5.3).

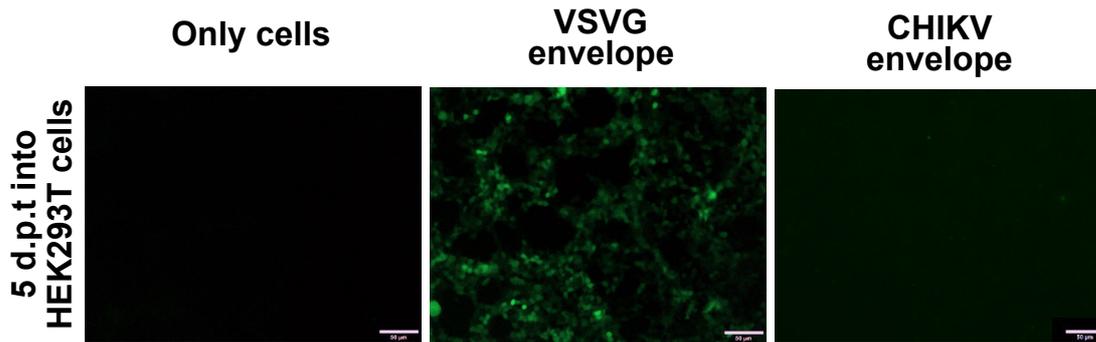


Figure 5.3 Lentiviral pseudo-typing of full-length CHIKV envelope proteins. Fluorescence images from transduction experiment from lentiviral packaging of VSV G (control) and CHIKV envelope proteins (test) samples are shown. Only cells - zero transduction control is also shown. Scale bar is 50 μ m.

This result was consistent between two of the mammalian cell lines that we tested, both HEK293T and Hela cells. Our literature analyses on CHIKV pseudo-typing show that CHIKV envelope proteins though are used earlier for production of VLPs (141), lentiviral packaging or for that matter packaging of CHIKV envelope with any other packaging methods was not thoroughly established. A recent study where CHIKV envelope proteins were used towards packaging into lentiviral system (102), have reported use of complex and multiple constructs with IRES sites for successful expression and packaging of CHIKV envelope proteins into pseudo-lentiviruses.

5.3.2 Development of gp64 knockout and Cm-GFP knock in bacmid genome

We developed a Δ gp64-Cm-GFP bacmid by utilizing the pRED/ET based homologous recombination approach. By following a detailed procedure for recombination (see materials and methods section, and Figure 5.4) we were able to create a deletion of gp64 coding region and incorporation of Cm and GFP coding regions in frame with gp64 promoter in gp64 coding region (Figures 5.5a and 5.5b).

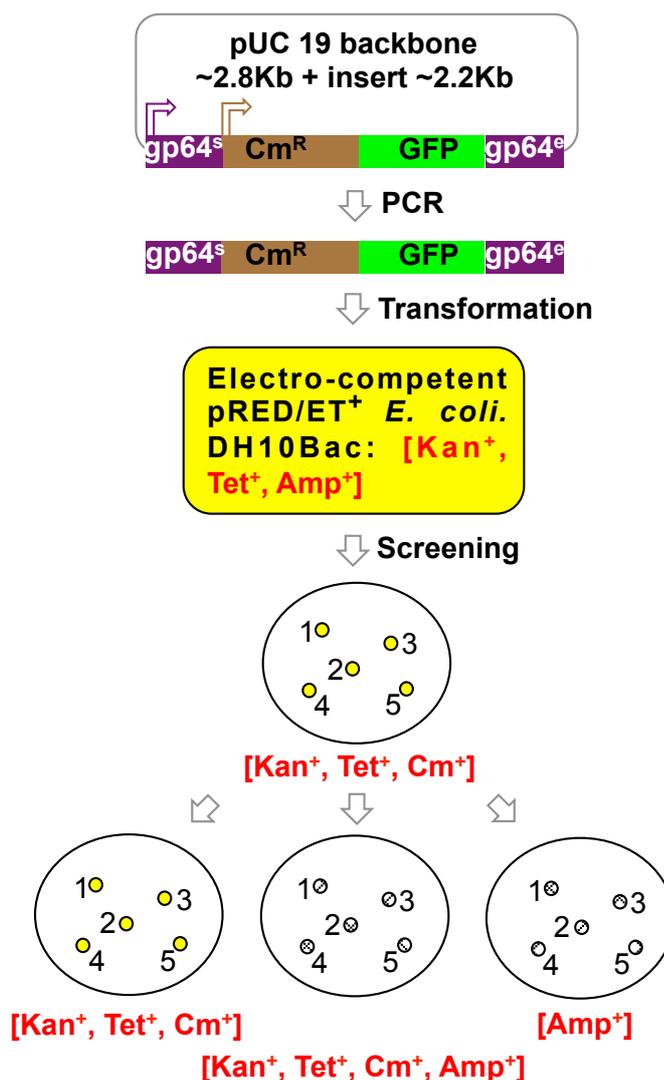


Figure 5.4 Steps for generation of recombinant bacmids through homologous recombination approach.

Our results from PCR verification using gp64 forward and reverse primers from bacmid that had undergone recombination (based on growth on Cm + LB agar plates as well as GFP fluorescence of these Cm resistant bacteria) revealed that a ~2.2 Kb (size of the recombination cassette) band could be seen in place of a ~1.5 Kb (size of gp64 coding region) band (Figure 5.5b).

This confirmed a successful homologous recombination event that led to Δ gp64-Cm-GFP bacmid generation. Further, these bacmid containing *E. coli* DH10Bac cells grew as blue colonies indicating intact MCS region where transposition can be performed later. We also observed GFP expression both in *E. coli* as well as sf9 cells suggesting the integrity and incorporation of the recombination cassette at right location in the bacmid.

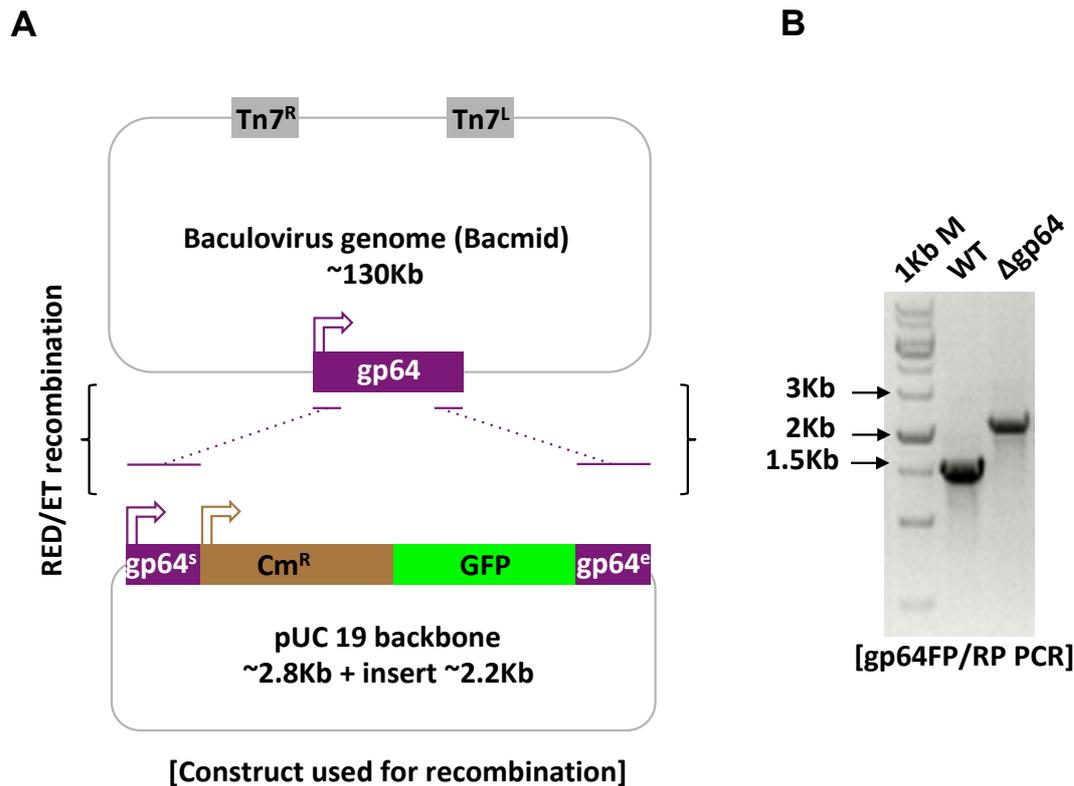


Figure 5.5 Homologous recombination verification. (A) scheme for homologous recombination process. (B) PCR verification of recombination of gp64-start-Cm-GFP-gp64-end cassette in place of gp64 using gp64 F/R primers with WT non-recombination control.

5.3.3 Generation of pseudo viruses with BV gp64 and VSV G proteins

We prepared clones with viral envelope protein coding sequences from baculovirus gp64, VSV G and CHIKV E3-E2-6K-E1 (Figure 5.6). We used these clones for

transposition inside *E. coli* DH10Bac cells and isolated the recombinant bacmids from these cells post transposition. Through PCR verification we confirmed that all envelope protein coding sequences are successfully transposed into the Δ gp64-Cm-GFP bacmid (Figure 5.7, left lane).

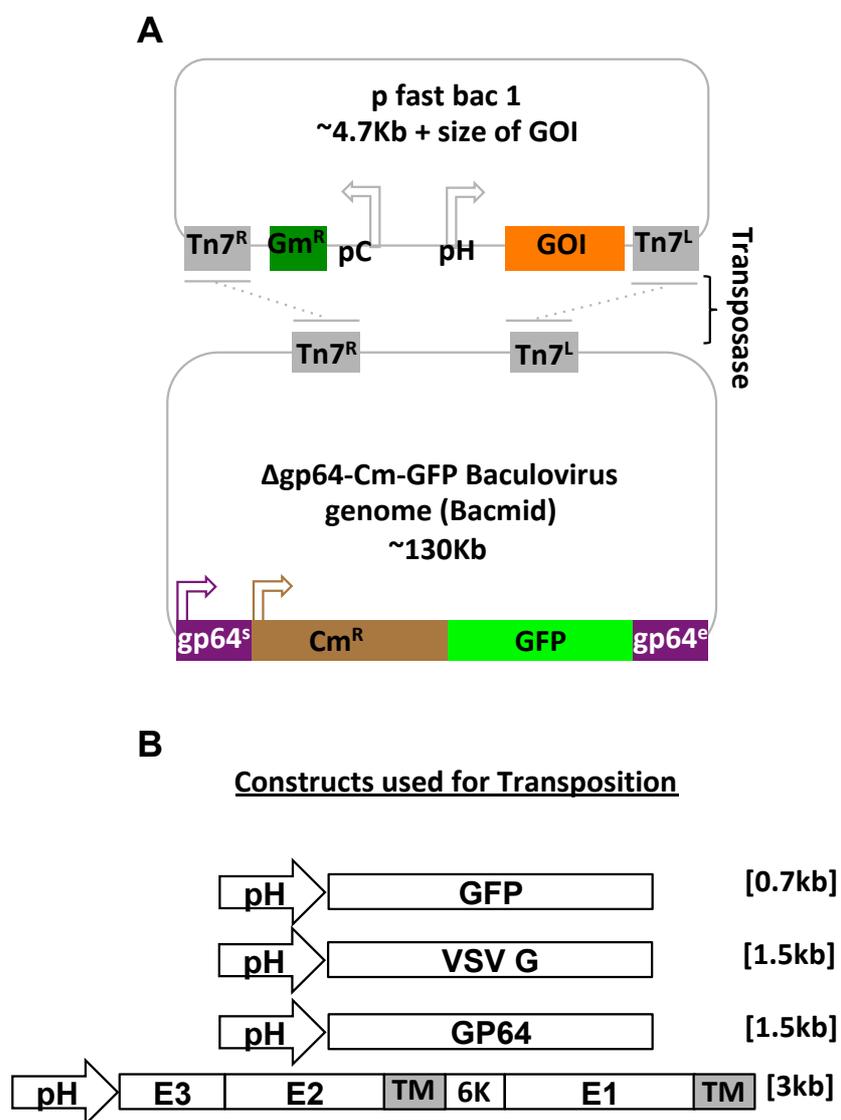


Figure 5.6 Transposition reaction procedure. (A) Scheme for transposition process. (B) Constructs for transposition reaction with sizes for coding regions mentioned to their right.

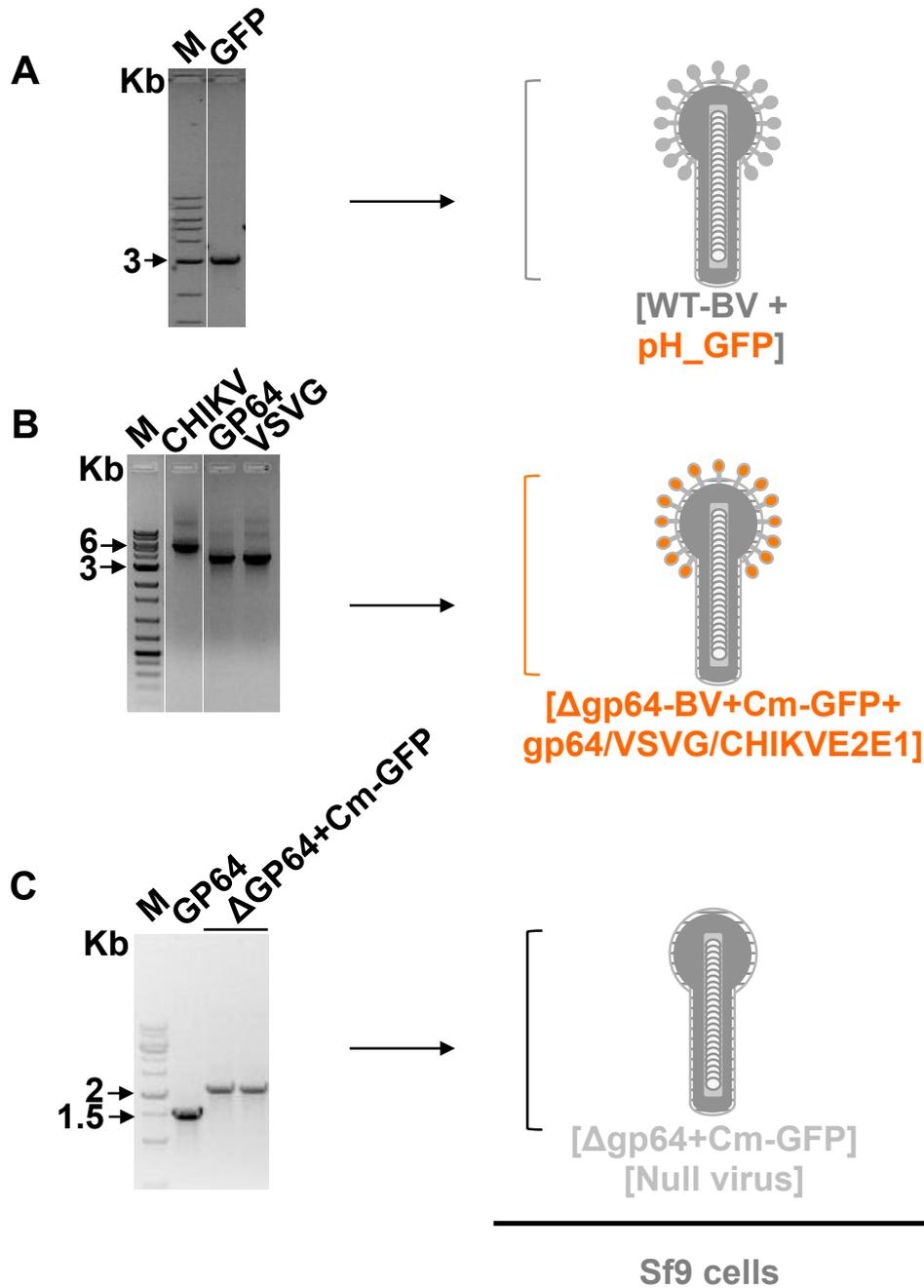


Figure 5.7 Verification of transposition reaction. (A) M13F/R PCR amplicon from GFP transposed bacmid on left, scheme of representative pseudovirion to right. (B) M13F/R PCR amplicons from CHIKV E3-E1, BV gp64 and VSV G transposed bacmid on left, scheme of representative pseudovirion to right. (C) gp64 F/R PCR amplicon from WT bacmid and Δ gp64-Cm-GFP bacmid on left, scheme of representative pseudovirion to right.

Now using these purified bacmids we performed transfections into sf9 cells for generation of pseudovirions (scheme in Figure 5.7, right lane). After 5-6 days post transfection in case of BV gp64 and VSV G expressing cells, spent medium from infected cells (with phenotype of dark and increased cell nucleus, granular cell morphologies, cessation of growth etc.) were collected. We removed the cell debris and used the clear supernatant fraction that contained the pseudotyped virus particles. We made several aliquots, some of which were kept at -20°C (for long term storage) and some of which were kept at 4°C (for regular use). We used these supernatants to infect fresh round of cells and observed that even with 1000X dilutions the new round of cells were ~100% infected with these inoculums in another 2-3 days (as observed through infected cell morphologies and GFP fluorescence as well).

5.3.4 Testing of BV gp64 and VSV G pseudotyped virions in functional assays

To test the functionality of complementation with heterologous viral envelope proteins (pseudo tying) we used the purified pseudo viruses expressing BV gp64 and VSV G on their envelope in syncytia (Figure 5.8) and cell entry/transduction assays (Figure 5.9) in sf9 cells. Our results from Syncytia assay shows that the gp64 knock out virus (null virus) failed to induce syncytia (Figure 5.8, third lane) in cell culture assay when triggered with acidic pH.

However, when we used the BV gp64 and VSV G complemented pseudotyped virus, both these viruses restored the syncytia forming ability (Figure 5.8, fifth and sixth lane) as seen with the WT gp64 containing virus (used as a positive control) (Figure 5.8, second lane). These results not only confirmed the expression of complemented BV gp64

and VSV G from the polyhedrin promoter but also showed functional packaging and surface expression of these proteins, which is essential for inducing syncytia formation.

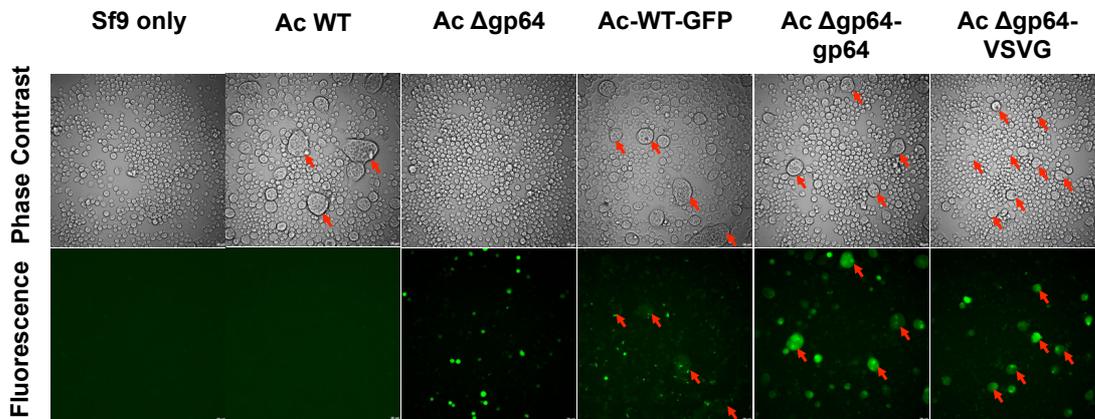


Figure 5.8 Syncytia assay. Labelling on top shows the name of bacmid used for transfection. Ac WT - WT bacmid, Ac Δ gp64 - gp64 knockout bacmid, Ac-WT-GFP – WT bacmid with GFP transposed, Ac Δ gp64-gp64 - gp64 transposed into Δ gp64 bacmid, Ac Δ gp64-VSVG - VSVG transposed into Δ gp64 bacmid.

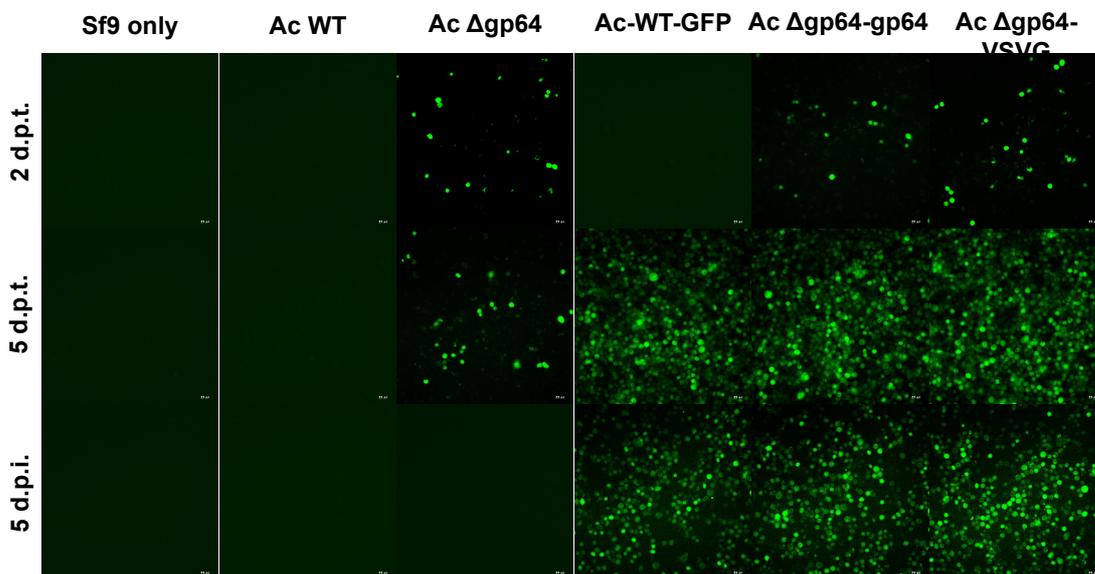


Figure 5.9 Cell entry assay. Labelling on top as in figure 6.6. Labelling on side shows time point and type of experiment. 2 d.p.t. and 5 d.p.t. - 2 days and 5 days post transfection. 5 d.p.i. - 5 days post infection.

Further, we used these pseudotyped viruses in cell entry assays (without any trigger) to observe if these can successfully do transduction into a fresh batch of sf9 cells. From our results it is clear that the null virus failed to perform cell entry (as evident from no GFP fluorescence) (Figure 5.9, third lane).

However, the BV gp64 and VSV G pseudotyped viruses could successfully do cell-transduction (Figure 5.9, fifth and sixth lane) followed by cell to cell infection (as evident from GFP fluorescence obtained at different time intervals post transfection and infection). This not only proved functionality of the pseudotyped particles but also their capability to replicate their genome inside infected sf9 cells and then package them further into progeny viruses for cell-cell infection. Taken together, our results from both syncytia and cell entry assays suggest that the pseudovirion system that we developed in this study is functional and also validates the proof of principle of the system development strategy.

5.3.5 Checking surface expression of CHIKV envelope proteins in sf9 cells

We checked surface expression of CHIKV envelope proteins from the Δ gp64-Cm-GFP bacmid genome in sf9 cells (Figure 5.10). We did this through immunofluorescence approach using the E1 protein specific polyclonal antibody developed in this study (refer to chapter 1, materials and methods section). From our IF results we could see surface expression of the CHIKV E1 protein (Figure 5.10, fourth lane). More fluorescence at cell surface was observed in case of CHIKV E1 (Figure 5.10, fourth lane) as compared to diffused GFP fluorescence throughout the inside of the transfected cells (Figure 5.10, third lane). We confirmed the fluorescence coming from E1, by taking control samples (one where E1 expression was absent, and one where we skipped the incubation step with E1-pAb and used only fluorophore tagged secondary antibody, data not shown). All these

results suggest that the CHIKV envelope proteins were successfully expressed at cell surface (a pre-requisite for successful pseudo typing by newly budded virus particles) from the polyhedrin promoter and from the Δ gp64-Cm-GFP bacmid backbone in sf9 cells.

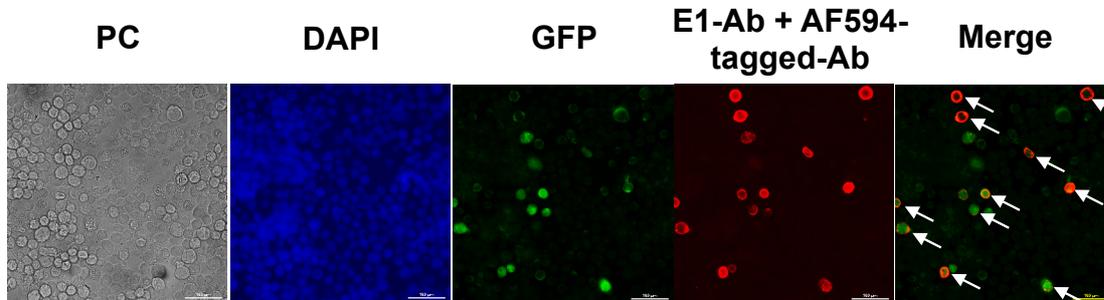


Figure 5.10 Immunofluorescence assay using E1-pAb on CHIKV envelope protein expressing sf9 cells. Labelling on top are as follows, PC – phase contrast image, DAPI – nuclear staining image, GFP – GFP fluorescence image, E1-Ab+AF594-tagged-Ab – AF594 tagged secondary Ab staining image with E1-pAb used as primary, Merge – Merge of GFP and AF594 fluorescence images.

5.3.6 Optimization of BV pseudo typing with CHIKV envelope proteins

Though we could successfully verify the surface expression of CHIKV envelope proteins from the Δ gp64-Cm-GFP bacmid genome in sf9 cells, we found it difficult to isolate the functional CHIKV E3-E2-6K-E1 pseudotyped viruses. Supernatants collected in the same manner as that of the BV gp64 or VSV G pseudotyped viruses in case of CHIKV did not result in syncytia formation or cell entry in cell culture based assays (Figure 5.11, also refer to figure 5.8 and 5.9).

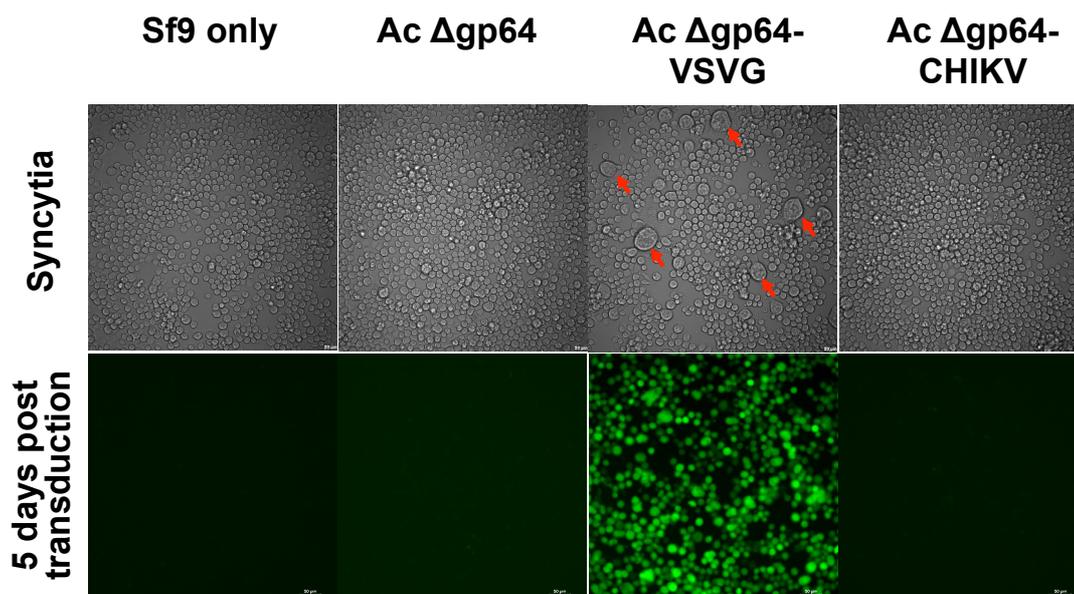


Figure 5.11 Testing of CHIKV pseudo type in syncytia and transduction assay. Top panel represents results from syncytia assay and bottom panel represents results from transduction assay. Labelling on top shows the name of bacmid used for transfection and virus generation. Ac Δ gp64 - gp64 knockout bacmid, Ac Δ gp64-VSVG - VSVG transposed into Δ gp64 bacmid, and Ac Δ gp64-CHIKV – CHIKV E3-E2-6K-E1 transposed into Δ gp64 bacmid.

Further, since CHIKV envelope fusion protein represents a complex of multiple associated proteins (unlike in BV gp64 and VSVG, where it's a single proteins from the class III fusion protein group) representative of class II fusion proteins (a different class than that of gp64 and VSV G), some problem might have occurred in functional assembly/packaging of these proteins inside sf9 cells towards making up of infectious virus particles. Assuming that even non-infectious viral particles were not produced (we have not tested this in the current study), we addressed the assembly problem further. We have tried expressing these proteins from different promoters (instead of the default polyhedrin promoter) from the bacmid genome. We used OPIE2 and even the WT gp64/Op166 promoter but in none of the cases any functional pseudo typing (based on our results from syncytia and cell entry assays, not shown) was observed. Some other approach has to be

tested (may be tagging these proteins with a C-terminal anchor region from BV gp64 itself that can interact with the BV genome and help in functional packaging) further to functionally pseudo type viral envelope proteins from different classes of viruses into the baculovirus pseudovirion system to make it a simple and useful tool for functional viral cell entry based studies.

5.4 DISCUSSION:

Pseudo viruses are non-infectious forms of actual virions and in general carry few viral components from the actual virus such as envelope entry proteins and can be utilized to study many processes of viral biology without the involvement of the infectious virion. Few well known pseudo virus making systems that are utilized in the field are Lentiviral packaging system (207) and VSVG helper virus mediated packaging system (206). Use of both these systems require use of attenuated/inactive components from pathogenic viruses with risk of re-activation. Also, since these are animal viruses chances of false positive signals coming from replication of the reporter coding sequence inside mammalian cells is also there. In addition, use of either a helper virus along with other envelope protein coding plasmids or use of multiple plasmids carrying different viral components required together for making a functional pseudovirion, is necessary (206, 207). All these requirements in addition to differential maintenance and use of mammalian cells (differences in results obtained with different batches of cells/different cell passages) leads to very poor efficiency of pseudo virus production. We planned and generated a simple pseudovirion system using an insect virus (Baculovirus) engineered to carry any viral envelope protein with ease. Our results are discussed in line with other available literature below.

We successfully generated the Δ gp64-Cm-GFP baculovirus genome through homologous recombination approach. We used the pRED/ET based recombination approach, which was extensively used earlier for manipulation of bacterial genomes to modify and engineer the baculovirus genome (a virus genome) for pseudovirion development (226–228). We manipulated the BV genome and using this modified genome, developed functional pseudotyped viruses that carry BV gp64 and VSV G proteins on their envelope. Our results from both syncytia and cell entry assays using these pseudo viruses suggest that the pseudovirion system that we developed in this study is functional and also validates the proof of principle of our system development strategy. However, though we could verify the surface expression of CHIKV envelope proteins from the Δ gp64-Cm-GFP bacmid genome in sf9 cells, we did not observe any functional CHIKV E3-E2-6K-E1 pseudotyped viruses (based on our results from functional assays). Earlier using different approaches, CHIKV envelope proteins were tested in pseudo typing or VLP production in sf9 cells, involving baculoviruses, in couple of studies (136, 229). In one case, only CHIKV 6K-E1 expressed from a recombinant baculovirus in Sf21 cells showed to have syncytia activity (136). In this study the WT envelope protein (gp64) was still intact in the recombinant baculovirus used to make CHIKV E1 pseudotyped viruses and the only distinction used to separate the fusion activity of gp64 versus E1 was a pH threshold of 5.0 for gp64 and 6.4 for E1 in the syncytia assay. The VLPs were produced as secreted out particles in Sf21 cells by expressing the complete structural cassette of the S27 strain of CHIKV from a recombinant baculovirus and were tested to be highly immunogenic and also showed protection against CHIKV infection in animal models (229). In this case also, the WT gp64 was still intact in the baculovirus and the structural cassette included the

capsid plus envelope protein coding sequences for VLP production. Further, since CHIKV envelope fusion protein represents a complex of multiple associated proteins from class II fusion proteins group that is different from the class III group of BV gp64 and VSV G proteins, we tested expressing CHIKV envelope proteins from different promoters (instead of the default polyhedrin promoter) from the bacmid genome. Role of such promoters in regulation of expression of different viral envelope proteins in insect cells using baculovirus genome has been discussed in earlier studies (230, 231). However, our trials involving OPIE2 and even the WT gp64/Op166 promoter (early promoters) in place or in combination with polyhedrin promoter did not result in functional pseudo typing. We believe some other approach involving tagging/expressing of these viral envelope proteins with a C-terminal anchor region (as a fusion protein) from BV gp64 itself may lead to WT BV like genome packaging and may lead to functional pseudo typing (probably through inducing native like viral envelope protein-viral genome interaction for packaging). This approach however need to be tested further in order to make the currently developed BV pseudovirion system an useful universal tool (capable of packaging different classes of viral envelope proteins) for use in functional viral cell entry assays.

SUMMARY

Chikungunya virus (CHIKV) is an aedes mosquito-transmitted alphavirus that is a public health threat. Similar to other enveloped viruses, CHIKV enters cells after fusion of its membrane with that of the host endosome. CHIKV cell entry is facilitated by its surface envelope-anchored proteins, E1 and E2, which exist as tightly associated heterodimer complex. And, three E2-E1 heterodimers assemble to form a trimeric spike-like structure on virion surface. In a trimeric spike, E2 is at the center, and E1 to the periphery. As first step in entry, E2 binds to a cell surface proteoglycan such as heparan sulfate (HS) or a protein receptor (for example mxra8). Receptor binding triggers endocytosis of the virion. In maturing endosome, upon endosome acidification, E1 performs viral – endosome membrane fusion.

Function of E2 and E1 are tightly regulated. E2 structurally masks E1 and thus regulates E1 function till cell-surface receptor is bound. Dissociation of E2-E1, upon receptor binding to E2, is a required step in cell entry. How receptor binding triggers E2-E1 dissociation is not known. In this thesis study, I characterized interaction of HS with E2. Sequence analysis, complemented with molecular docking of HS structure on to CHIKV E2-E1 crystal structure, identified a novel HS binding motif – XBXXBX (X is any residue; B is a basic residue) - on domain A of alphavirus E2. HS binding site identified in CHIKV E2 is structurally conserved amongst other alphaviruses. Then, using purified recombinant E2 protein, I biochemically characterized interaction with HS and heparin. HS and/or heparin bind to E2 protein through charge-charge interactions. Mutagenesis of basic residues from the predicted binding pocket resulted in significant loss of HS binding to E2. In order to explain E2-E1 dissociation upon receptor binding, I performed MD simulation of HS-bound E2 structure – HS binding results in allosteric movement of domain C. Fluorescence resonance energy transfer experiments and other biochemical studies with E2 protein and HS validated simulation predications. Based on these results, I explained a possible mechanism of E2-E1 dissociation upon receptor binding: E2 domain C movement, upon receptor binding, pushes E1 away. This breaks E2-E1 interface, leading to heterodimer dissociation.

Another regulatory feature in CHIKV (and other alphaviruses as well) entry is

endosomal acidic pH triggering conformational changes in E1, the membrane fusion protein. As per the unified structural mechanism of viral membrane fusion proteins (proposed based on structures of different fusion proteins in pre- and post-membrane fusion states) viral fusion proteins undergo pre- to post-fusion conformation switch during membrane fusion. The three β -sheet domains (dIII-dI-dII, in viral membrane proximal to distal order) of E1 rearrange to form an extended conformation (all domains arranged linearly) upon endosome acidification. Mechanism of acid sensing, and conformational changes to make the extended intermediate conformation are not known. Based on sequence and structure analysis, and molecular dynamics simulations on E1 at acidic pH, I explained the role of the linker region connecting the domain I and III of E1. Also, I designed cysteine stapling mutations in E1, and using large unilamellar vesicle (LUV) fusion assays, showed that when dIII and dI are cysteine-stapled, fusion is affected. Based on the results, I proposed that dIII-dI separation and a swiveling motion over the hinge connecting dIII-dI as a structural mechanism of acid-pH triggered E1 extended intermediate formation.

In alphaviruses all structural proteins are made as a single polypeptide in infected cell and upon proteolytic processing mature to E2, E1 and E3. E3 is retained on virion surface, associated with E2, only in few alphaviruses – CHIKV being one of those. Role of E3 in entry is not known. From structure of CHIKV E3-E2-E1 complex it is suggested that E3 may have a regulatory role, as it is positioned as a ‘brace’ on E2. E3 should move out for E2 and E1 to function. I hypothesized that E3 association with membrane may result in loss of E3-E2 interaction. I tested if E3 protein interacts with membrane using LUVs. In the experimental conditions that we used, E3 does not interact with membrane.

Over all, studies presented in this thesis contributed to explain structural mechanism of regulation of cell entry proteins in alphaviruses, which will be helpful for therapeutic development. I also contributed several important resources for studying structure-function relationship of alphavirus cell entry proteins.

APPENDICES

Table a1.1: List of overlapping primers for E3-gene synthesis.

Primer initials	Primer sequences
E3-R0	GCAAGACTTCTAGAATAC
E3-F0 (+ XbaI site)	GTATTCTAGAAGTCTTGCCATCCCAGTTATGTG
E3-R18	ATTTGCCAACAGGCACATAACTGGGATG
E3-F33	CCTGTTGGCAAATACCACGTTCCCC
E3-R46	GGCTGGGAGCAGGGGAACGTGGT
E3-F58	TGCTCCCAGCCCCCTTGCACGC
E3-R69	CGTAGCAGCAGGGCGTGCAAGGG
E3-F80	CCTGCTGCTACGAAAAGGAACCGGA
E3-R92	GCGTAGGGTTTCCTCCGGTTCCTTTT
E3-F105	GGAAACCCTACGCATGCTTGAGGACA
E3-R118	GGTCTCATGACGTTGTCCTCAAGCAT
E3-F131	ACGTCATGAGACCTGGGTACTATCAGC
E3-R144	GGATGCTTGTAGCAGCTGATAGTACCCA
E3-F158	TGCTACAAGCATCCTTAACATGTTCTCCC
E3-R172	GCTGGCGGTGGGGAGAACATGTTAA
E3-F187	CACCGCCAGCGACGCAGCAGATCTAG
E3-R197 (+ BglII site)	CTAGATCTGCTGCGTC

Table a1.2: List of overlapping primers for 6K-gene synthesis.

Primer initials	Primer sequences
6KR0	TGGTATGTGGCGAATTCAT
6KF0 (+ EcoRI site)	ATGAATTCGCCACATACCAAGAGGCTGCGATATACC
6KR19	TGCTGCTCGTTCCACAGGTATATCGCAGCCTCT
6KF36	TGTGGAACGAGCAGCAACCTTTGTTTTGGCTACAA
6KR52	CCAGCGGAATAAGGGCTTGTAGCCAAAACAAAGGT
6KF71	GCCCTTATTCCGCTGGCAGCCCTGATTGTTCTATG
6KR87	GGTAAGAGTCTCAGACAGTTGCATAGAACAATCAGGGCTG
6KF106	CAACTGTCTGAGACTCTTACCATGCTGCTGTAACGTT
6KR127	CTCATTACGGCTAAAAAAGCCAACGTTTTACAGCAGCAT
6KF145	GGCTTTTTTAGCCGTAATGAGCGTCGGTGCCCACT
6KR166 (+ NdeI site)	TTATCATATGCGCGCTCACAGTGTGGGCACCGACG
6KF182	GTGAGCGCGCATATGATAA

Table a1.3: *E. coli*, insect cell and mammalian cell expression construct details.

Construct details	Primer details
[pET24b+ E3] with C-terminal 6His-tag	FP (+NdeI site) - GCTCATATGAGTCTTGCCATCCCAGTTATG RP (+XhoI site) - CTCTCGAGGCGTCGCTGGCGGTGGGG
[pET24b+ E3-E2] with C-terminal 6His-tag	FP (+NdeI site) - GCTCATATGAGTCTTGCCATCCCAGTTATG RP (+NotI site) - ATGCGGCCGCGTACGAACGACCTTCGATGTACAGCT CATAATAATACA
[pET24b+ E2] with C-terminal 6His-tag	FP (+NdeI site) - AGCATATGACCAAGGACAACCTTCAATGTCTATA RP (+NotI site) - ATGCGGCCGCGTACGAACGACCTTCGATGTACAGCT CATAATAATACA
[pET24b+ E1] with C-terminal 6His-tag	FP (+NdeI site) - AGGCATATGGAACACGCAACAGTGATCCCGAAC RP (+HindIII site) - ATAAGCTTTTAGTGATGGTGATGGTGATGGTACGAAC GACCTTCGATCTGCACCCATGACATCGCCGTAGCGG AG
[pFB1+Mel-E3-E2-GS-E1-ectodomain] with N-terminal melittin signal and C-terminal 6His-tag	<u>Multiple primers:</u> Melittin FP (+EcoRI site) - GCTGAATTCATGAAATTCTTAGTCAACGTTGCC Melittin RP (+XbaI site) - GTATTCTAGACGCATAGATGTAAGAAATGTAC E3 FP (+XbaI site) - GTATTCTAGAAGTCTTGCCATCCCAGTTATGTG E3 RP (+BglII site) - GCTGGCGGTGGGGAGAACATGTAA E2 FP (+BglII site) - AGAGATCTACCAAGGACAACCTTCAATGTCTATA E2 RP (+BamHI site) - GAGGATCCGTACAGCTCATAATAATACAGAAT GS FP (+BamHI overhang) - GATTCGGGAGCGGGTCCCACCGCCAGCGACGCAGC ACCAAGGGGAGCGGGTCCCA GS RP (+NdeI overhang) - TATGGGACCCGCTCCCCTTGGTGCTGCGTCGCTGGC GGTGGGACCCGCTCCCG E1 FP (+NdeI site) - AGGCATATGGAACACGCAACAGTGATCCCGAAC

	<p>E1 RP (+HindIII-Stop-6XHis-FactorX region) - ATAAGCTTTTAGTGATGGTGGTGGTGGTACGAAC GACCTTCGATCTGCACCCATGACATCGCCGTAGCGG AG</p>
<p>[pcDNA3.1+E3-E2- 6K-E1] without any tags</p>	<p><u>Multiple primers:</u></p> <p>E3 FP (+HindIII site) - CGAAAGCTTATGAGTCTTGCCATCCCAGTTAT E3 RP (+BglIII site) - GCTGGCGGTGGGGAGAACATGTAA E2 FP (+BglIII site) - AGAGATCTACCAAGGACAACCTCAATGTCTATA E2 RP (+EcoRI site) - GGTGGAAATCCGCTTTAGCTGTTCTGATGC 6K FP (+EcoRI site) - ATGAATTCGCCACATACCAAGAGGCTGCGATATACC 6K RP (+NdeI site) - TTATCATATGCGCGCTCACAGTGTGGGCACCGACG E1 FP (+NdeI site) - AGGCATATGGAACACGCAACAGTGATCCCGAAC E1 RP (+XbaI site) - CGATCTAGATTAGTGCCTGCTGAACGACACGC</p>

Table a1.4: Homologous recombination cassette and heterologous envelope protein coding construct details.

Construct details	Primer details
[pUC19+(gp64-start+Cm-GFP+gp64-end)]	<p><u>Multiple primers:</u></p> <p>gp64 FP (+SpeI site) - CGAACTAGTAATGGTAAGCGCTATTG</p> <p>gp64 RP (+XbaI site) - GGCTCTAGAATTAATATTGTCTATTACGG</p> <p>Cm FP (+XhoI site) - ATCTCGAGGACGTTGATCGGCACGTAAG</p> <p>Cm RP (+XbaI site) - ATTCTAGACGCCCCGCCCTGCCAC</p> <p>GFP FP (+XbaI site) - ACTCTAGAATGGCTAGCAAAGGAGAAGAAC</p> <p>GFP RP (+NcoI site) - ATCCATGGTCAGCCATGTGTAATCCCAG</p>
[pFB1+VSVG]	Cut-pasted from pLP-VSVG vector using EcoRI site
[pFB1+BV gp64]	<p>gp64 FP (+SpeI site) - CGAACTAGTAATGGTAAGCGCTATTG</p> <p>gp64 RP (+XbaI site) - GGCTCTAGAATTAATATTGTCTATTACGG</p>
[pFB1+(E3-E2-6K-E1)]	<p><u>Multiple primers:</u></p> <p>E3 FP (+HindIII site) - CGAAAGCTTATGAGTCTTGCCATCCCAGTTAT</p> <p>E3 RP (+BglII site) - GCTGGCGGTGGGGAGAACATGTAA</p> <p>E2 FP (+BglII site) - AGAGATCTACCAAGGACAACTTCAATGTCTATA</p> <p>E2 RP (+EcoRI site) - GGTGGAATTCCGCTTTAGCTGTTCTGATGC</p> <p>6K FP (+EcoRI site) - ATGAATTCGCCACATACCAAGAGGCTGCGATAT ACC</p> <p>6K RP (+NdeI site) - TTATCATATGCGCGCTCACAGTGTGGGCACCGA CG</p> <p>E1 FP (+NdeI site) - AGGCATATGGAACACGCAACAGTGATCCCGAA C</p> <p>E1 RP (+XbaI site) - CGATCTAGATTAGTGCCTGCTGAACGACACGC</p>

REFERENCES

1. Gelderblom HR. 1996. Structure and Classification of Viruses, p. 1–12. *In* Medical Microbiology.
2. Marsh M, Helenius A. 2006. Virus entry: Open sesame. *Cell* 124:729–740.
3. Pöhlmann S, Simmons G. 2013. Viral Entry into Host Cells, p. 1–197. *In* Advances in Experimental Medicine and Biology.
4. Harrison SC. 2008. Viral membrane fusion. *Nat Struct Mol Biol* 15:690–698.
5. Plemper KR. 2012. Cell Entry of Enveloped Viruses. *Curr Opin Virol* 1:92–100.
6. Agelidis AM, Shukla D. 2015. Cell entry mechanisms of HSV: What we have learned in recent years. *Future Virol* 10:1145–1154.
7. Granzow H, Klupp BG, Mettenleiter TC. 2005. Entry of Pseudorabies Virus: an Immunogold-Labeling Study. *J Virol* 79:3200–3205.
8. Chen J, Lee KH, Steinhauer DA, Stevens DJ, Skehel JJ, Wiley DC, Hill M, Nw L. 1998. Structure of the Hemagglutinin Precursor Cleavage Site , a Determinant of Influenza Pathogenicity and the Origin of the Labile Conformation. *Cell* 95:409–417.
9. Eggink D, Baldwin CE, Deng Y, Langedijk JPM, Lu M, Sanders RW, Berkhout B. 2008. Selection of T1249-Resistant Human Immunodeficiency Virus Type 1 Variants. *J Virol* 82:6678–6688.
10. McCune JM, Rabin LB, Feinberg MB, Lieberman M, Kosek JC, Reyes GR, Weissman IL. 1988. Endoproteolytic cleavage of gp160 is required for the activation of human immunodeficiency virus. *Cell* 53:55–67.
11. Kielian M. 2014. Mechanisms of Virus Membrane Fusion Proteins. *Annu Rev Virol* 1:171–189.
12. Chernomordik L V., Kozlov MM. 2008. Mechanics of membrane fusion. *Nat Struct Mol Biol* 15:675–683.
13. Cohen FS, Melikyan GB. 2004. The energetics of membrane fusion from binding, through hemifusion, pore formation, and pore enlargement. *J Membr Biol* 199:1–14.
14. Melancon P, Garoff H. 1986. Reinitiation of translocation in the Semliki Forest virus structural polyprotein: identification of the signal for the E1 glycoprotein. *EMBO J* 5:1551–1560.

15. Carleton M, Lee H, Mulvey M, Brown DT. 1997. Role of glycoprotein PE2 in formation and maturation of the Sindbis virus spike. *J Virol* 71:1558–1566.
16. Mercer J, Schelhaas M, Helenius A. 2010. Virus Entry by Endocytosis. *Annu Rev Biochem* 79:803–833.
17. Huotari J, Helenius A. 2011. Endosome maturation. *EMBO J* 30:3481–3500.
18. Mellman I. 1992. The importance of being acid: the role of acidification in intracellular membrane traffic. *J Exp Biol* 172:39–45.
19. Mellman I, Fuchs R, Helenius A. 1986. Acidification of the endocytic and exocytic pathways. *Annu Rev Biochem* 55:663–700.
20. Kielian M, Rey FA. 2006. Virus membrane-fusion proteins: more than one way to make a hairpin. *Nat Rev Microbiol* 4:67–76.
21. White JM, Delos SE, Brecher M, Schornberg K. 2008. Structures and Mechanisms of Viral Membrane Fusion Proteins: Multiple Variations on a Common Theme. *Crit Rev Biochem Mol Biol* 43:189–219.
22. Cao S, Zhang W. 2013. Characterization of an early-stage fusion intermediate of Sindbis virus using cryoelectron microscopy. *Proc Natl Acad Sci* 110:13362–13367.
23. Calder LJ, Rosenthal PB. 2016. Cryomicroscopy provides structural snapshots of influenza virus membrane fusion. *Nat Struct Mol Biol* 23:853–858.
24. Chernomordik L V., Zimmerberg J, Kozlov MM. 2006. Membranes of the world unite! *J Cell Biol* 175:201–207.
25. Kuzmin PI, Zimmerberg J, Chizmadzhev YA, Cohen FS. 2001. A quantitative model for membrane fusion based on low-energy intermediates. *Proc Natl Acad Sci U S A* 98:7235–7240.
26. Kozlovsky Y, Chernomordik L V., Kozlov MM. 2002. Lipid intermediates in membrane fusion: Formation, structure, and decay of hemifusion diaphragm. *Biophys J* 83:2634–2651.
27. Ivanovic T, Choi JL, Whelan SP, van Oijen AM, Harrison SC. 2013. Influenza-virus membrane fusion by cooperative fold-back of stochastically induced hemagglutinin intermediates. *Elife* 2:e00333.
28. Gibbons DL, Erk I, Reilly B, Navaza J, Kielian M, Rey FA, Lepault J. 2003. Visualization of the Target-Membrane-Inserted Fusion Protein of Semliki Forest Virus by Combined Electron Microscopy and Crystallography. *Cell* 114:573–583.
29. Schibli DJ, Weissenhorn W. 2004. Class I and class II viral fusion protein structures reveal similar principles in membrane fusion. *Mol Membr Biol* 21:361–371.

30. Backovic M, Jardetzky TS. 2009. Class III viral membrane fusion proteins. *Curr Opin Struct Biol* 19:189–196.
31. Wilen CB, Tilton JC, Doms RW. 2012. HIV: Cell Binding and Entry. *Cold Spring Harb Perspect Med* 2:1–13.
32. Shaik MM, Peng H, Lu J, Rits-Volloch S, Xu C, Liao M, Chen B. 2019. Structural basis of coreceptor recognition by HIV-1 envelope spike. *Nature* 565:318–323.
33. Moller-Tank S, Maury W. 2015. Ebola Virus Entry: A Curious and Complex Series of Events. *PLoS Pathog* 11:e1004731.
34. Grove J, Marsh M. 2011. The cell biology of receptor-mediated virus entry. *J Cell Biol* 195:1071–1082.
35. Flynn DC, William Meyer It J, Mackenzie JM, Johnston RE, Meyer WJ, Mackenzie JM, Johnston RE. 1990. A Conformational Change in Sindbis Virus Glycoproteins E1 and E2 Is Detected at the Plasma Membrane as a Consequence of Early Virus-Cell Interaction. *J Virol* 64:3643–3653.
36. Navaratnarajah CK, Oezguen N, Rupp L, Kay L, Leonard VHJ, Braun W, Cattaneo R. 2011. The heads of the measles virus attachment protein move to transmit the fusion-triggering signal. *Nat Struct Mol Biol* 18:128–134.
37. Sahoo B, Chowdary TK. 2019. Conformational changes in Chikungunya virus E2 protein upon heparan sulfate receptor binding explain mechanism of E2-E1 dissociation during viral entry. *Biosci Rep* 39:1–14.
38. Wang H, Shi Y, Song J, Qi J, Lu G, Yan J, Gao GF. 2016. Ebola Viral Glycoprotein Bound to Its Endosomal Receptor Niemann-Pick C1. *Cell* 164:258–268.
39. Morrison TG. 2003. Structure and function of a paramyxovirus fusion protein. *Biochim Biophys Acta* 1614:73–84.
40. Jardetzky TS, Lamb RA. 2014. Activation of Paramyxovirus Membrane Fusion and Virus Entry. *Curr Opin Virol* 0:24–33.
41. Skehel JJ, Wiley DC. 2000. Receptor binding and membrane fusion in virus entry: the influenza hemagglutinin. *Annu Rev Biochem* 69:531–569.
42. Babel AR, Bruce J, Young JAT. 2007. The hr1 and fusion peptide regions of the subgroup B avian sarcoma and leukosis virus envelope glycoprotein influence low pH-dependent membrane fusion. *PLoS One* 2:e171.
43. Lavillette D, Boson B, Russell SJ, Cosset F-L. 2001. Activation of Membrane Fusion by Murine Leukemia Viruses Is Controlled in cis or in trans by Interactions between the Receptor-Binding Domain and a Conserved Disulfide Loop of the Carboxy Terminus of the Surface Glycoprotein. *J Virol* 75:3685–3695.

44. Fenouillet E, Barbouche R, Jones IM. 2007. Cell entry by enveloped viruses: Redox considerations for HIV and SARS-coronavirus. *Antioxidants Redox Signal* 9:1009–1034.
45. Smith JG, Cunningham JM. 2007. Receptor-induced thiolate couples Env activation to retrovirus fusion and infection. *PLoS Pathog* 3:1971–1981.
46. Stiasny K, Kossel C, Lepault J, Rey FA, Heinz FX. 2006. Characterization of a Structural Intermediate of Flavivirus Membrane Fusion. *PLoS Pathog* 3:e20.
47. Wiley DC, Skehel JJ. 1987. The structure and function of the hemagglutinin membrane glycoprotein of influenza virus. *Annu Rev Biochem* 56:365–394.
48. Kampmann T, Mueller DS, Mark AE, Young PR, Kobe B. 2006. The Role of Histidine Residues in Low-pH-Mediated Viral Membrane Fusion. *Structure* 14:1481–1487.
49. Leiman PG, Shneider MM. 2012. Viral Entry Machines, p. 93–223. *In Advances in Experimental Medicine and Biology*.
50. Moore JP, Doms RW. 2003. The entry of entry inhibitors: A fusion of science and medicine. *Proc Natl Acad Sci* 100:10598–10602.
51. Stencel-Baerenwald JE, Reiss K, Reiter DM, Stehle T, Dermody TS. 2014. The sweet spot: defining virus–sialic acid interactions. *Nat Rev Microbiol* 12:739–749.
52. Baleux F, Loureiro-Morais L, Hersant Y, Clayette P, Arenzana-Seisdedos F, Bonnaffé D, Lortat-Jacob H. 2009. A synthetic CD4-heparan sulfate glycoconjugate inhibits CCR5 and CXCR4 HIV-1 attachment and entry. *Nat Chem Biol* 5:743–748.
53. Cianci C, Langley DR, Dischino DD, Sun Y, Yu K-L, Stanley A, Roach J, Li Z, Dalterio R, Colonna R, Meanwell NA, Krystal M. 2004. Targeting a binding pocket within the trimer-of-hairpins: Small-molecule inhibition of viral fusion. *Proc Natl Acad Sci* 101:15046–15051.
54. Russell RJ, Kerry PS, Stevens DJ, Steinhauer DA, Martin SR, Gamblin SJ, Skehel JJ. 2008. Structure of influenza hemagglutinin in complex with an inhibitor of membrane fusion. *Proc Natl Acad Sci* 105:17736–17741.
55. Frey G, Rits-Volloch S, Zhang X-Q, Schooley RT, Chen B, Harrison SC. 2006. Small molecules that bind the inner core of gp41 and inhibit HIV envelope-mediated fusion. *Proc Natl Acad Sci* 103:13938–13943.
56. Zhou Z, Khaliq M, Suk JE, Patkar C, Li L, Kuhn RJ, Post CB. 2008. Antiviral compounds discovered by virtual screening of small-molecule libraries against dengue virus E protein. *ACS Chem Biol* 3:765–775.
57. Lambert DM, Barney S, Lambert AL, Guthrie K, Medinas R, Davis DE, Bucy T,

- Erickson J, Merutka G, Petteway SR. 1996. Peptides from conserved regions of paramyxovirus fusion (F) proteins are potent inhibitors of viral fusion. *Proc Natl Acad Sci* 93:2186–2191.
58. Wild CT, Shugars DC, Greenwell TK, McDanal CB, Matthews TJ. 1994. Peptides corresponding to a predictive alpha-helical domain of human immunodeficiency virus type 1 gp41 are potent inhibitors of virus infection. *Proc Natl Acad Sci* 91:9770–9774.
 59. Hrobowski YM, Garry RF, Michael SF. 2005. Peptide inhibitors of dengue virus and West Nile virus infectivity. *Viol J* 2:1–10.
 60. Liao M, Kielian M. 2005. Domain III from class II fusion proteins functions as a dominant-negative inhibitor of virus membrane fusion. *J Cell Biol* 171:111–120.
 61. Smit JM, Moesker B, Rodenhuis-Zybert I, Wilschut J. 2011. Flavivirus cell entry and membrane fusion. *Viruses* 3:160–171.
 62. Kielian M, Chanel-Vos C, Liao M. 2010. Alphavirus entry and membrane fusion. *Viruses* 2:796–825.
 63. Tsetsarkin KA, Weaver SC. 2011. Sequential adaptive mutations enhance efficient vector switching by chikungunya virus and its epidemic emergence. *PLoS Pathog* 7:e1002412.
 64. Tsetsarkin KA, Chen R, Sherman MB, Weaver SC. 2011. Chikungunya virus: evolution and genetic determinants of emergence. *Curr Opin Virol* 1:310–317.
 65. Schwartz O, Albert ML. 2010. Biology and pathogenesis of chikungunya virus. *Nat Rev Microbiol* 8:491–500.
 66. Ganesan VK, Duan B, Reid SP. 2017. Chikungunya virus: Pathophysiology, mechanism, and modeling. *Viruses* 9:1–14.
 67. Lum FM, Ng LFP. 2015. Cellular and molecular mechanisms of chikungunya pathogenesis. *Antiviral Res* 120:165–174.
 68. Gérardin P, Barau G, Michault A, Bintner M, Randrianaivo H, Choker G, Lenglet Y, Touret Y, Bouveret A, Grivard P, Le Roux K, Blanc S, Schuffenecker I, Couderc T, Arenzana-Seisdedos F, Lecuit M, Robillard PY. 2008. Multidisciplinary prospective study of mother-to-child chikungunya virus infections on the island of La Réunion. *PLoS Med* 5:e60.
 69. Couderc T, Lecuit M. 2015. Chikungunya virus pathogenesis: From bedside to bench. *Antiviral Res* 121:120–131.
 70. Economopoulou A, Dominguez M, Helynck B, Sissoko D, Wichmann O, Quenel P, Germonneau P, Quatresous I. 2009. Atypical Chikungunya virus infections: Clinical

- manifestations, mortality and risk factors for severe disease during the 2005-2006 outbreak on Réunion. *Epidemiol Infect* 137:534–541.
71. Thiberville SD, Moyen N, Dupuis-Maguiraga L, Nougairede A, Gould EA, Roques P, de Lamballerie X. 2013. Chikungunya fever: Epidemiology, clinical syndrome, pathogenesis and therapy. *Antiviral Res* 99:345–370.
 72. Rougeron V, Sam I-CC, Caron M, Nkoghe D, Leroy E, Roques P. 2015. Chikungunya, a paradigm of neglected tropical disease that emerged to be a new health global risk. *J Clin Virol* 64:144–152.
 73. Martínez-Pulgarín DF, Chowdhury FR, Villamil-Gomez WE, Rodriguez-Morales AJ, Blohm GM, Paniz-Mondolfi AE. 2016. Ophthalmologic aspects of chikungunya infection. *Travel Med Infect Dis* 14:451–457.
 74. Chandak NH, Kashyap RS, Kabra D, Karandikar P, Saha SS, Morey SH, Purohit HJ, Taori GM, Dagainawala HF. 2009. Neurological complications of Chikungunya virus infection. *Neurol India* 57:177–180.
 75. Alvarez MF, Bolívar-Mejía A, Rodriguez-Morales AJ, Ramirez-Vallejo E. 2017. Cardiovascular involvement and manifestations of systemic Chikungunya virus infection: A systematic review. *F1000Research* 6:1–22.
 76. Kielian M. 2006. Class II virus membrane fusion proteins. *Virology* 344:38–47.
 77. van Duijl-Richter M, Hoornweg T, Rodenhuis-Zybert I, Smit J. 2015. Early Events in Chikungunya Virus Infection—From Virus Cell Binding to Membrane Fusion. *Viruses* 7:3647–3674.
 78. Leung JY-S, Ng MM-L, Chu JJH. 2011. Replication of alphaviruses: a review on the entry process of alphaviruses into cells. *Adv Virol* 2011:249640.
 79. Lobigs M, Zhao HX, 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.
 80. Li L, Jose J, Xiang Y, Kuhn RJ, Rossmann MG. 2010. Structural changes of envelope proteins during alphavirus fusion. *Nature* 468:705–708.
 81. Garoff H, Huylebroeck D, Robinson A, Tillman U, Liljeström P. 1990. The Signal Sequence of the p62 Protein of Semliki Forest Virus is Involved in Initiation but Not in Completing Chain Translocation. *J Cell Biol* 111:867–876.
 82. Voss JE, Vaney MC, Duquerroy S, Vonrhein C, Girard-Blanc C, Crublet E, Thompson A, Bricogne G, Rey FA. 2010. Glycoprotein organization of Chikungunya virus particles revealed by X-ray crystallography. *Nature* 468:709–712.

83. Uchime O, Fields W, Kielian M. 2013. The Role of E3 in pH Protection during Alphavirus Assembly and Exit. *J Virol* 87:10255–10262.
84. Fields W, Kielian M. 2015. Interactions involved in pH protection of the alphavirus fusion protein. *Virology* 486:173–179.
85. Sjoberg M, Lindqvist B, Garoff H. 2011. Activation of the Alphavirus Spike Protein Is Suppressed by Bound E3. *J Virol* 85:5644–5650.
86. Paredes AM, Heidner H, Thuman-Commike P, Prasad B V, Johnston RE, Chiu W. 1998. Structural localization of the E3 glycoprotein in attenuated Sindbis virus mutants. *J Virol* 72:1534–1541.
87. Wu S-R, Haag L, Sjöberg M, Garoff H, Hammar L. 2008. The Dynamic Envelope of a Fusion Class II Virus. *J Biol Chem* 283:26452–26460.
88. Gibbons DL, Vaney MC, Roussel A, Vigouroux A, Reilly B, Lepault J, Kielian M, Rey FA. 2004. Conformational change and protein-protein interactions of the fusion protein of Semliki Forest virus. *Nature* 427:320–325.
89. Zheng Y, Sanchez-San Martin C, Qin Z -l., Kielian M. 2011. The Domain I-Domain III Linker Plays an Important Role in the Fusogenic Conformational Change of the Alphavirus Membrane Fusion Protein. *J Virol* 85:6334–6342.
90. Basore K, Kim AS, Nelson CA, Zhang R, Smith BK, Uranga C, Vang L, Cheng M, Gross ML, Smith J, Diamond MS, Fremont DH. 2019. Cryo-EM Structure of Chikungunya Virus in Complex with the Mxra8 Receptor. *Cell* 177:1–13.
91. Zhang R, Hryc CF, Cong Y, Liu X, Jakana J, Gorchakov R, Baker ML, Weaver SC, Chiu W. 2011. 4.4 Å cryo-EM structure of an enveloped alphavirus Venezuelan equine encephalitis virus. *EMBO J* 30:3854–3863.
92. Lan Yap M, Klose T, Urakami A, Saif Hasan S, Akahata W, Rossmann MG. 2017. Structural studies of Chikungunya virus maturation. *Proc Natl Acad Sci* 114:13703–13707.
93. Wintachai P, Wikan N, Kuadkitkan A, Jaimipuk T, Ubol S, Pulmanausahakul R, Auewarakul P, Kasinrerk W, Weng W, Panyasrivanit M, Paemanee A, Kittisenachai S, Roytrakul S, Smith DR. 2012. Identification of Prohibitin as a Chikungunya Virus Receptor Protein. *J Med Virol* 84:1757–1770.
94. Moller-Tank S, Kondratowicz AS, Davey RA, Rennert PD, Maury W. 2013. Role of the phosphatidylserine receptor TIM-1 in enveloped-virus entry. *J Virol* 87:8327–8241.
95. Fongsaran C, Jirakanwisal K, Kuadkitkan A, Wikan N, Wintachai P, Thepparit C, Ubol S, Phaonakrop N, Roytrakul S, Smith DR. 2014. Involvement of ATP synthase β subunit in chikungunya virus entry into insect cells. *Arch Virol* 159:3353–3364.

96. Zhang R, Kim AS, Fox JM, Nair S, Basore K, Klimstra WB, Rimkunas R, Fong RH, Lin H, Poddar S, Crowe JE, Doranz BJ, Fremont DH, Diamond MS. 2018. Mxra8 is a receptor for multiple arthritogenic alphaviruses. *Nature* 557:570–574.
97. Song H, Zhao Z, Chai Y, Jin X, Li C, Yuan F, Liu S, Gao Z, Wang H, Song J, Vazquez L, Zhang Y, Tan S, Morel CM, Yan J, Shi Y, Qi J, Gao F, Gao GF. 2019. Molecular Basis of Arthritogenic Alphavirus Receptor MXRA8 Binding to Chikungunya Virus Envelope Protein. *Cell* 177:1–11.
98. Mettenleiter TC. 2002. Brief overview on cellular virus receptors. *Virus Res* 82:3–8.
99. Cardin AD, Weintraub HJ. 1989. Molecular modeling of protein-glycosaminoglycan interactions. *Arteriosclerosis* 9:21–32.
100. Capila I, Linhardt RJ. 2002. Heparin-protein interactions. *Angew Chem Int Ed Engl* 41:391–412.
101. Tanaka A, Tumkosit U, Nakamura S, Motooka D, Kishishita N, Priengprom T, Sa-
Ngasang A, Kinoshita T, Takeda N, Maeda Y. 2017. Genome-Wide Screening
Uncovers the Significance of N-Sulfation of Heparan Sulfate as a Host Cell Factor
for Chikungunya Virus Infection. *J Virol* 91:e00432-17.
102. Weber C, Berberich E, Von Rhein C, Henß L, Hildt E, Schnierle BS. 2017.
Identification of Functional Determinants in the Chikungunya Virus E2 Protein.
PLoS Negl Trop Dis 11:e0005318.
103. Silva L a, Khomandiak S, Ashbrook AW, Weller R, Heise MT, Morrison TE,
Dermody TS. 2014. A single-amino-acid polymorphism in Chikungunya virus E2
glycoprotein influences glycosaminoglycan utilization. *J Virol* 88:2385–2397.
104. Ciano KA, Saredy JJ, Bowers DF. 2014. Heparan Sulfate Proteoglycan: An
Arbovirus Attachment Factor Integral to Mosquito Salivary Gland Ducts. *Viruses*
6:5182–5197.
105. Hasan SS, Sun C, Kim AS, Watanabe Y, Chen CL, Klose T, Buda G, Crispin M,
Diamond MS, Klimstra WB, Rossmann MG. 2018. Cryo-EM Structures of Eastern
Equine Encephalitis Virus Reveal Mechanisms of Virus Disassembly and Antibody
Neutralization. *Cell Rep* 25:3136–3147.
106. Abell BA, Brown DT. 1993. Sindbis Virus Membrane Fusion Is Mediated by
Reduction of Glycoprotein Disulfide Bridges at the Cell Surface. *J Virol* 67:5496–
5501.
107. Doxsey SJ, Brodsky FM, Blank GS, Helenius A. 1987. Inhibition of Endocytosis by
Anti-Clathrin Antibodies. *Cell* 50:453–463.
108. Sieczkarski SB, Whittaker GR. 2003. Differential requirements of Rab5 and Rab7

- for endocytosis of influenza and other enveloped viruses. *Traffic* 4:333–343.
109. Kielian MC, Marsh M, Helenius A. 1986. Kinetics of endosome acidification detected by mutant and wild-type Semliki Forest virus. *EMBO J* 5:3103–3109.
 110. Glomb-Reinmund S, Kielian M. 1998. The role of low pH and disulfide shuffling in the entry and fusion of Semliki Forest virus and Sindbis virus. *Virology* 248:372–381.
 111. Van Duijl-Richter MKS, Blijleven JS, van Oijen AM, Smit JM. 2015. Chikungunya virus fusion properties elucidated by single-particle and bulk approaches. *J Gen Virol* 96:2122–2132.
 112. Fan DP, Sefton BM. 1978. The entry into host cells of Sindbis virus, vesicular stomatitis virus and Sendai virus. *Cell* 15:985–992.
 113. Irurzun A, Nieva JL, Carrasco L. 1997. Entry of Semliki forest virus into cells: Effects of concanamycin A and nigericin on viral membrane fusion and infection. *Virology* 227:488–492.
 114. DeTulleo L, Kirchhausen T. 1998. The clathrin endocytic pathway in viral infection. *EMBO J* 17:4585–4593.
 115. 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.
 116. Helenius A, Marsh M, White J. 1982. Inhibition of Semliki Forest virus penetration by lysosomotropic weak bases. *J Gen Virol* 58:47–61.
 117. Marsh M, Wellsted J, Kern H, Harms E, Helenius A. 1982. Monensin inhibits Semliki Forest virus penetration into culture cells. *Proc Natl Acad Sci* 79:5297–5301.
 118. Fuller SD, Berriman JA, Butcher SJ, Gowen BE. 1995. Low pH Induces Swiveling of the Glycoprotein Heterodimers in the Semliki Forest Virus Spike Complex. *Cell* 81:715–725.
 119. Zaitseva E, Mittal A, Griffin DE, Chernomordik L V. 2005. Class II fusion protein of alphaviruses drives membrane fusion through the pathway as class I proteins. *J Cell Biol* 169:167–177.
 120. Liu CY, Besanceney C, Song Y, Kielian M. 2010. Pseudorevertants of a Semliki Forest Virus Fusion-Blocking Mutation Reveal a Critical Interchain Interaction in the Core Trimer. *J Virol* 84:11624–11633.
 121. Zeng X, Mukhopadhyay S, Brooks CL. 2015. Residue-level resolution of alphavirus envelope protein interactions in pH-dependent fusion. *Proc Natl Acad Sci* 112:2034–2039.

122. Liu CY, Kielian M. 2009. E1 Mutants Identify a Critical Region in the Trimer Interface of the Semliki Forest Virus Fusion Protein. *J Virol* 83:11298–11306.
123. Yoshii K, Igarashi M, Ichii O, Yokozawa K, Ito K, Kariwa H, Takashima I. 2012. A conserved region in the prM protein is a critical determinant in the assembly of flavivirus particles. *J Gen Virol* 93:27–38.
124. Peng T, Ponce-de-Leon M, Jiang H, Dubin G, Lubinski JM, Eisenberg RJ, Cohen GH. 1998. The gH-gL complex of herpes simplex virus (HSV) stimulates neutralizing antibody and protects mice against HSV type 1 challenge. *J Virol* 72:65–72.
125. Binley JM, Sanders RW, Clas B, Schuelke N, Master A, Guo Y, Kajumo F, Anselma DJ, Maddon PJ, Olson WC, Moore JP. 2000. A Recombinant Human Immunodeficiency Virus Type 1 Envelope Glycoprotein Complex Stabilized by an Intermolecular Disulfide Bond between the gp120 and gp41 Subunits Is an Antigenic Mimic of the Trimeric Virion-Associated Structure. *J Virol* 74:627–643.
126. Lineberger JE, Danzeisen R, Hazuda DJ, Simon AJ, Miller MD. 2002. Altering Expression Levels of Human Immunodeficiency Virus Type 1 gp120-gp41 Affects Efficiency but Not Kinetics of Cell-Cell Fusion. *J Virol* 76:3522–3533.
127. Mohan GS, Ye L, Li W, Monteiro A, Lin X, Sapkota B, Pollack BP, Compans RW, Yang C. 2015. Less Is More: Ebola Virus Surface Glycoprotein Expression Levels Regulate Virus Production and Infectivity. *J Virol* 89:1205–1217.
128. Santiago FW, Emo KL, Fitzgerald T, Treanor JJ, Topham DJ. 2012. Antigenic and immunogenic properties of recombinant hemagglutinin proteins from H1N1 A/Brisbane/59/07 and B/Florida/04/06 when produced in various protein expression systems. *Vaccine* 30:4606–4616.
129. Gong Z, Kessans SA, Song L, Dörner K, Lee HH, Meador LR, Labaer J, Hogue BG, Mor TS, Fromme P. 2014. Recombinant expression, purification, and biophysical characterization of the transmembrane and membrane proximal domains of HIV-1 gp41. *Protein Sci* 23:1607–1618.
130. Talha SM, Nemani SK, Salminen T, Kumar S, Swaminathan S, Soukka T, Pettersson K, Khanna N. 2012. Escherichia coli-expressed near full length HIV-1 envelope glycoprotein is a highly sensitive and specific diagnostic antigen. *BMC Infect Dis* 12:1–11.
131. Bommakanti G, Citron MP, Hepler RW, Callahan C, Heidecker GJ, Najjar TA, Lu X, Joyce JG, Shiver JW, Casimiro DR, Meulen J Ter, Liang X, Varadarajan R. 2010. Design of an HA2-based escherichia coli expressed influenza immunogen that protects mice from pathogenic challenge. *Proc Natl Acad Sci* 107:13701–13706.
132. Khurana S, Verma S, Verma N, Crevar CJ, Carter DM, Manischewitz J, King LR, Ross TM, Golding H. 2010. Properly folded bacterially expressed H1N1

- hemagglutinin globular head and ectodomain vaccines protect ferrets against H1N1 pandemic influenza virus. *PLoS One* 5:1–11.
133. Smit JM, Bittman R, Wilschut J. 1999. Low-pH-dependent fusion of Sindbis virus with receptor-free cholesterol- and sphingolipid-containing liposomes. *J Virol* 73:8476–8484.
 134. Klimjack MR, Jeffrey S, Kielian M. 1994. Membrane and protein interactions of a soluble form of the Semliki Forest virus fusion protein. *J Virol* 68:6940–6946.
 135. Kielian M, Helenius A. 1985. pH-induced alterations in the fusogenic spike protein of Semliki Forest virus. *J Cell Biol* 101:2284–2291.
 136. Metz SW, Geertsema C, Martina BE, Andrade P, Heldens JG, Van Oers MM, Vlak JM, Pijlman GP. 2011. Functional processing and secretion of Chikungunya virus E1 and E2 glycoproteins in insect cells. *Virology* 418:1–12.
 137. Dudha N, Rana J, Gabrani R, Gupta A, Chaudhary VK, Gupta S. 2014. Small scale expression, solubilization, and characterization of Chikungunya virus structural proteins. *Asian J Pharm Clin Res* 7:268–271.
 138. Verma A, Chandele A, Nayak K, Kaja MK, Arulandu A, Lodha R, Ray P. 2016. High yield expression and purification of Chikungunya virus E2 recombinant protein and its evaluation for serodiagnosis. *J Virol Methods* 235:73–79.
 139. Sanchez-San Martin C, Nanda S, Zheng Y, Fields W, Kielian M. 2013. Cross-Inhibition of Chikungunya Virus Fusion and Infection by Alphavirus E1 Domain III Proteins. *J Virol* 87:7680–7687.
 140. Rouillard J-M, Lee W, Truan G, Gao X, Zhou X, Gulari E. 2004. Gene2Oligo: oligonucleotide design for in vitro gene synthesis. *Nucleic Acids Res* 32:W176–W180.
 141. Sun S, Xiang Y, Akahata W, Holdaway H, Pal P, Zhang X, Diamond MS, Nabel GJ, Rossmann MG. 2013. Structural analyses at pseudo atomic resolution of Chikungunya virus and antibodies show mechanisms of neutralization. *Elife* 2:e00435.
 142. Gibbons DL, Reilly B, Ahn A, Vaney M-C, Vigouroux A, Rey FA, Kielian M, Gibbons DL, Vigouroux A, Kielian M, Ahn A, Rey FA, Reilly B. 2004. Purification and Crystallization Reveal Two Types of Interactions of the Fusion Protein Homotrimer of Semliki Forest Virus. *J Virol* 78:3514–3523.
 143. Gibbons DL, Kielian M. 2010. Molecular Dissection of the Semliki Forest Virus Homotrimer Reveals Two Functionally Distinct Regions of the Fusion Protein. *J Virol* 76:1194–1205.
 144. Wahlberg JM, 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.
145. Wahlberg JM, Bron R, Wilschut J, Garoff H. 1992. Membrane fusion of Semliki Forest virus involves homotrimers of the fusion protein. *J Virol* 66:7309–7318.
 146. Martín CS-S, Liu CY, Kielian M. 2009. Dealing with low pH: entry and exit of alphaviruses and flaviviruses. *Trends Microbiol* 17:514–521.
 147. Waarts B-LL, Smit JM, Aneke OJC, McInerney GM, Liljestrom P, Bittman R, Wilschut J. 2005. Reversible acid-induced inactivation of the membrane fusion protein of Semliki Forest virus. *J Virol* 79:7942–7948.
 148. Hernandez R, Sinodis C, Horton M, Ferreira D, Yang C, Brown DT. 2003. Deletions in the Transmembrane Domain of a Sindbis Virus Glycoprotein Alter Virus Infectivity, Stability, and Host Range. *J Virol* 77:12710–12719.
 149. Sjoberg 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.
 150. Bron R, Wahlberg JM, Garoff H, Wilschut J. 1993. Membrane fusion of Semliki Forest virus in a model system: correlation between fusion kinetics and structural changes in the envelope glycoprotein. *EMBO J* 12:693–701.
 151. Trott O, Olson AJ. 2009. Software news and update AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J Comput Chem* 31:455–461.
 152. Kozakov D, Hall DR, Xia B, Porter KA, Padhorney D, Yueh C, Beglov D, Vajda S. 2017. The ClusPro web server for protein-protein docking. *Nat Protoc* 12:255–278.
 153. Bailey TL, Johnson J, Grant CE, Noble WS. 2015. The MEME Suite. *Nucleic Acids Res* 43:W39–W49.
 154. Abraham MJ, Murtola T, Schulz R, Páll S, Smith JC, Hess B, Lindahl E. 2015. Gromacs: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* 1–2:19–25.
 155. Chen C, Hasan SS, Klose T, Sun Y, Buda G, Sun C. 2020. Cryo-EM structure of eastern equine encephalitis virus in complex with heparan sulfate analogues. *Proc Natl Acad Sci U S A* 117:8890–8899.
 156. Richard B, Swanson R, Izaguirre G, Olson ST. 2018. Cooperative Interactions of Three Hotspot Heparin Binding Residues Are Critical for Allosteric Activation of Antithrombin by Heparin. *Biochemistry* 57:2211–2226.
 157. Peysselon F, Ricard-Blum S. 2014. Heparin-protein interactions: From affinity and

- kinetics to biological roles. Application to an interaction network regulating angiogenesis. *Matrix Biol* 35:73–81.
158. Sasmal DK, Pulido L, Kasal S, Huang J. 2016. Single-Molecule Fluorescence Resonance Energy Transfer in Molecular Biology. *Nanoscale* 8:19928–19944.
 159. Smit JM, Waarts B, Kimata K, Klimstra WB, Bittman R, Wilschut J. 2002. Adaptation of Alphaviruses to Heparan Sulfate : Interaction of Sindbis and Semliki Forest Viruses with Liposomes Containing Lipid-Conjugated Heparin. *J Virol* 76:10128–10137.
 160. Zhu W, Wang L, Yang Y, Jia J, Fu S, Feng Y, He Y, Li J, Liang G. 2010. Interaction of E2 Glycoprotein with Heparan Sulfate Is Crucial for Cellular Infection of Sindbis Virus. *PLoS One* 5:e9656.
 161. Heil ML, Albee A, Strauss JH, Kuhn RJ. 2001. An amino acid substitution in the coding region of the E2 glycoprotein adapts Ross River virus to utilize heparan sulfate as an attachment moiety. *J Virol* 75:6303–6309.
 162. Gardner CL, Ebel GD, Ryman KD, Klimstra WB. 2011. Heparan sulfate binding by natural eastern equine encephalitis viruses promotes neurovirulence. *Proc Natl Acad Sci* 108:16026–16031.
 163. Vaney M-C, Duquerroy S, Rey FA. 2013. Alphavirus structure: activation for entry at the target cell surface. *Curr Opin Virol* 3:151–158.
 164. Crublet E, Andrieu J-P, Vivès RR, Lortat-Jacob H. 2008. The HIV-1 Envelope Glycoprotein gp120 Features Four Heparan Sulfate Binding Domains, Including the Co-receptor Binding Site. *J Biol Chem* 283:15193–15200.
 165. Krusat T, Streckert HJ. 1997. Heparin-dependent attachment of respiratory syncytial virus (RSV) to host cells. *Arch Virol* 142:1247–1254.
 166. Chen Y, Maguire T, Hileman RE, Fromm JR, Esko JD, Linhardt RJ, Marks RM. 1997. Dengue virus infectivity depends on envelope protein binding to target cell heparan sulfate. *Nat Med* 3:866–871.
 167. Shukla D, Liu J, Blaiklock P, Shworak NW, Bai X, Esko JD, Cohen GH, Eisenberg RJ, Rosenberg RD, Spear PG. 1999. A novel role for 3-O-sulfated heparan sulfate in herpes simplex virus 1 entry. *Cell* 99:13–22.
 168. Kam YW, Lum FM, Teo TH, Lee WWL, Simarmata D, Harjanto S, Chua CL, Chan YF, Wee JK, Chow A, Lin RTP, Leo YS, Le Grand R, Sam IC, Tong JC, Roques P, Wiesmüller KH, Rénia L, Röttschke O, Ng LFP. 2012. Early neutralizing IgG response to Chikungunya virus in infected patients targets a dominant linear epitope on the E2 glycoprotein. *EMBO Mol Med* 4:330–343.
 169. Lescar J, Roussel A, Wien MW, Navaza J, Fuller SD, Wengler G, Rey FA. 2001.

- The Fusion glycoprotein shell of Semliki Forest virus: an icosahedral assembly primed for fusogenic activation at endosomal pH. *Cell* 105:137–148.
170. Vaney M-C, Duquerroy S, Rey FA. 2013. Alphavirus structure: activation for entry at the target cell surface. *Curr Opin Virol* 3:151–158.
 171. Kielian M, Rey FA. 2006. Virus membrane-fusion proteins: more than one way to make a hairpin. *Nat Rev Microbiol* 4:67–76.
 172. Gibbons DL, Vaney M-C, Roussel A, Vigouroux A, Reilly B, Lepault J, Kielian M, Rey FA. 2004. Conformational change and protein-protein interactions of the fusion protein of Semliki Forest virus. *Nature* 427:320–325.
 173. Li L, Jose J, Xiang Y, Kuhn RJ, Rossmann MG. 2010. Structural changes of envelope proteins during alphavirus fusion. *Nature* 468:705–708.
 174. Kielian M, Helenius A. 1985. pH-induced alterations in the fusogenic spike protein of Semliki Forest virus. *J Cell Biol* 101:2284–2291.
 175. Kielian M, Chanel-Vos C, Liao M. 2010. Alphavirus Entry and Membrane Fusion. *Viruses* 2:796–825.
 176. Sun S, Xiang Y, Akahata W, Holdaway H, Pal P, Zhang X, Diamond MS, Nabel GJ, Rossmann MG. 2013. Structural analyses at pseudo atomic resolution of Chikungunya virus and antibodies show mechanisms of neutralization. *Elife* 2:e00435.
 177. Cao S, Zhang W. 2013. Characterization of an early-stage fusion intermediate of Sindbis virus using cryoelectron microscopy. *Proc Natl Acad Sci U S A* 110:13362–13367.
 178. Roman-Sosa G, Kielian M. 2011. The interaction of alphavirus E1 protein with exogenous domain III defines stages in virus-membrane fusion. *J Virol* 85:12271–12279.
 179. Zheng Y, Sanchez-San Martin C, Qin Z, Kielian M. 2011. The domain I-domain III linker plays an important role in the fusogenic conformational change of the alphavirus membrane fusion protein. *J Virol* 85:6334–6342.
 180. Sánchez-San Martín C, Sosa H, Kielian M. 2008. A stable prefusion intermediate of the alphavirus fusion protein reveals critical features of class II membrane fusion. *Cell Host Microbe* 4:600–608.
 181. Sahoo B, Chowdary TK. 2019. Conformational changes in Chikungunya virus E2 protein upon heparan sulfate receptor binding explain mechanism of E2-E1 dissociation during viral entry. *Biosci Rep* 39.
 182. Fields W, Kielian M. 2013. A key interaction between the alphavirus envelope

- proteins responsible for initial dimer dissociation during fusion. *J Virol* 2013/01/16. 87:3774–3781.
183. Lescar J, Roussel A, Wien MW, Navaza J, Fuller SD, Wengler G, Wengler G, Rey FA. 2001. The Fusion glycoprotein shell of Semliki Forest virus: an icosahedral assembly primed for fusogenic activation at endosomal pH. *Cell* 105:137–148.
 184. Fuller SD, Berriman JA, Butcher SJ, Gowen BE. 1995. Low pH induces swiveling of the glycoprotein heterodimers in the Semliki Forest virus spike complex. *Cell* 81:715–725.
 185. Mukhopadhyay S, Zhang W, Gabler S, Chipman PR, Strauss EG, Strauss JH, Baker TS, Kuhn RJ, Rossmann MG. 2006. Mapping the Structure and Function of the E1 and E2 Glycoproteins in Alphaviruses. *Structure* 14:63–73.
 186. Voss JE, Vaney M-C, Duquerroy S, Vornrhein C, Girard-Blanc C, Crublet E, Thompson A, Bricogne G, Rey FA. 2010. Glycoprotein organization of Chikungunya virus particles revealed by X-ray crystallography. *Nature* 468:709–712.
 187. Mueller DS, Kampmann T, Yennamalli R, Young PR, Kobe B, Mark AE. 2008. Histidine protonation and the activation of viral fusion proteins. *Biochem Soc Trans* 36:43–45.
 188. Zeng X, Mukhopadhyay S, Brooks CL 3rd. 2015. Residue-level resolution of alphavirus envelope protein interactions in pH-dependent fusion. *Proc Natl Acad Sci U S A* 112:2034–2039.
 189. van Duijl-Richter MKS, Blijleven JS, van Oijen AM, Smit JM. 2015. Chikungunya virus fusion properties elucidated by single-particle and bulk approaches. *J Gen Virol* 96:2122–2132.
 190. Fritz R, Stiasny K, Heinz FX. 2008. Identification of specific histidines as pH sensors in flavivirus membrane fusion. *J Cell Biol* 183:353–361.
 191. Dombkowski AA, Sultana KZ, Craig DB. 2014. Protein disulfide engineering. *FEBS Lett* 588:206–212.
 192. Gibbons DL, Erk I, Reilly B, Navaza J, Kielian M, Rey FA, Lepault J. 2003. Visualization of the target-membrane-inserted fusion protein of Semliki Forest virus by combined electron microscopy and crystallography. *Cell* 114:573–583.
 193. Smit JM, Bittman R, Wilschut J. 1999. Low-pH-dependent fusion of Sindbis virus with receptor-free cholesterol- and sphingolipid-containing liposomes. *J Virol* 73:8476–8484.
 194. Stiasny K, Allison SL, Schalich J, Heinz FX. 2002. Membrane interactions of the tick-borne encephalitis virus fusion protein E at low pH. *J Virol* 76:3784–3790.

195. Liao M, Kielian M. 2005. Domain III from class II fusion proteins functions as a dominant-negative inhibitor of virus membrane fusion. *J Cell Biol* 171:111–120.
196. Kampmann T, Mueller DS, Mark AE, Young PR, Kobe B. 2006. The Role of histidine residues in low-pH-mediated viral membrane fusion. *Structure* 14:1481–1487.
197. Qin Z-L, Zheng Y, Kielian M. 2009. Role of conserved histidine residues in the low-pH dependence of the Semliki Forest virus fusion protein. *J Virol* 83:4670–4677.
198. Parrott MM, Sitarski SA, Arnold RJ, Picton LK, Hill RB, Mukhopadhyay S. 2009. Role of Conserved Cysteines in the Alphavirus E3 Protein. *J Virol* 83:2584–2591.
199. Cheng RH, Kuhn RJ, Olson NH, Rossmann MG, Hok-Kin Choi T, Smith Thomas J, Baker and TS. 1995. Nucleocapsid and Glycoprotein Organization in an Enveloped Virus. *Cell* 80:621–630.
200. Snyder AJ, Mukhopadhyay S. 2012. The Alphavirus E3 Glycoprotein Functions in a Clade-Specific Manner. *J Virol* 86:13609–13620.
201. Hopp TP, Woods KR. 1981. Prediction of protein antigenic determinants from amino acid sequences. *Proc Natl Acad Sci* 78:3824–3828.
202. Kyte J, Doolittle RF. 1982. A simple method for displaying the hydrophobic character of a protein. *J Mol Biol* 157:105–132.
203. Ladner CL, Yang J, Turner RJ, Edwards RA. 2004. Visible fluorescent detection of proteins in polyacrylamide gels without staining. *Anal Biochem* 326:13–20.
204. Roussel A, Lescar J, Vaney MC, Wengler GG, Wengler GG, Rey FA. 2006. Structure and Interactions at the Viral Surface of the Envelope Protein E1 of Semliki Forest Virus. *Structure* 14:75–86.
205. Zeltins A. 2013. Construction and Characterization of Virus-Like Particles : A Review. *Mol* 53:92–107.
206. Tani H, Morikawa S, Matsuura Y. 2012. Development and applications of VSV vectors based on cell tropism. *Front Microbiol* 2:1–7.
207. Cronin J, Zhang X-Y, Reiser J. 2005. Altering the Tropism of Lentiviral Vectors through Pseudotyping. *Curr Gene Ther* 5:387–398.
208. Siridechadilok B, Gomutsukhavadee M, Sawaengpol T, Sangiambut S, Puttikhunt C, Chin-inmanu K, Suriyaphol P, Malasit P, Sreaton G, Mongkolsapaya J. 2013. A Simplified Positive-Sense-RNA Virus Construction Approach That Enhances Analysis Throughput. *J Virol* 87:12667–12674.
209. ZIBERT A, MAASS G, STREBEL K, FALK MM, BECK E, Zentrum. 1990.

- Infectious foot-and-mouth disease virus derived from a cloned full-length cDNA of OH/CHA/99. *J Virol* 64:2467–2473.
210. Lin T, Dowd KA, Manhart CJ, Nelson S, Whitehead SS, Pierson TC. 2012. A Novel Approach for the Rapid Mutagenesis and Directed Evolution of the Structural Genes of West Nile Virus. *J Virol* 86:3501–3512.
 211. Rasmussen TB, Reimann I, Uttenthal Å, Leifer I, Depner K, Schirrmeier H, Beer M. 2010. Generation of recombinant pestiviruses using a full-genome amplification strategy. *Vet Microbiol* 142:13–17.
 212. Lundstrom K. 2018. Viral vectors in gene therapy. *diseases* 6:2–20.
 213. Escors D, Breckpot K. 2010. Lentiviral vectors in gene therapy : their current status and future potential. *Arch Immunol Ther Exp* 58:107–119.
 214. Merrington CL, Bailey MJ, Possee RD. 1997. Manipulation of baculovirus vectors. *Mol Biotechnol* 8:283–297.
 215. Clem RJ, Passarelli AL. 2013. Baculoviruses: Sophisticated Pathogens of Insects. *PLoS Pathog* 9:e1003729.
 216. GF R. 2019. Chapter 1, Introduction to the baculoviruses, their taxonomy, and evolution., p. 1–19. *In* *Baculovirus Molecular Biology* [Internet]. 4th edition.
 217. Blissard GW, Theilmann DA. 2018. Baculovirus Entry and Egress from Insect Cells. *Annu Rev Virol* 5:113–139.
 218. Hefferon KL, Oomens AGP, Monsma SA, Finnerty CM, Blissard GW. 1999. Host cell receptor binding by baculovirus GP64 and kinetics of virion entry. *Virology* 258:455–468.
 219. Spenger A, Grabherr R, Töllner L, Katinger H, Ernst W. 2002. Altering the surface properties of baculovirus *Autographa californica* NPV by insertional mutagenesis of the envelope protein gp64. *Eur J Biochem* 269:4458–4467.
 220. Plonsky I, Zimmerberg J. 1996. The initial fusion pore induced by baculovirus GP64 is large and forms quickly. *J Cell Biol* 135:1831–1839.
 221. Leikina E, Onaran HO, Zimmerberg J. 1992. Acidic pH induces fusion of cells infected with baculovirus to form syncytia. *FEBS Lett* 304:221–224.
 222. Oomens AGP, Blissard GW. 1999. Requirement for GP64 to drive efficient budding of *Autographa californica* multicapsid nucleopolyhedrovirus. *Virology* 254:297–314.
 223. Monsma SA, Oomens AGP, Blissard GW. 1996. The GP64 Envelope Fusion Protein Is an Essential Baculovirus Protein Required for Cell-to-Cell Transmission of

Infection. *J Virol* 70:4607–4616.

224. Lung O, Westenberg M, Vlak JM, Zuidema D, Blissard GW. 2002. Pseudotyping *Autographa californica* Multicapsid Nucleopolyhedrovirus (AcMNPV): F Proteins from Group II NPVs Are Functionally Analogous to AcMNPV GP64. *J Virol* 76:5729–5736.
225. Mangor JT, Monsma SA, Johnson MC, Blissard GW. 2001. A GP64-Null Baculovirus Pseudotyped with Vesicular Stomatitis Virus G Protein. *J Virol* 75:2544–2556.
226. Jensen SI, Lennen RM, Herrgård MJ, Nielsen AT. 2015. Seven gene deletions in seven days: Fast generation of *Escherichia coli* strains tolerant to acetate and osmotic stress. *Sci Rep* 5:1–10.
227. Muyrers JPP, Zhang Y, Testa G, Stewart AF. 1999. Rapid modification of bacterial artificial chromosomes by ET-recombination. *Nucleic Acids Res* 27:1555–1557.
228. Yu D, Ellis HM, Lee EC, Jenkins NA, Copeland NG, Court DL. 2000. An efficient recombination system for chromosome engineering in *Escherichia coli*. *Proc Natl Acad Sci* 97:5978–5983.
229. Metz SW, Pijlman GP. 2016. Production of Chikungunya Virus-Like Particles and Subunit Vaccines in Insect Cells. *Methods Mol Biol* 1426:297–309.
230. Sokal N, Nie Y, Willis LG, Yamagishi J, Blissard GW, Rheault MR, Theilmann DA. 2014. Defining the roles of the baculovirus regulatory proteins IE0 and IE1 in genome replication and early gene transactivation. *Virology* 468–470:160–171.
231. Wang M, Wang J, Yin F, Tan Y, Deng F, Chen X, Jehle JA, Vlak JM, Hu Z, Wang H. 2014. Unraveling the Entry Mechanism of Baculoviruses and Its Evolutionary Implications. *J Virol* 88:2301–2311.