Structural and Functional Insights into the Molecular Mechanisms of Virus-Cell and Cell-Cell Fusion

by

Halil Aydin

A thesis submitted in conformity with the requirements for the degree of Doctorate of Philosophy
Graduate Department of Laboratory Medicine and Pathobiology
University of Toronto

© Copyright by Halil Aydin 2016
Membrane fusion is an essential step in various developmental, physiological, and pathological processes in eukaryotes. Specific fusion proteins catalyze the merger of two lipid bilayers. The molecular mechanisms of many membrane fusion reactions still remain unclear. Here, we characterized the molecular details of various adhesion and fusion proteins involved in viral-cell, placental trophoblast, and sperm-egg fusion events to better understand their mechanisms. In total, nine crystal structures of these human and viral glycoproteins are presented. Furthermore, structural comparisons combined with mutagenesis, biochemical, biophysical, and cell culture experiments allowed us to investigate the implications of specific residues of these proteins in mediating membrane fusion. Structures of class I viral fusion proteins revealed that electrostatic interactions are critical for post-fusion stability in viruses that fuse at neutral pH, whereas the salt bridges do not play a stabilizing role in fusion proteins that proceed through low pH. Instead, hydrophobic residues stabilize the fusion core, and histidine/arginine residues in the chain reversal region stabilize a helix-dipole moment. Moreover, the structure of human syncytin-1 fusion protein unveiled a striking overall structural similarity to class I viral fusion proteins, and
the presence of salt bridges within the fusion subunit zipper the inner and outer helices. These electrostatic interactions stabilize the post-fusion structure and provide the energetics for trophoblast cell fusion in placentation. While class I fusion proteins share similar structural and functional features, the mechanism of sperm-egg fusion requires the involvement of sperm Izumo1 and egg Juno proteins. Atomic resolution structures of Izumo1 and Juno display an interface stabilized through extensive interactions and a major conformational change within Izumo1 upon Juno binding. Moreover, mutational studies at the Izumo1-Juno interface revealed the structural determinants required for binding. We provide a comprehensive analysis that now identifies general trends and signatures important in viral-cell and cell-cell fusion events.
Acknowledgments

First and foremost, I would like to thank my supervisor Dr. Jeffrey E. Lee for this opportunity and his supervision. I would also like to extend my sincere gratitude to the members of my advisory committee, Dr. Emil F. Pai and Dr. Mario Ostrowski, who graciously offered their time and provided valuable advice and support throughout my graduate studies. In addition, I would like to acknowledge Dr. Azmiri Sultana and Dr. George Bikopoulos for their critical contributions to the work presented in my thesis and their expertise and advice. A big thank you to my former undergrads Brianna M. Smrke, Dina Al-Khooly, Annoj Thavalingam for their support, and willingness to help in any way possible. I would like to thank Annoj Thavalingam again for critically reading parts of my thesis. My sincerest gratitude to all the past and present members of the Lee lab; especially the best, Farshad Azimi, and the rest, Jonathan D. Cook, Matthew W. Taylor, Brandon Ho, David Dai, Shira Elion Jourard, and Dr. Peter K. Quashie. I would also like to thank people on MSB 6th floor, both past and present, and all my friends and colleagues in Toronto. Finally, I must thank my family for their love and support. Your continued encouragement, support, advice (both scientific and life in general), and comic relief has been essential to my sanity.

"My Way"

And now, the end is near
And so I face the final curtain
My friend, I'll say it clear
I'll state my case, of which I'm certain
I've lived a life that's full
I traveled each and every highway
And more, much more than this, I did it my way

Regrets, I've had a few
But then again, too few to mention
I did what I had to do and saw it through without exemption
I planned each charted course, each careful step along the byway
And more, much more than this, I did it my way

Yes, there were times, I'm sure you knew
When I bit off more than I could chew
But through it all, when there was doubt
I ate it up and spit it out
I faced it all and I stood tall and did it my way
I've loved, I've laughed and cried
I've had my fill, my share of losing
And now, as tears subside, I find it all so amusing
   To think I did all that
   And may I say, not in a shy way
   Oh, no, oh, no, not me, I did it my way

For what is a man, what has he got?
   If not himself, then he has naught
To say the things he truly feels and not the words of one who kneels
   The record shows I took the blows and did it my way

Yes, it was my way

Dedicated to the strongest woman I know:

Emine Cakir Yildiz

(1928-2015)

I miss you dearly…
Table of Contents

Abstract ........................................................................................................................................... ii
Acknowledgments .......................................................................................................................... iv
Table of Contents ........................................................................................................................... vi
List of Tables .................................................................................................................................. ix
List of Figures ................................................................................................................................. x
List of Abbreviations .................................................................................................................... xiv
Chapter 1 Introduction .................................................................................................................... 1
  1.1 Biological Membranes .............................................................................................................. 2
  1.2 Membrane proteins .................................................................................................................... 2
  1.3 Membrane shaping .................................................................................................................... 4
  1.4 Membrane fusion ....................................................................................................................... 5
  1.5 Fusogens .................................................................................................................................... 8
  1.6 Virus-cell fusion ........................................................................................................................ 8
  1.7 Cell-cell fusion ........................................................................................................................ 23
  1.8 Thesis Overview ...................................................................................................................... 37

Chapter 2 Crystal structures of β- and γ-retroviral fusion proteins reveal a role for electrostatic
stapling in viral entry ..................................................................................................................... 39
  2.1 Summary ................................................................................................................................. 40
  2.2 Introduction ............................................................................................................................. 41
  2.3 Results and Discussion ............................................................................................................ 43
  2.4 Materials and Methods ............................................................................................................ 58
  2.5 Acknowledgements .................................................................................................................. 64
  2.6 Inter-Chapter Transition ......................................................................................................... 64

Chapter 3 Structural characterization of a fusion glycoprotein from a retrovirus that undergoes a
hybrid two-step entry mechanism ................................................................................................. 65
<table>
<thead>
<tr>
<th>Chapter</th>
<th>Title</th>
<th>Pages</th>
</tr>
</thead>
<tbody>
<tr>
<td>6.2</td>
<td>Introduction</td>
<td>144</td>
</tr>
<tr>
<td>6.3</td>
<td>Results and Discussion</td>
<td>145</td>
</tr>
<tr>
<td>6.4</td>
<td>Materials and Methods</td>
<td>183</td>
</tr>
<tr>
<td>6.5</td>
<td>Acknowledgements</td>
<td>197</td>
</tr>
<tr>
<td>6.6</td>
<td>Inter-Chapter Transition</td>
<td>197</td>
</tr>
<tr>
<td>7</td>
<td>Chapter 7 A convenient and general expression platform for the production of secreted proteins from human cells</td>
<td>199</td>
</tr>
<tr>
<td>7.1</td>
<td>Summary</td>
<td>200</td>
</tr>
<tr>
<td>7.2</td>
<td>Introduction</td>
<td>201</td>
</tr>
<tr>
<td>7.3</td>
<td>Protocol</td>
<td>203</td>
</tr>
<tr>
<td>7.4</td>
<td>Discussion</td>
<td>209</td>
</tr>
<tr>
<td>7.5</td>
<td>Materials and Methods</td>
<td>214</td>
</tr>
<tr>
<td>7.6</td>
<td>Acknowledgements</td>
<td>219</td>
</tr>
<tr>
<td>8</td>
<td>Chapter 8 General Discussion and Conclusions</td>
<td>220</td>
</tr>
<tr>
<td>8.1</td>
<td>Insights into the general mechanisms of viral fusion</td>
<td>221</td>
</tr>
<tr>
<td>8.2</td>
<td>Structural basis of cell-cell fusion mechanisms</td>
<td>228</td>
</tr>
<tr>
<td>8.3</td>
<td>Conclusions</td>
<td>232</td>
</tr>
<tr>
<td></td>
<td>References</td>
<td>234</td>
</tr>
</tbody>
</table>
List of Tables

Table 2.1. Summary of retroviral fusion protein mutations......................................................... 51
Table 2.2. Data collection and refinement statistics................................................................. 60
Table 3.1. Calculated dipole moments of various trimeric viral fusion TM proteins............... 83
Table 3.2. Data collection and refinement statistics................................................................. 86
Table 4.1. Summary of SARS-CoV fusion protein stability under various pH conditions. ....... 96
Table 4.2. Summary of wild-type and mutant SARS-CoV fusion protein stabilities............... 102
Table 4.3. Summary of SARS-CoV S2 HR2 peptide mimics.................................................. 106
Table 5.1 Summary of wild-type and mutant syncytin-1 fusion protein stabilities............... 124
Table 5.2. Data collection and refinement statistics................................................................. 137
Table 6.1. Izumo1-Junon binding interface mutations............................................................ 176
Table 6.2. Data collection and refinement statistics................................................................. 194
Table 7.1. Comparison of cell culture vessels used for protein expression.......................... 208
List of Figures

Figure 1.1. Schematic representation of membrane lipids and bilayer formation. .........................3
Figure 1.2. Curvature and basic membrane shapes. ........................................................................5
Figure 1.3. Pathway of membrane fusion. .....................................................................................7
Figure 1.4. Pre- and post-fusion structures of class I viral fusion proteins. .................................10
Figure 1.5. Overview of Ebola virus entry. ...................................................................................12
Figure 1.6. Model of HIV-1 Env-mediated viral-cell fusion. ..........................................................14
Figure 1.7. Model of Dengue virus E glycoprotein conformational changes. .............................17
Figure 1.8. Structural transitions of fusogenic VSV G protein. ....................................................19
Figure 1.9. Structural motifs and proposed fusion mechanism of FAST proteins. .......................22
Figure 1.10. Model of sperm-egg adhesion and fusion machinery. .............................................26
Figure 1.11. Syncytin-mediated cell-cell fusion in human placentation. ......................................29
Figure 1.12. Schematic representation of EFF-1-mediated homotypic cell-cell fusion. ...............31
Figure 1.13. Model of membrane fusion facilitated by SNAREs alone. ......................................33
Figure 1.14. Models of myoblast fusion and yeast mating. ..........................................................36
Figure 2.1. Structures of retroviral fusion glycoproteins. ............................................................46
Figure 2.2. Intra- and complex intermolecular retroviral salt bridges. ........................................48
Figure 2.3. Surface electrostatic potential of the HR1-HR2 interface. .........................................49
Figure 2.4. Characterization of MPMV and XMRV pseudotyped virions. .................................51
Figure 2.5. Sequence alignment of retroviral fusion subunits.......................................................... 57

Figure 3.1. Structure of ASLV TM.................................................................................................. 73

Figure 3.2. ASLV TM stability as a function of pH................................................................. 74

Figure 3.3. ASLV TM electrostatic interactions........................................................................ 78

Figure 3.4. ASLV TM hydrophobic interactions........................................................................ 79

Figure 3.5. ASLV TM chain reversal region histidine residues are important for stability at low pH....................................................................................................................................... 82

Figure 4.1. Structural description and biophysical characterization of the SARS-CoV S2 L2H protein........................................................................................................................................... 95

Figure 4.2. SARS-CoV S2 L2H stability at various pH values.................................................. 97

Figure 4.3. Biophysical characterization of SARS-CoV S2 electrostatic interactions.............. 100

Figure 4.4. SARS-CoV S2 hydrophobic and polar interactions.............................................. 101

Figure 4.5. Primary sequence alignment of CoV S2 fusion cores............................................ 104

Figure 4.6. The common HR1 SARS-CoV S2 substructure..................................................... 107

Figure 5.1. Structure of syncytin-1 TM...................................................................................... 119

Figure 5.2. Structural comparison and electrostatic potential of syncytin-1 TM............... 121

Figure 5.3. Single and complex salt bridges of syncytin-1 TM............................................... 125

Figure 5.4. Sequence alignment of syncytins and retroviral fusion proteins.......................... 126

Figure 5.5. Syncytin-1-mediated cell-cell fusion...................................................................... 128

Figure 5.6. Immunosuppressive properties of human syncytin-1.......................................... 133

Figure 6.1. Alignment of Izumo1 protein sequences from various mammals....................... 146
Figure 6.2. Alignment of Juno protein sequences from various mammals. ........................................ 148

Figure 6.3. Purification and characterization of Izumo1 and Juno. .................................................. 150

Figure 6.4. Overall structure of human Izumo1 and Juno. .............................................................. 151

Figure 6.5. Overall structural architecture of Izumo1 and Juno. ...................................................... 153

Figure 6.5. Electrostatic potential surface representation of Izumo122-254. ............................... 154

Figure 6.6. Izumo1 4HB is structurally similar to SPECT1, a Plasmodium protein involved in host cell traversal. ......................................................................................................................... 155

Figure 6.7. Comparison of the Izumo1 Ig-like domain and various Ig-superfamily (IgSF) subtypes. ........................................................................................................................................ 157

Figure 6.8. Izumo1 Ig-like domain. .................................................................................................. 158

Figure 6.9. Crystal structure of Izumo122-268. ............................................................................. 159

Figure 6.10. Structural comparison of Juno and the folate receptor family of proteins. ............. 160

Figure 6.11. Structural comparison of human and murine Juno. .................................................. 162

Figure 6.12. Izumo1-Juno heterotypic assembly. ........................................................................... 167

Figure 6.13. Izumo1-Juno interface. .............................................................................................. 170

Figure 6.14. Conformational changes in Izumo1 upon binding to Juno. ....................................... 170

Figure 6.15. Hybrid structural analyses of human Izumo1 and Juno in a solution state. ............ 171

Figure 6.16. SAXS data and analysis. ............................................................................................. 173

Figure 6.17. Comparative DXMS analysis. ..................................................................................... 174

Figure 6.18. Biolayer interferometry binding affinity and kinetic analysis of Izumo1-Juno interface residues. .................................................................................................................. 175

Figure 6.19. Specific conserved interactions at the Izumo1-Juno binding interface. ................. 177
Figure 6.20. SAXS data and analysis of Izumo1 and Juno mutants. ................................................ 178
Figure 6.21. SAXS reconstruction of Izumo1 and Juno mutants.................................................... 179
Figure 6.22. Conservation of residues at the Izumo1-Juno interface........................................... 180
Figure 6.23. Comparison of Izumo1 with selected viral fusogens............................................... 181
Figure 6.24. Izumo1_{22-254} and Izumo1_{22-254}-Juno_{20-228}...................................................... 188
Figure 6.25. Juno_{20-228} X-ray diffraction and twinning analysis.............................................. 189
Figure 6.26. Difference anomalous Fourier electron density map................................................. 193
Figure 7.1. Workflow schematic of small-scale transfections...................................................... 205
Figure 7.2. Corning 10-layer CellSTACK for large-scale protein expression............................... 207
Figure 7.3. Small-scale expression of various secreted proteins.................................................. 210
Figure 7.4. Purified human cellular proteins from large-scale HEK293T cultures....................... 211
Figure 8.1. Summary of pH-dependent and pH-independent fusion protein properties.............. 225
Figure 8.2. Schematic representation of the class I viral fusion protein prehairpin intermediate and its inhibition................................................................. 227
Figure 8.3. The molecular basis of sperm and egg fusion machinery........................................ 232
List of Abbreviations

4HB: Four-helix bundle

6HB: Six-helix bundle

ACE2: Angiotensin-converting enzyme 2

ACE3: Angiotensin-converting enzyme 3

ADAM: A disintegrin and a metalloproteinase domain-containing protein

ADP: Adenosine diphosphate

AFF-1: Anchor-cell fusion failure 1

APN: Aminopeptidase N

ASCT-2: Na\(^+\)-dependent neutral amino acid transporter 2

ASLV: Avian sarcoma leukosis virus

ATCC: American type culture collection

ATP: Adenosine triphosphate

AUC: Analytical ultracentrifugation

AX1: Annexin A1

BaEV: baboon endogenous virus

BCoV: bovine coronavirus

BLI: Biolayer interferometry

BLV: Bovine leukemia virus
BMDC: Bone marrow derived cells

BSA: Bovine serum albumin

BtCoV: Bat coronavirus

CAECAM1: Carcinoembryonic antigen adhesion molecule 1

CCD: Charge-coupled device

CCoV: Canine coronavirus

CCR5: C-C chemokine receptor type 5

CD: Circular Dichroism

CD: Cluster of differentiation

CHAPS: (3-((3-cholamidopropyl) dimethylammonio)-1-propanesulfonate)

CHO: Chinese Hamster Ovary

COS: CV-1 cells in Origin carrying the SV40 large T-antigen

CoV: Coronavirus

CR: Chain reversal

Cryo-EM: Cryo electron microscopy

CT: Cytoplasmic tail

CTH: C-terminal halves

CXCR4: C-X-C chemokine receptor type 4

DC-SIGN: Dendritic cell-specific intercellular adhesion molecule-3-grabbing non-integrin

DC-STAMP: Dendritic cell-specific transmembrane protein
DLS: Dynamic light scattering

DMEM: Dulbecco’s Modified Eagle Medium

DNA: Deoxyribonucleic acid

DPP4: Dipeptidyl peptidase 4

Duf: Dumbfounded

DXMS: Deuterium exchange mass spectrometry

E: Envelope glycoprotein

EBOV: Ebola virus

EBV: Epstein-Barr virus

EDTA: Ethylenediaminetetraacetic acid

EFF-1: Epithelial fusion failure-1

Ek: Enterokinase

ELISA: Enzyme-linked immunosorbent assay

Env: Envelope glycoprotein

EnvF(c)1: Envelope glycoprotein type F(c)1

EnvF(c)2: Envelope glycoprotein type F(c)2

EnvH1: Envelope glycoprotein type H1

EnvH2: Envelope glycoprotein type H2

EnvR: Envelope glycoprotein type R

EnvT: Envelope glycoprotein type T
ER: Endoplasmic reticulum

ERV: Endogenous retrovirus

FACS: Fluorescence-activated cell sorting

FAST: Fusion-associated small transmembrane

FBS: Fetal bovine serum

FCM: Fusion competent myoblast

FCoV: Feline coronavirus

FD: Fusion domain

FeLV: Feline leukemia virus

FENV: Feline endogenous virus

FF: Fusion failure

FGFR2: Fibroblast growth factor receptor-2

FIPV: Feline infectious peritonitis virus

FLV: Feline leukemia virus

FOL: Folate

FOLR: Folate receptor

FP: Fusion peptide

Ga-FeSV: Gardner-Arnstein feline sarcoma virus

Gag: Group specific antigen

GALV: Gibbon ape leukemia virus
GAPDH: Glyceraldehyde 3-phosphate dehydrogenase

GFP: Green fluorescent protein

GP: Glycoprotein

GP1: Glycoprotein 1 attachment domain

GP2: Glycoprotein 2 fusion domain

GPI: Glycophosphatidylinositol

HA: Hemagglutinin

Hbs: Hibris

HCl: Hydrogen chloride

hCoV: Human coronavirus

HD: High-definition

HEK: Human embryonic kidney

HEPES: (4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid)

HERV: Human endogenous retrovirus

HIV-1: Human immunodeficiency virus type-1

HKU: Hong Kong University

HR1: Heptad repeat region 1

HR2: Heptad repeat region 2

HSC: Hematopoietic stem cells

HTLV-1: Human T-lymphotropic virus type-1
IAP: Integrin associated protein
IAV: Influenza A virus
IFL: Internal fusion loop
Ig: Immunoglobulin
IgSF: Immunoglobulin superfamily
IL: Interleukin
IP: Immunoprecipitation
IPTG: Isopropyl β-D-1-thiogalactopyranoside
IS: Immunosuppressive
ISD: Immunosuppressive domain
JSRV: Jaagsiekte sheep retrovirus
KiMSV: Kirsten murine sarcoma virus
KORV: Koala retrovirus
L2H: Linked two-heptad construct
LCMV: Lymphocytic choriomeningitis virus
LD: Linker domain
LDLR: Low-density lipoprotein receptor
LIC: Ligation independent cloning
M: Membrane
M-CSF: Macrophage colony-stimulating factor
MALS: Multiangle Light Scattering
MARV: Marburg virus
MERS: Middle East respiratory syndrome
MES: 2-(N-morpholino)ethanesulfonic acid
MFR: Macrophage fusion receptor
MFSD2: Major facilitator superfamily domain containing 2
MGC: Multinucleated giant cell
MHV: Mouse hepatitis virus
MLV: Murine leukemia virus
MMTV: Mouse mammary tumor virus
MoMLV: Moloney murine leukemia virus
MPER: Membrane proximal external region
MPMV: Mason-Pfizer monkey virus
MR: Molecular replacement
MS: Mass spectrometry
MuCoV: Munia coronavirus
MW: Molecular weight
MyoII: Myosin II
NCS: Non-crystallographic symmetry
Ni-NTA: Nickel-Nitrilotriacetic acid
NP40: Nonidet P-40

NPC1: Niemann Pick C1

NSF: N-ethylmaleimide-sensitive factor

NTH: N-terminal halves

OD: Optical Density

PBS: Phosphate-buffered saline

PBS-T: Phosphate-buffered saline with tween-20

PCR: Polymerase chain reaction

PDB: Protein Data Bank

PDI: Protein disulphide isomerase

PEDV: Porcine epidemic bronchitis virus

PEG: Polyethylene glycol

PEI: Polyethyleneimine

PERV: Porcine endogenous retrovirus

PHD: Pleckstrin homology domain

PHEV: Porcine hemagglutinating encephalomyelitis virus

PMSF: Phenylmethanesulfonyl fluoride

Pol: Polymerase

PRCoV: Porcine respiratory coronavirus

prM: Precursor Membrane
Prm1: Pheromone-regulated membrane protein 1

PVDF: Polyvinylidene fluoride

RANKL: Receptor activator of nuclear kappa B ligand

RbCoV: Rabbit coronavirus

RBD: Receptor-binding domain

RBM: Receptor-binding motif

RD114: Feline endogenous retrovirus

REV: Reticuloendotheliosis virus

rMMS: Random microseed matrix screening

RMSD: Root mean square deviation

RNA: Ribonucleic acid

Rst: Roughest

RSV: Rous sarcoma virus

RSV: Respiratory syncytial virus

RT: Room temperature

RtCoV: Rat coronavirus

S: Coronavirus spike glycoprotein

S1: Coronavirus spike glycoprotein attachment subunit

S2: Coronavirus spike glycoprotein fusion subunit

SAMP14: Sperm acrosomal membrane-associated 14
SARS: Severe acute respiratory syndrome

SAXS: Small angle X-ray scattering

SC: Stem cell

SD: Standard deviation

SEC: Size Exclusion Chromatography

sGP: Secreted glycoprotein

SDS-PAGE: Sodium dodecyl sulfate polyacrylamide gel electrophoresis

SLLP1: Sperm Lyzozyme-Like Acrosomal Protein 1

SMRV: Squirrel monkey retrovirus

SNARE: N-ethylmaleimide-sensitive factor attachment protein receptor

SNAP: N-ethylmaleimide-sensitive factor attachment protein

Sns: Sticks and stones

SNV: Spleen necrosis virus

SP: Signal peptide

SPACA: Sperm acrosome-associated protein

SpCoV: Sparrow coronavirus

SPECT: Sporozoite micronemal protein essential for cell transversal

SPESP1: Sperm equatorial segment protein 1

SPR: Surface plasmon resonance

SRV: Simian retrovirus
SSM: Secondary structure matching

STLV: Simian T-lymphotropic virus type

Stx-1: Syntaxin-1

SU: Surface receptor-binding subunit

Syb2: Synaptobrevin 2

T: Tether

TAM: Tyro3/Axl/Mer

TBS: Tris-buffered saline

TCEP: (tris(2-carboxyethyl)phosphine)

TEV: Tobacco Etch Virus

TFZ: Translation function Z score

TGEV: Transmissible gastroenteritis virus

TGN: Trans-Golgi network

ThCoV: Thrush coronavirus

TIM-1: T-cell immunoglobulin and mucin domain 1

TLS: Translation/Libration/Screw

TM: Transmembrane fusion subunit

TMD: Transmembrane domain

TMEM190: Transmembrane protein 190

TPS: Translational pseudosymmetry
Tva: Tumor virus locus a

VSV: Vesicular stomatitis virus

WT: Wild-type

XMRV: Xenotropic murine leukemia virus-related virus

ZEBOV: Zaire Ebola virus

ZP: Zona pellucida
Chapter 1
Introduction

Portions of this chapter were modified from the following publications:


1.1 Biological Membranes

Biological membranes are the basic building blocks of cells, defining the external boundary and the spatial identity, as well as the interior compartments within cells. These membranes provide protection and a barrier to the transfer of metabolites, and allow the creation of ion gradients to generate energy and enable signal transduction. Cell membranes are composed of a lipid bilayer embedded with proteins. In addition, sugars can be covalently bound to lipids and proteins on the outer surface of the lipid bilayer. The most abundant types of lipids found in biological membranes are phospholipids, glycolipids, and sterols. These lipid molecules consist of two parts—a hydrophilic head group, which readily interacts with the aqueous environment, and a hydrophobic tail that is more stable in a lipid environment and self-associates to minimize contact with water. (Figure 1.1). The amphipathic nature of lipids naturally supports the formation of a membrane lipid bilayer where the hydrophilic head groups point outward toward the aqueous environment and the hydrophobic tail from the apposing monolayers face each other in a lipid environment. Different biological membranes have specific lipid compositions, and complex structural organization of amphipathic lipids can determine the shape of cells and organelles.

1.2 Membrane proteins

Membrane-embedded proteins are nanomachines, which can mediate diverse physiological functions within the complex landscape of cellular membranes and maintain structural integrity. Approximately one-third of all active genes in mammalian genomes encode membrane proteins, and many other proteins also interact with membrane proteins to carry out cellular processes. Membrane proteins facilitate intercellular communication and perceive changes in their environment. When extracellular ligands bind to their functional domains, membrane proteins transmit signals across the lipid bilayer. They also confer the transport of metabolites and other essential molecules into and out of cells and compartments, and maintain the concentration gradient of these molecules. Membrane proteins may also act as receptors for other proteins and involved in recognition and binding in adhesion process of cells. Many membrane proteins exhibit enzymatic activity and catalyze chemical reactions as well. For instance, adenosine triphosphate (ATP) synthases harness the electrochemical gradients to
produce ATP by adding an inorganic phosphate group to a molecule of adenosine diphosphate (ADP)\textsuperscript{20}. Lastly, membrane proteins may promote the dynamic rearrangement of membrane and cellular structures, allowing membrane fusion or separation\textsuperscript{21−23}. Physiological processes such as endocytosis, exocytosis, and cell-cell fusion all require specific membrane proteins to be present in a particular membrane to carry out the remodeling of the membrane bilayer.

**Figure 1.1. Schematic representation of membrane lipids and bilayer formation.** The figure 1.1 was reproduced from Watson et al., *Essays Biochem.* (2015)\textsuperscript{24}. (A) Glycerophospholipid (B) Glycolipid (C) Sterol (D) The hydrophilic head domains (brown circles) face the aqueous environment while the hydrophobic tail domains (yellow) face inward in lipid bilayers\textsuperscript{24}. Glycolipids are colored in blue.
1.3 **Membrane shaping**

Biological membranes display an array of shapes for their function and undergo radical shape transformations during physiological processes. Membrane shaping is directly related to the formation of membrane curvatures. The molecular mechanisms that generate membrane curvatures require a complex interplay between membrane proteins and lipids that constitute the membrane bilayer (Figure 1.2). In general, there are two principal curvatures, c₁ and c₂, that describe the shape of the membrane at a given point of its surface. Different combinations of the principal curvatures allow biological membranes to adopt four basic shapes (flat, cylindrical, saddle-like and spherical) (Figure 1.2). Membrane shapes can be altered by remodeling. However, the generation of curvature is largely based on membrane elasticity and it requires energy. Although lipids alone can self-assemble into curved monolayers and/or generate various membrane shapes, the transformation requires large asymmetry of membrane lipid composition, as well as substantial energy input. Proteins, on the other hand, can apply physical forces to membrane surfaces and provide a persistent energy input that is required to change membrane topology. Proteins utilize two distinct mechanisms, pulling and bending, to generate membrane curvatures in a direct manner. The pulling mechanism involves proteins that exert a localized pulling force on a lipid bilayer in order to provide the energy needed for membrane reshaping, while proteins that bind to the membrane surface apply local bending forces to exceed the membrane-bending energy and produce membrane curvatures. Regardless of shaping mechanism, lipids and proteins work together to facilitate the generation of membrane curvatures.
Figure 1.2. Curvature and basic membrane shapes. The figures 1.2A and 1.2C were reproduced from Shibata et al., *Annu. Rev. Cell Dev. Biol.* (2009)28 and the figure 1.2B was reproduced from Zimmerberg et al., *Nat. Rev.* (2006)27. (A) Curvature of a lipid bilayer. (B) A flat shape with no curvature27. (C) Spherical, cylindrical, and saddle-like curved shapes. Electron microscopy images under each curved shape demonstrate actual biological membranes adopting curved shapes25,28.

1.4 Membrane fusion

Membrane fusion is an ubiquitous process in cell biology whereby two separate lipid bilayers come in close contact with one another and merge into a single membrane30. Membrane fusion can occur between cells, between viruses and cells, and between intracellular compartments30,31. These fusion events can either be homotypic, such as epithelial cell fusion in *C. elegans,* or heterotypic, such as the fusion of viral and host cell membranes31. Similar to membrane shaping, the fusion reaction between two lipid bilayers is generally mediated by proteins that can locally disturb the lipid bilayers32,33. The kinetic energy barriers related to membrane deformations are very high, and membrane fusion-mediating proteins are required to overcome these energy barriers and catalyze the merger of the two lipid bilayers34,35. The pathway of membrane fusion, from the close juxtaposition of two lipid bilayers to the expansion of a fusion pore, is composed
of a series of intermediate stages (hemifusion-stalk, hemifusion, and pore opening) by which proteins facilitate remodelling\textsuperscript{34,35} (Figure 1.3). The membrane fusion process is initiated by a protein bridge between the two membranes, followed by a point-like protrusion and lipid connection of the fusing bilayers\textsuperscript{36}. Hemifusion intermediate stages represent the local merger of contacting outer leaflets while inner leaflets remain intact. First, the outer leaflets of apposed membranes merge into a stalk-like hemifusion connection where the lipid mixing occurs only between the contacting outer leaflets. Then, the hemifusion-stalk radially expands into a half-merged hemifusion intermediate stage in which a transient fusion pore opens between the fusing bilayers\textsuperscript{34,35}. This hemifusion stage gives rise to the complete merger of two lipid bilayers into a single merged membrane and formation of an irreversible fusion pore\textsuperscript{37–39} (Figure 1.3). Finally, a persistent energy input provided by the fusion machinery drives the expansion of the nascent fusion pore and completes the membrane fusion reaction\textsuperscript{32,34,40}. Fusion pore opening and its expansion at the late fusion stages represent the most energy-demanding stages of fusion\textsuperscript{32,40}.
Figure 1.3. Pathway of membrane fusion. The figure 1.3A was reproduced from Chernomordik et al., Nat. Struct. Mol. Biol. (2008)\textsuperscript{34} and the figure 1.3B was reproduced from Martens et al., Nat. Rev. Mol. Cell Biol. (2008)\textsuperscript{30}. (A) Lipid bilayer fusion through hemifusion pathway. A pre-fusion contact between the two membranes is mediated by membrane proteins. A point-like membrane protrusion brings the two membranes into closer proximity. The high energy barrier must be lowered in order to catalyze the merger of two membranes. A hemifusion stalk allows the merger of outer leaflets, whereas inner leaflets remain unfused. The hemifusion stalk radially expands into a hemifusion diaphragm bilayer, followed by the merger of the inner leaflets and the formation of a fusion pore. Dashed lines represent the hydrophobic interfaces of the lipid bilayer\textsuperscript{34}. (B) A schematic representation of synaptic vesicle fusion with the plasma membrane. The fusion of vesicles with the plasma membrane is mediated through the same hemifusion pathway and is essential for efficient intracellular communication\textsuperscript{30}. 
1.5 Fusogens

Membrane fusion is an essential step for various processes throughout fertilization, placentation, intracellular communication, development of skeletal muscle, bone and the immune defense system, as well as cancer, stem cell-dependent tissue regeneration, and yeast mating\textsuperscript{41,42}. Entry of viruses into cells also involves fusion between the viral membrane and the target cell membrane\textsuperscript{31,35}. Fusion proteins (fusogens) are necessary for lowering the energy barrier during membrane fusion and sufficient to catalyze the fusion of two membranes\textsuperscript{35}. The majority of fusogens are anchored to a lipid bilayer via a transmembrane domain and contain a distinct hydrophobic patch (internal fusion loop(s) or fusion peptide) that facilitates interactions with the target membrane\textsuperscript{31,43}. In response to triggering mechanisms, such as binding to a specific cell surface receptor, low pH environment in the endosome or proteolytic cleavage, fusogens undergo drastic conformational changes that result in the exposure and extension of the buried hydrophobic patch, followed by its insertion into the target cell membrane\textsuperscript{43}. Concomitantly, the protein bridge established between the two membranes collapses into a more stable postfusion conformation, thereby bringing the two membranes into close proximity and leading to their fusion\textsuperscript{35,43}. There is great diversity in the structures and functions of known fusion proteins\textsuperscript{30}. Although the three-dimensional structures of many fusogens in pre- and/or post-fusion conformation(s) have been determined, the intermediate conformations of membrane fusion have not been characterized structurally but rather investigated through functional studies\textsuperscript{31}. Common structural and functional features exist in distinct fusion processes; however, the molecular mechanisms for specialized fusion events remain unclear for many biological processes\textsuperscript{44}. Nonetheless, the rich diversity of fusogens suggests distinct mechanisms during the membrane fusion reaction.

1.6 Virus-cell fusion

Viruses are obligate intracellular parasites, and membrane fusion is an essential step in the life cycle of every enveloped virus\textsuperscript{35}. Virus-cell fusion is mediated by one or more surface glycoproteins, which facilitate the recognition of and subsequent penetration into target cells\textsuperscript{35,43}. Structural insights into the fusion mechanism of many viruses suggest that the molecular architecture of surface glycoproteins from distinct viruses share clear structural similarities. To
date, there are four identified structural classes (class I, II, III, and IV) for viral fusion proteins\textsuperscript{31}. Class I, II, and III belong to enveloped viruses, and structures are available for a number of viruses in both pre- and post-fusion conformations\textsuperscript{35,43}. Class IV viral fusion proteins, fusion-associated small transmembrane (FAST) proteins, are identified in non-enveloped reoviruses and are the smallest known viral fusogens\textsuperscript{45}.

1.6.1 Class I viral fusion proteins. Enveloped viruses from different viral families including orthomyxoviridae, retroviridae, paramyxoviridae, coronaviridae, filoviridae, and arenaviridae employ class I surface glycoproteins to mediate cellular attachment and catalyze the merger of viral and host cell membranes\textsuperscript{43}. These surface glycoproteins share very little sequence similarity with each other. However, all class I fusion proteins are trimers in their pre- and post-fusion states. They all share a characteristic $\alpha$-helical trimer-of-hairpin structure in their post-fusion conformations (Figure 1.4). Some enveloped viruses, such as the human immunodeficiency virus type-1 (HIV-1)\textsuperscript{46,47} and Ebola virus (EBOV)\textsuperscript{48}, contain a single envelope glycoprotein on the viral surface, while other class I fusion machineries, such as respiratory syncytial virus (RSV)\textsuperscript{49}, are composed of two separately encoded envelope glycoproteins for attachment and fusion. Nascent viral glycoproteins are initially synthesized as inactive precursors, which are then cleaved by a cellular protease into two subunits (attachment and fusion) and assemble into a heterotrimeric metastable pre-fusion conformation on the viral surface\textsuperscript{43}. For some viral glycoproteins such as the EBOV glycoprotein (GP)\textsuperscript{48}, the attachment and fusion subunits are covalently linked with a disulfide bond, while others, including the HIV-1 Envelope (Env or gp160)\textsuperscript{21}, are non-covalently associated (Figure 1.4). In many class I fusion proteins, there is a hydrophobic fusion peptide or fusion loop near or at the N-terminus of the fusion protein, followed by two heptad repeat (HR) regions, a transmembrane domain and a C-terminal cytoplasmic tail\textsuperscript{31,43}. Upon triggering by receptor binding at neutral pH (retroviridae and paramyxoviridae), low pH (orthomyxoviridae and arenaviridae), and/or proteolytic cleavage (coronaviridae and filoviridae), the hydrophobic segment of class I fusion proteins is relieved from a buried position to engage with the target cell membrane\textsuperscript{43}. Then, the fusion protein undergoes irreversible structural rearrangements to draw the two membranes (virus and cell) together and facilitate their fusion\textsuperscript{31,35}. The energy released upon refolding of the fusion protein is adequate to cross the high kinetic energy barrier required for membrane fusion\textsuperscript{35,50}.
For more than three decades, the atomic resolution structures of influenza A virus (IAV) HA in pre- and post-fusion states provided insights into the fusion mechanism of class I viral fusion proteins\textsuperscript{51,52}. In recent years, atomic resolution structures of many class I viral fusion proteins were determined in both pre- and post-fusion states\textsuperscript{21,48,53–58}, thereby illustrating the beginning and end points of the fusion process. The class I viral fusion glycoproteins can be further classified into covalent- and non-covalent-type GP. In Sections 1.6.1.1 and 1.6.1.2, the mechanism of class I viral glycoprotein-mediated fusion is presented using Ebola virus GP and HIV-1 Env/gp160 structural models as examples.

**Figure 1.4. Pre- and post-fusion structures of class I viral fusion proteins.** The figure 1.4 was reproduced from Pancera et al., Nature (2014)\textsuperscript{21}. (A) Ribbon representation of pre-fusion and
post-fusion structures of HIV-1 Env, IAV HA, RSV F, EBOV Env (GP). Pre-fusion structures are shown as monomers for clarity. The HIV-1 gp120 and other attachment subunits are depicted in red, while HIV-1 gp41 and equivalent fusion subunits are colored in rainbow (blue to orange). Each monomer in the trimeric post-fusion structures are either shown in rainbow (blue to orange), dark gray and light gray. (B) Ribbon diagram of pre- and post-fusion structures of HIV-1 Env, IAV HA, RSV F, EBOV GP HR1 regions.

1.6.1.1 Covalent-type class I viral glycoproteins. An example of a covalent-type viral glycoprotein is the EBOV GP. The cellular entry of EBOV is a multistep process. The EBOV GP is synthesized as a 676 amino acid long precursor polypeptide, which is then cleaved by furin at a R-T-R-R cleavage site (residues 501-502) to yield an attachment subunit (GP1, residues 33-501) and a typical class I fusion subunit (GP2, residues 502-676). Processed GP forms a trimer on the viral surface. Each protomer consists of GP1 and GP2 subunits that are covalently attached through a disulfide bond between C53 in GP1 and C609 in GP2. GP1 contains a receptor binding domain (residues 54-201) capped by a heavily glycosylated mucin-like domain (residues 305-485) and a glycan-cap, while GP2 is comprised of a hydrophobic internal fusion loop (residues 534-539) located near its N-terminus followed by two HR regions, a transmembrane domain (residues 651-671), and a short cytoplasmic tail (residues 672-676). EBOV GP2 has a CX6CC motif between its HR1 (residues 554-595) and HR2 (residues 615-634) regions, where the first two cysteines form a disulfide bond within the GP2 subunit and the third cysteine is linked to a cysteine in the GP1 subunit to form the covalent interaction between the two subunits.

EBOV enters target cells via macropinocytosis and is trafficked to the endosome. GP is further processed by endosomal cathepsins B and L to remove the mucin-like domain and the glycan cap to expose a receptor-binding site in the bowl of the GP trimer. Both cathepsin B and L are endosomal proteases that are active at low pH; hence, EBOV requires trafficking through late endosomes and lysosomes for receptor binding. The host-primed GP binds the Niemann Pick C1 (NPC1) endosomal receptor (Figure 1.5). The precise trigger for fusion is not known, but is thought to be a combination of receptor-binding and low pH of the endosome. Upon
triggering, the GP1 constraints on GP2 are released to allow the insertion of the internal fusion loop into the target cell membrane to form the pre-hairpin intermediate\textsuperscript{48,59,71} (Figure 1.5). Then, the HR2 region of GP2 undergoes a fold-back to form the stable post-fusion trimer-of-hairpins conformation which drives the fusion of viral and host cell membranes\textsuperscript{59,64,72}.

![Figure 1.5. Overview of Ebola virus entry.](image)

**Figure 1.5. Overview of Ebola virus entry.** The figure 1.5 was reproduced from Hunt et al., *Viruses* (2012)\textsuperscript{73}. Trimeric EBOV GP makes contact with its receptor (TIM-1) at the cell surface,
and enters into the endosomal compartments via macropinocytosis. There are also other attachment factors, such as C-type lectins and TAM family proteins, that may help concentrate virions on the cell surface prior to receptor binding and internalization. Acidification of endosomal vesicles triggers cathepsins to cleave the mucin-like domain and glycan cap of EBOV GP, thereby rendering the GP capable of binding the endosomal portion of the NPC1 protein in the late endosome/lysosome. Whether EBOV GP still interacts with the TIM-1 within the endosomes remain unknown. (i) Binding of the NPC1 C-loop by the GP triggers the release of the fusion loop, followed by its insertion into the target cell membrane. (ii) Subsequent structural transitions of the EBOV GP2 into a post-fusion six-helix bundle conformation bring the host and viral membrane together, leading to fusion. (iii) Upon fusion pore formation, the viral nucleoprotein is released into the cytoplasm to initiate virus replication.

1.6.1.2 Non-covalent type class I viral glycoproteins. The recent structures of HIV-1 Env (gp140) at the pre-fusion state, along with the known structure of the post-fusion conformation, provide both structural and functional insights into the fusion mechanism of non-covalently attached class I viral fusion proteins. The HIV-1 Env glycoprotein (residues 33-856) is comprised of a receptor binding subunit (gp120, residues 33-511) and a transmembrane fusion subunit (gp41, residues 512-856). The proteolytic processing of precursor gp160 at a furin cleavage site (residues 511-512) gives rise to non-covalently attached heterodimeric gp120/gp41 that forms a metastable trimer on the viral surface. The Env trimer is mostly coated with high-mannose glycans and primed to catalyze the merger of viral and host cell membranes. Similar to other retroviruses, paramyxoviruses, and some coronaviruses, HIV-1 fusion machinery is activated upon binding to host cell receptor at neutral pH. The HIV-1 Env sequentially interacts with CD4 (receptor) first and then a member of the chemokine family of G-protein-coupled receptors, CXCR4 or CCR5 (co-receptors), prior to virus-cell fusion (Figure 1.6). The receptors CXCR4 or CCR5 are indispensible for HIV-1 infection. The extracellular region of CD4 contains four immunoglobulin-like domains and HIV-1 gp120 recognizes the outmost domain for cellular attachment. Binding of CD4 to a cavity that contains the CD4 binding loop (residues 364-374) triggers major conformational changes within gp120 and results in the formation of the co-receptor binding site. HIV-1 gp120 contains five variable loops (V1-V5), which are flexible and heavily glycosylated, and they can tolerate high
mutation rates\textsuperscript{21,46,47,78}. The gp120 V3 loop residues (296-330) interact directly with the co-receptor and are the major determinants of co-receptor usage (either CXCR4 or CCR5) during viral attachment to the target cell\textsuperscript{46,47,84}. The domain organization of gp41 fusion subunit is similar to the other class I viral fusion proteins\textsuperscript{46,47,74}. The gp41 has a conserved N-terminal fusion peptide (residues 512-532), an N-terminal HR1, a disulfide linked chain reversal (CR) region, a C-terminal HR2, a tryptophan rich membrane proximal external region (MPER, residues 662-683), a transmembrane domain (residues 685-705) and a long cytoplasmic tail (residues 706-856)\textsuperscript{75,76}. In the trimeric pre-fusion gp140 structure, the hydrophobic fusion peptide is buried in a hydrophobic pocket at the gp120-gp41 interface, while the HR1 $\alpha$-helices form a three-helix bundle at the center of the trimeric gp140 and the HR2 $\alpha$-helices extend diagonally to the membrane\textsuperscript{21,46,47}. CD4 engagement triggers conformational changes in gp41 that lead to the exposure and extension of the N-terminal fusion peptide toward the host cell membrane\textsuperscript{50,75,76} (Figure 1.6). The gp41 then folds back on itself and three HR2 helices pack against the central coiled-coil HR1 helices in an antiparallel fashion, forming an energetically stable six-helix bundle structure\textsuperscript{50,74-76}. The post-fusion conformation brings the N-terminal fusion peptide inserted in the target cell membrane and C-terminal transmembrane domain embedded in the viral membrane into close proximity, leading to the merger of two membranes and fusion pore opening\textsuperscript{50,74} (Figure 1.6). A final structural rearrangement is required for zipping up the six-helix bundle into a more stable conformation and expansion of the fusion pore\textsuperscript{35,50}.

![Figure 1.6. Model of HIV-1 Env-mediated viral-cell fusion.](image-url) The figure 1.6 was adapted from Wilen et al., Cold Spring Harb. Perspect. Med. (2012)\textsuperscript{85}. (i) The HIV-1 Env glycoprotein is comprised of three heterodimers of receptor binding gp120 and gp41 fusion subunits. (ii) HIV-1
Env glycoprotein first binds to CD4, and (iii) this interaction triggers conformational changes within Env, resulting in the formation of the co-receptor binding site. The gp120 variable loop 3 directly interacts with the co-receptor. (iv) Subsequently, the fusion peptide of the gp41 exposes itself from a buried place and inserts into the target membrane, followed by structural rearrangements of the gp41, leading to the formation of six-helix bundle post-fusion conformation and the merger of viral and host cell membranes.

1.6.2 Class II viral fusion proteins. Viruses in togaviridae, flaviviridae, and bunyaviridae families contain structurally unrelated class II fusion proteins to catalyze the process of membrane fusion. Class II fusion proteins consist almost entirely of β-stands and they share the same structural fold despite a low sequence similarity among divergent envelope (E) proteins. Typical class II fold include a three-part ectodomain (domains I-III), followed by a flexible stem region and a C-terminal transmembrane anchor. Each ectodomain is comprised of a β-barrel domain I, an elongated domain II, and an immunoglobulin-like domain III at the C-terminal end. The domain II bears the internal fusion peptides formed by tightly packed hydrophobic loops, while the domain III is responsible for cellular tropism and attachment. However, the receptor binding mechanism is not clear for many class II fusion proteins. These proteins exist as anti-parallel homo- or heterodimers in their pre-fusion state and assemble into canonical (alphaviruses) or non-equivalent (flaviviruses) icosahedral outer shells on the viral surface. The fusion mechanism is conserved in all class II fusion proteins and triggered by the low pH of endosomal compartments. Upon activation, the pre-fusion homo- or heterodimer realigns as an extended intermediate, where the hydrophobic fusion loops at the tips of domain II interacts with the target cell membrane. Subsequently, the extended intermediate folds back on itself and the domains I, II, and III associate into the trimeric core of the post-fusion conformation. These conformational changes bring the transmembrane anchor and the fusion loops together, leading to the fusion of viral and cellular membranes.

1.6.2.1 Dengue virus E glycoprotein. Dengue virus, a member of the flaviviridae family, is an arthropod-borne enveloped virus which encodes two proteins, membrane (M) and envelope (E), on the surface. In the endoplasmic reticulum (ER) of an infected cell, newly assembled
immature dengue virus particles contain asymmetric trimers of precursor M (prM) and E heterodimers\textsuperscript{92,93}. The prM tightly interacts with the E while passing through the low pH environment of the trans-Golgi network (TGN), and prevents premature fusion-inducing conformational changes\textsuperscript{94,95}. Cleavage by furin at a furin cleavage site (residues 91-92) releases the N-terminal fragment of prM and activates the E proteins to undergo structural rearrangements in the final step of maturation\textsuperscript{94–96}. Mature dengue particles possess an icosahedral scaffold of 90 E glycoprotein homodimers on the surface that are primed for low-pH induced membrane fusion\textsuperscript{97} (Figure 1.7). Similar to some other class II viral fusion proteins, Dengue virus E protein (residues 1-495) is an antiparallel homodimer of elongated finger-like molecules in its pre-fusion state\textsuperscript{98} (Figure 1.7). The ectodomain (residues 1-394) consist of three globular domains (domain I, II, and III) essentially constituted by β-sheets\textsuperscript{98,99}. The N-terminal domain I is a β-barrel and structurally positioned at the center of the molecule\textsuperscript{98,99}. Domain II has an elongated, finger-like structure and comprises the fusion loops at the tip of the molecule\textsuperscript{98,99}. The fusion loops are buried at the dimer interface in the pre-fusion conformation\textsuperscript{98}. At the C-terminal end of the polypeptide chain, domain III has an immunoglobulin-like fold stabilized by conserved disulfide bonds. It serves to connect the E protein ectodomain to the transmembrane anchor (residues 450-495) through the flexible stem region (residues 395-449)\textsuperscript{92,98,99}. The receptor binding function of the E protein relies on interactions with the C-type DC-SIGN at the cell surface\textsuperscript{100}. Immediately after Dengue virus particles encounter the low pH environment of endosomes, the E protein homodimer dissociates, resulting in the exposure of the fusion loops\textsuperscript{94}. The exposed fusion loops at the tip of domain II insert into the target endosomal membrane, and facilitate the quaternary reorganization of the membrane attached E protein into homotrimers\textsuperscript{101–103} (Figure 1.7). The core trimer is comprised of domains I and II, and the collapse of the extended trimeric intermediate brings the domain III closer to the core trimer\textsuperscript{90,103}. Subsequent stabilizing interactions between domain I of each monomer at the center, and the final folding back of domain III at the C-terminal end of the molecule complete the zipping up of the post-fusion E trimer to bring the fusion loops and the transmembrane anchor next to each other, leading to membrane merger and pore formation\textsuperscript{99} (Figure 1.7).
Figure 1.7. Model of Dengue virus E glycoprotein conformational changes. The figure 1.7 was reproduced from Harrison, S.C., Nat. Struct. Mol. Biol. (2008)35. (i) 90 E glycoprotein homodimers pack in an icosahedral array on the surface of a dengue virus particle. Each monomer in the following structures are either shown in color (blue, red, and yellow), dark gray or light gray. (ii) The mature pre-fusion dengue virus E glycoproteins form dimers and lie flat on the viral surface. (iii) Monomeric transition between dimeric pre-fusion state and trimeric extended intermediate state. The three monomers project outward and interact with the target cell membrane, but they are not yet in contact. (iv) The low pH environment of endosomes triggers structural rearrangements, and the domains I (red) and II (yellow) form the trimeric core. The lower and upper arrows indicate the subsequent conformational changes necessary for the formation of the post-fusion trimer. (v) Domain III (blue) and the stem region (blue random coil region) are flipped over and zip back along the trimeric core, bringing together the transmembrane anchor, inserted in the viral cell membrane, and the fusion loop, inserted into the target cell membrane, into closer proximity35. The post-fusion conformation facilitates membrane merger and pore formation.

1.6.3 Class III viral fusion proteins. Enveloped viruses from rhabdoviridae, herpesviridae, and baculoviridae contain structurally and functionally distinct viral fusion proteins that belong to a unique class43,104. These class III viral fusion proteins share overlapping features with both class I and class II fusion proteins31,35,104. In their pre-fusion state, class III viral fusion proteins adopt a trimeric conformation with a central α-helical coiled-coil that is similar to class I viral fusion
proteins\textsuperscript{105}. On the other hand, the primarily $\beta$-strand fusion domain of class III viral fusion proteins resembles its counterparts in class II viral fusion proteins. However, the topological organization of the $\beta$-strands is different than those observed in class II viral fusion proteins\textsuperscript{105-108}.

1.6.3.1 Vesicular stomatitis virus (VSV) G glycoprotein. VSV is a member of the rhabdovirus family, which possesses a single G glycoprotein (residues 17-511) on the viral surface to mediate cellular attachment and membrane fusion\textsuperscript{109}. Atomic structures of the VSV G have been determined in both pre- and post-fusion states\textsuperscript{105,106} (Figure 1.8). The VSV G ectodomain (residues 17-467) folds into four distinct domains: a fusion domain (FD, residues 68-197), a pleckstrin homology domain (PHD, residues 53-63 and 199-275), a central trimerization domain (residues 35-52, 276-326, and 400-422), and a $\beta$-sheet lateral domain (residues 18-34 and 327-399)\textsuperscript{105,106}. These four domains are connected to the transmembrane domain (TMD, residues 468-488) and the cytoplasmic tail (CT, residues 489-511) via a C-terminal segment (residues 423-430)\textsuperscript{104,109}. VSV utilizes the low-density lipoprotein receptor (LDLR) for cellular attachment, and the G protein-mediated membrane fusion is triggered by low pH in endosomal compartments\textsuperscript{110}. Unlike class I and class II viral fusion proteins, the conformational changes of VSV G from its pre-fusion to post-fusion state is reversible, and the G protein can be reactivated to the prefusion state by raising the pH of the environment to neutral or above\textsuperscript{111-113} (Figure 1.8). The G protein does not associate with other proteins on the viral surface\textsuperscript{104,109}. Hence, the reversible conformational changes allow the G protein to refold back into its pre-fusion state after passing through the low pH environment of the TGN\textsuperscript{112,113}. The G protein is trimeric in both pre- and post-fusion states, and the structural transitions are pH-dependent\textsuperscript{105,106} (Figure 1.8). At pH 7.5 or above, the G protein is in its trimeric pre-fusion state and it is ready to catalyze the fusion of viral and host cell membranes\textsuperscript{105}. As the pH of the environment becomes slightly acidic (around pH 6.7), the G protein transitions to a rod-shaped monomeric structure\textsuperscript{112,113}. When the pH of the environment drops below pH 6.5, the monomeric intermediates undergo structural rearrangements and form the trimeric post-fusion conformation\textsuperscript{106} (Figure 1.8). The G protein has two hydrophobic fusion loops within the fusion domain that are not buried and face toward the viral membrane in the pre-fusion state\textsuperscript{105,114}. Upon exposure to low pH, the fusion domain elongates and the hydrophobic fusion loops at the tip of the fusion domain position themselves
toward the target cell membrane\textsuperscript{112,113} (Figure 1.8). As the fusion loops interact with the target cell membrane under low pH, the C-terminal segment that connects the ectodomain to the TMD also undergoes drastic conformational changes\textsuperscript{112,113}. In the postfusion state, the C-terminal segment transitions loop-to-helix, and positions itself between the fusion domain and the central trimerization domain\textsuperscript{106} (Figure 1.8). These conformational changes allow VSV G monomers to adopt a more elongated conformation, which is similar to the loop-to-helix transition observed in IAV HA\textsuperscript{106,115}. In addition, three lateral helices pack alongside the central trimeric coiled-coil in the post-fusion state, resulting in a six-helix bundle conformation that is similar to the post-fusion state of some class I viral fusion proteins\textsuperscript{43,106} (Figure 1.8). The final post-fusion conformation of the trimeric G protein brings viral and host cell membranes to close proximity and promotes fusion.

\textbf{Figure 1.8. Structural transitions of fusogenic VSV G protein.} The figure 1.8 was reproduced from Baquero \textit{et al.}, \textit{Curr. Opin. Virol.} (2013)\textsuperscript{113}. At pH 7.5, VSV G protein forms a pre-fusion trimer on the viral surface\textsuperscript{105,113}. As the virus enters low pH endosomal compartments (pH 6.7 and lower), the trimer dissociates and elongated monomeric structures are observed. It is currently unknown whether the G protein trimerization occurs after (\textbf{i}) or before (\textbf{ii}) hairpin
formation. The post-fusion structure of the G protein suggests that the trimer-of-hairpins at the site of contact promotes fusion and pore expansion\textsuperscript{106,113}.

1.6.4 Class IV viral fusion proteins. Non-enveloped orthoreoviruses and aquareoviruses encode nonstructural viral proteins named fusion-associated small transmembrane (FAST) proteins to mediate cell-cell membrane fusion and syncytium formation\textsuperscript{45,116}. However, FAST proteins are not essential for the viral life cycle\textsuperscript{45}. There are at least six recognized FAST proteins (p10, p13, p14, p15, p15, p22) encoded by fusogenic reoviruses and these are the smallest known fusogens (95-198 residues) with very low to no sequence identity between them\textsuperscript{45,117–124}. FAST proteins catalyze membrane merger stages of cell-cell fusion and promote direct cell-cell virus transmission\textsuperscript{118}. Members of the FAST protein family have distinct structural and functional features compared to the fusion proteins of enveloped viruses, and thus define a new class of viral fusogens\textsuperscript{31,117,118}.

1.6.4.1 FAST protein structural features. FAST proteins have a small ectodomain (~19-37 residues), a transmembrane domain (TMD), and a large endodomain (~36-141 residues), all of which participate in the membrane fusion process\textsuperscript{120,125–133}. In contrast to the class I, II, and III viral fusion proteins, the majority of their mass is embedded in the lipid bilayer or cytoplasm\textsuperscript{118,130,133} (Figure 1.9). Although all three domains play a direct role in cell membrane fusion, active actin remodeling is still required for catalyzing the merger of two cell membranes\textsuperscript{120,126–134}. Unlike enveloped virus fusion proteins, the fusion mechanism of FAST proteins is not controlled by a specific triggering event such as receptor binding, low pH or proteolytic cleavage\textsuperscript{122,123,135}. Instead, FAST proteins catalyze the fusion of a variety of cell types at neutral pH\textsuperscript{118} (Figure 1.9). In the plasma membrane, liquid-ordered membrane microdomains facilitate the oligomeric assembly of the FAST proteins\textsuperscript{136–138}. Oligomerization of FAST proteins is a reversible step and it is required for cell-cell fusion\textsuperscript{137} (Figure 1.9). FAST protein ectodomains do not exert cellular attachment capabilities and they depend on cell surface adhesion molecules to mediate cell-cell attachment\textsuperscript{134}. Different FAST proteins utilize a diverse range of fusion peptides (FPs), such as a proline-hinged loop, cysteine loop, or polyproline type II helix, to interact with the outer leaflet of target membrane bilayers\textsuperscript{127,130,131,133} (Figure 1.9).
The insertion of the FP into the lipid bilayer induces membrane curvature, and the juxtaposition of two membranes result in lipid mixing and formation of a hemifusion stalk\textsuperscript{126,128,130}. Then, the TM, the hydrophobic, and the amphipathic regions within the endodomain participate in nascent fusion pore formation and post-fusion pore expansion\textsuperscript{126,128,130}. The large FAST protein endodomains play a greater role in facilitating fusion pore expansion leading to syncytium formation\textsuperscript{128,129}. In addition, the endodomain C-terminal tails may recruit cellular cofactors such as the calcium-dependent membrane binding protein annexin A1 (AX1) for syncytium formation\textsuperscript{139,140} (Figure 1.9). In summary, FAST proteins are virus-encoded cellular fusogens that exert a minimalistic approach to mediate the merger of two cell membranes.
Figure 1.9. Structural motifs and proposed fusion mechanism of FAST proteins. The figure 1.9 was reproduced from Ciechonska et al., Trends Microbiol. (2014)\textsuperscript{118}. (A) Cartoon representation of the orthoreovirus (top) and aquareovirus (bottom) FAST proteins embedded in the plasma membrane. Color code is described in the legend. Total number of residues in each protein is indicated after the C-terminal endodomain\textsuperscript{118}. (B) FAST protein-mediated membrane fusion has three stages. Cell surface adhesion molecules such as calcium-dependent cadherins (colored brown) mediate the initial interactions between the two cell membranes. Actin polymerization (red arrows) may then facilitate the oligomerization of FAST proteins (N-terminal TMD residues colored yellow and the rest colored green) in lipid rafts and subsequent insertion of the fusion peptide into the target cell membrane. Further structural rearrangements promote lipid mixing and the formation of a hemifusion intermediate. The FAST proteins then
catalyze the merger of two cell membranes and lower the energy barrier for stable pore formation. The C-terminal tails of the endodomain may recruit AX1 and/or other factors to promote pore expansion and syncytium formation\textsuperscript{118}.

1.7 **Cell-cell fusion**

Cell-cell fusion is fundamental to the physiology of uni- and multi-cellular organisms ranging from yeast to humans. Malfunction in cellular fusion events in humans might be associated with diseases such as infertility, preeclampsia, and myopathies\textsuperscript{41,42}. Various cellular fusogens and their binding partners display diverse structural and functional features, and execute distinct fusion mechanisms to mediate the merger of two cell membranes\textsuperscript{41,141}. Despite the diversity of molecules that take part in cell-cell fusion events, some common principles still apply to many cellular fusogens\textsuperscript{41,42,141}. A comparison of fusion events in unrelated cellular processes from early embryonic development to synaptogenesis would provide key insights into the molecular mechanisms of homotypic and heterotypic cell fusion. Here, we describe major cellular fusion processes involved in mammalian fertilization, placentation, organ shaping, muscle development, intracellular communication, immune response, carcinogenesis, tissue regeneration, and yeast mating in order to understand the general aspects of cell-cell recognition, adhesion, and fusion.

1.7.1 **Sperm and oocyte fusion.** Mammalian fertilization begins with sperm deposition, where active sperm migrate through the female reproductive tract in order to reach the site of fertilization in the oviduct\textsuperscript{142,143}. Sperm and egg fusion is the culminating event in fertilization, resulting in the development of a new, genetically distinct individual\textsuperscript{144–146}. Both sperm and egg undergo a series of morphological and functional changes before the fusion of two plasma membranes. During oocyte growth, primordial follicles develop into secondary oocytes by the processes of oocytogenesis, meiosis I, and meiosis II, until the oocyte growth is halted at the metaphase II stage prior to fertilization\textsuperscript{147,148} (Figure 1.10). On the other hand, the sperm released from testis is not fusion competent and it undergoes further maturation steps (capacitation) as it passes through the male epididymis and female reproductive tract\textsuperscript{149,150}. Capacitation alters the
motility of the sperm and renders it capable of fusion. However, the capacitated sperm still needs to penetrate through the two extracellular layers of oocyte, namely the multicellular cumulus oophorus and the glycoprotein rich zona pellucida, before it can fuse with the egg plasma membrane\textsuperscript{144,150–152}. First, the active sperm penetrates through the cumulus layer and binds the zona pellucida\textsuperscript{150,151}. The initial interactions with zona pellucida glycoproteins (ZP1-4) trigger the final maturation step (acrosome exocytosis) of capacitated sperm where the outer acrosomal membrane fuses with the overlaying plasma membrane, leading to the exposure of the inner acrosomal membrane as well as the equatorial segment\textsuperscript{143,151,153} (Figure 1.10). Both exposed components are critical for interactions with the egg plasma membrane\textsuperscript{143}. In addition, the initial interactions between an unknown sperm protein and the zona pellucida glycoproteins are species-specific, and serve as a gatekeeper for cross-species fertilization\textsuperscript{151,153}. Upon acrosome exocytosis, sperm penetrates through the zona pellucida and reaches the perivitelline space prior to its fusion with egg plasma membrane\textsuperscript{143,145,146}. Despite the central importance of sperm-egg fusion for fertilization, the molecular details of the gamete interaction are not well understood.

Over the past decades, many proteins have been implicated in sperm and egg fusion, such as Sperm Lyzozyme-Like Acrosomal Protein 1 (SLLP1)\textsuperscript{154,155}, integrins, a disintegrin and a metalloproteinase domain-containing (ADAM) proteins\textsuperscript{156}, Equatorin, sperm equatorial segment protein 1 (SPESP1)\textsuperscript{157,158}, sperm acrosome-associated (SPACA) family proteins\textsuperscript{159,160}, transmembrane protein 190 (TMEM190)\textsuperscript{161,162}, sperm acrosomal membrane-associated 14 (SAMP14)\textsuperscript{163} and angiotensin-converting enzyme 3 (ACE3)\textsuperscript{164} on the sperm surface, and CD9, CD81 and integrins on the egg surface\textsuperscript{165–167} (Figure 1.10). However, targeted gene deletion experiments producing knockout mice models revealed that none of these proteins are essential factors for the sperm and egg recognition/fusion machinery. Only in 2005 was a sperm surface protein, Izumo1, identified through screening antisperm monoclonal antibodies and shown to be essential by the generation of \textit{Izumo1}-deficient mice\textsuperscript{168} (Figure 1.10). Izumo1\textsuperscript{1−/−} male mice were characterized to be healthy without any developmental abnormalities but infertile\textsuperscript{168}. Subsequent studies showed that Izumo1\textsuperscript{1−/−} sperm can penetrate through the zona pellucida, but cannot fuse with the egg plasma membrane and accumulate in the perivitelline space. Izumo1 is a membrane-anchored cysteine-rich glycoprotein expressed on the sperm inner acrosomal membrane and it is conserved in mammals\textsuperscript{168,169}. Human Izumo1 (residues 1-350) has a cysteine-
rich N-terminal α-helical/β-strand domain (residues 22-166) followed by an immunoglobulin-like C-terminal domain (residues 167-254), a long linker region (residues 255-292), a transmembrane domain (residues 293-313), and a cytoplasmic tail (residues 314-350). The Izumo protein family consists of four members (Izumo1-4), three of which contain a transmembrane domain and exhibit testis/sperm specific expression (Izumo1-3) whereas Izumo4 does not have a transmembrane anchor and is expressed in testis and other tissues. Members of the Izumo family share very low sequence identity (<3%) with one another. The equatorial segment of sperm is the initial binding site to the egg plasma membrane (Figure 1.10). Izumo1 is predicted to relocate from the anterior part of the sperm head to the equatorial segment and participate in sperm and egg fusion. Yet, it is not known whether Izumo1 alone is capable of catalyzing the merger of two membranes or if it rather recruits other proteins to facilitate the fusion process as a multiprotein complex at the site of fusion.

Although the N-terminal domain of Izumo1 has been shown to play a critical role in sperm and egg fusion, the receptor of Izumo1 on the egg plasma membrane was not known for many years. In 2014, a glycophosphatidylinositol (GPI)-anchored cysteine-rich glycoprotein, Juno (residues 1-250), on the egg surface was shown to be the receptor of Izumo1 (Figure 1.10). Juno is previously known as folate receptor-δ (FOLR-δ) and it shares high sequence identity with folate receptor-α and -β (FOLR-α and FOLR-β). However, six key folate-binding residues observed in FOLR-α and FOLR-β are not conserved in Juno. Therefore, Juno is not able to coordinate folate, and is chiefly important in fertilization. In addition, Juno molecules on the egg surface are rapidly shed approximately 40 minutes after fertilization suggesting that this may be a possible mechanism for an effective block to polyspermy in mammalian eggs. Similar to Izumo1, female Juno-null mice show normal breeding behavior but the Juno-/- eggs are infertile, with an apparent defect of gamete fusion. Thus far, Izumo1 and Juno are the only two proteins believed to be essential for sperm and egg recognition/fusion, and the interactions between these proteins provide a strong adhesion between a sperm and egg cell, leading to their fusion.
Figure 1.10. Model of sperm-egg adhesion and fusion machinery. The figures 1.10A and 1.10B were reproduced from Evans, J.P., *Annu. Rev. Physiol.* (2012) and the figure 1.10C was adapted from Sutovsky, P., *Expert Rev. Mol. Med.* (2009). (A) Schematic diagram of a mammalian egg arrested in the metaphase II stage. (B) Schematic representation of the capacitated sperm head before and after acrosome exocytosis. The acrosome-reacted sperm binds to the egg plasma membrane via its equatorial segment, leading to heterotypic membrane fusion. (C) Sperm and egg surface proteins that may be relevant to sperm-oocyte adhesion and fusion.
1.7.2 Role of the captured syncytins in placentation. Endogenous retroviruses (ERVs) occupy approximately 8% of the human genome, display the typical genome organization of infectious retroviruses and encode intact functional proteins\textsuperscript{13,176,177}. Over time, most ERVs have been subjected to mutations, insertions or large deletions, and have lost some of their functional capabilities\textsuperscript{178,179}. However, in rare situations, single genes from ERVs have been conserved in the host genome, suggesting that those genes may provide a selective advantage to their hosts. Syncytins are fusogenic \textit{envelope (env)} genes of retroviral origin that have been independently captured by their host to fulfill critical physiological functions in placentation\textsuperscript{180–182}. The \textit{syncytin} genes were first discovered in primates (\textit{syncytin-1} and -2), exhibiting fusogenic properties and placenta-specific expression\textsuperscript{182,183}. Other homologous counterparts were also identified within the orders of eutherian mammals, including Rodentia (\textit{syncytin-A}, -B and -Mar1)\textsuperscript{184,185}, Lagomorpha (\textit{syncytin-Ory1})\textsuperscript{186}, Carnivora (\textit{syncytin-Car1})\textsuperscript{187}, Ruminantia (\textit{syncytin-Rum1})\textsuperscript{188} and Tenrecidae (\textit{syncytin-Ten1})\textsuperscript{189}, as well as in marsupial mammals (\textit{syncytin-Opo1})\textsuperscript{190}, which are the closest living relatives of placental eutherian mammals. All identified \textit{syncytin} genes have completely distinct origins and genomic locations in each lineage\textsuperscript{180,191}. Yet, they display the same placenta-specific expression profile and fusogenic properties, in addition to their sequence conservation, since the time of their insertion into the host genome\textsuperscript{180,191}. Hence, these findings suggest that the \textit{syncytin} genes have been co-opted by their host to facilitate important functions in placentation.

The placenta is an autonomous and transient organ that enables the efficient exchange of nutrients, gases and hormones during intrauterine fetal growth. The placenta also plays an essential role in mitigating the maternal immune response against the allogeneic fetoplacental graft. In humans, invasive placentation is initiated by fetal cytotrophoblast cells that differentiate into villous and extravillous trophoblasts for normal placental function. In the villous trophoblast, the mononuclear cytotrophoblast cells differentiate and fuse together to form a continuous, multinucleated syncytiotrophoblast layer that is highly invasive towards the maternal endometrium, and critical for the implantation of the embryo (Figure 1.11). The syncytiotrophoblast layer has numerous placental functions including absorption, reciprocal metabolic exchanges between maternal and fetal blood, and synthesis of hormones specific for fetal development (Figure 1.11). Investigations on the molecular mechanisms underlying
trophoblast differentiation and cell-cell fusion have led to the identification of the potential physiological relevance of the syncytin proteins in placental development. In situ analysis of the human placenta revealed that syncytin-1 is broadly expressed in all villous cytotrophoblast, syncytiotrophoblast and extravillous cytotrophoblast cell lineages, whereas syncytin-2 expression is restricted to villous cytotrophoblast cells. It was also demonstrated that the β-catenin/BCL9L/TCF4 signaling pathway regulates the fusion of trophoblast cells through targeting of the placenta-specific transcription factors, GCM1 (glial cell missing 1) and GLI2 (Glioma-associated oncoogene), as well as that induced expression of GCM1 along with GLI2 transactivates syncytin-1 and -2 expression in trophoblast cells. Furthermore, generating the homozygous syncytin-A knockout mice resulted in the death of embryos in utero between 11.5 and 13.5 days post coitum. The embryonic deaths were attributed to the failure of trophoblast cells to fuse into a syncytiotrophoblast layer as well as structural defects at the feto-maternal interface. Clinically, decreased gene expression and aberrant localization of syncytins may be related to pre-eclamptic pregnancies and other disorders; however, further studies are required to understand their functions completely.

The protein syncytin-1 is encoded by the env gene of the ERVWE1 proviral locus, which belongs to the human endogenous retrovirus-W (HERV-W) family. Retroviral Env glycoproteins belong to the class I viral glycoprotein family. Syncytin-1 Env (1-538) is initially expressed as a single polypeptide, which is then cleaved by cellular proteases into two subunits—a receptor binding surface (SU) subunit (21-313) that is responsible for interacting with specific cellular receptors, and a membrane-anchored fusion-active transmembrane (TM) subunit (318-538) that drives the fusion between the two cell membranes. The mature syncytin-1 Env is incorporated on the trophoblast cell surface as a metastable trimer of SU-TM heterodimers. In addition, syncytins are endowed with an immunosuppressive function by virtue of a specific immunosuppressive domain (ISD) within their TM subunit (Figure 1.11). Through their ISDs, certain syncytins contribute to maternal immune tolerance in a localized, placenta-specific manner. In summary, syncytins are a unique example of convergent evolution in which retrovirus env genes have been captured independently by different mammalian clades to fulfill physiological function in placental morphogenesis.
Figure 1.11. Syncytin-mediated cell-cell fusion in human placentation. The figure 1.11 was reproduced from Lavialle et al., Philos. Trans. R. Soc. Lond. B Biol. Sci. (2013)180. (A) Schematic representation of syncytin-1 Env180. (B) Homotypic cell-cell fusion-mediated by syncytins results in syncytium formation. Human syncytin-1 and -2 bind their cell surface receptors ASCT-2 (Na⁺-dependent neutral amino acid transporter 2) and MFSD2 (major facilitator superfamily domain containing 2), respectively, and subsequent conformational changes within the fusion subunit induce cell-cell fusion180. (C) Schematic representation of human placental villus180.

1.7.3 The fusion failure (FF) proteins in epidermal cell fusion. The nematode C. elegans is a powerful system to study cell-cell fusion, because stereotypic fusion of its 959 somatic cells generates 44 syncytia that are essential for the organogenesis of epidermis, pharynx, uterus, vulva, hymen, tail, excretory duct, and glands209,210. Through genetic screens for fusion-defective mutants, a novel family of integral membrane proteins named fusion failure (FF) proteins was identified in C. elegans and shown to be sufficient for cell fusion211,212. The epithelial fusion failure 1 (EFF-1) is the most common fusogen responsible for cell fusion events in the epidermis, pharynx, and vulva, whereas the anchor-cell fusion failure 1 (AFF-1) is involved in the formation of hymen, pharyngeal muscles, and epidermal syncytia in C. elegans212–214. Based on sequence similarities, other FF family proteins were also identified in arthropods, ctenophores, hemichordates, and protists215,216. The wide distribution of FF family proteins is likely due to the genome contamination of these organisms with nematode sequences31. The fusogenic EFF-1 protein is structurally homologous to class II viral fusion proteins, and it mediates the fusion of
two cell membranes through the hemifusion intermediate stage similar to the class I, II, and III viral fusion proteins\textsuperscript{211,214}. However, EFF-1 or AFF-1 has to be present in both cell membranes in order to catalyze membrane fusion. The homotypic arrangement of FF-mediated fusion is distinct from syncytins and viral fusion proteins, which are only present on one fusing membrane\textsuperscript{212,217}. The EFF-1 protein (residues 1-658) is a heavily glycosylated protein containing mostly $\beta$-sheets stabilized by eight disulfide bonds\textsuperscript{22}. It is comprised of three extracellular domains (domain I, II, and III, residues 23-509), followed by a C-terminal stem region (residues 510-555), a transmembrane domain (residues 556-576), and a cytoplasmic tail (residues 577-658)\textsuperscript{22}. In the pre-fusion state, EFF-1 proteins are predicted to be in a metastable monomeric form on the cell surface\textsuperscript{22} (Figure 1.12). At the site of fusion, EFF-1 monomers assemble into trimers with two other EFF-1 monomers on the surface of adjacent cells, where domains I and II facilitate the core trimer formation\textsuperscript{22}. The homotypic EFF-1 fusion machinery then undergoes structural rearrangements and extends into the hemifusion intermediate state\textsuperscript{217} (Figure 1.12). Subsequently, the direction of the polypeptide changes and domain III relocates in between subunits of the extended trimer and initiates the membrane curvature of both fusing membranes\textsuperscript{218}. Similar to the class II viral fusion proteins, the three globular domains of EFF-1 form a hairpin-like conformation in the post-fusion state and bring the two membranes together, leading to membrane mixing and fusion pore formation\textsuperscript{22} (Figure 1.12). Although the structural and mechanistic features of EFF-1 proteins are similar to the fusion mechanism of the class II viral fusion proteins, the triggering mechanism and the molecular details of the trans-trimerization process remain unknown for FF protein-induced membrane fusion.
Figure 1.12. Schematic representation of EFF-1-mediated homotypic cell-cell fusion. The figure 1.12 was reproduced from Pérez-Vargas et al., *Cell* (2014)\textsuperscript{22}. EFF-1 monomer is colored red, yellow, and blue for domains I, II, and III, respectively. C-terminal stem region is shown in magenta. (A) EFF-1 is a monomer in the pre-fusion state. (B) EFF-1 monomers cluster at the site of fusion, and (C) three monomers come together to form an extended intermediate trimer. (D) Domain III relocates and binds to the core trimer. (E) Then, the C-terminal stem region flips and make contact with the membrane proximal part of domain II in the trimer. (F) Positioning of the three stem regions alongside the trimeric core brings the two membranes into close proximity, leading to fusion. (G) Zippering of the post-fusion trimer completes the fusion of the two membranes and facilitates fusion pore formation and expansion\textsuperscript{22}. 
1.7.4 Fusion of intracellular vesicles. In multi-compartment eukaryotic cells, soluble N-ethylmaleimide-sensitive factor (NSF) attachment protein receptor (SNARE) proteins play a key role in many membrane fusion events, including membrane trafficking, hormone secretion and neurotransmitter release\textsuperscript{219–222}. Specific combinations of SNARE proteins, which are located on opposing membranes, form SNARE complexes and the free energy released upon complex formation is enough to drive the fusion of two opposing membranes\textsuperscript{223,224}. However, other proteins, such as synaptotagmin in neurotransmitter release, are also involved in the regulation of SNARE-mediated membrane fusion\textsuperscript{221,223–225}. SNARE complexes are recyclable and the disassembly of these complexes is mediated by ATPase NSF and soluble NSF attachment proteins (SNAPs)\textsuperscript{226,227}. The disassociation of SNARE complexes renders a pool of individual proteins that are fusion-competent for the next cycle of SNARE-mediated membrane fusion\textsuperscript{228,229}. The core synaptic fusion complex formed during the release of neurotransmitter consists of three proteins: the target membrane SNAREs (t-SNAREs) syntaxin-1 (stx-1) and SNAP-25, and vesicle SNARE (v-SNARE) synaptobrevin 2 (syb2). Both stx-1 (residues 1-288) and syb-2 (residues 1-116) are anchored in the plasma membrane via a single transmembrane (TM) helix, whereas SNAP-25 (residues 1-206) has lipid anchors in the plasma membrane\textsuperscript{220,226,230,231} (Figure 1.13). Similar to the class I fusion proteins, the cytoplasmic portions of all three proteins form a long $\alpha$-helix that includes heptad repeat motifs ($a$, $b$, $c$, $d$, $e$, $f$, $g$) where positions $a$ and $d$ are occupied by hydrophobic residues\textsuperscript{231}. During vesicle fusion, a v-SNARE (syb2) interacts with the t-SNAREs (stx-1 and SNAP-25) to form a highly stable complex that contains one copy of syb2 and stx-1 proteins and two copies of the SNAP-25 protein\textsuperscript{21,220,224,231,232} (Figure 1.13). Finally, zipperring up into a four-helix bundle brings the two membranes into close proximity and induces fusion\textsuperscript{231,232} (Figure 1.13). The reversible SNARE complex is then disassembled by NSF and SNAPs in order to reuse the SNAREs for another round of membrane fusion\textsuperscript{233,234}. Recently, the structural studies of NSF-mediated SNARE complex disassembly suggested that one NSF/SNAP species could facilitate the disassociation of many different SNARE complexes by recognizing the electrostatic pattern of these complexes\textsuperscript{234}. Upon binding, NSF and SNAPs could apply shear and/or pulling forces to twist and unwind the SNARE complex into individual SNARE proteins\textsuperscript{234}. 
**Figure 1.13. Model of membrane fusion facilitated by SNAREs alone.** The figure 1.13 was reproduced from Rizo et al., *Annu. Rev. Biophys.* (2015)\textsuperscript{224}. (A) Schematic representation of SNARE motifs, linker domains (LDs), and transmembrane (TM) regions of SNAREs. (i) Two copies of SNAP-25, stx-1, and syb2 are colored in green, yellow, and red, respectively. All TM regions are shown in gray. (ii and iii) The current model postulates that the complex formation between v- and t-SNAREs assembles from the N- to the C-terminus in multiple stages. There could be additional stages in the assembly pathway, facilitating membrane fusion. (iv) Final four-helix bundle conformation induces membrane fusion and pore opening\textsuperscript{224}. (B) Crystal structure of the neuronal SNARE complex in four-helix bundle conformation\textsuperscript{235}. The TM regions are shown in different shades of gray. NTH and CTH indicate the N- and C-terminal halves of the four-helix bundle SNARE complex\textsuperscript{224}.

1.7.5 Myoblast Fusion. The fusion event between two types of muscle cells, muscle founder cells and fusion competent myoblasts (FCMs), results in the formation of multinucleate muscle fibers\textsuperscript{236–238}. Myoblast fusion is an essential step in skeletal muscle differentiation because skeletal muscle is the product of thousands of myoblasts fused into multinucleated muscle fibers\textsuperscript{236–238}. At least three consecutive steps are required for the fusion of FCMs and muscle founder cells\textsuperscript{42,238}. The first step is the muscle cell adhesion, which is mediated by immunoglobulin (Ig) domain-containing transmembrane proteins. In *Drosophila*, the Dumbfounded (Duf) and Roughest (Rst) proteins in founder cells, and Sticks and stones (Sns)
and Hibris (Hbs) proteins in FCMs mediate recognition and adhesion of the two types of muscle cells via direct interactions\textsuperscript{239–244}. Structural and functional studies of \textit{C. elegans} homologs of Duf and Sns, SYG-1 and SYG-2, demonstrated that these adhesion molecules mostly use their membrane-distal Ig domains to interact with one another and form an L-shaped structure in order to bring the two apposing membranes \textasciitilde{}45 nm of one another\textsuperscript{245}. The second step toward myoblast fusion is to bring the two cell membranes into close proximity. Myoblast cell fusion is an asymmetric process in which the FCM cells insert F-actin-enriched finger-like protrusions into the muscle founder cells\textsuperscript{246,247}. On the other hand, Myosin II (MyoII) in the muscle founder cells acts as a mechanosensor and accumulates at the fusion site\textsuperscript{248} (Figure 1.14). The presence of MyoII at the site of fusion increases cortical tension and facilitates fusion pore formation\textsuperscript{248}. The last step is the disruption of the lipid bilayer and the merger of two apposing membranes\textsuperscript{238}. However, the molecular players involved in this final step of myoblast fusion have yet to be discovered. The high mechanical tension resulting from protrusive and resisting forces of FCMs and muscle founder cells, respectively, likely play a role in promoting cell membrane fusion\textsuperscript{238} (Figure 1.14).

\subsection*{1.7.6 Macrophage fusion.} Homotypic fusion of macrophages can form osteoclasts in bone for skeletal stability or multinucleated giant cells (MGCs) during an immune response\textsuperscript{249,250}. Macrophages have to acquire fusion competence through extracellular and intracellular stimuli such as receptor activator of nuclear kappa B ligand (RANKL), macrophage colony-stimulating factor (M-CSF), and IL-4, all of which upregulate fusion mediators on the cell surface and induce macrophage fusion\textsuperscript{251–254}. In order to fuse, macrophages must attach to each other and facilitate the juxtaposition of the two membranes. Two members of the immunoglobulin superfamily (IgSF), CD47 (Integrin associated protein (IAP)) and Sirp-\(\alpha\) (Macrophage fusion receptor (MFR)), have been implicated in macrophage adhesion\textsuperscript{255–257}. Experiments using knockout mice indicate that CD47 and Sirp-\(\alpha\) on apposing cell membranes act as a receptor pair and facilitate macrophage fusion\textsuperscript{257,258}. The adhesion event between CD47 and Sirp-\(\alpha\) is in a manner analogous to that of SYG-1 and SYG-2 in \textit{C. elegans}, and Duf and Sns in \textit{Drosophila} myoblast fusion\textsuperscript{238,239,241,245}. Additionally, a dendritic cell-specific transmembrane protein (DC-STAMP) was identified as an important cell-cell fusion regulator of macrophages\textsuperscript{259,260}. Multinucleation of osteoclasts was abrogated in DC-STAMP knockout mouse models, and DC-
STAMP-deficient macrophages lacked the formation of multinucleate osteoclasts and MGCs, indicating that DC-STAMP is essential for cell-cell fusion. Even though DC-STAMP is clearly involved in macrophage fusion, identification of its binding partner(s) on the apposing cell membrane and the precise mechanism of DC-STAMP-mediated cell-cell fusion still requires further investigation. As novel molecular mediators of the macrophage fusion machinery are identified, our understanding of any malfunction in homotypic macrophage fusion leading to forms of osteopetrosis will improve and pave the way for potential therapies.

1.7.7 Stem cell and cancer cell fusion. Membrane fusion events are not only restricted to viral cell and cell-cell fusion mechanisms. Cell fusion has also been implicated in tissue regeneration and genetic repair by stem cells (SCs). The role of cell fusion in tissue regeneration of liver, muscle, brain, and lung by stem cells has only emerged in the past years. Specifically, circulating hematopoietic stem cells (HSCs) and bone marrow derived cells (BMDCs) are able to fuse with several cell types including hepatocytes, cardiomyocytes and oligodendrocytes under physiological conditions or following an injury. It is tempting to speculate that similar to other cell-cell fusion mechanisms, specific fusion proteins may induce the fusion event between the SCs and the target cells. However, we do not fully understand the mechanisms regulating the SC fusion process. Cancer cells are also fusion competent and can merge with stromal cells or other cancer cells, leading to tumor progression and metastasis. Relatively little is known about the proteins involved in these cell-cell fusion events and the mechanisms by which cancer cells fuse with other cells. For instance, human oncogenic viruses, such as human T-lymphotropic virus type 1 (HTLV-1) and Epstein-Barr virus (EBV), are all fusogenic, yet it is unknown whether their fusogenic activity plays a role in carcinogenesis. Nevertheless, studies focusing on fused cancer cells may provide the first insights into the mechanisms of cancer cell fusion.
Figure 1.14. Models of myoblast fusion and yeast mating. The figure 1.14A was reproduced from Kim et al., Dev. Cell (2015)\textsuperscript{248} and the figure 1.14B was reproduced from Oren-Suissa et al., Trends Cell Biol. (2007)\textsuperscript{42}. (A) Mechanical tension in myoblast membrane fusion. FCM generates F-actin enriched podosomes and invade the opposing founder cell membrane. MyoII acts as a mechanosensor and mechanical tension drives cell membrane fusion\textsuperscript{248}. (B) Prm1 (red) protein on the surface of mating haploid yeast cells localizes at the site of membrane fusion, and it may play a critical role in catalyzing the merger of two plasma membranes\textsuperscript{42}.

1.7.8 Yeast mating. In \textit{Saccharomyces cerevisiae}, the pheromone exchange between opposite mating type haploid cells (MAT-α and MAT-a) initiates the mating process\textsuperscript{278,279}. Pheromone detection by the opposing mating type results in activation of gene transcription, cell cycle arrest, and cell polarization, followed by the cell wall degradation of mating partners and the membrane fusion process\textsuperscript{276,277} (Figure 1.14). Bioinformatic screens for membrane proteins, whose expression is induced by pheromones, led to the identification of a previously uncharacterized pheromone-regulated membrane protein 1 (Prm1) in \textit{S. cerevisiae}\textsuperscript{278}. Prm1 (residues 1-661) is a multispanning transmembrane protein, which forms disulfide-linked homodimers and localizes at the site of membrane fusion upon pheromone induction\textsuperscript{276,278–280} (Figure 1.14). However, Prm1 is not essential for \textit{S. cerevisiae} mating, because only 50\% of Δprm1 mating pairs result in fusion failure\textsuperscript{278}. Furthermore, Δprm1 mating pairs show normal membrane juxtaposition and locally break down their cell walls but they arrest at the membrane fusion stage\textsuperscript{278}. Prm1 does not have properties predictive of viral and cellular fusogens such as HIV-1 Env and \textit{C. elegans} EFF-
Thus, it is unlikely that Prm1 is a *bona fide* fusogen in the yeast mating process, but it could be a fusion facilitator acting upstream of the fusion process\textsuperscript{276,277}. Alternatively, yeast fusion machinery may require the formation of a multiprotein fusion complex in order to catalyze the merger of two plasma membranes, and Prm1 might be one of the core components of this complex\textsuperscript{277}. The critical fusion protein mediating the yeast plasma membrane fusion and the underlying mechanisms remain largely unknown.

### 1.8 Thesis Overview

**1.8.1 Rationale and Hypothesis.** Membrane fusion is an essential process for the progression of numerous developmental, physiological and pathological events in living organisms. Many factors appear to regulate the fusion process, including receptors and ligands, proteases, signaling molecules and fusogenic proteins. Most fusion processes start with sequential cell-cell recognition, alignment and adhesion events, forming the initial interactions between the two membranes. Subsequently, the fusion machinery orchestrates various steps during the fusion event and catalyzes the merging of the two lipid bilayers. Fusogenic proteins are important in the life cycle of many organisms and are one of the key elements of the membrane fusion process.

My thesis work has focused on characterizing the general features of various human and viral glycoproteins required for cellular attachment and fusion. Given the central importance of viral and cellular glycoproteins in membrane fusion processes, we hypothesized that there may be common structural and functional features that are utilized by disparate glycoproteins to catalyze the merger of two apposing membranes.

Although atomic structures of many class I fusion proteins involved in viral-cell fusion events have been determined in the past, molecular mechanisms stabilizing these fusogens under various pH environments and key elements of many fusion processes remain unclear. The first part of my work described in chapters 2, 3, and 4 attempts to characterize the general structural features of class I viral fusion proteins further in detail in order to uncover determinants required for the stabilization of class I viral fusion glycoproteins.
Class I fusion proteins also play an important role in critical cell-cell fusion events during development. Human syncytins are class I fusion proteins of retroviral origin that fulfill critical physiological functions in placentation. However, there are no complete structures of any syncytin proteins and molecular details of synctin-mediated cell-cell fusion are not well-understood. The second part of the work described in chapter 5 investigates the structural and functional details of synctin-mediated cell-cell fusion to gain insights into the fusion mechanism of the human syncytin-1 glycoprotein in placentation.

Cell fusion is a process that is crucial at many crossroads in cell biology. In sexually reproducing species, the fusion of sperm and egg plasma membranes results in the creation of a new genetically distinct diploid organism termed zygote. Izumo1 and Juno are the only known proteins truly essential for sperm and egg fusion. In chapter 6, we determine molecular architecture of the human sperm Izumo1 and egg Juno adhesion/fusion complex and characterize the molecular details of their interactions. These are the first atomic-resolution structures of any protein complex between sperm and egg at the point of conception for any organism. Our results now provide the foundation to further characterize the interactions of sperm and egg and allow us to better understand the fundamental principles of the mammalian fertilization and fusion processes.

This work will allow us to uncover the molecular details of viral and developmental membrane fusion. Ultimately, our goal is not only to understand the mechanisms of membrane fusion but also to illuminate how viruses initiate infection and how the malfunction of physiological processes such as placentation and sperm-egg fusion eventually result in human disease. These studies will potentially open new avenues for the development of therapeutics targeting these processes.
Chapter 2

Crystal structures of β- and γ-retroviral fusion proteins reveal a role for electrostatic stapling in viral entry

Halil Aydin¹, Jonathan D. Cook¹ and Jeffrey E. Lee¹

¹Department of Laboratory Medicine and Pathobiology, Faculty of Medicine, University of Toronto, Toronto, ON M5S 1A8, Canada

Published Manuscript
The Journal of Virology January 2014 vol. 88 no. 1 143-153
Copyright © 2014, American Society for Microbiology

- Author Contributions:

Halil Aydin: Designed the project, performed cloning, mutagenesis, expression, purification, CD spectroscopy and thermal melts, immunoblotting, pseudovirus production, viral titration and viral infectivity assays, performed bioinformatics and structural analysis.

Jonathan D. Cook: Determined the crystal structure of MPMV TM and refined XMRV TM and MPMV TM structures.

Jeffrey E. Lee: Designed the project, performed cloning, crystallization experiments, collected synchrotron X-ray diffraction data, determined and refined the crystal structures of XMRV TM and MPMV TM, supervised the research.
2.1 Summary

Membrane fusion is a key step in the life cycle of all envelope viruses, but this process is energetically unfavorable; the transmembrane fusion subunit (TM) of the virion-attached glycoprotein actively catalyzes the membrane merger process. Retroviral glycoproteins are the prototypical system to study pH-independent viral entry. In this study, we determined crystal structures of extramembrane region of the TMs from Mason-Pfizer monkey virus (MPMV) and xenotropic murine leukemia virus-related virus (XMRV) at 1.7 Å and 2.2 Å resolution, respectively. The structures comprise of trimer-of-hairpins that is characteristic to class I viral fusion proteins and now completes a structural library of retroviral fusion proteins. Our results allowed us to identify a series of intra- and interchain electrostatic interactions in the heptad repeat and chain reversal regions. Mutagenesis reveals that charge neutralizing salt bridge mutations significantly destabilize the postfusion six-helix bundle and abrogate retroviral infection, demonstrating that electrostatic stapling of the fusion subunit is essential for viral entry. Our data indicates that salt bridges are a major stabilizing force on the MPMV and XMRV retroviral TMs and likely provides the key energetics for viral and host membrane fusion.
2.2 Introduction

The fusion of the viral lipid bilayer with the host membrane is indispensable for the obligate intracellular life cycle of all enveloped viruses. Viral-host cell fusion is typically catalyzed by one or more virion-associated surface envelope glycoproteins, which are structurally organized into three classes: class I, II, and III (reviewed in 43,281). Regardless of the class, the glycoprotein is displayed on the surface of the virus in a metastable, prefusion conformation, and depending on the virus, the glycoprotein is activated for fusion through one of three principal mechanisms: (i) receptor-binding in a pH-independent fashion, (ii) low-pH, or (iii) through a combination of both receptor-binding and low-pH (reviewed in35,43).

Retroviruses are a distinct family of enveloped RNA viruses that are classified into seven genera: α-, β-, δ-, ε-, γ-retroviruses, lentivirus, and spumavirus. The retroviral envelope glycoproteins (termed Env) belong to the class I viral fusion protein family. Env is synthesized as a single-chain polypeptide precursor and post-translationally cleaved to yield a receptor-binding surface (SU) and fusion-active transmembrane (TM) domain. The mature Env is incorporated as a metastable, trimer of SU-TM heterodimers on the surface of the virus. Initial attachment to host cells is mediated through the interaction of the SU domain with a host cell receptor. In many retroviruses, this is thought to trigger a conformational change that allows the TM subunit to form an extended prehairpin conformation that facilitates the insertion of its hydrophobic fusion peptide into the host plasma membrane. Subsequently, two α-helical, heptad repeat regions (HR1 and HR2) in the TM fold back together in a structural rearrangement that brings the two bilayers together. The mixing of the outer leaflets of the viral and host membranes forms a hemifusion stalk. The hemifused bilayers further open into a fusion pore as the TM stabilizes into its final postfusion six-helix bundle conformation. Although membrane fusion is a thermodynamically feasible process, it has a large activation barrier. The formation of the highly stable six-helix bundle is thought to provide the free energy required for membrane fusion.

The β-, δ-, γ-retroviruses are the prototypical system for understanding pH-independent entry. Crystal structures of two δ-retrovirus fusion subunits, HTLV-1282 and bovine leukemia virus
(BLV)\textsuperscript{283}, have been determined previously, and both TM subunits were shown to adopt a similar trimeric coiled coil structure at the postfusion conformation. Surprisingly, there are not any complete fusion subunit structures for β- and γ-retroviruses, as the previously reported Moloney murine leukemia virus (MoMLV) TM structure displays only the heptad repeat 1 (HR1) and chain reversal regions of the fusion core; its entire C-terminal HR2 region is missing\textsuperscript{284}. Additional structural information from β- and γ-retrovirus fusion glycoproteins will complete a representative structural library for the retrovirus family and allow for a detailed cross-comparison with other viral fusion proteins to identify common features important for pH-independent entry.

Here, we determined the structure of the complete extramembrane region of the fusion protein from both the β- and γ-retroviral genera. Moreover, comparison of the MPMV and XMRV TM structures unveiled the presence of key electrostatic interactions, clustered at two unique regions of the postfusion six-helix bundle structure. Site-directed mutagenesis of these electrostatic staples abrogated viral entry, and reduced the overall stability of the fusion proteins. Our findings propose a common biochemical strategy utilized by pH-independent retroviruses to stabilize their postfusion glycoprotein conformation to overcome the energy barrier for fusion.
2.3 Results and Discussion

2.3.1 Structure determination. We determined the crystal structures of the fusion core regions of the transmembrane proteins of Mason-Pfizer monkey virus (MPMV), a β-retrovirus, and xenotropic murine leukemia virus-like virus (XMRV), a γ-retrovirus. The MPMV and XMRV fusion cores lack the hydrophobic N-terminal fusion peptide and the membrane-associated transmembrane anchor (Figure 2.1A). In addition, to prevent non-specific disulfide bond formation, the third cysteine in the CX_6CC motif was mutated to serine (MPMV: C483S; XMRV: C538S). This third cysteine is thought to mediate a disulfide linkage between the SU and TM subunits, similar to Ebola virus glycoprotein (GP)^48. Both MPMV and XMRV TM purify as a clear trimer by size exclusion chromatography. The resulting MPMV and XMRV TM structures (Figure 2.1B) were refined against 1.7 Å and 2.2 Å resolution data, respectively. Electron density maps revealed clear density for residues 421-512 and 476-563 in MPMV and XMRV TM, respectively (Figure 2.1C). Weak electron density was observed for the first 16 residues in XMRV TM. MPMV has a predicted N-linked glycosylation site at N487 in its fusion subunit. However, our crystallized MPMV TM was expressed in bacteria, and therefore lacks N-linked glycosylation. This is not expected to affect the overall structure of MPMV TM given that N487 resides in the tether loop region, rather than the more structured HR1 or HR2 regions. The MPMV TM described here is the first fusion protein structure from the β-retrovirus family, and the XMRV TM structure now provides a complete picture of fusion cores from γ-retroviruses, as the previously determined MoMLV TM structure^284 was missing the entire C-terminal region of the molecule.

2.3.2 Structure of retroviral fusion glycoproteins. All retroviral fusion glycoprotein subunits that have been biochemically characterized adopt a trimeric structure. Each monomer contains one long central α-helical region and a second shorter α-helix at its C-terminus (Figure 2.1B). These two helices are termed heptad repeat regions 1 and 2 (HR1 and HR2, respectively), and each consists of a repeating pattern of seven amino acids (a,b,c,d,e,f,g) with hydrophobic residues at the first and fourth positions. Hydrophobic leucine and isoleucine residues on the HR1 α-helical region align in regular knobs-into-hole packing to form a long, central trimeric core. In all of the retroviral TMs, a single asparagine layer in HR1 disrupts the leucine/isoleucine
knob-in-hole packing\textsuperscript{285} to make contact with a putative chloride ion that may act as a switch to allow conformational changes necessary for fusion with the host cell membrane. In the $\delta$-retroviral TMs, a second asparagine layer coordinates an ordered water molecule\textsuperscript{282}. At the base of the HR1 core, a short $3_{10}$ helix and a disulfide bond-stabilized, single turn $\alpha$-helix form the chain reversal region where the polypeptide chain completes a $180^\circ$ turn (Figure 2.1B). The polypeptide chain then extends into a random coil that tethers the HR2 helix. The helical content of the HR2 varies from six turns in MPMV to five turns in XMRV TM. The HR2 helix packs into the grooves formed between two HR1 helices to form the six-helix bundle that is widely considered to be the postfusion conformation. The packing of HR2 into HR1 brings the amino and carboxy termini together, positioning the virus and the host membrane-interacting segments of the full-length viral glycoprotein into close proximity. The MPMV and XMRV TMs have a high degree of overall structural similarity with the HTLV-1 and BLV fusion subunits\textsuperscript{282,283} (Figure 2.1D) (root mean squared deviation=\textasciitilde1.6 Å).
Figure 2.1. Structures of retroviral fusion glycoproteins. (A) General diagram of a CX6CC-containing retroviral envelope glycoprotein. MPMV and XMRV Env are synthesized as a single polypeptide precursor that is post-translationally cleaved to form the SU and TM subunits. The MPMV and XMRV SU subunit typically contains 6 and 10 N-linked glycans (red Y symbols), respectively, and is the subunit responsible for attachment to the host receptor. The MPMV Env N-linked glycosylation pattern is shown on the schematic. The TM subunit contains a single N-linked glycan, a hydrophobic fusion peptide (FP) and two α-helical heptad repeat regions (HR1 and HR2). The two heptad repeat regions are separated by a chain reversing CX6CC motif and a short tether region. The first and second cysteines in the CX6CC motif form an intrasubunit disulfide bond, and the third cysteine forms an intersubunit disulfide linkage with a cysteine in the SU subunit. The carboxy terminus contains a transmembrane (TM) anchor and a cytoplasmic tail (CT) that interacts with the retroviral gag protein. Fusion subunit regions included in the construct for structural studies are shown in color. (B) Ribbon representation of the β- and γ-retrovirus fusion cores. The intrasubunit disulfide bonds are indicated as sticks. (C) Representative σA-weighted 2Fo-Fc electron density map contoured at 1σ, superimposed with the final MPMV TM and XMRV TM refined model (N-terminal HR1 portion shown). Overall, the electron density is well defined throughout the HR1, HR2 and chain reversal regions. (D) Structural similarity of retroviral fusion subunits. Superimposition of the fusion subunit from HTLV-1 (PDB-1MG1, blue), MPMV (PDB-4JF3, green), and XMRV (PDB-4JGS, burgundy).

2.3.3 Identification of intra- and inter-subunit salt bridges. Structural comparison of the MPMV and XMRV TMs revealed an intricate network of intra- and inter-chain electrostatic interactions (Figure 2.2A). These interactions were not previously identified, as the HR2 region in previously crystallized MoMLV TM was missing. Retrovirus electrostatic interactions cluster into two distinct regions on the TM fusion core: the heptad repeat and chain reversal regions.

2.3.3.1 Heptad repeat region: The binding of HR2 to HR1 is stabilized through multiple single and complex-type salt bridges (Figure 2.2A). In MPMV TM, an intramolecular electrostatic interaction (D435-R506) is formed at the top of the fusion subunit, and the central HR region is
anchored by a series of complex, interchain electrostatic interactions. In the central HR1, three sets of complex salt bridges are formed (K495-D445-K488, D445-K488-D448 and K488-D448-R493). In XMRV TM, three separate pairs of salt bridges span the heptad repeat region (D490-R557, E495-R555 and E502-R548); it lacks the complex salt bridge in the middle of the HR1 core observed in the MPMV TM. In all of the retroviral fusion proteins, the HR1 trimer is rich in negatively charged aspartate or glutamate residues, and these residues cluster to form layers of anionic charge (Figure 2.3). On the counterpart HR2 helical surface in retroviral TMs, a cationic surface complementary to the negative charge is found. These positively and negatively charged residues are highly conserved within their own retroviral family.

2.3.3.2 Chain reversal region: Both MPMV and XMRV fusion proteins contain a single, structurally conserved, intersubunit arginine-glutamate electrostatic pair (MPMV TM: R462-E480 and XMRV TM: R517-E535) in the chain reversal region (Figure 2.2A). The arginine residue is present at the base of the HR1 region, whereas a glutamate resides on a loop after the single turn α-helix in the CX6CC motif.
Figure 2.2. Intra- and complex intermolecular retroviral salt bridges. (A) Ribbon diagram of the MPMV and XMRV TM. Electrostatic interactions are shown in the inset boxes. Residues labeled with asterisks are contributed by a neighboring subunit. (B) Relative infectivities of wild-type and salt-bridge mutant pseudotyped retroviruses. Data represent the average of at least six independent experiments. (C) Thermal denaturation of indicated wild-type and mutant retroviral fusion subunits. Data points were fitted to a non-linear biphasic sigmoidal curve in GraphPad.
Figure 2.3. Surface electrostatic potential of the HR1-HR2 interface. A negatively charged surface is displayed on the MPMV and XMRV HR1 trimer and is complementary to a positively charged surface on HR2. The footprint of the HR2 helix on the HR1 trimeric core is shown by a dashed line. Red and blue colored regions denote negative and positive charges, respectively.

2.3.4 Electrostatic stapling of HR1-HR2 is necessary for retroviral entry. The role of each retroviral salt bridge was quantified in a cell-based viral entry assay. Single-cycle infectious viral particles containing MPMV or XMRV full-length wild-type or mutant envelope glycoprotein were produced. Wild-type MPMV and XMRV Env pseudovirions were tested against human embryonic kidney (HEK) 293T cells. HEK 293T cells exhibited high infectivity levels, consistent with previous studies\textsuperscript{286,287}, and were used for all viral infectivity studies. Immunoprecipitation and immunoblot analysis of the producer cells and pseudotyped viruses revealed that mutations to the full-length retroviral glycoproteins did not affect processing or incorporation into virions (Figure 2.4). Electron microscopy of the pseudovirions revealed a
typical retrovirus particle structure of approximately 100 nm diameter, and proper formation of the retroviral capsid and envelope (data not shown). Infectivity levels measured by flow cytometry demonstrated that alanine point mutations to any of the salt bridges in the central parts of the HR or chain reversal regions completely abolished the entry of retroviral particles into host cells or reduced it to very low levels (Figure 2.2B and Table 2.1). Mutations to salt bridges located at the membrane-proximal ends of HR1 and HR2 (MPMV TM: D435-R506) inhibited but did not completely block entry. Inter- and intra-subunit salt bridges that anchor the membrane-distal and central portions of the heptad repeats and those within the chain reversal region were highly sensitive to mutagenic disruption and thus these electrostatic interactions are critical to the viral entry mechanism. The striking all-or-none effect suggests that the loss of a single electrostatic interaction, which typically contributes 3 to 5 kcal/mol of energy for binding288,289, may be enough to prevent crossing the activation barrier threshold for the formation of the postfusion conformation.

2.3.5 Electrostatic staples stabilize the structure of the heptad repeat regions. Circular dichroism (CD) spectroscopy is an excellent tool to investigate protein secondary structure, and the folding and unfolding of macromolecules as a function of temperature. α-helices display negative CD spectral peaks at 208 nm and 222 nm, whereas β-sheet or random coil conformation is detected as single minima at 215 nm or below 210 nm, respectively. CD spectra of the wild-type and mutant MPMV and XMRV fusion proteins exhibited the typical double minima at 208 nm and 222 nm, characteristic of their predominantly α-helical secondary structures. The CD signatures for all retroviral fusion mutants were similar (data not shown), suggesting that the relative α-helical content did not change as a result of the targeted mutations.
Table 2.1. Summary of retroviral fusion protein mutations.

<table>
<thead>
<tr>
<th>Virus/Mutant</th>
<th>Protein expression/ Viral incorporation</th>
<th>Relative infectivity (%)</th>
<th>$Tm_1(°C)$</th>
<th>$Tm_2(°C)$</th>
<th>Location</th>
</tr>
</thead>
<tbody>
<tr>
<td>MPMV (β-retrovirus)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wild type</td>
<td>++++/+++++</td>
<td>100.0</td>
<td>–</td>
<td>78</td>
<td>–</td>
</tr>
<tr>
<td>D435A</td>
<td>++++/+++++</td>
<td>12.7</td>
<td>55*</td>
<td>78</td>
<td>HR1 top</td>
</tr>
<tr>
<td>D445A</td>
<td>++++/+++++</td>
<td>0.0</td>
<td>54*</td>
<td>74</td>
<td>HR1 central</td>
</tr>
<tr>
<td>D448A</td>
<td>++++/+++++</td>
<td>0.9</td>
<td>53*</td>
<td>73</td>
<td>HR1 bottom</td>
</tr>
<tr>
<td>R462A</td>
<td>++++/+++++</td>
<td>0.0</td>
<td>–</td>
<td>71</td>
<td>chain reversal region</td>
</tr>
<tr>
<td>E480A</td>
<td>++++/+++++</td>
<td>0.0</td>
<td>51*</td>
<td>70</td>
<td>chain reversal region</td>
</tr>
<tr>
<td>K488A</td>
<td>++++/+++++</td>
<td>0.0</td>
<td>51*</td>
<td>70</td>
<td>HR2 bottom</td>
</tr>
<tr>
<td>R493A</td>
<td>++++/+++++</td>
<td>0.0</td>
<td>–</td>
<td>69</td>
<td>HR2 bottom</td>
</tr>
<tr>
<td>K495A</td>
<td>++++/+++++</td>
<td>0.0</td>
<td>–</td>
<td>78</td>
<td>HR2 central</td>
</tr>
<tr>
<td>R506A</td>
<td>++++/+++++</td>
<td>10.0</td>
<td>–</td>
<td>73</td>
<td>HR2 top</td>
</tr>
<tr>
<td>XMRV (γ-retrovirus)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wild type</td>
<td>++++/+++++</td>
<td>100.0</td>
<td>–</td>
<td>93</td>
<td>–</td>
</tr>
<tr>
<td>D490A</td>
<td>++++/+++++</td>
<td>0.0</td>
<td>62</td>
<td>89</td>
<td>HR1 top</td>
</tr>
<tr>
<td>E495A</td>
<td>++++/+++++</td>
<td>0.0</td>
<td>66</td>
<td>89</td>
<td>HR1 central</td>
</tr>
<tr>
<td>E502A</td>
<td>++++/+++++</td>
<td>30.5</td>
<td>64</td>
<td>89</td>
<td>HR1 bottom</td>
</tr>
<tr>
<td>R517A</td>
<td>++++/+++++</td>
<td>0.0</td>
<td>–</td>
<td>83</td>
<td>chain reversal region</td>
</tr>
<tr>
<td>E535A</td>
<td>++++/+++++</td>
<td>5.4</td>
<td>–</td>
<td>83</td>
<td>chain reversal region</td>
</tr>
<tr>
<td>R548A</td>
<td>++++/+++++</td>
<td>0.0</td>
<td>67</td>
<td>85</td>
<td>HR2 bottom</td>
</tr>
<tr>
<td>R555A</td>
<td>++++/+++++</td>
<td>0.0</td>
<td>67</td>
<td>86</td>
<td>HR2 central</td>
</tr>
<tr>
<td>R557A</td>
<td>++++/+++++</td>
<td>0.0</td>
<td>–</td>
<td>86</td>
<td>HR2 top</td>
</tr>
</tbody>
</table>

* $T_m$ value is an estimate

Figure 2.4. Characterization of MPMV and XMRV pseudotyped virions. WT and mutant MPMV or XMRV pseudovirions were normalized by MLV p30 ELISA, and separated by SDS-
Expression of WT or mutant Env were detected in pseudovirions and producer cells by immunoblot analysis using an anti-HA primary antibody. All pseudovirion samples were non-reduced, and the mature covalently attached MPMV and XMRV Env (MPMV: gp86 and XMRV: gp85) were detected. In the producer cell lysate, the MPMV and XMRV Env do not dissociate into the SU and TM subunits under reducing conditions, suggesting the presence of the uncleaved precursor glycoprotein (MPMV: pr86 and XMRV: pr85). The mock is a separate experiment consisting of co-transfection of the MLV backbone and gag-pol vectors; no envelope glycoprotein vector was transfected. β-actin was used as a loading control.

The stabilities of retroviral fusion glycoproteins and mutants were determined by monitoring the decrease in helical content at 222 nm as a function of temperature over the range from 20°C to 99°C (Table 2.1). The thermal unfolding of all retroviral fusion proteins was irreversible. Although thermodynamic properties cannot be obtained from an irreversible thermal melt, the estimation of an apparent melting temperature ($T_m$) serves as a simple measure of protein stability. All wild-type retroviral fusion subunits unfolded cooperatively and were highly stable, with apparent $T_m$s above 75°C. The high melting temperatures are consistent with the established stabilities of other viral fusion proteins.$^{290-292}$

Mutations to salt bridges in the heptad repeat regions in XMRV TM resulted in pronounced biphasic transitions in the thermal denaturation curves, whereas wild-type proteins and the chain reversal region salt bridge mutations displayed only single transitions. We hypothesize that in the first transition, $T_{m1}$, the HR1 trimeric core remained intact, and the loss of CD signal resulted from unfolding of the HR2 helix (Figure 2.2C). The second, higher temperature (>80°C) transition ($T_{m2}$) is likely due to unfolding of the HR1 helix core. In XMRV TM, salt bridge mutations in the HR region resulted in 27°C-31°C decreases in $T_{m1}$ as compared to wild type. Specifically, mutations to residues on the cation pocket (R548, R555 and R557 in XMRV TM) revealed a decrease of ~30% in CD signal, which corresponded well with the total α-helical content in the HR2 region. In addition, this was consistent with the structural models, as these residues were only predicted to destabilize HR2 binding to HR1. It is interesting to note that
mutations to the HR1 anion pocket (XMRV TM D490A, E495A and E502A) resulted in a larger loss in CD signal (50-60%). This suggests that mutations to the negatively charged aspartate or glutamate residues in the HR1 helix destabilize both the binding of HR2 and also likely the HR1 trimeric core. In MPMV TM, the two-phase transitions of HR mutants were not as pronounced as those seen in XMRV TM mutants (Figure 2.2C); however there was a decrease in CD signal resulting in a downward shift of the apparent \( T_m \). Given the poorly defined melting transition, we were not able to accurately determine the apparent \( T_{m1} \) and \( T_{m2} \) values for MPMV TM. The lack of a clear biphasic transition in MPMV TM was likely due to the higher number of interchain complex salt bridges in the HR region, as compared to XMRV TM. Overall, our results suggest that extensive ion-pair networks in the HR region provide a general strategy to influence the thermal stability of retroviral fusion subunits.

In the chain reversal region, all salt bridge mutations had a single transition thermal denaturation profile (Figure 2.2C). A single transition was expected, as mutations to the chain reversal region salt bridges should not significantly affect the folding or assembly of HR2 onto the HR1 core. In general, mutations to chain reversal salt bridges had more modest effects on the melting temperature (\( \Delta T_m \sim 10^\circ C \)), with the exception of MPMV TM E480A, than mutations in the HR regions (\( \Delta T_m > 25^\circ C \)). Our results suggest that the chain reversal ion-pair plays a role in stabilizing the fusion subunit, but not as large an extent as the electrostatics found in the HR1 or HR2 region.

2.3.6 Implications of electrostatic interactions to retroviral fusion. Retroviral envelope glycoproteins may be divided into two types: ones that contain a disulfide bond between the SU and TM subunits, and ones that are non-covalently associated\(^{293}\). Covalent-type retroviral Env contains a conserved immunosuppressive domain (ISD) and CX\(_6\)CC motif in the chain reversal region of its fusion subunit. CX\(_6\)CC-containing glycoproteins are commonly observed within the \( \alpha-, \delta-, \) and \( \gamma- \)retroviral genera, but not lentiviruses or spumaviruses (Figure 2.5). The \( \beta- \)retrovirus genus is unique from the others, as examples of both covalent-type and non-covalent-type viral glycoproteins are observed\(^{293}\). For instance, MPMV Env, which contains a CX\(_6\)CC motif, is thought to be derived from a \( \gamma- \)retrovirus through a recombination event. On the other hand, mouse mammary tumor virus (MMTV) and Jaagsiekte sheep retrovirus (JSRV) belong to the \( \beta- \)
retrovirus genus, but the SU and TM subunits are non-covalently associated\textsuperscript{293}, as similar to the lentiviral envelope glycoproteins (HIV-1 gp160).

Electrostatic interactions are of fundamental importance for protein folding and function. The MPMV and XMRV retroviral fusion subunits are rich with salt bridges that form “electrostatic staples”. Each retroviral fusion subunit contains at least four sets of salt bridges, and each electrostatic interaction plays an important role in the formation of the postfusion state. Our sequence alignment of retroviral TMs demonstrated that all covalent-type fusion subunits contain a high degree of sequence similarity, in particular residues at the C-terminal end of the HR1 helix and those within the chain reversal region (Figure 2.5A). Sequence alignments with the non-covalent type Env, such as HIV-1, JSRV and MMTV, reveal poor overall sequence conservation (Figure 2.5B). For instance, MMTV TM has a 30% sequence identity with JSRV TM, but only ~6% sequence identity with the covalent type MPMV TM. Moreover, the non-covalent type fusion subunits are different in size and physiochemical properties to the CX\textsubscript{6}CC-containing fusion subunits. Our crystal structure of MPMV TM is a representative model for other covalent-type β-retrovirus TM proteins, but it is difficult to predict the position or roles of salt bridges from non-covalent-type β-retroviral TMs without a crystal structure.

The intricate network of intra- and intersubunit electrostatic interactions found at the HR1/HR2 and chain reversal regions of MPMV and XMRV TM are strictly conserved within their respective retroviral families. However, an electrostatic salt bridge formed at the membrane proximal end of the HR1/HR2 core (HTLV-1 TM: D353-R416, MPMV TM: D435-R506 and XMRV TM: D490-R557) and an intersubunit electrostatic pair at the chain reversal region (HTLV-1 TM: R380-E398, MPMV TM: R462-E480 and XMRV TM: R517-E535) are structurally conserved in all covalent-type β-, δ-, γ-retrovirus fusion subunits (Figure 2.5A). Given the structural conservation, this suggests that electrostatic interactions may be required at both ends of the fusion protein for entry.
Due to the long-range attractive effects of electrostatic interactions, we suggest that the retroviral salt bridges help locate, position and stabilize the interactions of HR1 and HR2, acting in a manner analogous to a zipper. Cryo-EM reconstructions of the prefusion Moloney murine leukemia virus Env revealed a tripod-like structure that implies the HR2 regions of the fusion subunit are splayed out from the central glycoprotein core\textsuperscript{294}. In order to form the postfusion six-helix bundle, the HR2 region must undergo significant conformational changes to allow its packing into the HR1 trimeric core. The chain reversal region is the point at which the fusion subunit polypeptide chain reverses direction. Thus, after triggering of the retroviral glycoprotein, it is logical that the CR region is important in initiating the conformational changes to form the postfusion state. We hypothesize that the CR region salt bridge (MPMV: R462-E480, XMRV: R517-E535) directs the initial alignment and anchors the turn in order to facilitate HR2 to pack into the prehairpin HR1 intermediate. In addition, the CR region salt bridge is intersubunit in nature, thus the formation of one salt bridge will have the potential to confer a concerted conformational change to the other subunits. The retroviral CR region salt bridge is analogous to the role of a slider in a zipper in fastening the base and aligning the teeth. This interpretation accounts for the comparably low infectivity levels of both the mutated CR and HR1/HR2 core pseudotyped virions, despite significant differences in thermal stability. Thus, stabilization of the chain reversal region during the fusion event appears to be just as important to infectivity as stabilization of the final post-fusion conformation. Previous work on HTLV-1 gp21 also suggested that the R380-E398 chain reversal salt bridge may stabilize the disulfide-bonded chain reversal region\textsuperscript{295}. The electrostatic interactions located between the HR1 and HR2 helices are equivalent to the teeth of the zipper to fasten and stabilize the postfusion six-helix bundle once HR2 packs into the HR1 trimeric core.

The extensive use of electrostatic interactions to stabilize the postfusion structure is consistent with the biology of enveloped viruses. For retroviruses entering the plasma membrane at neutral pH, electrostatic interactions would have a large effect, as acidic and basic residues would be fully ionized to stabilize the six-helix bundle. Notably, a conserved salt bridge (K574-D632) between the N- and C-terminal heptad repeat regions of the HIV-1 gp41 was shown to be critical for stability and viral entry\textsuperscript{296,297}. Moreover, synthetic peptides and small molecule inhibitors targeting the gp41 salt bridge demonstrated potent antiviral activity against HIV-1 mediated
membrane fusion\textsuperscript{298,299}. Many members of the paramyxovirus family, such as respiratory syncytial virus and parainfluenza virus, are also thought to undergo a pH-independent mechanism of entry\textsuperscript{300,301}. The postfusion structures of paramyxovirus family F glycoproteins also reveal the presence of electrostatic interactions in their heptad repeat region\textsuperscript{56,302}, supporting a trend that electrostatic interactions may be an important feature in fusion subunits that function in a pH-independent manner.

In contrast, the role of electrostatic interactions in class I fusion proteins from viruses that enter through an endosomal pathway is not clearly defined. In influenza A virus (IAV) and lymphocytic choriomeningitis virus (LCMV), a number of salt bridges are observed in its fusion protein subunit\textsuperscript{54,303}. It is likely that these salt bridges play a role in viral entry, as mutation of the LCMV-equivalent salt bridge in another arenavirus (Junin virus) resulted in reduced viral infectivity\textsuperscript{304}. No thermal denaturation studies on these electrostatic interactions are available and thus its role in stabilizing the postfusion six-helix bundle is unclear. It is tempting to speculate that electrostatic interactions in fusion subunits from pH-dependent viruses are not involved in stabilizing the postfusion state; the precise role of these interactions await further studies.

In this study, we determined the structures of the β- and γ-retrovirus fusion subunits, thus completing a structural library of pH-independent retrovirus fusion proteins. Structural analysis of the fusion proteins reveals an abundance of electrostatic interactions in the heptad repeat and chain reversal regions. The conserved nature of the ion pairs suggests that electrostatic stapling is critical to stabilize the postfusion six-helix bundle. The elimination of the MPMV and XMRV TM salt bridges destabilized the fusion core, and fully blocked entry in a cell-based assay. Disruption of these electrostatic staples presumably imbalance the energetics required to overcome the large activation barrier to fuse the viral and host lipid bilayers. Our results describe a key feature of the class I retroviral fusion machinery and improves our overall understanding of mechanisms of viral entry.
Figure 2.5. Sequence alignment of retroviral fusion subunits. Sequence alignments are separated into (A) covalent-type and (B) non-covalent-type retroviral envelope glycoproteins. Secondary structural definitions and residue numbering shown above the alignments are based on HTLV-1 gp21 and HIV-1 gp41 for the covalently and non-covalently attached Env, respectively. Residues involved in electrostatic interactions are labeled with an asterisk, and those that are conserved within the β-, δ-, and γ-retroviral families are highlighted in green. Abbreviations are as follows: HTLV-1/2: human T-lymphotropic virus; STLV-1/2/3: simian T-cell leukemia virus; BLV: bovine leukemia virus; MPMV: Mason-Pfizer monkey virus; SMRV: squirrel monkey retrovirus; SRV-1/2: simian retrovirus; RD114: feline endogenous retrovirus; REV; Reticuloendotheliosis virus; SNV: spleen necrosis virus; GALV: gibbon ape leukemia virus; KORV: koala retrovirus; PERV: porcine endogenous retrovirus; XMRV: xenotropic murine leukemia virus-like virus; MoMLV: Moloney murine leukemia virus; KiMSV: Kirsten murine sarcoma virus; FeLV: feline leukemia virus; Ga-FeSV: Gardner-Arnstein feline sarcoma virus; ASLV: avian sarcoma leukosis virus; RSV: Rous sarcoma virus; HIV-1: human immunodeficiency virus-1; MMTV: mouse mammary tumor virus; JSRV: Jaagsiekte sheep retrovirus.
2.4 Materials and Methods

2.4.1 Expression and protein purification. DNA corresponding to Mason-Pfizer monkey virus (MPMV) TM (residues 412-513; GenBank: NP_056894.1) and xenotropic murine leukemia virus-like virus (XMRV) TM (residues 472-568; GenBank: ACY30460.1) were codon-optimized, and synthesized with a 6-histidine tag followed by a TEV cleavage site into a pJexpress414 vector (DNA2.0 Inc). Cysteine to serine mutations (MPMV: C483S; XMRV: C538S) were engineered to abrogate non-specific intermolecular disulfide bond mediated aggregation. All wild-type (WT) or mutant MPMV and XMRV fusion glycoproteins were expressed in BL21 (DE3) E. coli cells. Cell cultures were grown to an OD$_{600}$=0.6 at 37°C and induced with 0.5 mM IPTG for 18 hours at 25°C. Cells were lysed using a hydraulic cell disruption system (Constant Systems). Protein purification was performed stepwise using Ni-NTA metal affinity followed by size exclusion chromatography on a prep grade Superdex-75 10/300 column equilibrated in 1X TBS (10 mM Tris-HCl pH 7.5 and 150 mM NaCl). Protein concentration was quantified by A$_{280}$.

2.4.2 Crystallization and structure determination. Prior to crystallization, the poly-histidine tag was removed using proteases. For MPMV TM, TEV protease (1:100 protease to protein molar ratio), purified as previously described$^{305}$, was used to cleave the protein at 22°C for 24 hours. For XMRV TM, TEV protease did not efficiently cleave the tag, thus a chymotrypsin digest (1:3000 protease to protein molar ratio) was performed at 22°C for 1.5 hours. Cleavage reactions were stopped with a final concentration of 1 mM PMSF and subsequently purified on a Superdex-75 10/300 column equilibrated in 1X TBS. MPMV and XMRV TM were concentrated to 15 mg/mL and crystallized by hanging drop vapor diffusion. Crystals for MPMV TM were obtained in 18% (w/v) PEG 3400 and 0.2 M sodium thiocyanate, while plate-like crystals were grown for XMRV TM in 23% (w/v) PEG 3350, 0.1 M Bis-Tris pH 5.5, 0.2 M NaCl and 4% (v/v) 2,2,2-trifluoroethanol. Both MPMV and XMRV TM crystals were soaked in a cryoprotectant containing the mother liquor components plus 25% (v/v) glycerol, prior to being flash cooled in liquid nitrogen. Data for XMRV and MPMV TMs were collected on a Rigaku FR-E Superbright X-ray generator and Saturn A200 HD CCD detector. All data were reduced and scaled using d*TREK$^{306}$. The MPMV and XMRV TM structures were determined by molecular replacement
using the program Phaser\textsuperscript{307} and human syncytin-2 (PDB code: 1Y4M)\textsuperscript{308} and Moloney murine leukemia virus TM (PDB: 1MOF)\textsuperscript{284} as the initial search models, respectively. Iterative rounds of model rebuilding were performed using the program Coot\textsuperscript{309}, followed by simulated annealing torsion angle refinement using the program PHENIX.refine\textsuperscript{310}. Data collection and refinement statistics are presented in Table 2.2.

2.4.3 Circular dichroism spectroscopy and thermal melts. Circular dichroism (CD) spectral scans and thermal melts of fusion proteins (concentrations ranging from 0.2-2 mg/ml) were acquired on a Jasco J-810 spectropolarimeter using 1 mm quartz cuvettes (Helma). CD wavelength scans were collected between 190-250 nm and averaged over five accumulations. Thermal denaturation assays were carried out at a single wavelength (222 nm) by increasing the temperature from 20°C-99°C in 2°C/min increments. The heating of all retroviral fusion proteins to 99°C in a 10 mM potassium phosphate pH 7.5 and 150 mM NaCl buffer was insufficient to denature the predominantly α-helical fusion proteins. 1M Guanidine-HCl was added to all samples in order to fully denature MPMV TM and XMRV TM within a temperature range of 20°C-99°C. The data were baseline corrected and plotted as a function of folded and unfolded states versus temperature. $T_m$ values were calculated from thermal melt curves fitted using a biphasic sigmoidal non-linear regression algorithm in GraphPad.

2.4.4 Viral entry assays

2.4.4.1 Cells and expression constructs. All cell culture media, additives and FBS were purchased from Life Technologies/Gibco. Human embryonic kidney (HEK) 293T cells (ATCC #CRL-11268) were cultured in 1X Dulbecco’s Modified Eagle Medium (DMEM) supplemented with 10% (v/v) heat inactivated fetal bovine serum (FBS), 1X antibiotic-antimycotic, and propagated at 37°C in a 5% CO₂ incubator. A single-cycle murine leukemia virus (MLV) pseudovirus was generated by a multi-plasmid expression system consisting of a shuttle vector for MLV, a MLV gag-pol encoding plasmid, and a plasmid encoding the appropriate XMRV or MPMV envelope glycoprotein\textsuperscript{311,312}. 
Table 2.2. Data collection and refinement statistics.

<table>
<thead>
<tr>
<th><strong>Data collection statistics</strong></th>
<th><strong>MPMV TM</strong></th>
<th><strong>XMRV TM</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>Space group</td>
<td>P3</td>
<td>P2₁,2₁,2₂</td>
</tr>
<tr>
<td>Unit cell dimensions (Å)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>a=</td>
<td>44.0</td>
<td>100.0</td>
</tr>
<tr>
<td>b=</td>
<td>44.0</td>
<td>148.4</td>
</tr>
<tr>
<td>c=</td>
<td>81.7</td>
<td>58.1</td>
</tr>
<tr>
<td>α–β–</td>
<td>90°</td>
<td>90°</td>
</tr>
<tr>
<td>γ=</td>
<td>120°</td>
<td>90°</td>
</tr>
<tr>
<td>Wavelength (Å)</td>
<td>1.54</td>
<td>1.54</td>
</tr>
<tr>
<td>Resolution range (Å)</td>
<td>34.5-1.7</td>
<td>47.6-2.2</td>
</tr>
<tr>
<td>Total reflections</td>
<td>101,292</td>
<td>295,524</td>
</tr>
<tr>
<td>Unique reflections</td>
<td>18,731 (1,694)</td>
<td>44,752 (4,394)</td>
</tr>
<tr>
<td>Completeness (%)</td>
<td>96.2 (85.4)</td>
<td>99.9 (99.5)</td>
</tr>
<tr>
<td>Rₘₐₓ,ₑₙₑₑ (%)ᵃ</td>
<td>4.6 (26.5)</td>
<td>8.8 (50.3)</td>
</tr>
<tr>
<td>Multiplicity</td>
<td>5.4 (4.1)</td>
<td>6.6 (5.7)</td>
</tr>
<tr>
<td>I/s(I)</td>
<td>19.3 (4.4)</td>
<td>9.0 (2.2)</td>
</tr>
</tbody>
</table>

| **Refinement statistics**     |              |              |
| No. of molecules in asymmetric unit | 2 | 9       |
| No. of residues                | 182         | 757         |
| No. of waters                  | 197         | 301         |
| No. of chloride ions           | 3           | 3           |
| Rₑₓ,ₑₑₑ / Rₑₓ,ₑₓₑ (%)ᵇ         | 15.4/19.3   | 21.4/26.6   |
| Average B-factor (Å²)          |             |             |
| Overall                       | 17.1        | 31.4        |
| Protein                       | 15.8        | 31.3        |
| Solvent                       | 26.4        | 32.9        |
| RMSD bonds (Å)                 | 0.010       | 0.013       |
| RMSD angles (°)                | 1.1         | 1.3         |
| Ramachandran plot statistics (%)|        |              |
| Most favored region            | 100.0       | 99.7        |
| Additional allowed region      | 0.0         | 0.3         |
| Disallowed region              | 0.0         | 0.0         |

ᵃ $R_{\text{merge}} = \Sigma \left| I_1 - \langle I_1 \rangle \right| / \Sigma \langle I_1 \rangle$, where $I_1$ and $\langle I_1 \rangle$ represent the diffraction intensity values of the individual measurements and the corresponding mean values. The summation is over all unique measurements.

ᵇ $R_{\text{work}} = \left( \frac{\sum_{\text{obs}} \left| F_{\text{obs}} - k F_{\text{calc}} \right|}{\sum_{\text{obs}} |F_{\text{obs}}|} \right)$, where $F_{\text{obs}}$ and $F_{\text{calc}}$ are the observed and calculated structure factors, respectively. For $R_{\text{free}}$, the sum is extended over a subset of reflections (5%) excluded from all stages of refinement.
The MLV shuttle vectors are self-inactivating, packageable plasmids containing an eGFP reporter gene and a Spleen focus forming virus 5'-LTR promoter. The MLV gag-pol plasmid contains the full gene. The XMRV env gene was synthesized and cloned into pDISPLAY (Invitrogen). The expression plasmid for MPMV Env (pTMO) and MLV backbone were kindly provided by Dr. Eric Hunter (Emory University) and Dr. Marc-André Langlois (University of Ottawa). XMRV and MPMV Envs were fused with a hemagglutinin A (HA) epitope tag at their N- and C-terminus, respectively, for detection. Viral glycoprotein mutants were generated using the QuikChange II site-directed mutagenesis protocol.

2.4.4.2 Pseudovirus production. 1.5 x 10⁶ HEK 293T cells were seeded in each 10 cm cell culture dish (Corning) and incubated overnight at 37°C with 5% CO₂. At 40% confluency, growth medium was replaced with serum-free 1X DMEM. Expression vectors (2.4 µg packaging, 1.8 µg gag-pol, and 1.8 µg Env plasmids) were mixed with Genejuice transfection agent (Millipore) at a ratio of 3:1 transfection agent to plasmid DNA. The transfection mixture was added dropwise onto HEK 293T cells and incubated for 6 hours. Subsequently, cells were washed with 1X PBS and overlaid with 10 ml of 1X DMEM with 10% (v/v) FBS. Supernatants containing infectious viral particles were collected 48 hours post-transfection and filtered through 0.45 µm pore PVDF syringe filters and concentrated 20-fold using Amicon Ultra-15 centrifugal concentrators (100 kDa MWCO; Millipore).

2.4.4.3 Virus titration. The MLV Core Antigen ELISA kit (Cell Biolabs) was used, according to manufacturer’s protocols, to determine the number of viral particles in the supernatant. Mutant pseudotyped viruses were measured by ELISA for viral particle amounts only and normalized to WT levels for infectivity assays. In addition, for WT Env pseudotyped viruses, biological titration assays were performed to measure the number of transducing viral particles. Briefly, 3 x 10⁵ HEK 293T cells were seeded per well in a 6-well plate 24 hours prior to infection. The concentrated virus-containing supernatants were serially diluted 1, 10⁻¹, 10⁻², 10⁻³, and 10⁻⁴ fold. 50 µl of XMRV or 100 µl of MPMV pseudovirus-containing samples were mixed with 500 µl of 1X DMEM, added to HEK 293T cells and incubated for 90 minutes. Cells were then washed
with 1X PBS and supplemented with and additional 2 ml of 1X DMEM with 10% (v/v) FBS. The percentage of eGFP-positive cells was analyzed by flow cytometry (FACSCalibur, BD Biosciences) 24 hours post-infection. The titre (in transducing unit/ml) is calculated by ((% of eGFP positive cells/100) x number of cells transduced)/volume of virus)\(^{315}\).

2.4.4.4 Detection of GP expression on producer cells. The expression of WT and mutant viral glycoproteins on HEK 293T producer cells were analyzed by immunoblots. Here, Env-transfected cell monolayers were washed with phosphate-buffered saline (PBS) and cells lysed with NP40 lysis buffer (50 mM Tris-HCl pH 8.0, 120 mM NaCl and 0.5% (v/v) NP40) supplemented with 2X protease inhibitor cocktail (Roche). Cell lysates were normalized with Bradford assays and analyzed by anti-HA immunoblot.

2.4.4.5 Detection of Env incorporation into pseudotyped viruses. The incorporation of WT or mutant viral Env into pseudotyped viruses were analyzed by immunoprecipitation (IP). The amount of pseudotyped virus particles was estimated in the supernatant by MLV p30 ELISA as described above. Equivalent amounts of WT and mutant pseudotyped-virus particles were mixed with anti-HA antibody protein A-agarose beads at 4°C for two hours. Subsequently, the beads were washed five times with 1 ml TNEN buffer (20 mM Tris-HCl pH 8.0, 100 mM NaCl, 1 mM EDTA, 0.5% (v/v) NP40) and subjected to immunoblot analysis, as described above.

2.4.4.6 Viral infectivity assay. HEK 293T cells (3.0 x 10^5) were seeded in six-well plates and grown overnight at 37°C in 5% CO\(_2\). These target cells were overlaid with 0.5 ml of pseudovirus containing DMEM media (normalized titres, as described above) and incubated for 90 minutes. Subsequently, cells were washed with 1X PBS and overlaid with 2 ml of 1X DMEM with 10% (v/v) FBS. Samples were analyzed 24 hours post-infection by flow cytometry (FACSCalibur, BD Biosciences). Flow cytometry data were processed using the FlowJo software package. Expression of eGFP was used to determine the efficiency of viral fusion. Experimental results are shown as the number of cells infected by virions containing the mutant viral glycoprotein.
divided by the number of cells infected by virions containing the WT glycoprotein, multiplied by 100 to obtain percentages. Each experiment was done in triplicate and repeated twice.

2.4.5 Accession codes. Atomic coordinates and structure factors for MPMV and XMRV TM have been deposited in the Protein Data Bank (PDB) with the accession code 4JF3 and 4JGS, respectively. All plasmids, along with sequences, have been deposited with Addgene (Addgene ID: 49087 (MPMV) and 49088 (XMRV)).
2.5 Acknowledgements

This work was supported by a Canadian Institutes of Health Research (CIHR) Operating Grant (MOP-115066), Canada Research Chair in Structural Virology, and a CIHR New Investigator Award (MSH-113554) to JEL. Moreover, support for stipends were provided by an University of Toronto Graduate Fellowship to HA and JDC, and an Ontario Graduate Scholarship to JDC. The authors would like to thank the Molecular Structure & Function program at the Hospital for Sick Children Research Institute (Toronto, ON) for access to a circular dichroism spectrometer, Aiping Dong and Drs. Cheryl Arrowsmith and Aled Edwards for access to the Structural Genomics Consortium X-ray diffraction facility; and Farshad Azimi for technical discussions and critical reading of the manuscript. In addition, the authors are indebted for the generosity of Dr. Eric Hunter (Emory University) for the MPMV Env expression construct and Dr. Marc-André Langlois (University of Ottawa) for the MLV backbone.

2.6 Inter-Chapter Transition

Retroviral glycoproteins are the prototypical system to study pH-independent viral entry\textsuperscript{316}. However, avian sarcoma leukosis virus ASLV is an α-retrovirus that undergoes a two-step entry mechanism with receptor binding occurring at the plasma membrane followed by trafficking and fusion at the low pH endosome\textsuperscript{317}. Therefore, structural and functional studies on the ASLV TM may provide insights into the role of fusion protein salt bridges from viruses that are low pH-dependent. In chapter 3, we determined the crystal structure of the ALSV TM and showed that ASLV TM is lined with at least three pairs of salt bridges. We then asked whether these charge-charge interactions play a role in the two-step entry mechanism of the virus. Alanine-scanning mutations of salt bridges failed to significantly reduce postfusion TM stability. Instead, hydrophobic residues were identified as the key to stabilizing the ASLV TM six-helix bundle. The use of non-ionizable residues to stabilize the ASLV postfusion glycoprotein makes biochemical sense, as hydrophobic interactions will be unaffected by pH changes when the virus migrates to the endosome, whereas electrostatic interactions will have reduced effects at acidic pH.
Chapter 3

Structural characterization of a fusion glycoprotein from a retrovirus that undergoes a hybrid two-step entry mechanism

Halil Aydin¹, Brianna M. Smrke¹ and Jeffrey E. Lee¹

¹Department of Laboratory Medicine and Pathobiology, Faculty of Medicine, University of Toronto, Toronto, ON M5S 1A8, Canada

Published Manuscript
The FASEB Journal December 2013 vol. 27 no. 12 5059-5071
Copyright © 2013 FASEB

- Author Contributions:

Halil Aydin: Designed the project, performed cloning, mutagenesis, expression, purification, Circular Dichroism (CD) spectroscopy, thermal melt and crystallization experiments, determined and refined the crystal structure of ASLV TM, performed bioinformatics and structural analysis.

Brianna M. Smrke: Assisted with protein expression and purification, performed crystallization experiments, determined and refined the crystal structure of ASLV TM.

Jeffrey E. Lee: Performed structural analysis and supervised the research.
3.1 Summary

Entry of enveloped viruses into host cells is mediated by their surface envelope glycoproteins (Env). On the surface of the virus, Env is in a metastable, prefusion state, primed to catalyze the fusion of the viral and host membranes. An external trigger is needed to promote the drastic conformational changes necessary for the fusion subunit to fold into the low-energy, six-helix bundle. These triggers typically facilitate pH-independent entry at the plasma membrane or pH-dependent entry in a low pH endosomal compartment. The α-retrovirus avian sarcoma leukosis virus (ASLV) has a rare, two-step entry mechanism with both pH-dependent and pH-independent features. Here, we present the 2.0 Å resolution crystal structure of the ASLV TM fusion protein. Our structural and biophysical studies indicated that unlike other pH-dependent or pH-independent viral TMs, the ASLV fusion subunit is stable irrespective of pH. Two histidine residues (H490 and H492) in the chain reversal region confer stability at low pH. A structural comparison of class I viral fusion proteins suggests that the presence of a positive charge, either histidine or arginine amino acids, stabilizes a helical dipole moment and is a signature of fusion proteins active at low pH. The structure now reveals key residues and features that explain its two-step mechanism, and we discuss the implications of the ASLV TM structure in the context of general mechanisms required for membrane fusion.
3.2 Introduction

Invasion of a cell by enveloped viruses requires the attachment and fusion of the host and viral membranes\textsuperscript{35}. This process is mediated by one or more virion-attached glycoproteins (Env)\textsuperscript{35,318}, classified into three groups (class I, II, and III) based on its structural characteristics\textsuperscript{281,319}. Many of the world’s greatest viral threats, such as paramyxo-, orthomyxo-, filo-, corona-, arena- and retroviruses, display glycoproteins belonging to the class I family. Class I Envs are initially translated as a single polypeptide chain with a type I transmembrane anchor\textsuperscript{43,320}. Proteolytic processing of Env forms separate surface attachment (SU) and fusion (TM) subunits\textsuperscript{43,321} that protrude on the viral surface as a fusion-competent, metastable, trimeric peplomer.

Class I viral glycoproteins are triggered through two general mechanisms: (1) pH-independent entry at the plasma membrane or (2) pH-dependent entry at a low pH endosome (reviewed in\textsuperscript{322}). In the pH-independent mechanism, such as those employed by retro-, corona-, or paramyxoviruses, receptor binding by the attachment subunit triggers conformational changes in the Env to mediate the fusion event\textsuperscript{323}. In contrast, for pH-dependent viral entry, such as those employed by flavi-, arena- or influenza A viruses, the acidic environment of the endosome triggers the Env conformational changes\textsuperscript{324,325}. Following triggering, structural constraints on the hydrophobic fusion peptide or loop is released to allow its insertion into the target host membrane to form an extended prehairpin intermediate. Class I viral fusion proteins are thought to adopt a central trimeric $\alpha$-helical structure at the prehairpin intermediate state where its N-terminal end is anchored to the target host lipid bilayer and the C-terminal end embedded in the viral membrane. Multiple fusion proteins may be necessary to anchor and complete the membrane merger process. Subsequently, conformational changes to the fusion subunit allow the prehairpin intermediate to collapse. The C-terminal hydrophobic helical region folds back and packs against the central trimeric $\alpha$-helical structure. As the extended prehairpin intermediate folds up, this brings both the viral and host membranes together to promote mixing of the two outer leaflets and the formation of the hemifusion intermediate. The formation of the final low energy, postfusion six-helix bundle then completes the formation of the fusion pore. The fusion of two lipid bilayers is a thermodynamically feasible process, but comes with a high energetic
price. The free energy released from the conformational changes of the prefusion glycoprotein to the low energy six-helix bundle drives the fusion reaction.

Interestingly, the \( \alpha \)-retrovirus avian sarcoma leukemia virus (ASLV) enters cells through a rare, two-step mechanism that has characteristics of the entry mechanisms of disparate viruses such as HIV-1 and influenza A (IAV)\(^{317,326} \). The unique mechanism of ASLV entry makes it an excellent model system to study the general viral entry processes\(^{317} \). ASLV Env contains an N-terminal SU that is involved in receptor binding and a C-terminal TM subunit that contains the fusion machinery\(^{327-329} \). The receptor for the ASLV subgroup A is Tva, a protein that belongs to the low-density lipoprotein receptor family\(^{330} \). The ASLV Env hydrophobic internal fusion loop is sequestered in the prefusion, metastable state, but upon receptor binding at neutral pH and at a temperature >22°C, conformational changes are triggered to expose and insert the fusion loop into the cell plasma membrane\(^{331-333} \). This triggering of the fusion subunit by receptor binding is pH-independent, and is analogous to the effects of conformational changes on HIV-1 gp160 upon CD4 and chemokine receptor binding. However, previous studies show that in the presence of bafilomycin, an agent that raises the pH of endosomes, viral reverse transcripts were not detected in infected cells, thus suggesting a key role for low pH in entry\(^{317} \). Additional studies on ASLV support the role of low pH being required to induce hemifusion and stabilization of the six-helix bundle\(^{334,335} \). The second step in ASLV entry is reminiscent of the low pH requirements of influenza A virus (IAV) and other pH-dependent viruses.

To explore the rare, two-step entry mechanism of ASLV, we determined the crystal structure of the ASLV TM core. While a number of class I viral fusion proteins have been determined structurally\(^{52,54,74,282,336} \), the ASLV TM is the first crystal structure of a viral fusion subunit that undergoes a hybrid entry mechanism. Interestingly, the ASLV TM is stable to all pH conditions, as opposed to other class I viral fusion proteins. Moreover, the ASLV TM has a strong negatively charged helix dipole at the chain reversal region that is stabilized by two histidine residues at low pH. A structural comparison of class I viral fusion proteins suggests that the presence of a positive charge from a histidine or arginine residue is a signature of fusion proteins that function at low pH. We also show that hydrophobic interactions between HR1 and HR2 and
a single arginine residue (R473) located at the center of the molecule are important for the stability of six-helix bundle. This study identified the residues that are important to the two-step entry mechanism and implications on our general understanding of viral fusion processes are discussed.
3.3 Results and Discussion

3.3.1 Overall structure of ASLV TM. The ASLV TM core is the first structure determined from the α-retrovirus family. Each ASLV TM monomer consists of a long α-helical N-terminal segment (residues 455-489), a disulfide-linked chain reversal (CR) region (residues 490-502), a tether (residues 503-508), and a short C-terminal α-helical segment (residues 509-525) (Figure 3.1A and 3.1B). Three ASLV TM monomers form a 24 Å x 24 Å x 60 Å trimeric protomer that adopts the typical six-helix bundle conformation (Figure 3.1C) seen in the postfusion state of other class I viral fusion proteins (reviewed in\textsuperscript{43,87,281}).

3.3.1.1 N-terminal heptad repeat region 1. The N-terminal segment of each ASLV TM monomer consists of a 10-turn α-helix; the polypeptide chain changes direction after this segment. The N-terminal helix displays the typical heptad repeat motif \((a, b, c, d, e, f, g)\) where \(a\) and \(d\) positions are apolar residues that are located at the coiled coil interface and \(e\) and \(g\) are solvent exposed polar residues\textsuperscript{337,338}. The repeating 3-4-3-4-3-4 periodicity of the heptad repeat allows the hydrophobic leucine, isoleucine, and valine residues to form the “knobs-into-holes” packing with a neighboring chain to form the trimeric core\textsuperscript{337,338}. The hydrophobic knob-into-holes packing is interrupted by a single asparagine layer (N479) that interacts with a chloride ion within the inner core (Figure 3.1C). In addition, the ASLV TM N-terminal heptad repeat (HR1) has a stutter phase shift that corresponds to a four-residue insertion resulting in core packing discontinuities and underwinding of the coiled-coil core (residues 458-461). Similar stutters and chloride binding sites have also been noted in other members of the class I fusion proteins such as those from lymphocytic choriomeningitis virus (LCMV), Ebola virus (EBOV), Marburg virus (MARV), and severe acute respiratory syndrome (SARS) virus\textsuperscript{54,284,336,339}. Another feature of the ASLV TM is the similarity of its HR1 base region (residues 478-504) with a putative immunosuppressive domain found in EBOV and MARV (~50% sequence identity), and β-, δ-, and γ-retroviruses (~36% sequence identity)\textsuperscript{340}. Synthetic peptides of the immunosuppressive motif from EBOV or MARV modulated the expression of interferon-γ, IL-2 and IL-10 cytokines, lower CD4+ and CD8+ T-cell activation and increase immune cell apoptosis\textsuperscript{341}. 
3.3.1.2 **Chain reversal region.** The chain reversal region (residues 490-502) allows the protein to change direction to complete a 180° turn that places the polypeptide chain anti-parallel to the N-terminal HR1 helix. The CR region contains a random coil H-G-H-G loop, a 3_10 helix, and a highly conserved CX_6CC motif (residues 494-502). The CX_6CC motif forms an internal disulfide bond (C494-C501) that stabilizes the loop containing the 3_10 helix. In the δ- and γ-retroviruses, such as HTLV-1 and MoMLV, the SU subunit carries a CXXC thiol disulfide exchange motif that is involved in disulfide bond formation to the third cysteine in the CX_6CC motif^{342,343}. A thiol-disulfide exchange induces isomerization of the SU-TM disulfide bond to initiate fusion. In ASLV Env and filovirus glycoproteins, the SU subunit lacks the CXXC disulfide exchange motif, but the attachment subunit remains covalently linked to the TM. In ASLV, in lieu of the CXXC thio disulfide exchange motif^{344}, it was proposed that the low pH might trigger the structural rearrangements observed during viral fusion^{43,317,344}.

3.3.1.3 **Tether and C-terminal heptad repeat region 2.** Six residues (residues 503-508) in a random coil conformation act as a tether between the CR and the C-terminal heptad repeat region 2 (HR2). The C-terminal HR2 region forms four helical turns that pack between two HR1 helices. The HR1-HR2 interface is stabilized by hydrophobic interactions with residues I512, F516, M519, and V523. Moreover, K515 forms an intersubunit electrostatic ion-pair with D464 in the N-terminal HR1.

3.3.2 **Structural similarity of ASLV TM with other retroviral and filoviral fusion subunits.** Postfusion class I viral fusion subunits all adopt a six-helix bundle conformation. Side-by-side analysis of these structures reveals significant variation in the lengths of the HR1 and HR2 helices, presence of stutter regions, and surface electrostatic potential. Structural analysis previously demonstrated that the fusion machineries of β-, δ-, and γ-retroviruses are similar to EBOV and MARV^{282,336,345}. This led to the hypothesis that the retroviral and filoviral fusion subunits diverged from a common ancestral fusion protein. Not surprisingly, the ASLV TM also shares a strong overall structural similarity with transmembrane subunits of the β-, δ-, and γ-retroviruses and the glycoproteins of EBOV and MARV (Figure 3.1D). A closer inspection, however, reveals differences in the chain reversal region. In all characterized β-, δ-, and γ-
retroviral and filoviral fusion subunits, a $3_{10}$ helix is formed directly after the central HR1 helix which is followed by a critical glycine-glycine (G-G) linker and a disulfide bond stabilized single turn $\alpha$-helix in the CX$_6$CC motif. Mutations to the G-G motif in HTLV-1 Env render the virus defective in fusion and prevent the formation of the six-helix bundle in the maltose-binding protein-gp21 chimera$^{282,295}$. In the ASLV TM, there is no $3_{10}$ helix directly after the HR1 core and the G-G pair is interrupted by His490 and His492 to give a H-G-H-G motif. Moreover, the short $\alpha$-helix observed in other retroviruses is replaced by a $3_{10}$ helix. These features of the ASLV chain reversal region, in particular the H-G-H-G motif, are important for viral fusion$^{346}$, and are conserved in all ASLV subtypes (Figure 3.1E).

3.3.3 ASLV TM is stable over a wide pH spectrum, unlike other retroviral TMs. The two-step entry mechanism of ASLV requires that the envelope glycoprotein function at both neutral and low pH. We performed CD thermal denaturation assays to investigate the stability of the ASLV fusion subunit as a function of pH (pH 5.0 to pH 8.5). At neutral pH, the denaturation of ASLV TM was irreversible with a melting temperature ($T_m$) of 67°C. This is consistent with the melting temperatures of other class I viral fusion proteins$^{290-292}$. Although thermodynamic parameters (i.e. $\Delta G^\circ$ values) cannot be calculated from an irreversible thermal denaturation study, the apparent $T_m$ provides a simple measure of protein stability. The ASLV TM was stable over a wide pH range (Figure 3.2A), with a slightly increased apparent melting temperature ($\Delta T_m=2-5^\circ$C) at pH values corresponding to early (pH 6.0-6.5) and late (pH 5.0-6.0) endosomal stages of the viral life cycle. Considering the nature of the ASLV two-step entry mechanism, our results are in alignment with the biology of the virus. Initially, receptor binding of ASLV SU triggers conformational changes at the cell surface where the ASLV TM structure is rearranged to form the extended prehairpin intermediate at neutral pH. Then, low pH conditions trigger additional conformational changes to form the hemifusion intermediate and postfusion six-helix bundle conformation. Given that ASLV TM must interact at both the plasma and endocytic membranes, the TM must be stable over a wide range of pH.
Figure 3.1. Structure of ASLV TM. (A) Schematic of the ASLV Env. Abbreviations are as follows: SP, signal peptide; SU, attachment subunit; IFL, internal fusion loop; HR1, heptad repeat region 1; CR, chain reversal region; HR2, heptad repeat region 2; TM, transmembrane domain; CT, cytoplasmic tail. The colored regions correspond to the ASLV TM core that was...
crystallized. Red Y shaped symbols denote N-linked glycans. The TM fusion subunit contains three disulfide linkages: one within the hydrophobic internal fusion loop, one in the chain reversal region, and an intermolecular covalent linkage between the SU and TM. (B) Monomer of the ASLV TM. Features in the TM are colored coded to the regions found in the schematic in panel (A). (C) Trimeric ASLV TM postfusion peplomer. The three ASLV TM monomers are shown in red, blue, and green. The inset box displays the ASLV TM down the three-fold axis and shows the chloride ion bound by three aspargine residues within the inner HR1 core. (D) Structural superimposition of ASLV TM and other retroviral and filoviral fusion subunits. (E) Primary sequence alignment of ASLV TM subtypes A, B, C, and D, Rous sarcoma virus (RSV) TM, Ebola virus (EBOV) GP2, and Marburg virus (MARV) GP2. The 3-4 periodicity of the heptad repeats are shown below the alignment. The stutter region, immunosuppressive domain (ISD), and CX6CC motif are highlighted within labeled green boxes.

Figure 3.2. ASLV TM stability as a function of pH. Circular dichroism thermal denaturation profiles of (A) ASLV TM and (B) HTLV-1 gp21 at pH values between 5.0 and 8.5. The CD signal was normalized between 0 (folded) and 1 (unfolded). Plot of $T_m$ vs. pH for (C) ASLV TM
and (D) HTLV-1 gp21. ASLV TM is stable from pH 5.0 to pH 8.5, whereas HTLV-1 gp21 is highly stable at pH >7.0.

The stability of ASLV TM over a wide range of pH values is unique to this viral fusion protein. In contrast, fusion proteins of viruses that traffic to the endosome, such as influenza A, EBOV and MARV, are most stable only at low pH\textsuperscript{290,291,347}. Previous studies of EBOV GP\textsubscript{2} revealed a dramatic decrease in melting temperature (\(\Delta T_m=37^\circ\text{C}\)) upon a pH change from 5.3 to 6.1\textsuperscript{290}. For viruses that fuse at the plasma membrane, such as the \(\beta\)-, \(\delta\)- and \(\gamma\)-retroviruses, we have shown that the fusion subunits are most stable at neutral pH (Figure 3.2B). Thermal denaturation analysis of HTLV-1 TM revealed that at pH values lower than 7.0, the \(T_m\) decreased by over 40\(^\circ\text{C}\) relative to that at neutral pH, suggesting a dramatic destabilization of the fusion protein (Figure 3.2B). At pH values above 7.0, HTLV-1 gp21 was very stable and did not fully denature even at 99\(^\circ\text{C}\). In addition, the thermal melts above pH 7.0 had a biphasic nature, which was not seen in ASLV TM. The first transition (\(T_{m1}\)) likely corresponds to the denaturation of the HR2 helix and the second and larger transition (\(T_{m2}\)) to denaturation of the HR1 trimeric core. At pH values below than 7.0, a cooperative single transition was observed, suggesting that the HR2 region is not as tightly associated to the HR1 core. One explanation for the neutral pH-dependent stability of HTLV-1 gp21 is electrostatic interactions. Previous structural analysis of retroviral fusion subunits revealed that HTLV-1 gp21 contains an extensive series of inter- and intra-chain salt bridges between HR1 and HR2 regions\textsuperscript{282,283}. The relative contribution of electrostatic interactions to protein stability predominates at neutral pH as both side chains are fully ionized (reviewed in\textsuperscript{348}). However, as the virus is taken up into the low pH endosomes, the strength of the ion-pair is thus reduced. Moreover, HTLV-1 gp21 has three histidine residues (H348, H365, His409) positioned in a hydrophobic cavity at the HR1-HR2 interface. At endosomal pH, these histidine residues are protonated and the placement of an unpaired positive charge in a hydrophobic pocket may explain why HTLV-1 gp21 is unstable at low pH. Taken together, our results are in excellent agreement with a neutral pH environment being required to mediate the host cell fusion, as low pH entry pathways are not a viable route for HTLV-1 entry.
3.3.4 Electrostatic interactions are not critical for ASLV TM postfusion stability. The ASLV fusion protein contains three sets of salt bridges (Figure 3.3A): 1) between the HR1 and HR2 helices, 2) between neighboring HR1-HR1 interfaces at the trimeric core, and 3) in the chain reversal region. Specifically, at the HR1-HR2 interface, a single inter-chain electrostatic interaction is present between K515 on the HR2 helix and D464 on the HR1 trimeric core. In the HR1 trimeric core, an inter-chain salt bridge is located between two neighboring HR1-HR1 helices. R473 is located centrally on HR1 and makes an electrostatic interaction with D468 in a neighboring HR1 helix. A third intra-chain salt bridge is situated in the chain reversal region between H492 and D496.

In order to understand the role of the ASLV TM salt bridges on protein stability, the positively and negatively charged residues were mutated to alanine residues, and the effects on protein stability were monitored by CD spectroscopy as a function of temperature. The CD wavelength scan indicated that the alanine mutations did not change the overall spectral characteristics of the protein, suggesting that the secondary structural elements remained intact. The $T_m$ of the wild-type ASLV TM was 72.4°C (Figure 3.3B). Interestingly, alanine mutations to the salt bridge pair that link the HR1 and HR2 helices (D464A and K515A) or the ion-pair in the chain reversal region (H492A and D496A) had little to no effect on the stability of the protein ($\Delta T_m=3.0°C$) (Figure 3.3B). This is in stark contrast to glycoproteins from other retroviruses, such as HTLV-1 gp21, XMRV TM, and MPMV TM that have intricate networks of inter- and intra-chain electrostatic interactions in the fusion subunit. Mutations to residues involved in electrostatic interactions between the HR1 and HR2 regions have profound effects ($\Delta T_m >30°C$) on fusion protein stability and viral entry$^{295}$.

An D468-R473 ion-pair located between HR1 helices within the ASLV TM trimeric core is unique. To our knowledge, no other class I fusion protein has an ion-pair at the HR1 trimer interface. Typically, the interactions between HR1-HR1 helices are largely mediated by hydrophobic forces. An alanine mutation to R473 resulted in a 10°C decrease in the apparent $T_m$ compared to wild-type protein, suggesting R473 plays a strong role in ASLV TM stability.
However, an alanine mutation to D468, the ion-pair partner to R473, stabilized the fusion protein with a 5°C increase in melting temperature (Figure 3.3B). We believe that the increase in the melting temperature is due to the formation of a new salt bridge to D507 located on the outer HR2 helix. This new ion-pair will stabilize the inner core (HR1) to the outer layer (HR2) of the six-helix bundle, thus leading to the observed increased $T_m$. An alanine mutation to D507, as expected, did not affect the $T_m$ indicating that this residue is not involved in stabilizing the fusion protein. We decided to make a double mutation encompassing both D468A and D507A. The removal of all the negative charged partners to R473 should decrease the stability of the fusion subunit. However, the D468A-D507A double mutant revealed no major changes in apparent $T_m$ compared to wild-type (Figure 3.3B). This strongly suggests that the HR1-HR1 ion-pair does not play a role in stabilizing the inner trimeric core. Instead, R473 is likely involved in stabilizing the HR1-HR2 interface through a network of hydrogen bonds made between R473 to S471, S506 and S509 (Figure 3.3A). Overall, our data suggests that electrostatic interactions between HR1-HR1 and HR1-HR2 regions of the ASLV TM fusion subunit are not critical for the stability of the postfusion six-helix bundle structure.

3.3.5 Hydrophobic interactions between the HR1 and HR2 regions stabilize the postfusion conformation. Considering the minimal contribution of ASLV TM electrostatic interactions to the stability of the fusion subunit, we wanted to investigate the role of hydrophobic residues that line the HR1-HR2 interface (Figure 3.4A). We mutated L461, I512, F516 and M519 residues at the HR1-HR2 interface to an alanine. Mutations on three hydrophobic residues within the HR2 helical region (I512, F516 and M519) decreased the stability by 5°C to 7°C compared to wild-type, whereas the L461A mutation on HR1 destabilized ASLV TM substantially by reducing the melting temperature by >10°C (Figure 3.4B). L461 is located on HR1 and makes van der Waals interactions with M510 and F516 from HR2. Previous viral entry assays identified that a single alanine mutation to L461 significantly hinders the infectivity of the virus. The importance of hydrophobic residues in stabilizing the fusion protein is consistent with the nature of the ASLV lifecycle. Hydrophobic interactions are independent of pH, and thus are able to stabilize the postfusion state in a broad range of pH environments.
Figure 3.3. ASLV TM electrostatic interactions. (A) Ribbon diagram of ASLV TM with ion-pair interactions shown as green sticks. Each TM monomer is shown in a different shade of pink. The inset box provides a zoomed view of electrostatic interactions between the HR1-HR2 and HR1-HR1 interfaces and the chain reversal region. Distances between residues are shown in Angstroms, and residues labeled with an asterisk are contributed by a neighboring molecule. (B) Circular dichroism thermal denaturation profiles of wild-type and salt bridge ASLV TM mutants. The CD signals were normalized between 0 (folded) and 1 (unfolded), and thermal melting temperatures ($T_m$) in degrees Celsius are shown next to the graph.
3.3.6 Histidine residues in the chain reversal region stabilize a negative helix dipole at low pH. There are two conserved histidine residues (H490 and H492) at the chain reversal region of the ASLV TM. H490 is positioned immediately after the HR1 region following a short stretch of hydrophobic residues (FLLLA). The side chain of the H490 faces into the center of the hydrophobic inner trimeric TM core (Figure 3.5A). In contrast, H492 faces out into the solvent and forms an intra-chain ion pair with D496. In order to investigate the role of these histidine residues for the postfusion conformation of the protein, we mutated H490 and H492 to alanine residues, and monitored the melting temperatures of the ASLV TM mutants as a function of pH (pH 5.0 to 8.0). Neither H490A nor H492A had significantly different thermal stabilities from

Figure 3.4. ASLV TM hydrophobic interactions. (A) Ribbon diagram of ASLV TM with hydrophobic interactions between the HR1-HR2 regions. Each TM monomer is shown in a different shade of red. The inset box provides a zoomed view of the hydrophobic interactions (blue sticks). (B) Circular dichroism thermal denaturation profiles of wild-type and hydrophobic ASLV TM mutants. The CD signals were all normalized between 0 (folded) and 1 (unfolded), and thermal melting temperatures ($T_m$) in degrees Celsius are shown next to the graph.
that of the wild-type protein at pH 6.5, 7.5, or 8.0 (Figure 3.5B). However, at pH 5.0, mutation of either H490 or H492 to alanine significantly destabilized the ASLV TM by ~20°C compared to the wild-type TM. Mutation of H490 had the larger effect on the stability of the protein, decreasing the $T_m$ below 50°C at pH 5.0, whereas mutation of H492 reduced the melting temperature to 56°C. Delos et al. previously showed that alanine substitutions to H490 or to H492 did not hinder the membrane association of the TM during the pH-independent, receptor-mediated first step of viral fusion, inhibited the entry of infectious ASLV particles for complete fusion. Subsequently, their pH profiling studies on CR histidines demonstrated that although H490A was not significantly different from wild-type, but H492Ala shifted the pH required for activation by 0.6 units. Delos et al. suggested that H492 acts as a trigger to release structural constraints on the fusion subunit upon encountering a low pH environment.

Our results show that both H490 and H492 play a major role for the stability of the ASLV TM postfusion conformation. We hypothesize that H490 and H492 are important in stabilizing a helix dipole moment on the fusion subunit. From prefusion and postfusion crystal structures of the EBOV and IAV envelope glycoproteins, it is clear that, upon triggering, the HR1 region remodels into a single long $\alpha$-helix. As the HR1 $\alpha$-helix extends, the alignment of individual dipoles from the carbonyl groups along the helical axis place increasing net positive and negative charges at the N- and C-terminal poles of the helix, respectively. The negatively and positively charged helix dipole moments are stabilized by oppositely charged end-capping residues. Unbalanced helix dipole moments have been shown to destabilize helical bundles. Calculation of the electrostatic potential on the surface of the ASLV TM reveals negatively and positively charged regions at the poles of the six-helix bundle. At low pH, the ASLV TM histidine residues will have a positive charge and, due to their positions at the base of the central HR1 trimeric helix, are in excellent position to stabilize the negative dipole moment (Figure 3.5A). It is not clear if there are any residues that cap the positively charged dipole moment, as most crystal structures lack the fusion peptide and transmembrane anchor regions of the fusion subunit.
3.3.7 Helix dipole moments are strongest in pH-dependent viral fusion proteins. All class I viral fusion proteins adopt an α-helical postfusion structure to overcome the large energy barrier of the viral and host cell membrane merger. To our knowledge, the role of a helix dipole moment on postfusion protein stability has not been investigated for viral class I fusion proteins. Currently, there are representative models in the Protein Data Bank (PDB) of viral fusion proteins from viruses that undergo either a pH-dependent or pH-independent entry mechanism. ASLV, LCMV and IAV are well-characterized viruses that require low pH for activation, whereas retroviruses are prototypical viruses for pH-independent entry. Assessment of these structures revealed that the helix dipole moments of the viral fusion subunits from three different viral families (ASLV, LCMV and IAV) appear to have a three- to four-fold larger dipole moment compared to other retroviral (HTLV-1, BLV, MoMLV and HERV-FRD) and filoviral (EBOV and MARV) fusion proteins (Table 3.1). From our analysis, a trend exists where viruses undergoing a pH-dependent activation (ASLV, LCMV and IAV) have a stronger helix dipole moment than those that enter through a pH-independent mechanism (MoMLV, HERV-FRD, HTLV-1 and BLV). The filovirus (EBOV and MARV) GP2 structures are an exception to this helix dipole trend. However, Ebola virus is a rather unique pH-dependent virus. Whereas other pH-dependent viruses utilize the low pH environment of the endosome to activate its viral glycoprotein for fusion, the low pH requirement for EBOV entry is linked to the activation of a cellular protease to cleave EBOV GP for fusion. Moreover, the EBOV and MARV GP2 structures are highly similar to the retroviral fusion subunits (Figure 3.1D). Therefore, while EBOV GP2 is classified as a pH-dependent virus, its viral fusion protein may have characteristics, including its helix dipole properties, more similar to the pH-independent viruses. It has been shown that electrostatic interactions between protonated histidine and a negative helix dipole are more effective at low pH. Stabilization of a significant helix dipole moment may provide a contribution to the energetics for host and viral membrane fusion.
Figure 3.5. ASLV TM chain reversal region histidine residues are important for stability at low pH. (A) Ribbon diagram of ASLV TM, HTLV-1 gp21, IAV HA2, and EBOV GP2. The molecules are viewed down the three-fold axis with the chain reversal region facing toward the
viewer. Beside each ribbon diagram, the electrostatic potential mapped onto the molecular surface is shown. Red and blue colored regions denote negative and positive charges, respectively. Positively charged residues that stabilize the negative helix dipole moment are highlighted as green sticks. (B) Circular dichroism thermal denaturation profiles of ASLV TM H490A, H490R, H492A, H492E, and HTLV-1 gp21 K394H at pH 8.0, 7.5, 6.5, and 5.0. The CD signal was normalized between 0 (folded) and 1 (unfolded).

Table 3.1. Calculated dipole moments of various trimeric viral fusion TM proteins.

<table>
<thead>
<tr>
<th>PDB code</th>
<th>Virus</th>
<th>Viral family</th>
<th>No. of atoms&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Residue length of HR1/HR2</th>
<th>Dipole moment&lt;sup&gt;a&lt;/sup&gt; (Debye)</th>
</tr>
</thead>
<tbody>
<tr>
<td>4JPR</td>
<td>ASLV</td>
<td>α–retroviridae</td>
<td>687</td>
<td>30/16</td>
<td>621</td>
</tr>
<tr>
<td>3MK0</td>
<td>LCMV</td>
<td>arenaviridae</td>
<td>875</td>
<td>48/14</td>
<td>850</td>
</tr>
<tr>
<td>1HTM</td>
<td>IAV</td>
<td>orthomyxoviridae</td>
<td>989</td>
<td>65/17</td>
<td>802</td>
</tr>
<tr>
<td>2EBO</td>
<td>EBOV</td>
<td>filoviridae</td>
<td>601</td>
<td>36/14</td>
<td>261</td>
</tr>
<tr>
<td>4G2K</td>
<td>MARV</td>
<td>filoviridae</td>
<td>662</td>
<td>41/15</td>
<td>252</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pH-dependent fusion</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1MG1</td>
<td>HTLV-1</td>
<td>δ–retroviridae</td>
<td>663</td>
<td>46/8</td>
<td>176</td>
</tr>
<tr>
<td>2XZ3</td>
<td>BLV</td>
<td>δ–retroviridae</td>
<td>746</td>
<td>52/8</td>
<td>232</td>
</tr>
<tr>
<td>1Y4M</td>
<td>HERV-FRD</td>
<td>γ–retroviridae</td>
<td>399</td>
<td>31/0&lt;sup&gt;b&lt;/sup&gt;</td>
<td>275</td>
</tr>
<tr>
<td>1MOF</td>
<td>MoMLV</td>
<td>γ–retroviridae</td>
<td>418</td>
<td>33/0&lt;sup&gt;b&lt;/sup&gt;</td>
<td>210</td>
</tr>
</tbody>
</table>

<sup>a</sup>calculated for a single fusion TM monomer  
<sup>b</sup>the HR2 helix is disordered in the crystal structure
3.4 Materials and Methods

3.4.1 Expression and purification of ASLV TM. The DNA sequence of the ASLV TM (residues 455-525) was codon-optimized for expression in *E. coli*, and the gene was synthesized and cloned into pET-46 EK/LIC (EMD Millipore) for expression of an N-terminal 6-His tagged protein with a TEV cleavage site. C506 was mutated to a serine to prevent non-specific intersubunit disulfide bond formation. The expression vector was transformed into the *E. coli* SHuffle T7 expression cell line (New England Biolabs). A 20 ml overnight culture was used to inoculate 1 L of LB, and cells were grown at 37°C to an OD$_{600}$ of 0.6. Expression of the ASLV TM was induced with a final concentration of 0.5 mM isopropyl β-D-1-thiogalactopyranoside (IPTG). The temperature was lowered to 18°C, and cells were harvested 18 hours post-induction by centrifugation at 4000 rpm (Sorvall SLC-4000 rotor) for 20 minutes. Cells were resuspended in 25 mL 1X Ni-NTA binding buffer (50 mM Tris-HCl pH 7.5, 300 mM NaCl, and 20 mM imidazole) supplemented with 0.05% (w/v) CHAPS and 1X EDTA-free Complete protease inhibitor cocktail (Roche). Cells were then lysed at 30 kpsi using a hydraulic cell disruption system (Constant Systems), and centrifuged at 17000 rpm (Sorvall SS-34 rotor) for 45 minutes to remove the cell debris. The supernatant was then applied to a 2 ml Ni-NTA column (Thermo Pierce) by gravity flow at 1 ml/min. Protein was eluted in sequential steps using increasing concentrations of imidazole (125 mM imidazole, 250 mM imidazole, 375 mM imidazole, or 500 mM imidazole in 1X Ni-NTA buffer). Samples containing ASLV were concentrated and further purified by size exclusion chromatography on a Superdex-75 10/300 GL column equilibrated with 10mM Tris-HCl pH 7.5, 150mM NaCl, and 0.05% (w/v) CHAPS. The concentration of purified protein was determined from the absorbance at 280 nm, and purity was confirmed by SDS-PAGE and mass spectrometry.

3.4.2 Crystallization and structure determination of ASLV TM. The protein was concentrated to 10 mg/ml using an Amicon Ultra-0.5 concentrator and initial sparse matrix crystallization screens were performed in sitting drop 96-well plates using an Art Robbins Phoenix liquid handling system. Rod-shaped crystals were obtained by hanging drop vapor diffusion in 20% (w/v) 2-methyl-2,4-pentanediol, 5% (w/v) PEG 8000, and 0.1 M sodium cacodylate, pH 6.5. Crystals were soaked in a mother liquor solution containing 30% (v/v)
glycerol for cryoprotection and then flash cooled in liquid nitrogen. ASLV TM diffraction data were collected on a Rigaku FR-E Superbright X-ray generator and Saturn A200 HD CCD detector at the Structural Genomics Consortium (Toronto, ON). Data were reduced and scaled using d*TREK (Rigaku Corp.)\textsuperscript{306}. The structure of ASLV TM was determined by molecular replacement using the program PHASER\textsuperscript{307} and a truncated, poly-alanine search model of the Moloney murine leukemia virus (MoMLV) fusion glycoprotein (PDB code: 1MOF)\textsuperscript{284}. A clear solution was found with one monomer in the asymmetric unit. Alternating rounds of manual model rebuilding in COOT\textsuperscript{309} followed by simulated annealing torsion angle refinement with TLS refinement using PHENIX.refine\textsuperscript{310} were performed until the $R_{work}/R_{free}$ converged. Data collection and refinement statistics are presented in Table 3.2. Analysis of protein dipole moments were performed using the Protein Dipole Moments Server\textsuperscript{359}. ASLV TM crystal structure and structure factors were deposited into the Protein Data Bank with accession code: 4JPR.

3.4.3 Circular dichroism spectroscopy and thermal melts. Human T-cell leukemia virus (HTLV-1) gp21 was expressed and purified as previously described\textsuperscript{282,295}. ASLV TM and HTLV-1 fusion proteins were buffer exchanged by size exclusion chromatography on a Superdex-75 10/300 column into 10 mM potassium phosphate, 150 mM NaCl, and 0.05% (w/v) CHAPS at pH 6.5 and 7.5, respectively. Circular dichroism (CD) spectra of wild-type and mutant ASLV TM and HTLV-1 gp21 (concentrations ranging from 0.2-2 mg/ml) were acquired on a Jasco J-810 spectropolarimeter using 1 mm quartz cuvettes at 25°C. Data were collected between 190 nm and 250 nm, and averaged over five accumulations. Raw data were then converted to molar ellipticity (degrees cm$^2$ dmol$^{-1}$) for analysis. Thermal denaturation assays were carried out at a single wavelength (222 nm) by monitoring the change in ellipticity, as a function of temperature (20-99°C) at a rate of 5°C/min. Thermal melts as a function of pH were performed under the following buffer conditions: pH 5.0-5.5, sodium acetate (NaOAc) buffer (10 mM NaOAc, 150 mM NaCl, and 0.05% (w/v) CHAPS); pH 6.0-7.5, potassium phosphate buffer (10 mM K$_2$HPO$_4$/KH$_2$PO$_4$, 150 mM NaCl, and 0.05% (w/v) CHAPS); pH 7.5-8.5, Tris-HCl buffer (10 mM Tris-HCl, 150 mM NaCl, and 0.05% (w/v) CHAPS). All thermal denaturation data were normalized between 0 (folded) and 1 (unfolded) and were fit to a non-linear biphasic sigmoidal curve in Graphpad.
Table 3.2. Data collection and refinement statistics.

**Data collection statistics**

<table>
<thead>
<tr>
<th>Parameter</th>
<th>ASLV TM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Space Group</td>
<td>H3</td>
</tr>
<tr>
<td>Unit cell dimensions (Å)</td>
<td></td>
</tr>
<tr>
<td>(a=b)</td>
<td>42.5</td>
</tr>
<tr>
<td>(c)</td>
<td>120.0</td>
</tr>
<tr>
<td>(\alpha=\beta)</td>
<td>90°</td>
</tr>
<tr>
<td>(\gamma)</td>
<td>120°</td>
</tr>
<tr>
<td>Wavelength (Å)</td>
<td>1.54</td>
</tr>
<tr>
<td>Resolution range (Å)</td>
<td>35.2-2.0</td>
</tr>
<tr>
<td>Total reflections</td>
<td>28,380</td>
</tr>
<tr>
<td>Unique reflections</td>
<td>5,371 (444)</td>
</tr>
<tr>
<td>Completeness (%)</td>
<td>97.9 (81.8)</td>
</tr>
<tr>
<td>(R_{\text{merge}}) (%)</td>
<td>2.6 (30.8)</td>
</tr>
<tr>
<td>Redundancy</td>
<td>5.3 (2.5)</td>
</tr>
<tr>
<td>(I/s(I))</td>
<td>28.6 (2.7)</td>
</tr>
</tbody>
</table>

**Refinement statistics**

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. of molecules in asymmetric unit</td>
<td>1</td>
</tr>
<tr>
<td>No. of residues</td>
<td>76</td>
</tr>
<tr>
<td>No. of waters</td>
<td>7</td>
</tr>
<tr>
<td>No. of chloride atoms</td>
<td>1</td>
</tr>
<tr>
<td>(R_{\text{work}}/R_{\text{free}}) (%)</td>
<td>20.7/23.7</td>
</tr>
<tr>
<td>Average B-factor (Å²)</td>
<td></td>
</tr>
<tr>
<td>Overall</td>
<td>38.0</td>
</tr>
<tr>
<td>Protein</td>
<td>38.1</td>
</tr>
<tr>
<td>Solvent</td>
<td>38.0</td>
</tr>
<tr>
<td>RMSD Bonds (Å)</td>
<td>0.014</td>
</tr>
<tr>
<td>RMSD Angles (°)</td>
<td>1.36</td>
</tr>
<tr>
<td>Ramachandran plot statistics (%)</td>
<td></td>
</tr>
<tr>
<td>Most favored region</td>
<td>97.3</td>
</tr>
<tr>
<td>Additional allowed region</td>
<td>2.7</td>
</tr>
<tr>
<td>Disallowed region</td>
<td>0.0</td>
</tr>
</tbody>
</table>
3.5 Acknowledgements

The research was supported in part by University of Toronto start-up funds, Canada Research Chair in Structural Virology, a Canadian Institutes of Health Research (CIHR) Open Operating Grant (MOP-115066), and New Investigator Award (MSH-113554) to JEL. HA was supported by a University of Toronto Graduate Fellowship and BMS was funded by an Ontario Genomics Institute Summer Research Fellowship and Canadian Society for Immunology Summer Internship Award. We thank the Molecular Structure & Function program at the Hospital for Sick Children Research Institute for generous access to a circular dichroism spectropolarimeter, and Aiping Dong, Drs. Aled Edwards and Cheryl Arrowsmith at the Structural Genomics Consortium for access to the X-ray diffraction facility.

3.6 Inter-Chapter Transition

All class I viral fusion proteins are different in size and physiochemical properties but they all proceed through similar structural rearrangements to adopt a trimer of antiparallel α-helices in the postfusion conformation. Therefore, we hypothesize that not only retrovirus fusion proteins, but all class I fusion proteins share common structural features that are used by disparate viruses to stabilize fusion proteins in particular pH environments. Severe acute respiratory syndrome (SARS) is an acute respiratory disease caused by the SARS coronavirus (SARS-CoV) and was first identified in 2002-2003. In 2012, the Middle East respiratory syndrome (MERS)-CoV was discovered in the Arabian Peninsula and remains a major public health threat. Host and viral membrane fusion is a key event in the life cycle of all coronaviruses. SARS-CoV S2 fusion subunit is an important player for drug development, as this protein not only plays a key role in entry, but is highly conserved. SARS-CoV undergoes promiscuous entry at the cell surface and/or low pH at the endosome; however, the residues important for stabilizing the postfusion state remain unclear. In chapter 4 we performed comprehensive biophysical analysis of the SARS-CoV S2 fusion protein to elucidate key features necessary for pH-dependent viral fusion. Similar to ASLV TM, the SARS-CoV S2 fusion protein is stable over a wide pH range and hydrophobic residues at the HR1-HR2 interface may play a major role in protein stability and SARS-CoV S2-mediated membrane fusion. The presence of conserved hydrophobic interactions may be a signature of fusion proteins active at low pH.
Chapter 4

Influence of hydrophobic and electrostatic residues on SARS-coronavirus S2 protein stability: Insights into mechanisms of general viral fusion and inhibitor design

Halil Aydin¹, Dina Al-Khooly¹ and Jeffrey E. Lee¹

¹Department of Laboratory Medicine and Pathobiology, Faculty of Medicine, University of Toronto, Toronto, ON M5S 1A8, Canada

Published Manuscript
Protein Science May 2014 vol. 23 no. 5 603-617
Copyright © 2014 The Protein Society

- Author Contributions:

Halil Aydin: Designed the project, performed cloning, mutagenesis, expression, purification, analytical ultracentrifugation, CD spectroscopy and thermal melt experiments, performed bioinformatics and structural analysis.

Dina Al-Khooly: Assisted with mutagenesis, protein expression and purification.

Jeffrey E. Lee: Supervised the research.
4.1 Summary

Severe acute respiratory syndrome (SARS) is an acute respiratory disease caused by the SARS-coronavirus (SARS-CoV). SARS-CoV entry is facilitated by the spike protein (S), which consists of an N-terminal domain (S1) responsible for cellular attachment and a C-terminal domain (S2) that mediates viral and host cell membrane fusion. The SARS-CoV S2 is a potential drug target, as peptidomimetics against S2 act as potent fusion inhibitors. In this study, we performed site-directed mutagenesis and thermal stability experiments on electrostatic, hydrophobic and polar residues to dissect their roles in stabilizing the S2 postfusion conformation. We show that unlike the pH-independent retroviral fusion proteins, SARS-CoV S2 is stable over a wide pH range, supporting its ability to fuse at both the plasma membrane and endosome. A comprehensive SARS-CoV S2 analysis show that specific hydrophobic positions at the C-terminal end of the HR2, rather than electrostatics are critical for fusion protein stabilization. Disruption of the conserved C-terminal hydrophobic residues destabilizes the fusion core and reduces the melting temperature by 30°C. The importance of the C-terminal hydrophobic residues led us to identify a 42-residue substructure on the central core that is structurally conserved in all existing CoV S2 fusion proteins (RMSD=0.4 Å). This is the first study to identify such a conserved substructure and likely represents a common foundation to facilitate viral fusion. We discuss the role of key residues in the design of fusion inhibitors and the potential of the substructure as a general target for the development of novel therapeutics against coronavirus infections.
4.2 Introduction

Coronaviruses (CoVs) are enveloped, positive-strand RNA viruses responsible for enteric and respiratory diseases in avian and mammalian species\(^{360}\). In 2002, the severe acute respiratory syndrome coronavirus (SARS-CoV) emerged in Southeast Asia and rapidly spread worldwide, resulting in more than 8,000 cases and almost 800 deaths\(^{361–363}\). The unexpected emergence of the highly pathogenic human SARS-CoV revealed the potential for cross-species transmission from circulating strains of CoVs in zoonotic reservoirs\(^{364,365}\). Recently, a novel beta-coronavirus, termed Middle East respiratory syndrome (MERS) CoV, was discovered in the Arabian Peninsula\(^{366}\). Since then, the virus has now migrated to the United Kingdom, France, Italy, and Africa through infected travelers, and is considered a threat to global health with 42.5% case fatality rate among infected individuals\(^{367}\). Genetic sequence analyses show that MERS-CoV belongs to the betacoronavirus genus, along with the BtCoVs HKU4 and HKU5\(^{366}\). Currently, bats host more than 60 CoV species and a number of other SARS-like CoVs were identified from bats in Eurasia, Africa and North America\(^{360,368,369}\). Although much has been discovered in the ten years since the SARS-CoV discovery, emerging zoonotic CoVs continue to cause deadly outbreaks and threaten human health.

CoV infection is initiated by the spike (S) protein on the viral surface\(^{370}\). The SARS-CoV S is synthesized as a 1255-amino acid glycoprotein precursor and is classified as a class I viral fusion protein\(^{371}\). Upon proteolytic activation\(^{371}\), the S protein is cleaved into a S1 domain (residues 12-680) that is responsible for tropism and cellular attachment, and the S2 domain (residues 681-1255) that facilitates virus and host cell membrane fusion\(^{371,372}\). The SARS-CoV S1-S2 heterodimer assembles as a metastable trimer on the viral surface. Similar to other class I viral fusion proteins, such as HIV-1 gp41, Ebola virus glycoprotein (GP2), and influenza A hemagglutinin (HA2), conformational changes in three functional elements of S2: the putative fusion peptide, heptad repeat 1 (HR1) and heptad repeat 2 (HR2) are critical for facilitating the fusion process\(^{371,373}\). Upon activation, the fusion peptide unfolds and inserts into the target cell membrane forming the pre-hairpin intermediate\(^{35,373}\). Subsequently, the HR2 region that anchors the viral membrane folds back near the HR1 trimeric core and triggers the collapse of the pre-hairpin intermediate state\(^{35}\). These conformational changes in the HR1 and HR2 regions draw the
viral and host cell membranes together and mediate the merger of the two outer leaflets into a hemifusion stalk intermediate\textsuperscript{35,373}. A final conformational step results in refolding of both HR1 and HR2 into a low energy postfusion state and allows the fusion pore to form\textsuperscript{35,373}. There is a high kinetic barrier for the fusion of the two bilayer membranes; the free energy released during conformational changes of the fusion protein S2 provides the energetics to overcome the kinetic barriers for fusion pore formation\textsuperscript{35}.

For class I viral fusion proteins, there are three types of fusion triggers: low pH, receptor binding, and proteolytic cleavage\textsuperscript{43}. Some viruses, such as avian sarcoma leucosis virus\textsuperscript{317}, SARS-CoV\textsuperscript{374}, and perhaps Ebola virus\textsuperscript{375,376}, utilize combinations of these triggers. The fusion of SARS-CoV is complex but is thought to require both receptor binding and proteolytic cleavage\textsuperscript{370,371}. The proteolytic cleavage event that separates the receptor binding and fusion domains into non-covalently associated fragments depends on the species of CoV. Some CoVs are proteolytically cleaved at the S1-S2 boundary\textsuperscript{377}, whereas others remain uncleaved, yet are still infectious\textsuperscript{370,378}. For SARS-CoV, a primary proteolytic cleavage at the S1-S2 boundary, followed by a secondary cleavage at a S2' position is often required to mediate membrane fusion\textsuperscript{378}. Protease activation by trypsin-like, thermolysin, elastase, and factor Xa proteases on the plasma membrane\textsuperscript{379–381} and cathepsin L proteolytic cleavage in the low pH endosomes were shown to enhance the SARS-CoV infection\textsuperscript{381–383}. Regardless of the route of entry, viral fusion proteins require structural rearrangements in the S2 domain to mediate the merger of the virus and host cell lipid bilayers\textsuperscript{380,384}.

Recent studies have identified key features that contribute to the function of viral fusion proteins from viruses that enter either at the plasma membrane or low pH endosomes. Although the atomic resolution structures of the SARS-CoV S1 domain and S2 fusion core in the postfusion hairpin conformation have been determined previously, characterization of specific residues involved in the stabilizing the SARS-CoV S2 during membrane fusion remain unclear. Any functional information regarding the molecular identities of these residues is important for the development of novel antiviral therapeutics. In order to understand the structural determinants involved in stabilizing the SARS-CoV S2 fusion protein, we performed site-directed mutagenesis
to investigate the roles of electrostatic, polar, and hydrophobic residues on the SARS-CoV S2 extracellular fusion core and the key features necessary for pH-dependent viral fusion. Our results revealed that the SARS-CoV S2 fusion core is stable over a wide pH range and that specific hydrophobic residues at the HR1-HR2 interface play a major role in stabilizing the six-helix bundle. In contrast, three ion-pairs and chloride-binding site residues were shown to play minor roles in stabilizing the postfusion conformation. Specifically, interhelix interactions between the trimeric coiled-coil HR1 inner core and C-terminal portion of the HR2 helices are important determinants of SARS-CoV fusion, whereas those between the tether and inner HR1 core regions are less important in stabilizing the postfusion state. We also identified a 42-residue conserved substructure within the central heptad repeat region of the SARS-CoV S2 fusion core that we hypothesize will provide a structural foundation for fusion. Our biophysical thermal stability data now explains the inhibition profiles of an array of SARS-CoV S2 peptidomimetics. The results presented here provide insights into the general mechanisms of viral fusion and identify an attractive site for coronavirus fusion inhibitor design.
4.3 Results

4.3.1 Generation and characterization of a linked HR1-HR2 trimeric SARS-CoV S2. The SARS-CoV S2 domain contains an extramembrane helical region that transforms into a coiled-coil six-helix bundle structure in the postfusion state. The N-terminus (residues 890-973) forms a long helical strand, often termed the heptad repeat 1 (HR1) region, with twenty-two helical turns. The HR1 region contains the typical heptad repeat motif of coiled-coil structures. Each repeat consists of a seven-residue \textit{abcdefg} motif, where hydrophobic residues (leucine, isoleucine, phenylalanine, and valine) are displayed in the \textit{a} and \textit{d} positions. The C-terminal segment of each protomer extends alongside the inner core in an antiparallel manner. These residues first form the random coil tether (residues 1142-1160) and then second heptad repeat region (HR2) between residues 1161-1179. The HR2 helices make five full turns and pack into the central HR1 trimer to form a highly stable six-helix bundle conformation that coordinates the fusogenic events between the virus and host cell membranes.

Recombinant expression of the full-length SARS-CoV S2 for structural or biophysical studies is challenging. There are no existing structural models of the entire SARS-CoV S2 protein. The SARS-CoV S2 fusion core, consisting of only the HR1 and HR2 helical regions, has been crystallized, and the structure solved. Commonly, the SARS-CoV S2 fusion core is reconstituted through the addition of synthesized peptides corresponding to HR1 and HR2 regions.\textsuperscript{339,385,386} Here, we designed a linked recombinant SARS-CoV S2 fusion protein (SARS-CoV S2 L2H) using N-terminal residues 896-972 and C-terminal residues 1142-1183, with a six-residue glycine-serine linker (GGS-GGS) between the two regions (Figure 4.1A). This construct is similar to a previously designed construct\textsuperscript{387}. When expressed using an \textit{E. coli} SHuffle T7 expression system, we were able to obtain multi-milligram quantities of protein. The SARS-CoV S2 L2H is soluble and migrates as a stable trimer on size exclusion chromatography. Sedimentation equilibrium analytical ultracentrifugation confirmed the trimeric nature of the SARS-CoV S2 L2H (Figure 4.1B). Furthermore, the circular dichroism (CD) spectrum of the protein was characterized by double minima at 208 and 222 nm, the typical CD signature for a predominantly \textit{\alpha}-helical protein (Figure 4.1C). The CD wavelength scans for all SARS-CoV S2 L2H mutants were similar (data not shown), suggesting that the relative \textit{\alpha}-helical content did not
change as a result of the mutations. Moreover, SARS-CoV S2 L2H contains an estimated 50% α-helical content, in line with the secondary structural composition seen in the X-ray crystal structure of SARS-CoV S2.

4.3.2 SARS S2 fusion core is stable over a wide pH range. Some coronaviruses such as hCoV-229E enter cells via the low pH endosomal environment, whereas others, like mouse hepatitis virus (MHV)-4, directly fuse at the plasma membrane\textsuperscript{388,389}. Interestingly, SARS-CoV can enter cells through either pH-dependent or pH-independent entry pathways depending on the presence of proteases\textsuperscript{379,380}. In order to investigate the pH dependence of the SARS-CoV S2 fusion core structure, we performed CD thermal denaturation assays in buffers ranging from pH 4.0 to pH 8.5. The wild-type (WT) SARS-CoV S2 L2H denatured irreversibly with a melting temperature ($T_m$) of 97.0 °C at neutral pH. This is consistent with the previously reported $T_m$ for SARS-CoV S2 of greater than 90°C\textsuperscript{390}. The Gibbs free energy of unfolding cannot be calculated from an irreversible denaturation curve; however, the apparent melting temperature may provide a simple measure of protein stability. At all pH levels, the SARS-CoV S2 L2H fusion core was highly stable (Figure 4.2A and Table 4.1). As pH was increased, the $T_m$ values remained unchanged (~95°C). At lower pH conditions that correspond to early (6.0-6.5) and late (5.0-6.0) endosomal environments, slightly lower melting temperatures from 92.9-94.6°C were observed. Varying pH did not appear to have drastic effects on protein stability (Figure 4.2B); the postfusion SARS-CoV S2 L2H is stable from pH 4.0 to 8.5.
Figure 4.1. Structural description and biophysical characterization of the SARS-CoV S2 L2H protein. (A) Schematic diagram of the SARS-CoV S protein. The S protein exhibits the characteristic domain organization of class I viral proteins. Abbreviations are as follows: S1, CoV attachment subunit; S2, CoV fusion subunit; SP, signal peptide; RBD, receptor binding domain; RBM, receptor binding motif; FP, fusion peptide; HR1, heptad repeat 1 region; HR2, heptad repeat 2 region; T, tether region; TM, transmembrane domain; CT, cytoplasmic tail; L2H, linked two-heptad construct. The positions of the S1 domain (residues 14-667), S2 domain (residues 668-1255), SP (residues 1-14), RBD (residues 306-527), RBM (residues 424-494), HR1 (residues 890-973), tether and HR2 (residues 1142-1184), TM and CT (residues 1196-1255) are shown above the schematic. Red arrows indicate the S1-S2 and S’ proteolytic cleavage sites at residues R667 and R797, respectively. SARS-CoV S2 L2H construct was generated by using HR1 residues 896 to 972 and tether/HR2 residues 1142 to 1183 connected by a six amino acid linker at the HR1 C-terminal and HR2 N-terminal ends (colored in orange). (B) Sedimentation equilibrium data for a 20 µM sample at 4°C and 22,000 rpm in TBS buffer. The curve indicates the distribution of a 48.4-kDa protein. The data fit closely to a trimeric model for SARS-CoV S2 L2H. The deviation in the data from the linear fit for a trimeric model is plotted.
in the upper panel. (C) Experimental CD wavelength scan of SARS-CoV S2 L2H (blue) at 25°C reveals minima at 208 and 222 nm, indicative of strong α-helical secondary structural characteristics. The SARS-CoV S2 L2H is calculated to contain 50% α-helical content. A reconstructed CD wavelength scan (red) shows the quality of the fit used in the calculation of secondary structural content.

Table 4.1. Summary of SARS-CoV fusion protein stability under various pH conditions.

<table>
<thead>
<tr>
<th>Buffered pH</th>
<th>$T_m$ (°C)*</th>
</tr>
</thead>
<tbody>
<tr>
<td>10 mM NaOAc, pH 4.0</td>
<td>90.6 ± 0.3</td>
</tr>
<tr>
<td>10 mM NaOAc, pH 4.5</td>
<td>92.7 ± 0.3</td>
</tr>
<tr>
<td>10 mM NaOAc, pH 5.0</td>
<td>92.9 ± 0.3</td>
</tr>
<tr>
<td>10 mM NaOAc, pH 5.5</td>
<td>93.4 ± 0.3</td>
</tr>
<tr>
<td>10 mM NaOAc, pH 6.0</td>
<td>93.5 ± 0.3</td>
</tr>
<tr>
<td>10 mM NaOAc, pH 6.5</td>
<td>94.6 ± 0.3</td>
</tr>
<tr>
<td>10 mM Tris-HCl, pH 7.0</td>
<td>95.0 ± 0.5</td>
</tr>
<tr>
<td>10 mM Tris-HCl, pH 7.5</td>
<td>97.0 ± 0.2</td>
</tr>
<tr>
<td>10 mM Tris-HCl, pH 8.0</td>
<td>94.9 ± 0.2</td>
</tr>
<tr>
<td>10 mM Tris-HCl, pH 8.5</td>
<td>95.3 ± 0.2</td>
</tr>
</tbody>
</table>

*The midpoint thermal denaturation ($T_m$) value was estimated from fraction unfolded ($F_{unf}$) and plotted as a function of temperature. Error values indicate 95% confidence intervals from fitting to a non-linear biphasic sigmoidal curve.
Figure 4.2. SARS-CoV S2 L2H stability at various pH values. (A) Thermal denaturation of SARS-CoV S2 L2H monitored by CD molar ellipticity at 222 nm in sodium acetate buffer (between pH 4.0 and 6.5), and TBS buffer (between pH 7.0 and 8.5). The CD signal was baseline corrected, normalized between 0 (folded) and 1 (unfolded), and fit to a non-linear biphasic sigmoidal curve. The $T_m$ values correspond to the temperature where 50% of the protein has unfolded. (B) Plot of SARS-CoV S2 L2H stability as a function of pH. The trimeric SARS-CoV S2 L2H is stable between pH values 4.0 and 8.5.

4.3.3 Electrostatic interactions play a minor role in the stability of the fusion core. Salt bridges are long-range electrostatic interactions typically formed between an anionic carboxylate
(-COO\textsuperscript{-}) functional group of aspartate or glutamate and the cationic ammonium of lysine (-\text{NH}_3^+), the guanidinium group of arginine, or imidazole ring of histidine. Electrostatic interactions can contribute up to 10 kcal/mol in free energy\textsuperscript{348}, and thus are important factors for stabilizing protein structures. The free energy contribution of a salt bridge is pH-dependent, as the ionization of the side chain is affected by the pH of the local environment\textsuperscript{348}. The SARS-CoV S2 fusion core structure revealed three sets of electrostatic interactions clustered in two regions of the protein\textsuperscript{339}. The HR1-HR2 region contains two salt bridge pairs: between R1167 and E918 in the middle of the HR2, and between E1164 and K929 residues at the membrane-distal end of the HR2 (Figure 4.3A). A third complex inter-subunit electrostatic interaction is formed between R965, D967 and E970 at the base of the HR1 helical core. Alanine mutations to abrogate these salt bridges led to a modest decrease in melting temperatures ($\Delta T_m$ < 8°C) (Figure 4.3B and Table 4.2). Interestingly, charge reversal mutations (K929E, R965E and R1167E) destabilized the SARS-CoV S2 L2H six-helix bundle to a similar extent ($\Delta T_m$ < 8°C). Double and triple charge reversal mutations (K929E-R965E, R965E-R1167E, and K929E-R965E-R1167E) resulted in an additive reduction in stability ($\Delta T_m$ ~ 13°C), however the changes were still modest (Figure 4.3C and Table 4.2). In conclusion, our results demonstrated that salt bridges play a small role on the SARS-CoV S2 fusion core stability.

4.3.4 Hydrophobic residues are important for postfusion stability. Given our findings that electrostatic interactions play only a minor role in the stability of the SARS-CoV S2 L2H fusion subunit, we focused on the role of the hydrophobic residues in maintaining structural integrity. The postfusion structure of SARS-CoV S2 reveals a series of hydrophobic residues positioned at the interface of the HR1-tether (L1148 and I1151) and HR1-HR2 (I1161, L1168 and L1175) (Figure 4.4A). We hypothesized that the hydrophobic interactions between the HR1-tether and HR1-HR2 regions play an essential role in stabilizing the outer layer (HR2 and tether) to the inner core (HR1) in the postfusion state. To test our hypothesis, we mutated all five hydrophobic residues to an alanine residue and performed CD thermal denaturation assays. Mutations to hydrophobes at the HR1-tether interface (L1148A and I1151A) destabilized SARS-CoV S2 L2H on the order of $\Delta T_m$ ~ 10°C (Figure 4.4B and Table 4.2). Strikingly, alanine mutations to the hydrophobic residues at the HR1-HR2 interface (I1161A, L1168A and L1175A) had drastic effects on protein stability, with > 20°C decrease each on the apparent $T_m$, as compared to WT
SARS-CoV S2 L2H. Specifically, a single alanine mutation to L1168 or L1175 residue led to a \(\sim 30^\circ C\) decrease on the melting temperature of SARS-CoV S2 L2H (Figure 4.4B and Table 4.2). Our results indicate that hydrophobic interactions between HR1 and HR2, specifically L1168 and L1175, are critical for the stability of postfusion SARS-CoV S2. The hydrophobic HR2 residues involved in postfusion stability are well conserved across the coronavirus family (Figure 4.5).

### 4.3.5 Putative chloride binding site reinforces the structural stability of the postfusion core.

In many postfusion viral fusion proteins, the heptad repeats in the HR1 helix are broken up by a layer of asparagine or glutamine residues to coordinate a putative chloride ion\(^ {284}\). This phenomenon is seen in all CX\(_6\)CC-containing retrovirus and filovirus fusion proteins, and is suggested to be important as a conformational switch between the prefusion and postfusion states\(^ {282,284,336}\). The SARS-CoV S2 extramembrane helical fusion core contains two chloride-binding sites\(^ {339}\) (Figure 4.4C). The first chloride-binding site of HR1 is located at the membrane-proximal end of the protein and is coordinated by Q902, whereas the second site is at the center of the trimeric HR1 coiled coil and is coordinated by N937. In order to investigate the significance of these chloride-binding sites, we mutated both Q902 and N937 to an alanine residue and monitored the changes on the stability of the protein. Q902A and N937A resulted in the decrease of the apparent \(T_m\) values by 8°C and 14°C, respectively (Figure 4.4D and Table 4.2). Thermal unfolding of the Q902A-N937A double mutant revealed an additive effect in the change of melting temperature (\(\Delta T_m = \sim 23^\circ C\)). As a control, we mutated an asparagine residue (N951) located outside of the chloride-binding sites and assessed its contribution on postfusion stability. As expected, N951A did not result in a change on the apparent \(T_m\) value (Figure 4.4D and Table 4.2). Taken together, this suggests that conserved polar interactions with the chloride ion, in particular the central chloride-binding site, are important for the postfusion stability of the protein.
Figure 4.3. Biophysical characterization of SARS-CoV S2 electrostatic interactions. (A) Ribbon diagram of SARS-CoV S2 fusion core (PDB code: 2BEZ) shows electrostatic interactions between the HR1-HR1 and HR1-HR2 regions. The HR1 and tether/HR2 regions are depicted in gray and green, respectively. The side chains of the ion-pair interactions are colored in magenta. The zoomed views of ion-pairs are shown in the inset boxes and the distances between the residues are indicated in Ångstroms. (B) Thermal denaturation profiles of wild-type (WT), single, (C) double and triple mutant of electrostatic residues in the SARS-CoV S2 fusion subunit. Thermal stability was recorded at 222 nm. All data were baseline corrected, normalized between 0 (folded) and 1 (unfolded) and plotted as a function of temperature. The $T_m$ values indicate the midpoint melting temperatures for WT and mutant proteins.
Figure 4.4. SARS-CoV S2 hydrophobic and polar interactions. (A) Ribbon diagram of SARS-CoV S2 fusion core structure (PDB Code: 2BEZ). The HR1 region forms a long helical strand with twenty-two helical turns (colored in gray). The tether and HR2 regions extend alongside the HR1 inner core in an antiparallel manner (colored in green). The hydrophobic residues at the HR1-tether and HR1-HR2 interface are depicted in orange. The inset boxes show the zoomed view of critical hydrophobic residues positioned at the interfaces. (B) Thermal denaturation of wild-type (WT) and mutant hydrophobic residues in the SARS-CoV fusion subunit. (C) Ribbon diagram of an extended SARS-CoV S2 fusion core structure (PDB Code:
1WYY) displaying two putative chloride binding sites. Chloride ions observed in the crystal structure of the HR1 inner core are shown in red. The polar residues interacting with chloride ion (Q902 and N937) and a single polar residue (N951) at the HR1-tether interface are shown as blue sticks. The HR1 and tether/HR2 regions are depicted in gray and green, respectively. (D) Thermal denaturation profiles of wild-type (WT) and chloride binding site mutants. All thermal denaturation profiles are plotted as described in Figure 4.3.

Table 4.2. Summary of wild-type and mutant SARS-CoV fusion protein stabilities.

<table>
<thead>
<tr>
<th>SARS-CoV S2 mutant</th>
<th>$T_m$ (°C)*</th>
<th>Location of mutation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wild type</td>
<td>96.2 ± 0.2</td>
<td>–</td>
</tr>
<tr>
<td>Q902A</td>
<td>89.0 ± 0.2</td>
<td>HR1 top</td>
</tr>
<tr>
<td>E918A</td>
<td>91.8 ± 0.4</td>
<td>HR1 top</td>
</tr>
<tr>
<td>K929A</td>
<td>88.9 ± 0.3</td>
<td>HR1 top</td>
</tr>
<tr>
<td>K929E</td>
<td>88.7 ± 0.2</td>
<td>HR1 top</td>
</tr>
<tr>
<td>N937A</td>
<td>82.7 ± 0.3</td>
<td>HR1 central</td>
</tr>
<tr>
<td>N951A</td>
<td>94.4 ± 0.2</td>
<td>HR1 central</td>
</tr>
<tr>
<td>R965A</td>
<td>92.3 ± 0.4</td>
<td>HR1 bottom</td>
</tr>
<tr>
<td>R965E</td>
<td>93.3 ± 0.8</td>
<td>HR1 bottom</td>
</tr>
<tr>
<td>D967A</td>
<td>91.3 ± 0.3</td>
<td>HR1 bottom</td>
</tr>
<tr>
<td>E970A</td>
<td>92.3 ± 0.1</td>
<td>HR1 bottom</td>
</tr>
<tr>
<td>L1148A</td>
<td>88.2 ± 0.2</td>
<td>Tether region</td>
</tr>
<tr>
<td>I1151A</td>
<td>85.4 ± 0.2</td>
<td>Tether region</td>
</tr>
<tr>
<td>I1161A</td>
<td>79.5 ± 0.3</td>
<td>HR2 bottom</td>
</tr>
<tr>
<td>E1164A</td>
<td>88.5 ± 0.3</td>
<td>HR2 bottom</td>
</tr>
<tr>
<td>R1167A</td>
<td>92.5 ± 0.4</td>
<td>HR2 middle</td>
</tr>
<tr>
<td>R1167E</td>
<td>92.1 ± 0.3</td>
<td>HR2 middle</td>
</tr>
<tr>
<td>L1168A</td>
<td>67.0 ± 0.3</td>
<td>HR2 top</td>
</tr>
<tr>
<td>L1175A</td>
<td>66.0 ± 0.4</td>
<td>HR2 top</td>
</tr>
<tr>
<td>Q902A-N937A</td>
<td>74.3 ± 0.2</td>
<td>HR1 top and central</td>
</tr>
<tr>
<td>K929A-R965A</td>
<td>92.8 ± 0.3</td>
<td>HR1 top and bottom</td>
</tr>
<tr>
<td>K929E-R965E</td>
<td>85.9 ± 0.2</td>
<td>HR1 top and bottom</td>
</tr>
<tr>
<td>K929E-R1167A</td>
<td>91.7 ± 0.3</td>
<td>HR1 top and HR2 middle</td>
</tr>
<tr>
<td>K929E-R1167E</td>
<td>84.0 ± 0.2</td>
<td>HR1 top and HR2 middle</td>
</tr>
<tr>
<td>R965A-R1167A</td>
<td>-89.1**</td>
<td>HR1 bottom and HR2 middle</td>
</tr>
<tr>
<td>R965E-R1167E</td>
<td>88.1 ± 0.2</td>
<td>HR1 bottom and HR2 middle</td>
</tr>
<tr>
<td>K929A-R965A-R1167A</td>
<td>90.5 ± 0.2</td>
<td>HR1 top, HR1 bottom and HR2 middle</td>
</tr>
<tr>
<td>K929E-R965E-R1167E</td>
<td>83.1 ± 0.2</td>
<td>HR1 top, HR1 bottom and HR2 middle</td>
</tr>
</tbody>
</table>

* The midpoint thermal denaturation ($T_m$) value was estimated from fraction unfolded ($F_{unf}$) and plotted as a function of temperature. Error values indicate 95% confidence intervals from fitting to a non-linear biphasic sigmoidal curve.
** The $T_m$ value for this double mutant is an estimate; errors were not calculated.
4.4 Discussion

Coronaviruses are capable of animal-to-human transition, and CoVs that infect pets or animals that frequent urban centers are human health threats due to the potential for mutations that will allow the virus to cross the interspecies barrier. The development of CoV inhibitors will provide a weapon against existing, emerging, or re-emerging CoV outbreaks. The coronavirus S protein plays key roles in facilitating the attachment of the virus to the host receptor, and catalyzing the fusion of the virus and host lipid bilayers. While drugs have been developed against the attachment subunit (S1 equivalent) for other viruses, this may not be a good target for coronaviruses, as they utilize a diverse range of cellular receptors for host attachment. For example, SARS-CoV and hCoV-NL63 use ACE2 as a receptor for infection of target cells, whereas MHV and hCoV-229E utilize carcinoembryonic antigen adhesion molecule 1 (CAECAM1) and aminopeptidase N (APN) as receptors, respectively. Lastly, Raj et al. showed that the recently identified MERS-CoV binds to an exopeptidase, dipeptidyl peptidase 4 (DPP4), as a functional receptor for entry. Studies conducted by Lu et al., Du et al., and Wang et al. revealed that a 286-amino acid fragment within the S1 domain of MERS-CoV interact with the DPP4 receptor. These analyses highlighted notable differences between coronavirus S1 protein-receptor interactions. Furthermore, primary sequence analysis reveals that there is only a <2% sequence identity between all the S1 domains of CoV. Antiviral therapeutics targeting this critical region will likely result in species-specific drugs.

In contrast, the CoV S2 protein is likely an excellent target for the design of more general CoV inhibitors. CoV S2 plays an indispensable role in catalyzing the fusion of the virus and host lipid bilayers and residues involved in fusion are relatively well conserved across all family members (Figure 4.5). Targeting the viral fusion subunit is a proven strategy, as demonstrated by the efficacy of the FDA-approved HIV-1 gp41 HR2 mimic enfuvirtide (T-20). SARS-CoV S2 HR2 peptides are also effective entry inhibitors based on pseudovirus and cell-cell fusion assays. The peptides have traditionally been designed blindly by systematical additions of residues to the core HR2 region. Based on our biophysical fusion protein stability data, we are able to rationalize the trends of effectiveness of peptides. Our thermal denaturation data clearly shows that I1161, L1168, and L1175 are critical to the stability of the postfusion six-
helix bundle structure, whereas hydrophobic residues that belong to the tether region have modest effects on stability. Hydrophobic residues at the C-terminal end of the HR2 region are more critical to the stability of the postfusion state than those at the tether region. The importance of these residues correlates well with the SARS-CoV HR2 peptide inhibition studies\(^3\) (Table 4.3). Peptides that encompass the HR2 hydrophobic residues (I1161, L1168, and L1175) and those corresponding to the HR2 C-terminal ends had better IC\(_{50}\) values than other peptides tested. Peptides that contain residues from the N-terminal tether region were less effective. Our data suggest that effective SARS-CoV HR2 peptide inhibitors should encompass the HR2 region and residues C-terminal to HR2.

**Figure 4.5. Primary sequence alignment of CoV S2 fusion cores.** Multiple sequence alignment of various human and animal CoV fusion proteins. Abbreviations are as follows:
SARS-CoV, severe acute respiratory syndrome-coronavirus; MERS-CoV, middle east respiratory syndrome-coronavirus; hCoV, human coronavirus; HKU, Hong Kong University strain; BCoV, bovine coronavirus; BtCoV, bat coronavirus; CCoV, canine coronavirus; FCoV, feline coronavirus; FIPV, feline infectious peritonitis virus; MHV, mouse hepatitis virus; MuCoV, munia coronavirus; PEDV, porcine epidemic bronchitis virus; PRCoV, porcine respiratory coronavirus; PHEV, porcine hemagglutinating encephalomyelitis virus; TGEV, transmissible gastroenteritis virus; RbCoV, rabbit coronavirus; RtCoV, rat coronavirus; SpCoV, sparrow coronavirus; ThCoV, thrush coronavirus. Sequence boundaries for HR1 and tether/HR2 regions are depicted with gray and green lines, respectively. Residue numbers corresponding to the SARS-CoV S2 fusion subunit numbering are indicated above the alignment. Strictly conserved residues are outlined in red and residues that are important for the stability of the SARS-CoV S2 fusion core are highlighted in yellow and marked with an asterisk (*). Residues (911 to 924) involved in the formation of the common HR1 substructure are shown in a black box. The heptad repeat register \((a, b, c, d, e, f, g)\) of the SARS-CoV S2 fusion core is indicated below the alignment.

Recently, crystal structures of MERS-CoV S2 fusion core have been determined at high atomic resolution\(^{410,411}\) and structural comparison between MERS-CoV and SARS-CoV fusion cores revealed a high degree of overall structural similarity with an RMSD of ~0.9 Å for 204 C\(\alpha\) atoms \(^{410,411}\). We hypothesized that interacting residues on the HR1 may form a conserved interface to accommodate the HR2 hydrophobic residues for fusion. Analysis of available coronavirus S2 structures revealed strong structural conservation of the HR1 region that interacts with the key hydrophobic HR2 residues. Superimposition of a 42-residue HR1 region surrounding the HR2 binding site reveals an average RMSD of 0.4 Å between the SARS-CoV, MERS-CoV, MHV, and hCoV-NL63 S2 fusion subunits (Figure 4.6A). This is in contrast to an overall superimposition of the entire SARS-CoV, hCoV-NL63, and MERS-CoV, MHV S2 six-helix bundle structures or of the inner HR1 trimeric structures which showed C\(\alpha\) atom RMSDs of ~1.3 Å, ~0.6 Å and ~0.6 Å, respectively. The surface of the substructure has a long groove (16-Å long x 9-Å wide x 7-Å deep) and a pocket (7-Å long x 9-Å wide x 7-Å deep) at the HR2 interface. The rim of the pocket is surrounded with polar (glutamine and asparagine) and hydrophobic residues (leucine, isoleucine, serine, and alanine) whereas the bottom of the pocket is lined with
isoleucine residues (Figure 4.6B). HR2 residues that pack into the HR1 pocket are conserved in all fusion core structures with the exception of a conservative isoleucine substitution (I1279) in the hCoV-NL63 fusion core for the leucine in MHV, MERS-CoV and SARS-CoV S2 (Figure 4.5). The conservation of the 42-residue substructure among coronavirus S2 fusion subunits may represent a common foundation to facilitate viral fusion.

Table 4.3. Summary of SARS-CoV S2 HR2 peptide mimics.

<table>
<thead>
<tr>
<th>Peptide</th>
<th>Residue region</th>
<th>Number of residues</th>
<th>IC₅₀ (assay)</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>N-term HR2 extensions</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>sHR2-1</td>
<td>1126-1189</td>
<td>63</td>
<td>43 ± 6.4 µM</td>
<td>Bosch et al., 2004</td>
</tr>
<tr>
<td>sHR2-2</td>
<td>1130-1189</td>
<td>60</td>
<td>24 ± 2.8 µM</td>
<td>Bosch et al., 2004</td>
</tr>
<tr>
<td>sHR2-8</td>
<td>1126-1193</td>
<td>68</td>
<td>17 ± 3.0 µM</td>
<td>Bosch et al., 2004</td>
</tr>
<tr>
<td>sHR2-9</td>
<td>1126-1185</td>
<td>60</td>
<td>34 ± 4.0 µM</td>
<td>Bosch et al., 2004</td>
</tr>
<tr>
<td><strong>C-term HR2 extensions</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HIR2-18</td>
<td>1161-1187</td>
<td>27</td>
<td>3.63 ± 1.5 µM</td>
<td>Yuan et al., 2004</td>
</tr>
<tr>
<td>CP-1</td>
<td>1153-1189</td>
<td>37</td>
<td>19 µM</td>
<td>Liu et al., 2004</td>
</tr>
<tr>
<td>HR2-38</td>
<td>1149-1186</td>
<td>38</td>
<td>66.2 nM</td>
<td>Zhu et al., 2004</td>
</tr>
<tr>
<td>HR2-38*</td>
<td>1149-1186</td>
<td>38</td>
<td>0.5-5 nM</td>
<td>Zhu et al., 2004</td>
</tr>
<tr>
<td>HR2-44</td>
<td>1149-1192</td>
<td>44</td>
<td>500 nM</td>
<td>Zhu et al., 2004</td>
</tr>
<tr>
<td>HR2-38</td>
<td>1149-1186</td>
<td>38</td>
<td>1.02 ± 0.02 µM</td>
<td>Ni et al., 2005</td>
</tr>
<tr>
<td>SR9</td>
<td>1151-1185</td>
<td>35</td>
<td>100 nM</td>
<td>Ujike et al., 2008</td>
</tr>
<tr>
<td>HR2</td>
<td>1151-1185</td>
<td>35</td>
<td>0.34 µM</td>
<td>Chu et al., 2008</td>
</tr>
<tr>
<td>P1</td>
<td>1153-1189</td>
<td>37</td>
<td>3.04 µM</td>
<td>Liu et al., 2009</td>
</tr>
<tr>
<td>P4</td>
<td>1153-1182</td>
<td>30</td>
<td>3.17 µM</td>
<td>Liu et al., 2009</td>
</tr>
<tr>
<td>P6</td>
<td>1153-1175</td>
<td>23</td>
<td>2.28 µM</td>
<td>Liu et al., 2009</td>
</tr>
</tbody>
</table>
Figure 4.6. The common HR1 SARS-CoV S2 substructure. (A) Ribbon diagram of the structural alignments of the HR1 substructure. The SARS-CoV HR1 substructure (colored in green) was superimposed with MHV (colored in purple), MERS-CoV (colored in magenta) and hCoV-NL63 (colored in orange) HR1 substructures (root mean squared deviation=−0.4 Å). Conserved side chains within the HR1 substructure are shown for each virus. SARS-CoV S2 residue numbering was used in both structural alignments. (B) Characterization of the groove on the surface of the HR1 inner core. The electrostatic surface potential of the HR1 region at the HR2 interface was depicted for the SARS-CoV S2, MHV S2, MERS-CoV S2 and hCoV-NL63 S2 fusion cores. HR2 helical region residues extending alongside the HR1 inner core are shown as yellow sticks. Conserved HR2 residues interacting with the HR1 inner core are labeled accordingly. The structurally conserved HR1 core boundaries are indicated between the black dashed lines. Using computational solvent mapping, a hydrophobic pocket was identified, as shown by the green small molecules, on the HR1 substructure.
Our discovery of a structurally conserved 42-residue substructure on the HR1 region in SARS-CoV, hCoV-NL63, MERS-CoV, and MHV S2 provides another target for drug development. The conserved HR1 substructure corresponds to the site of interaction for residues L1168 and L1175 from the HR2 helix (Figure 4.6A); these are the key HR2 hydrophobic residues involved in stabilizing the postfusion conformation. More importantly, a large amphipathic cavity with hydrophobic and polar character is present within the HR1 core substructure (Figure 4.6B) that allows interaction with the critical L1175 hydrophobic residue. This HR1 substructure also coincides with the deep hydrophobic grooves previously identified on the HR1 coiled-coil regions of the SARS-CoV and MERS-CoV S2 fusion core crystal structures\textsuperscript{386,411}. We rationalize that the conserved core maintains the structural integrity of the viral glycoprotein and acts as a foundation for conformational changes necessary for the fusion of the viral and host cell membranes. Inhibitors designed against this site have the potential to block formation of the fusion core complex of many coronaviruses.

In conclusion, SARS-CoV entry requires binding to specific host cell receptors, and fusion of the viral and host cell membranes in a protease-dependent manner. Biophysical and structural analyses of the S2 protein showed that hydrophobic amino acids occupying the ‘d’ position in the C-terminal HR2 region are highly critical for the stability of the six-helix bundle. These residues interact with a conserved substructure within the N-terminal HR1 region and maintain the integrity of the fusion core in the postfusion state. The HR1 substructure is a common feature among many characterized CoV fusion proteins and provides a tangible target for small molecule and peptide fusion inhibitor design. A general strategy targeting this site could be used to combat disease caused by emerging or re-emerging coronaviruses.
4.5 Materials and Methods


4.5.2 Expression and purification of the SARS-CoV S2 linked core. Plasmids containing SARS-CoV S2 L2H WT and mutants were transformed into the *E. coli* SHuffle T7 expression cell line (New England Biolabs). A single colony was inoculated into LB media supplemented with 100 µg/ml ampicillin and grown overnight at 37°C. 20 ml of the overnight culture was then used to inoculate 1 L of LB media supplemented with 100 µg/ml ampicillin, and grown to OD_{600} of 0.6 at 37°C. Protein expression was induced by the addition of isopropyl β-D-1-thiogalactopyranoside (IPTG) to a final concentration of 0.5 mM. The temperature was lowered to 18°C and harvested 18 hours post-induction by centrifugation at 3,000xg. The bacterial cell pellet was resuspended in 25 mL 1X Ni-NTA binding buffer (50 mM Tris-HCl pH 7.5, 300 mM NaCl, and 20 mM imidazole) supplemented with 0.05% (w/v) CHAPS and 1X EDTA-free protease inhibitor cocktail (Bioshop). A hydraulic cell disruption system (Constant Systems) was used to lyse the homogenized cell suspension at 30 kpsi, and the lysate was clarified by centrifugation at 36,500xg for 45 minutes at 4°C. The supernatant was then loaded onto a 2 ml Ni-NTA column (Thermo Pierce), and washed with 10 bed volumes of 1X Ni-NTA binding buffer. The SARS-CoV S2 L2H protein was eluted in sequential steps using increasing concentrations of imidazole (125 mM imidazole, 250 mM imidazole, 375 mM imidazole, or 500 mM imidazole in 1X Ni-NTA buffer). The eluted SARS-CoV S2 L2H protein was then
concentrated and further purified by size exclusion chromatography on a prep grade Superdex-75 10/300 column equilibrated with 10 mM K₂HPO₄/KH₂PO₄ pH 7.5, 150 mM NaCl, and 0.05% (w/v) CHAPS. Protein concentration was quantified by absorbance at 280 nm, and protein purity was analyzed by SDS-PAGE and electrospray mass spectrometry.

4.5.3 Analytical ultracentrifugation (AUC). Sedimentation equilibrium experiments were performed in a Beckman Coulter Optima XL-A analytical ultracentrifuge equipped with an An-60 Ti rotor (Beckman-Coulter, Palo Alto, CA, USA) at the Analytical Ultracentrifugation Facility in the Department of Biochemistry at the University of Toronto. The experiments were carried out at 4°C using purified SARS-CoV S2 L2H fusion core in 10 mM Tris-HCl, 150 mM NaCl pH 7.5. 20 µM protein sample was centrifuged at three different speeds (18,000, 20,000 and 22,000 rpm) and migration of SARS-CoV S2 L2H was monitored by absorbance at 230 and 280 nm over 60 hours. Data analysis was performed using the Origin MicroCal XL-A/CL-I Data Analysis Software Package Version 4.0.

4.5.4 Circular dichroism spectroscopy. Purified SARS-CoV S2 L2H proteins were characterized by circular dichroism (CD) spectroscopy on a Jasco J-810 spectropolarimeter using 1 mm quartz cuvettes (Helma).

4.5.4.1 Wavelength scans. Wavelength spectra were recorded from 190 to 250 nm for wild-type and mutant SARS-CoV S2 L2H in 10 mM K₂HPO₄/KH₂PO₄ pH 7.5, 150 mM NaCl, and 0.05% (w/v) CHAPS buffer at 20°C to determine overall protein secondary structural changes due to the mutation. Five spectra were acquired and averaged, with the results reported as molar ellipticity [θ] (units of deg•cm²•dmol⁻¹). The α-helical content of wild-type SARS-CoV S2 L2H was calculated from the experimental CD wavelength scans using the SELCON3 algorithm in the program DichroWeb⁴¹².
4.5.4.2 **Thermal denaturation scans.** The relative thermal stabilities of SARS-CoV S2 L2H mutants were performed by heating the sample from 20°C to 99°C at 0.2°C intervals and monitoring the loss of CD signal at 222 nm. Heating alone was insufficient to denature the predominantly helical SARS-CoV S2 L2H fusion core. Similar to previous studies\(^{413}\), 4 M guanidine hydrochloride was added to all samples to facilitate unfolding within a temperature range of 20°C-99°C. Ellipticity readings for thermal denaturation data were baseline corrected, normalized between 0 (folded) and 1 (unfolded) and fit to a non-linear biphasic sigmoidal curve in Graphpad. Values of midpoint unfolding transitions \((T_m)\) were calculated from thermal melt curves and they correspond to the temperature where 50% of the protein has unfolded.

4.5.4.3 **pH scans.** In order to study the effects of pH on the stability of the SARS-CoV S2 L2H fusion core, the purified proteins were buffer exchanged using an Amicon Ultra-0.5 centrifugal concentrator (10 kDa molecular weight cut off) into the following buffer conditions: pH 4.0-6.5, sodium acetate (NaOAc) buffer (10 mM NaOAc, 150 mM NaCl, and 0.05% (w/v) CHAPS); pH 7.0-8.5, Tris-HCl buffer (10 mM Tris-HCl, 150 mM NaCl, and 0.05% (w/v) CHAPS). CD wavelength scans and thermal melts as a function of pH were performed as described above. The pH may exhibit fluctuations within the buffering range during the CD thermal melt experiments due to temperature-dependent phenomenon of the buffer.

4.5.5 **Structural comparison and analysis.** Coronavirus fusion protein structures used for structural comparisons were obtained from the Protein Data Bank (1WNC, 1WYY, 1ZV7, 1ZV8, 1ZVA, 1ZVB, 2BEQ and 2BEZ for SARS-CoV S2; 4MOD for MERS-CoV S2; 1WDF and 1WDG for MHV S2; and 2IEQ for hCoV-NL63 S2). Superimpositions of 3D structures and root mean squared deviation (RMSD) calculations were obtained using the pair-wise alignment of the CLICK server\(^{414}\). Identification of pockets on the surface of the SARS-CoV conserved substructure was performed by computational solvent mapping using the FTMap server\(^{415}\).
4.6 Acknowledgements

This work was supported by a Canadian Institutes of Health Research (CIHR) Open Operating Grant (MOP-115066), Canada Research Chair in Structural Virology and a CIHR New Investigator Award (MSH-113554) to JEL. HA was supported by a University of Toronto Graduate Fellowship. The authors would like to thank the Molecular Structure & Function program at the Hospital for Sick Children Research Institute (Toronto, ON) for access to the circular dichroism spectrometer, Drs. Walid Houry and Yoshito Kakihara for assistance with the sedimentation equilibrium ultracentrifugation experiments.

4.7 Inter-Chapter Transition

Cell fusion is a process that is crucial at many crossroads in cell biology. Although the molecular mechanisms of viral-cell fusion are relatively well-established for all four structural classes of viral fusion proteins, molecular details of many cell-cell fusion events remain unclear. After studying the structural and functional features of class I fusion proteins involved in viral-cell fusion, I decided to shift my focus towards the mechanisms of cell-cell fusion in developmental and physiological processes in eukaryotes. An envelope glycoprotein known as syncytin-1, encoded by the ERVWE1 locus of the HERV-W family, has been found to fulfill vital physiological functions in human placenta. Syncytin-1 is involved in the fusion of mononucleate trophoblasts cells to form the syncytiotrophoblast layer at the fetomaternal interface. Syncytin genes encode typical class I fusion proteins that may share similar structural and functional features with other class I viral fusion proteins of exogenous viruses. However, the mechanisms of syncytin-mediated cell-cell fusion have not been unveiled. In chapter 5 we determined the crystal structure of the syncytin-1 TM. The structure reveals key surface-exposed charge-charge interactions that may stabilize the post-fusion structure and facilitate cell-cell fusion. These charge-charge interactions may also contribute to the alignment of heptad repeat regions during the conformational changes required for the merger of two cell membranes. Hence, we asked whether these electrostatic interactions are involved in syncytin-1 Env-mediated cell-cell fusion in the formation of the syncytiotrophoblast layer at the materno-fetal interface of human placenta.
Chapter 5

Structural and functional insights into the fusion mechanism of the human Syncytin-1

Halil Aydin¹, Annoj Thavalingam¹, Geroge Bikopoulosp¹, Azmiri Sultana¹, Shira Elion-Jourard¹ and Jeffrey E. Lee¹

¹Department of Laboratory Medicine and Pathobiology, Faculty of Medicine, University of Toronto, Toronto, ON M5S 1A8, Canada

- Author Contributions:

**Halil Aydin**: Designed the project, performed cloning, mutagenesis, expression, purification, crystallization experiments, determined and refined the crystal structure of syncytin-1 TM, performed CD spectroscopy and thermal melts, immunoblotting, and syncytium assays, performed bioinformatics and structural analysis.

**Annoj Thavalingam**: Assisted with mutagenesis and performed syncytium assays.

**George Bikopoulos**: Assisted with CD spectroscopy and thermal melts and optimized immunoblotting experiments.

**Azmiri Sultana**: Provided guidance for crystallization, data processing and refinement of syncytin-1 TM crystal structure.

**Shira Elion-Jourard**: Assisted with immunoblotting experiments.

**Jeffrey E. Lee**: Supervised the research.
5.1 Summary

*Syncytins* are envelope (env) genes of retroviral origin that have been captured by placental mammals for a physiological function in placentation. In humans, two such genes, *syncytin-1* and *syncytin-2*, display placenta-specific expression and encode fusogenic Env glycoproteins to promote cell-cell fusion of trophoblasts in the formation of the syncytiotrophoblast layer at the materno-fetal interface. Syncytins are essential for placentation, as the generation of *syncytin* knockout mice resulted in impaired cell-cell fusion and death of the null embryos *in utero*. However, little is known about the structural and functional details of syncytin-mediated cell-cell fusion. Here, we determined the crystal structure of the human syncytin-1 transmembrane (TM) fusion subunit at 2.0-Å resolution. The syncytin-1 TM structure reveals the same fold and quaternary organization of post-fusion class I fusion proteins, and unveils multiple surface-exposed single and complex electrostatic interactions within the heptad repeat and chain reversal regions. Site-directed mutagenesis of these inter- and intrachain salt bridges obliterates the fusogenic activity of the syncytin-1 Env and reduces the overall stability of the fusion protein. The importance of the electrostatic interactions is reminiscent of those observed in exogenous β-, γ- and δ-retroviral TM subunits. Our analyses provide insights into the structural determinants required for cell-cell fusion of trophoblast cells by syncytin-1 in placental morphogenesis.
5.2 Introduction

Human Endogenous Retroviruses (HERVs) display the typical genome organization of infectious retroviruses while encoding intact functional proteins\textsuperscript{176}. An envelope glycoprotein known as syncytin-1 is encoded by the \textit{env} gene of the ERVWE1 proviral locus, which belongs to the human endogenous retrovirus W (HERV-W) family\textsuperscript{205}. Retroviral Env glycoproteins belong to the class I viral glycoprotein family\textsuperscript{31,44}. Syncytin-1 Env (1-538) is initially expressed as a single polypeptide, which is then cleaved by cellular proteases into two subunits—a receptor binding surface (SU) subunit (21-313) that is responsible for interacting with the Na\textsuperscript{+}-dependent neutral amino acid transporter-2, and a membrane-anchored fusion-active transmembrane (TM) subunit (318-538) that catalyzes the merger of two cell membranes\textsuperscript{206}. The mature syncytin-1 Env is incorporated on the trophoblast cell surface as a metastable trimer of SU-TM heterodimers, and fulfills vital physiological functions in human placenta\textsuperscript{205,206}. Syncytin-1 is a \textit{bona fide} fusogen that is involved in the fusion of mononucleate trophoblasts cells to form the syncytiotrophoblast layer at the fetomaternal interface\textsuperscript{182}. In addition, syncytin-2 is endowed with an immunosuppressive function by virtue of a specific immunosuppressive domain (ISD) within their TM subunit\textsuperscript{207,208}. Through their ISDs, certain syncytins contribute to maternal immune tolerance in a localized, placenta-specific manner\textsuperscript{207}.

Syncytin-1 mediated trophoblast fusion is triggered by receptor binding at the cell surface under neutral pH\textsuperscript{188,316}. The fusion-competent syncytin-1 undergoes a cascade of irreversible conformational rearrangements to orchestrate the fusion of two cell membranes. Following receptor binding, the hydrophobic fusion peptide near or at the N-terminus of the TM subunit exposes itself from a buried position and inserts into the target cell membrane\textsuperscript{31,42}. The refolding of the fusion peptide gives rise to a transient prehairpin intermediate conformation, in which the C- and N-terminal ends of the trimeric protein are embedded in two separate cell membranes\textsuperscript{31,42}. The extended prehairpin conformation then folds back on itself, forming a trimer of hairpins that brings the C-terminal transmembrane region, anchored in the host cell membrane, and N-terminal fusion peptide, inserted into the target cell membrane\textsuperscript{31,42}. When the two anchored membranes initially interact, the outer leaflets of the two lipid bilayers merge into a hemifusion stalk, promoting the formation of a small, transient fusion pore\textsuperscript{35,50}. A final conformational step
during the refolding of the fusion protein irreversibly induces the enlargement of the fusion pore, propelling the formation of an energetically stable postfusion conformation.\textsuperscript{35,50} The energy released upon refolding of the syncytin proteins into a low energy postfusion hairpin conformation is critical in order to overcome the high kinetic barrier required for the fusion of the two cell membranes.\textsuperscript{35,50}

There are no structural models of any intact, trimeric, prefusion syncytin proteins, or any complete structures of the receptor binding or fusion subunits. A syncytin-2 TM structure was previously reported, but it only depicts a part of the HR1 and CR regions, and is missing the N-terminal portion of the central coiled-coil HR1, as well as the entire C-terminal HR2 region.\textsuperscript{308} Accordingly, the molecular mechanisms underlying the fusion of trophoblast cells into multinuclei syncytiotrophoblasts remain elusive. Any additional structural and functional information regarding these proteins will help reveal the molecular determinants of syncytium formation and deliver key model principles that can be applied towards other syncytins. Here, we have determined the 2.0-Å resolution crystal structure of the human syncytin-1 fusion subunit. Interestingly, structural analysis of the syncytin-1 TM revealed several electrostatic interactions clustered between the HR1 and HR2 regions, and the chain reversal (CR) region. We demonstrate that charge-neutralizing mutations of these residues significantly destabilize the fusion protein and diminish cell-cell fusion. These findings uncover the molecular determinants involved in syncytin-mediated cell-cell fusion during placenta formation.
5.3 Results and Discussion

5.3.1 Crystal structure of syncytin-1 TM. Syncytin-1 TM was expressed in *E. coli* and purified via Ni²⁺ affinity and size-exclusion chromatography. The resulting protein was crystallized in the space group *H*32 with two protomers (chain A and chain B) in the crystallographic asymmetric unit. Diffraction to 2.0-Å resolution was obtained from a single crystal of the syncytin-1 TM, and its structure was determined by molecular replacement. Electron density maps revealed clear density for residues 343 to 434 in chain A, but weak electron density was observed throughout chain B. The CR region was also disordered in chain B resulting in a total of 14 unmodelled CR residues (391-404). Structural analysis of syncytin-1 TM was based on chain A.

Each syncytin-1 TM monomer contains an N-terminal helical HR1 region (residues 343-390) followed by a disulfide-linked CR region (residues 391-405), a random coil tether region (residues 406-410) and a C-terminal helical HR2 region (residues 411-434) (Figure 5.1A and 1B). Syncytin-1 TM monomers assemble into a ~30- X 30- X 70-Å coiled-coil trimer, which together adopt a trimer-of-hairpins post-fusion conformation that is characteristic of class I fusion proteins (Figure 5.1B and 1C). The syncytin-1 TM structure is the first complete fusion subunit structure for syncytins. The previously described syncytin-2 TM structure is missing a part of its N-terminal HR1, as well as the entire C-terminal HR2 region.

The trimer interface is composed of hydrophobic residues from the N-terminal HR1 helices of each monomer, which mediate homotrimeric interactions between monomers and form a buried surface area of approximately 1520-Å² from each molecule at the engaged interfaces. Interestingly, the hydrophobic core of the coiled-coil trimer is interrupted by a single layer of asparagine residues to coordinate a putative chloride ion (Figure 5.1C). Both the N-terminal HR1 and the C-terminal HR2 helices consist of a periodicity of seven amino acids, namely heptad repeats (*a*, *b*, *c*, *d*, *e*, *f*, *g*), where positions *a* and *d* are occupied by hydrophobic residues, while *b*, *c*, and *f* are commonly polar amino acids and *e* and *g* represent typically solvent exposed charged residues. The heptad repeat motif is important for stabilizing α-helix trimerization through “knobs-into-holes” packing of hydrophobic residues at the *a* and *d* positions and
removing hydrophobic residues from the aqueous environment during assembly\textsuperscript{337,338}. At the base of the coiled coil, the polypeptide changes direction at residue T390 to form the CR region. The CR region consists of a short 3_10-helix, followed by a disulfide bond stabilized short $\alpha$-helix, and then completes a 180° turn to continue anti-parallel to the HR1 coiled-coil (Figure 5.1B). The C-terminal portion of the HR1 and the CR region also contain the ISD (residues 380-396) of the syncytin-1 TM. Following the CR region, the polypeptide extends into a short random coil, and then forms a six-turn HR2 $\alpha$-helix that packs into the grooves on the surface of the trimeric HR1 coiled-coil in an anti-parallel manner (Figure 5.1B and 1C). The packing of the C-terminal HR2 $\alpha$-helices against the N-terminal HR1 coiled-coil results in the assembly of a six-helix bundle that brings the N-terminal fusion peptide and C-terminal transmembrane region into close proximity to mediate the fusion of two cell membranes. The syncytin-1 TM structure provides a framework for characterizing the fusogenic activity of syncytins in placenta formation.

In addition to the structural studies of the syncytin-1 TM in crystalline state, we utilized CD spectroscopy to determine the content of secondary structural features of the syncytin-1 TM in solution. A typical CD spectrum for predominantly $\alpha$-helical proteins displays two negative bands at 208 and 222 nm. $\beta$-sheet and random coil proteins produce only a single negative band at 218 nm and 195 nm, respectively\textsuperscript{417}. The CD spectra of syncytin-1 TM demonstrates a characteristic inverted saddle-shaped curve with a calculated $\sim$55% $\alpha$-helical content, thereby confirming the high $\alpha$-helical secondary structural characteristics of syncytin-1 TM in solution (Figure 5.1D).
Figure 5.1. Structure of syncytin-1 TM. (A) Schematic representation of the syncytin-1 Env. Colored regions within the TM subunit correspond to the crystallized region of the syncytin-1 TM presented in this study. Syncytin-1 Env contains a CX6CC motif similar to retro- and filovirus Env glycoproteins. Three cysteine residues within the syncytin-1 TM CR region form one intramolecular disulfide linkage, and an intermolecular bond between the SU and TM. Abbreviations are as follows: SU, surface subunit; TM, transmembrane subunit; SP, signal peptide; RBD, receptor binding domain; FP, fusion peptide, HR1, heptad repeat 1; CR, chain reversal; T, tether; HR2, heptad repeat 2; TM, transmembrane domain; CT, cytoplasmic tail. Red Y-shaped symbols denote N-linked glycans. Note the schematic is not drawn to scale. (B) Ribbon diagram of the syncytin-1 TM monomer. The HR1 helix, disulfide-bonded CR region, random coil tether, and C-terminal HR2 helix are colored in green, purple, blue and red,
respectively. The intrasubunit disulfide linkage is indicated as yellow sticks. (C) Ribbon diagram of the crystallized syncytin-1 TM trimer-of-hairpins structure. Each syncytin-1 TM monomer has an α-helical hairpin conformation and is colored in green, blue or orange. The chloride ion coordinated by three N382 residues (shown as green, blue and orange sticks) is shown as a grey sphere. (D) CD spectra of syncytin-1 TM (blue curve), presented as mean residue ellipticity vs. wavelength, reveals strong α-helical secondary structural characteristics. The secondary structural content is calculated from the CD spectra using the program K2D and is estimated at 55% α-helical content. The calculated reconstructed CD spectra (red curve) demonstrate the agreement of the fit with the measured experimental CD spectra.

5.3.2 Comparison to other viral fusion protein structures. All class I viral and human fusion proteins adopt a trimer of antiparallel α-helices in the postfusion conformation, but these proteins vary in terms of size and physiochemical properties. Although fusion proteins containing the CX6CC motif share a conserved architecture, retroviral fusion proteins share very little sequence similarity with glycoproteins of other viral families. Even between retroviral subfamilies, sequence similarities are quite low. Nonetheless, syncytin-1 TM has a high degree of overall structural similarity with the syncytin-2 TM (RMSD=~1.2 Å) and with other CX6CC-containing viral fusion proteins, such as HTLV-1 gp21 (RMSD=~1.4 Å), Ebola virus (EBOV) GP2 (RMSD=~1.5 Å) and Marburg virus (MARV) GP2 (RMSD=~1.2 Å). We have previously suggested that the CX6CC-containing syncytin-1 TM, syncytin-2 TM and other TM subunits from the retroviral and filoviral families may have diverged from a common ancestral origin. Within the CX6CC motif, the first and second cysteines form an internal disulfide bond, while the third cysteine forms a covalent linkage between the fusion and attachment subunits. The physical conformational constraints resulting from the disulfide linkage within the fusion protein likely provide rigidity to the CR region and contribute to the overall notion of structural similarities between human syncytins and viral glycoproteins. We are interested in investigating the fusion mechanism of the human syncytin-1 in order to determine the functional role of these common structural features. The identification of common features shared between various fusion proteins may provide models for further analysis of the evolutionarily retained Env functions of syncytins leading to cell-cell membrane fusion, and
improve our understanding of the role of syncytins in placental physiology and materno-fetal immune tolerance.

Figure 5.2. Structural comparison and electrostatic potential of syncytin-1 TM. (A) Ribbon diagram of the structural alignments of syncytin-1, syncytin-2 and exogenous retro- and filovirus fusion proteins. The syncytin-1 TM (green) was superimposed with syncytin-2 (pink; PDB: 14YM), MPMV (orange; PDB: 4JF3), EBOV (yellow; PDB: 2EBO), MARV (gray; PDB: 4G2K), HTLV-1 (peach; PDB: 1MG1), BLV (magenta; PDB: 2XZ3), ASLV (blue; PDB: 4JPR), MoMLV (purple; PDB: 1MOF), and XMRV (forest green; PDB: 4JGS) fusion proteins. The root-mean-square deviation (RMSD) values are indicated in Angstroms (Å) for each structural superimposition with syncytin-1 TM. (B) Surface electrostatic potential of the syncytin-1 HR1 trimeric central core and HR2 regions. The electrostatic surface of the syncytin-1 HR1 core is highly negatively charged and the HR2 helix displays a complementary positively charged surface. The footprint of the HR2 region on the HR1 trimeric core is shown by a dashed line. Red and blue regions denote negative and positive charges, respectively.
5.3.3 Syncytin-1 TM structure is lined with electrostatic interactions in the postfusion state.
The syncytin-1 TM contains a total of twelve charged residues forming single or complex electrostatic interactions between the HR1-HR2 and CR regions (Figure 5.3). These charge-charge interactions form an extensive negatively charged surface on HR1, and a complementary positively charged surface on HR2 (Figure 5.2B). Two-third of all residues forming electrostatic interactions at the syncytin-1 TM interfaces were shown to be a part of complex network of salt bridges: one set is located towards the membrane-proximal end of the fusion subunit (E353-R427-R428-R424-D357) and a second set in the central region of the HR1-HR2 interface (K417-D363-R422) (Figure 5.3A). A single inter-chain salt bridge (R360-E420) is also in the vicinity of the central complex salt bridge network (Figure 5.3A). However, the charged side chains of the R360 and E420 residues are ~3.8 Å apart, suggesting a weaker electrostatic contribution. At the CR region, syncytin-1 TM contains a single inter-chain salt bridge (R384-E402) between the base of the HR1 helix and the disulfide-bonded loop of the CX6CC motif in the adjacent monomer (Figure 5.3A). Together, these intra- and inter-chain ion pairs along the length of syncytin-1 TM are hypothesized to strengthen the interactions of outer HR2 α-helices and CR region with the inner HR1 coiled-coil to provide the net free energy necessary for cell-cell fusion.

5.3.4 Syncytin-1 bears resemblance to exogenous β-, γ- and δ-retroviral TMs. Our primary sequence analysis demonstrated that there is <2% sequence identity between the env genes of various mammalian syncytins, thus supporting the independent domestication of these genes by distant eutherian mammals and marsupials. Although the sequence identity among syncytins is higher within the TM subunit (<8%), only a few salt bridge-forming residues within this subunit are conserved between human (syncytin-1 and -2) and mouse (syncytin-A and -B) genes (Figure 5.4A). This also holds true between human syncytins and those of all other major placental mammalian clades (Figure 5.4A). Surprisingly, many of the syncytin-1 intra- and inter-chain electrostatic residues are conserved in the Envs from the exogenous β-, γ- and δ-retrovirus families (Figure 5.4B). For example, nine out of twelve electrostatic residues are conserved within the syncytin-1 and γ-retroviral TMs. These retroviral TM fusion subunit electrostatic pairs
have been characterized to be important in stabilizing its TM subunit for fusion\textsuperscript{416}. This suggests that the syncytin-1 fusion machinery may bear functional resemblance to the exogenous retroviruses.

5.3.5 Syncytin-1 TM salt bridges are required for cell-cell fusion. In order to examine the role of the electrostatic interactions in the cell-cell fusion process, we mutated positively and negatively charged residues to alanine and performed cell-based syncytium assays. Representative images of syncytium formation in HEK293 cells transiently transfected with the full-length syncytin-1 Env wild-type and mutants are shown in Figure 5.5A. Immunoblot analysis confirmed that all mutants were properly expressed and post-translationally processed in the cells in a manner similar to the wild-type syncytin-1 Env (Figure 5.5B).

A proven cell-cell fusion assay was used to analyze the effects of salt bridge mutations\textsuperscript{183}. Alanine substitutions of the syncytin-1 Env complex salt bridge (E353A, D357A, R424A, R427A, or R428A) residues located on the membrane-proximal end of the fusion subunit resulted in varying degrees of reduction in syncytium formation. The E353A and R424A mutants promoted syncytium formation 5 to 7-fold less efficiently than wild-type syncytin-1 Env, whereas the D357A mutant completely eliminated syncytin-1 Env-mediated syncytium formation (Figure 5.3B). On the other hand, R427A and R428A mutations promoted syncytium formation similarly compared to the wild-type syncytin-1 Env (Figure 5.3B). To determine whether the abundance of arginine residues within the HR2 region allow syncytin-1 Env to compensate for the inhibitory effects of single alanine mutations on its fusogenic activity, we generated double (R424A-R427A, R424A-R428A, R427A-R428A) and triple (R424A-R427A-R428A) mutants. Although the R427A-R428A double mutant still promoted syncytium formation similarly to the wild-type syncytin-1 Env, R424A-R427A resulted in an ~85% decrease in cell-cell fusion. Additionally, the R424A-R428A and R424A-R427A-R428A mutants obliterated the fusogenic activity of the syncytin-1 Env (Figure 5.3D). These results suggest that the HR1 E353 and D357 residues along with the HR2 R424 residue play a larger role in syncytin-1-mediated cell-cell fusion, whereas the HR2 R427 and R428 residues have minimal effects on its fusogenic activity. Moreover, alanine substitutions of the complex salt
bridge located at the central region of the HR1-HR2 interface (D363A, K417A, and R422A) severely impaired the fusogenic activity of syncytin-1 (Figure 5.3B). On the other hand, alanine mutations to the R360-E420 salt bridge revealed only a partial loss of fusogenicity (Figure 5.3B). R360 and E420 are 3.8 Å apart thus supporting a weaker electrostatic interaction and fusogenic effect. In the CR region, disruption of the single inter-chain salt bridge (R384-E402) led to significant defects in syncytium formation. More precisely, the E402A mutation decreased syncytium formation by 80%, while R384A completely abolished the cell-cell fusion (Figure 5.3B). Our results demonstrate that multiple electrostatic interactions along the entire length of the fusion subunit are important for its high fusogenic activity.

Table 5.1 Summary of wild-type and mutant syncytin-1 fusion protein stabilities.

<table>
<thead>
<tr>
<th></th>
<th>$T_m$ (°C)*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wild-type</td>
<td>88.3±0.2</td>
</tr>
<tr>
<td>E353A</td>
<td>84.3±0.2</td>
</tr>
<tr>
<td>D357A</td>
<td>78.8±0.2</td>
</tr>
<tr>
<td>R360A</td>
<td>86.2±0.2</td>
</tr>
<tr>
<td>D363A</td>
<td>85.5±0.2</td>
</tr>
<tr>
<td>R384A</td>
<td>83.9±0.2</td>
</tr>
<tr>
<td>E402A</td>
<td>79.6±0.2</td>
</tr>
<tr>
<td>K417A</td>
<td>86.5±0.2</td>
</tr>
<tr>
<td>E420A</td>
<td>82.7±0.2</td>
</tr>
<tr>
<td>R422A</td>
<td>81.1±0.2</td>
</tr>
<tr>
<td>R424A</td>
<td>85.3±0.2</td>
</tr>
<tr>
<td>R427A</td>
<td>84.4±0.2</td>
</tr>
<tr>
<td>R428A</td>
<td>77.5±0.2</td>
</tr>
<tr>
<td>R424A-R427A</td>
<td>83.1±0.1</td>
</tr>
<tr>
<td>R424A-R428A</td>
<td>74.2±0.1</td>
</tr>
<tr>
<td>R427A-R428A</td>
<td>76.7±0.2</td>
</tr>
<tr>
<td>R424A-R427A-R428A</td>
<td>73.0±0.1</td>
</tr>
</tbody>
</table>

* The midpoint thermal denaturation ($T_m$) value was estimated from fraction unfolded ($F_{unf}$) and plotted as a function of temperature. Error values indicate 95% confidence intervals from fitting to a non-linear biphasic sigmoidal curve.
Figure 5.3. Single and complex salt bridges of syncytin-1 TM. (A) Ribbon diagram of syncytin-1 TM. The three syncytin-1 TM protomers are depicted in orange, blue and green and side chains of the ion-pair interactions are shown in the same color of the protomer to which they are associated. The inset boxes show zoomed images of each ion-pair with the distances between the residues indicated above dashed lines in Angstroms (Å). (B) Relative cell fusion activity of wild-type and single alanine mutations of syncytin-1 Env in HEK293 cells. Fusion indexes were determined as indicated in Materials and Methods. All experiments were performed independently three times (n=3) and counted over 6 random field of views. Values are reported as the mean ± standard error of the mean. (C) Thermal denaturation profiles of wild-type and single alanine mutations of electrostatic residues in the syncytin-1 TM. Thermal stability was
recorded at 222 nm between 20°-99°C. All data were baseline corrected, normalized between 0 (folded) and 1 (unfolded) and plotted as a function of temperature. Data points were fitted to a nonlinear biphasic sigmoidal curve in GraphPad. (D) Relative cell-cell fusion activity and thermal denaturation profiles of wild-type, double and triple alanine mutations of syncytin-1 Env.

**Figure 5.4.** Sequence alignment of syncytins and retroviral fusion proteins. (A) Sequence alignment of syncytin fusion proteins from various mammals. Salt bridge-forming residues are poorly conserved among all the syncytins. Abbreviations are as follows: Ory1 - rabbit; Mar1- squirrel; Car1 - carnivora; Rum1- ruminant; Ten1- Afrotherian tenrec; Opo1- opossum. (B) Sequence alignment of the human syncytin-1 and α-, β-, δ-, and γ-retroviral fusion proteins. Secondary structural definitions and residue numbering shown above the alignments are based on syncytin-1 TM. Residues involved in electrostatic interactions in syncytin-1 are labeled with an asterisk, and those that are conserved among syncytins or within the α-, β-, δ-, and γ-retroviral
families are highlighted in yellow. Sequence boundaries for HR1, CR, Tether (T) and HR2 regions are depicted above the sequence alignments in green, purple, blue and red lines, respectively. Abbreviations are as follows: MPMV: Mason-Pfizer monkey virus; SMRV: squirrel monkey retrovirus; SRV-1/2: simian retrovirus; BaEV: baboon endogenous virus; RD114: feline endogenous retrovirus; REV: reticuloendotheliosis virus; SNV: spleen necrosis virus; GALV: gibbon ape leukemia virus; KORV: koala retrovirus; PERV: porcine endogenous retrovirus; XMRV: xenotropic murine leukemia virus-like virus; MoMLV: Moloney murine leukemia virus; KiMSV: Kirsten murine sarcoma virus; FLV: feline leukemia virus; Ga-FeSV: Gardner-Arnstein feline sarcoma virus; FENV: feline endogenous virus; ASLV: avian sarcoma leukosis virus; RSV: Rous sarcoma virus; HTLV-1/2: human T-lymphotropic virus; STLV-1/2/3: simian T-cell leukemia virus; BLV: bovine leukemia virus.
Figure 5.5. Syncytin-1-mediated cell-cell fusion. (A) Representative images of cell-cell fusion mediated by wild-type or mutant syncytin-1 Env. Human HEK293 cells were transfected with an expression vector driving the expression of syncytin-1 Env or with an empty vector (mock) as a control. The fusion activity of the transfected syncytin-1 Env glycoproteins was determined 20 hours post-transfection. All samples were stained with May-Grünwald and Giemsa solutions to visualize the nuclei of the cells at 100X magnification. (B) Expression of wild-type and mutant syncytin-1 Env were detected in transfected cells by immunoblotting. The Env transfected cells were lysed and separated by 12% SDS-PAGE. The blots were probed with antibodies against the N-terminal HA epitope of syncytin-1 Env and GAPDH protein, as indicated. The numbers on the left indicate the molecular mass of the protein standard in kDa.
5.3.6 Electrostatic interactions stabilize the syncytin-1 TM postfusion conformation. CD is commonly used to measure changes in the folding of proteins as a function of temperature. In this regard, CD thermal denaturation is a valuable tool to analyze the stability of mutant proteins in comparison to the wild-type. To test the folding stability of each wild-type and mutant syncytin-1 TM, we performed CD thermal denaturation assays by monitoring the signal at 222 nm as the temperature was slowly raised from 20 to 99°C. The wild-type syncytin-1 TM protein underwent a cooperative unfolding transition upon increasing the temperature, but the unfolding of the protein was irreversible. Therefore, we were not able to obtain the thermodynamic parameters of unfolding and the estimation of thermal midpoint temperatures ($T_m$) was used as a simple measure of protein stability.

In the thermal denaturation assays, the wild-type syncytin-1 TM was highly stable, with its apparent $T_m$ exceeding 85°C. The high melting temperature of syncytin-1 TM is consistent with the observed $T_m$ from other class I viral fusion proteins. All single alanine syncytin-1 TM salt bridge mutations displayed a single sigmoidal transition in their melting curves and exhibited reductions in $T_m$ values by 2-10°C relative to wild-type (Figure 5.3C and Table 5.1). At the membrane-proximal end of the fusion subunit, the mutations D357A and R428A revealed the largest differences ($\Delta T_m=\sim10°C$) in apparent $T_m$ compared to the wild-type (Figure 5.3C). Mutations of surrounding residues (E353A, R424A, R427A) only led to a modest decrease in melting temperatures ($\Delta T_m=\leq5°C$) (Figure 5.3C). Double and triple alanine mutations (R424A-R427A, R424A-R428A, R427A-R428A and R424A-R427A-R428A) resulted in additive reductions in stability. The combination of the R428A mutation with the R424A and R427A significantly destabilized the syncytin-1 TM by $\sim15°C$ relative to the wild-type (Figure 5.3D). Alanine substitutions to the other complex salt bridge in the central region of the HR1-HR2 interface (D363, K417, R422) also decreased the apparent $T_m$ of the syncytin-1 TM to a considerable extent ($\Delta T_m=\leq8°C$) (Figure 5.3C). However, alanine mutations to the long-distance R360-E420 salt bridge residues had only modest effects ($\Delta T_m=2-5°C$) (Figure 5.3C). In the CR region, alanine mutations of the R384-E402 ion pair resulted in the destabilization of syncytin-1 TM. The single E402A mutation reduced the apparent $T_m$ values by 9°C, while R384A had a
smaller effect on the melting temperatures ($\Delta T_m = \leq 5^\circ C$) (Figure 5.3C). Overall, most mutations studied here revealed a modest decrease in stability ($\Delta T_m = \leq 10^\circ C$) suggesting that the combined effects of these charge-charge interactions stabilize the post-fusion state.

5.3.7 Significance of syncytin-1 TM salt bridges on fusogenic activity and protein stability. The syncytin-1 TM is lined with twelve charged residues along the length of the protein forming single or complex electrostatic interactions. For syncytin-1, we observe that at least one key electrostatic interaction at the top (E353-R424-D357), middle (K417-D363-R422) and bottom (R384-E402) of the fusion subunit is important for stability and fusogenic activity. This is not surprising as salt bridges typically contribute 3-5 kcal mol$^{-1}$ of energy for stabilization$^{288,289}$, and the fusion requirement of at least three ion pairs along the post-fusion TM has also been observed in the retroviral TMs$^{295,416}$.

The network of electrostatic interactions observed in syncytin-1 is the most extensive for any class I fusion protein. While retroviral TM fusion subunits also display an intricate network of electrostatic interactions, there are some key differences to syncytin-1. First, the $\beta$- and $\gamma$-retroviral TMs (MPMV and XMRV) salt bridges are commonly single paired electrostatics and alanine mutations to any of these salt bridges abolish viral entry. In syncytin-1, mutations to the salt bridges were less sensitive; only three mutations (D357A, R384A and R422A) fully eliminated cell-cell fusion. Secondly, each salt bridge in the retroviral TMs played a key role in overall stability of the post-fusion protein, with large changes ($\sim 25^\circ C$) in apparent $T_m$$^{416}$. In contrast, the melting temperature of syncytin-1 ion pair mutants were more modest with $<10^\circ C$ change. It is likely given the abundance of salt bridges in syncytin-1, the overall stability is the result of combined effects of these charge-charge interactions. This additive effect of the salt bridges can also be seen in the double and triple arginine syncytin-1 mutations.

In general, two-third of all residues forming electrostatic interactions at the interfaces were shown to be a part of complex salt bridges$^{421}$. Specifically, arginine residues, which can form
interactions from three possible directions, participate in complex salt bridges and are also involved in hydrogen bond interactions. Hence, they provide more complex interactions and serve as key connectors at the interfaces. The syncytin-1 TM HR2 residues R424, R427 and R428 participate in a total of ten salt bridge and five hydrogen bond interactions with the HR1 residues E353 and D357 at the HR1-HR2 interface (Figure 5.3A). The HR2 arginine residues tightly anchor the outer HR2 α-helices to the inner HR1 coiled-coil. The abundance of electrostatic and hydrogen bond interactions attenuates the damaging effects of single arginine-to-alanine substitutions, as the neighboring arginine residues compensate the destabilizing effects of single mutations (Figure 5.3B).

The CR region R384-E402 electrostatic pair is the only ion pair that is conserved throughout syncytin-1, syncytin-2 and the β-, γ- and δ-retroviral fusion proteins. Previously, we and others have shown that mutations to the CR region salt bridge in retroviral TMs abolishes the entry of infectious viral particles. Given the long-range effects of electrostatic interactions, it is hypothesized that the CR region salt bridge is evolutionarily conserved to initiate the conformational changes to form the final post-fusion conformation. The formation of the R384-E402 salt bridge is proposed to align and anchor the chain reversal region to initiate the fold back of HR2 into the HR1 core. Hence, the R384-E402 and its counterpart salt bridges stabilize the CR region and are important for the activity of the fusion protein.

The requirement of electrostatic interactions to stabilize the post-fusion subunit of the β-, γ- and δ-retroviral and syncytin-1 fusion proteins is consistent with their biology. Trophoblast cell-cell fusion and the fusion of the retroviral and host cell plasma membrane are triggered via receptor binding at neutral pH. At neutrality, the acidic and basic amino acids lining the fusion subunits would be completely ionized to form the maximal stabilizing salt bridge interaction. In contrast, severe acute respiratory syndrome-coronavirus (SARS-CoV) and EBOV, which encounter low pH during membrane fusion in the endosomes, utilize hydrophobic residues to stabilize the post-fusion state. Hydrophobic interactions are more resilient to pH changes, whereas salt bridges will be weakened at low pH due to protonation of acidic residues. Hence, class I fusion proteins, such as SARS-CoV S2 and EBOV GP2, which require low pH to catalyze
the fusion of two membranes, may have eliminated the functional role of salt bridges to utilize alternate mechanisms to provide energy for structural transitions and membrane fusion.

5.3.8 Structural insights into the immunosuppressive activity of syncytins and other class I fusion proteins. Most retroviral Env glycoproteins are endowed with a highly conserved immunosuppressive domain (ISD) within the TM subunit that is critical for immunosuppressive (IS) activity and the propagation of the retrovirus. The syncytins contain a 17-amino acid ISD that shares ~30% sequence identity with human endogenous retroviral Env and exogenous β-, γ-, and δ-retroviral Env ISDs (Figure 5.6A). Interestingly, the ISDs in syncytin-2 and syncytin-B retain IS activity, whereas the syncytin-1 ISD is non-IS. Mutational studies combined with functional analyses of allogeneic animal models revealed that the IS activity of syncytins and other Env proteins is primarily dependent on amino acid position 14 of the ISD (syncytin-2 Q427, syncytin-B K495 and MoMLV E551). For instance, when the key residue at position 14 of the ISD is an arginine, as found in syncytin-1 and syncytin-A, the IS activity is switched off. The structure of syncytin-1 TM now provides an enhanced model with which the IS activity of syncytins can be examined (Figure 5.6). Structural comparison of the non-immunosuppressive syncytin-1 and immunosuppressive syncytin-2 ISDs revealed no major conformational changes (RMSD=0.4 Å) within its α-helical and random coil structure (Figure 5.6B). Comparisons with other retro- and filovirus ISDs also showed strong structural conservation (RMSD=0.9 Å) (Figure 5.6B). Our results suggest that the ISDs share a similar overall topology with the immunosuppressive properties defined at position 14 of the ISD. The target of the ISD likely does not interact via charge-charge interactions, as lysine, glutamine or glutamic acid at position 14 all result in immunosuppression. However, the length and size of the amino acid may be critical, as the presence of a bulkier arginine side chain at the same position hinders the IS function in syncytins (Figure 5.6).
Figure 5.6. Immunosuppressive properties of human syncytin-1. (A) Sequence alignment of syncytin, human endogenous retrovirus (HERV) and exogenous retro- and filovirus immunosuppressive domains (ISDs). Secondary structural definitions and residue numbering shown above the alignments are based on the syncytin-1 TM. Position numbers of the ISD are listed below the sequence alignment. The key residue at ISD position 14 is highlighted in yellow. Abbreviations are as follows: EnvH1: Envelope type H1; EnvH2: Envelope type H2; EnvT: Envelope type T; EnvR: Envelope type R; EnvF(c)1: Envelope type F(c)1; EnvF(c)2: Envelope type F(c)2; MPMV: Mason-Pfizer monkey virus; XMRV: xenotropic murine leukemia virus-like virus; MoMLV: Moloney murine leukemia virus; ASLV: avian sarcoma leukosis virus; HTLV-
1: human T-lymphotropic virus; BLV: bovine leukemia virus; ZEBOV: Zaire Ebola virus; MARV: Marburg virus. (B) Ribbon diagram of the structural alignments of syncytin-1 (green), syncytin-2 (purple) and exogenous retro- and filovirus ISDs. The key residue at ISD position 14 is highlighted with yellow boxes. The root-mean-square deviation (RMSD) values are indicated in Angstroms (Å) for each structural superimposition with syncytin-1 ISD. PDB coordinate files used in the superimposition are the same as in Figure 2. * Side chain density is missing for E392 in the crystal structure.

**5.3.9 Conclusions.** In this study, we determined the complete structure of the syncytin-1 TM and showed that the previously uncharacterized N-terminal part of the HR1 and HR2 regions are anchored with multiple charge-charge interactions that may underlie the fusion strategy of syncytin-1. Our functional analyses showed that the abundant electrostatic interactions are essential for its fusogenic activity. The elimination of a single salt bridge in syncytin-1 destabilized the post-fusion conformation and fully blocked syncytium formation. Disruption of these charge-charge interactions presumably imbalances the energetics required for cell-cell fusion. These electrostatic interactions are also conserved in the β-, γ- and δ-retroviral TM subunits, indicating that electrostatic stapling is commonly used by class I fusion proteins to stabilize the post-fusion conformation during cell-cell or viral-cell fusion events. Finally, it is noteworthy that the structural comparison of the IS syncytin-2 and non-IS syncytin-1 showed that the guanidinyl sidechain of an arginine in the ISD may be responsible for switching off the IS activity of syncytins. Overall, unraveling the molecular details of the fusogenic and IS properties of syncytin-1 Env glycoproteins may contribute to the elucidation of the biological significance of captured syncytin genes in placental physiology.
5.4 Materials and Methods

5.4.1 Construct design, protein expression and purification. The gene corresponding to the ectodomain of human syncytin-1 TM (residues 343-435) (GenBank accession number: AF072506.2) was codon-optimized, synthesized (BioBasic), and subsequently subcloned into the pET46 Ek/LIC vector with an N-terminal 6-Histidine (His₆) tag. Site-directed mutagenesis was performed to introduce a C405S mutation to prevent non-specific intermolecular disulfide bond formation and protein aggregation. *E. coli* SHuffle (DE3) cells (New England Biolabs) were transformed with the syncytin-1 TM expression plasmids and 100 ml overnight cultures were grown in LB supplemented with 100 µg ml⁻¹ ampicillin. 1 L of LB with a final concentration of 100 µg ml⁻¹ ampicillin was inoculated with 20 ml of overnight culture and grown to an OD₆₀₀=0.6 at 37°C. The expression of syncytin-1 TM was induced by the addition of a final concentration of 0.5 mM isopropyl β-D-1-thiogalactopyranoside (IPTG). The temperature was then lowered to 18°C and the cultures were incubated for an additional 18 hours post-induction. The cells were harvested by centrifugation at 3000xg for 20 minutes and resuspended in 25 ml 1X Ni-NTA binding buffer (50 mM Tris-HCl pH 7.5, 300 mM NaCl, and 20 mM imidazole) supplemented with 1X Complete EDTA-free protease inhibitor cocktail (Sigma-Aldrich/Roche). Cells were then lysed at 30 kpsi using a hydraulic cell disruption system (Constant Systems). The soluble fraction was separated by centrifugation at 34,500xg for 45 minutes. The supernatant was applied directly to a 2 ml Ni-NTA metal affinity column (ThermoPierce) and syncytin-1 TM was washed with 10-column volumes (CV) of 1X Ni-NTA buffer and eluted with 2 CV of increasing concentrations of imidazole (250, 375 and 500 mM in 1X Ni-NTA buffer). Subsequently, the purified protein was thrombin (EMD Millipore) digested at 22°C for 24 hours (1:2000 (w/w) thrombin to syncytin-1 ratio) in order to remove the N-terminal His₆ tag. The cleaved product was re-applied to a 2 ml Ni-NTA metal affinity column and the flow-through was collected. Final purification by size exclusion chromatography on a prep-grade Superdex-75 10/300 column equilibrated in 10 mM Tris-HCl pH 7.5 and 150 mM NaCl was performed. The protein concentration was determined by UV absorbance at 280 nm.

5.4.2 Crystallization and structure determination. Purified syncytin-1 TM was concentrated to 15 mg ml⁻¹ and crystallization trials were performed at 22°C by sitting-drop vapor diffusion
(500 nl protein and 500 nl reservoir solution) in 96-well low profile Intelliplates (Art Robbins) using the Oryx8 protein crystallization robot (Douglas Instruments). Initial sparse matrix screening of syncytin-1 TM yielded clusters of needle-like crystals in multiple conditions. Diffraction-quality crystals of syncytin-1 TM were obtained by hanging-drop vapor diffusion in 5% (w/v) PEG 3350 and 0.1 M sodium acetate-HCl pH 4.5. These crystals were further optimized by using cross-matrix microseeding with the Oryx8\textsuperscript{424}. Compact plate-shaped crystals were obtained in hanging drops containing 0.9 µl purified protein, 0.9 µl crystallant and 0.2 µl diluted seed stock. In order to slow down the equilibration rate between the reservoir and the crystallization drop, 20% (v/v) paraffin oil layer was introduced over the reservoir solution in a 24-well VDX plate. Crystals were cryo-protected by briefly soaking in the reservoir solution supplemented with increasing amounts of glycerol (up to 30% (v/v) in final solution), before flash-freezing in liquid nitrogen. Syncytin-1 TM diffraction data were collected on beamline 08ID-1 at the Canadian Light Source (CLS; Saskatoon, SK). Data were integrated and reduced using XDS\textsuperscript{425} and scaled using Aimless\textsuperscript{426}. The syncytin-1 TM structure was determined by molecular replacement using the program Phaser\textsuperscript{307}, and the Mason-Pfizer monkey virus (MPMV) fusion glycoprotein (PDB accession number 4JF3) was used as the initial search model\textsuperscript{416}. A clear solution (TFZ=15.5) was identified with two monomers in the asymmetric unit. Autobuild\textsuperscript{427} in the Phenix package\textsuperscript{428} was used to build the initial model. The structure was completed with multiple rounds of manual model rebuilding in Coot\textsuperscript{309}, followed by refinement using phenix.refine\textsuperscript{310} with torsion-angle simulated annealing refinement, individual atomic displacement and Translation/Liberation/Screw (TLS) refinement until the $R_{\text{work}}/R_{\text{free}}$ converged. The stereochemical quality of the refined model was assessed with MolProbity\textsuperscript{429} and Coot\textsuperscript{309}. Data collection and refinement statistics are presented in Table 5.2.

5.4.3 Structural comparison and analysis. PDB files used in the structural superimpositions were: (i) retroviruses: syncytin-2 TM (PDB: 1Y4M), bovine leukemia virus (BLV) TM (PDB: 2XZ3), human T-lymphotropic virus type-1 (HTLV-1) gp21 (PDB: 1MG1), MPMV TM (PDB: 4JF3), xenotropic murine leukemia virus (XMRV) TM (PDB: 4JGS), moloney murine leukemia virus (MoMLV) TM (PDB: 1MOF), avian sarcoma leukosis virus (ASLV) TM (PDB: 5H9C) (ii) filoviruses: EBOV GP\textsubscript{2} (PDB: 2EBO) and MARV GP\textsubscript{2} (PDB: 4G2K). Structural superimpositions and RMSD calculations were performed using the CLICK server\textsuperscript{414} and all structural figures generated using PyMOL (v. 1.7.4 Schrödinger, LLC.).
Table 5.2. Data collection and refinement statistics.

<table>
<thead>
<tr>
<th></th>
<th>Syncytin-1 fusion protein</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Data collection</strong></td>
<td></td>
</tr>
<tr>
<td>Wavelength (Å)</td>
<td>0.84592</td>
</tr>
<tr>
<td>Space group</td>
<td>H32</td>
</tr>
<tr>
<td>Cell dimensions</td>
<td></td>
</tr>
<tr>
<td>$a, b, c$ (Å)</td>
<td>48.1, 48.1, 378.9</td>
</tr>
<tr>
<td>$a, b, g$ (°)</td>
<td>90, 90, 120</td>
</tr>
<tr>
<td>Resolution (Å)</td>
<td>42.1 - 2.0 (2.05 - 2.0) *</td>
</tr>
<tr>
<td>Total number of reflections</td>
<td>126807</td>
</tr>
<tr>
<td>Number of unique reflections</td>
<td>12052</td>
</tr>
<tr>
<td>$R_{\text{meas}}$</td>
<td>8.2 (79.5)</td>
</tr>
<tr>
<td>&lt;$I/\sigma I&gt;$</td>
<td>20.9 (3.0)</td>
</tr>
<tr>
<td>CC$_{1/2}$ (%)</td>
<td>99.9 (78.0)</td>
</tr>
<tr>
<td>Completeness (%)</td>
<td>99.7 (96.1)</td>
</tr>
<tr>
<td>Redundancy</td>
<td>10.5 (8.8)</td>
</tr>
<tr>
<td><strong>Refinement</strong></td>
<td></td>
</tr>
<tr>
<td>Resolution (Å)</td>
<td>42.1 - 2.0 (2.07 - 2.0)</td>
</tr>
<tr>
<td>$R_{\text{work}}/R_{\text{free}}$</td>
<td>19.9/23.3</td>
</tr>
<tr>
<td>No. atoms</td>
<td></td>
</tr>
<tr>
<td>Protein</td>
<td>1325</td>
</tr>
<tr>
<td>Ligand/ion</td>
<td>2</td>
</tr>
<tr>
<td>Water</td>
<td>73</td>
</tr>
<tr>
<td>B-factors</td>
<td></td>
</tr>
<tr>
<td>Protein</td>
<td>40.6</td>
</tr>
<tr>
<td>Ligand/ion</td>
<td>38.0</td>
</tr>
<tr>
<td>Water</td>
<td>39.6</td>
</tr>
<tr>
<td>R.m.s deviations</td>
<td></td>
</tr>
<tr>
<td>Bond lengths (Å)</td>
<td>0.013</td>
</tr>
<tr>
<td>Bond angles (°)</td>
<td>1.2</td>
</tr>
<tr>
<td>Ramachandran plot statistics (%)</td>
<td></td>
</tr>
<tr>
<td>Most favored region</td>
<td>97.0</td>
</tr>
<tr>
<td>Additional allowed region</td>
<td>3.0</td>
</tr>
<tr>
<td>Disallowed region</td>
<td>0</td>
</tr>
</tbody>
</table>

*Highest resolution shell is shown in parenthesis.
5.4.4 Cell culture

5.4.4.1 Cells lines and plasmids. Human embryonic kidney (HEK)293 cells were cultured in 1X Dulbecco’s Modified Eagle Medium (DMEM) supplemented with 10% (v/v) heat inactivated fetal bovine serum (FBS), and 1X antibiotic-antimycotic, and propagated at 37°C in a 5% CO₂ incubator. The gene corresponding to the intact Env glycoprotein of human syncytin-1 (residues 1-538) (GenBank accession number: AF072506.2) was codon-optimized for expression in Homo sapiens, and synthesized (BioBasic, Toronto ON). The DNA encoding the gene for syncytin-1 Env with an N-terminal hemagglutinin (HA) epitope tag and Igκ signal peptide was subcloned into a pcDNA 3.4 expression vector. All cell culture media and serum were purchased from ThermoFisher Scientific/Gibco.

5.4.4.2 Cell-cell fusion assays. Briefly, 22 mm x 22 mm coverslips (VWR) were placed inside each well in 6-well tissue culture-treated plates (Corning), and HEK293 cells were seeded at a density of 5 x 10⁵ cells per well. Following an overnight incubation at 37°C with 5% CO₂, the cells in each well were transfected with 0.3 µg of syncytin-1 Env expression vector using GeneJuice transfection reagent (EMD Millipore). The transfection mixture was diluted in serum-free DMEM at a plasmid DNA:GeneJuice ratio of 1:3, as per the manufacturer instructions. Six hours post-transfection, the serum-free medium was removed from the wells and replaced with 2 ml of serum-supplemented DMEM. Following incubation for an additional 14 h, the cells were washed briefly with 1 ml of 1X PBS, and then fixed with 2 ml of 100% methanol (Caledon) for 15 minutes at -20°C. The cells were then incubated with pre-chilled May-Grünwald solution (Sigma-Aldrich) for 5 minutes at room temperature (RT), and subsequently washed with 1 ml 1X PBS. A pre-chilled working solution of Giemsa was prepared by diluting the stock with deionized water by a ratio of 1:20. Following the 1X PBS wash, the cells were incubated with 2 ml of working Giemsa solution for 20 minutes at RT. After the Giemsa solution was aspirated from the wells, the coverslip in each well was removed with forceps and air-dried for 10 minutes. Once the coverslips were dried, two droplets of Cytoseal 280 (Richard-Allan Scientific) were each dropped onto pre-labeled microscope slides (VWR), and the coverslips were immediately mounted onto a microscope slide. The cells were imaged with bright-field microscopy using a Nikon Eclipse E600 microscope, DS-Fi1 camera and NIS-Elements software at 100X total magnification. In order to randomize the fields of view chosen
for imaging, 5 x 5 grids were drawn on the coverslips, and a random number generator ([https://www.random.org/](https://www.random.org/)) was used to select the fields of view. A total of 6 images were taken per coverslip in technical replicates, and the experiment was carried out in biological triplicates per mutant (n=3). A fusion propensity index was then calculated for each image by counting the total number of cells in syncytia (N), the number of syncytia (S) and the total number of cells in the image (T). These values were used in the following equation: Fusion Index = (N-S)/T x 100%.

5.4.4.3 Detection of GP expression. Cellular expression of GPs was determined 48 hours posttransfection by anti-HA immunoblot. Transfected HEK293 cells were washed with phosphate-buffered saline (PBS) and treated with NP40 lysis buffer (50 mM Tris-HCl pH 8.0, 120 mM NaCl and 0.5% (v/v) NP40) supplemented with 2X complete protease inhibitor cocktail. Cell lysates were clarified by centrifugation at 21,100xg for 10 minutes at 4°C and normalized with Bradford assays prior to SDS-PAGE analysis in 12% polyacrylamide gels under reducing conditions.

5.4.5 Circular dichroism spectroscopy and thermal melts. Circular dichroism (CD) spectral scans and thermal melt measurements on syncytin-1 TM were performed on a Jasco J-810 spectropolarimeter using 1 mm quartz cuvettes (Helma). Protein concentrations ranged from 0.4 to 1.0 mg ml⁻¹ in a 10 mM potassium phosphate pH 7.5, 150 mM NaCl buffer. CD wavelength scans were collected between 190-250 nm at 25°C and averaged over five measurements. Thermal denaturation was monitored at a single wavelength (222 nm) by applying a temperature gradient of 5°C min⁻¹ from 20°C to 99°C. Increasing the temperature of syncytin-1 TM to 99°C in a 10 mM potassium phosphate pH 7.5 and 150 mM NaCl buffer was insufficient to unfold the protein. Similar to previous studies, the thermal melts were performed in the presence of 1 M guanidine hydrochloride (Gu-HCl) in order to facilitate the complete unfolding of all samples. The spectra were baseline corrected, normalized between 0 (folded) and 1 (unfolded), and plotted as a function of temperature. The melting curves were
fitted to a biphasic sigmoidal curve and the midpoint thermal unfolding transitions ($T_m$) were calculated using Prism 6 GraphPad software.

**5.4.6 Accession numbers.** Atomic coordinates and structure factors for syncytin-1 TM have been deposited in the Protein Data Bank (PDB) under accession number 5HA6.
5.5 Acknowledgements

This work was supported by a Canadian Institutes of Health Research (CIHR) Operating Grant (MOP-115066), Ontario Early Researcher Award (ER-13-09-116), and a Canada Research Chair (Tier II) to J.E.L. Moreover, support for stipends were provided by University of Toronto and Ontario Graduate Scholarships to H.A., and a Natural Science and Engineering Council of Canada (NSERC) Undergraduate Summer Research Award to A.T.. The authors would like to thank Dr. Walid Houry in the Department of Biochemistry (University of Toronto) for access to their circular dichroism spectrometer and Dr. Michelle Bendeck for access to their microscopy facility. This work is based upon X-ray data collected at beamline 08ID-1 at the Canadian Light Source (CLS). The authors would like to thank Marsel Lino for technical support, and Farshad C. Azimi for technical support and X-ray data collection.

5.6 Inter-Chapter Transition

In sexually reproducing species, the fusion of sperm and egg plasma membranes results in the creation of a new genetically distinct diploid organism termed zygote. In 2005, Inoue et al. identified a protein on the surface of the sperm that is essential for the recognition and fusion of sperm and egg membranes168. This 350-residue membrane-bound glycoprotein was named Izumo1. Only in 2014, the receptor of Izumo1, termed Juno, was discovered on the surface of the mature egg173. Izumo1 and Juno are the only known proteins truly essential for sperm and egg fusion. Mice models lacking Juno or Izumo1 are healthy but infertile, thus suggesting that a specific Izumo1-Juno interaction is universally required for fertilization. Yet, atomic resolution structures of both human Izumo1 and Juno are unknown. In chapter 6 we provide the first atomic resolution structures into the molecular recognition of Izumo1 and Juno. These are the first structures of any protein complex between sperm and egg at the point of conception for any organism. Through a hybrid approach combining structural and biophysical analyses, we show that Izumo1 undergoes a major conformational change and is stabilized into a “locked” upright position upon Juno binding. Moreover, mutational studies at the Izumo1-Juno interface revealed the structural determinants of required for binding. Our results suggest that human Izumo1-Juno axis may provide an additional barrier to cross-species fertilization and polyspermy, and may act as the foundation of the sperm-egg recognition and fusion machinery.
Chapter 6

Molecular architecture of the human sperm Izumo1 and egg Juno fertilization complex

Halil Aydin1, Azmiri Sultana1, Sheng Li2, Annoj Thavalingam1 and Jeffrey E. Lee1

1Department of Laboratory Medicine and Pathobiology, Faculty of Medicine, University of Toronto, Toronto, ON M5S 1A8, Canada

2Department of Medicine, University of California, San Diego, La Jolla, CA 92093, USA

Published Manuscript
Nature 2016 534, 562-565
Copyright © 2016, Rights managed by Nature Publishing Group

- Author Contributions:

Halil Aydin: Designed the project, performed all the cloning, mutagenesis, expression, purification, biophysical characterization and crystallization experiments, collected synchrotron X-ray diffraction data, determined the crystal structures and SAXS reconstructions, prepared the DXMS samples, performed bioinformatics and structural analysis.

Azmiri Sultana: Assisted with BLI and SPR experiments and provided crystallographic guidance for refinement and validation of the crystal structures.

Sheng Li: Performed and analyzed the DXMS data.

Annoj Thavalingam: Assisted with protein expression and purification.

Jeffrey E. Lee: Supervised the research, assisted with BLI experiments, performed bioinformatics and structural analysis.
6.1 Summary

Fertilization is an essential biological process in sexual reproduction and comprises a series of molecular interactions between the sperm and egg\textsuperscript{142,143}. The fusion of haploid spermatozoon and oocyte is the culminating event in mammalian fertilization, enabling the creation of a new genetically distinct diploid organism\textsuperscript{145,146}. The merger of two gametes is achieved through a two-step mechanism where the sperm antigen Izumo1 on the equatorial segment of the acrosome-reacted sperm recognizes its receptor Juno, on the egg surface\textsuperscript{145,168,173}. This is followed by the fusion of two plasma membranes. Izumo1 and Juno proteins are indispensable for fertilization as constitutive knockout of either Izumo1 or Juno result in mice that are healthy but infertile\textsuperscript{168,173}. Despite their central importance in reproductive medicine, the molecular architectures and the details of their functional roles in fertilization are not known. Here, we present the crystal structures of the human Izumo1 and Juno in unbound and bound conformations. The human Izumo1 structure exhibits a distinct boomerang shape and provides the first structural insights into the Izumo family of proteins\textsuperscript{170}. Human Izumo1 forms a high-affinity complex with Juno and undergoes a major conformational change within its N-terminal domain upon binding to the egg-surface receptor. No structure is yet known for any complex between a sperm protein and an egg protein in any organism. Our results provide new insights into the molecular basis of sperm-egg recognition, cross-species fertilization, and barrier to polyspermy, thus promising benefits for the rational development of novel non-hormonal contraceptives and fertility treatments for humans and other species of mammals.
6.2 Introduction

The journey of human sperm to an egg ends in the female oviduct, when the active sperm penetrates through the egg zona pellucida glycoprotein layer to reach the perivitelline space between the zona layer and the plasma membrane of the oocyte\textsuperscript{149,150,153}. The active sperm then fuses with the oocyte membrane to allow the formation of the zygote\textsuperscript{142,145}. At least two membrane-bound proteins, sperm Izumo1 and egg Juno, are essential in gamete recognition and/or the fusion process\textsuperscript{168,173}. Both \textit{izumo1} and \textit{juno} genes are conserved in other mammals (Figure 6.1 and 6.2) suggesting that a specific Izumo1-Juno interaction is required for mammalian fertilization\textsuperscript{169,173}. Structural characterization of sperm and egg surface proteins would provide key insights into the mechanisms of mammalian fertilization.
6.3 Results and Discussion

Although it has been more than a decade since the Izumo family of proteins were first described, structural and biochemical studies of these cysteine-rich glycoproteins are hampered by difficulties in recombinant protein expression. Using Drosophila melanogaster S2 cells, we expressed and purified the extracellular region of human Izumo1 (residues 22-254) by Ni$^{2+}$-affinity and gel filtration chromatography. Biophysical characterization of Izumo1 revealed a stable and monomeric protein with extensive mixed $\alpha/\beta$ secondary structural character (Figure 6.3). We obtained crystals of unbound Izumo1$_{22-254}$ and determined the structure at 3.1-Å resolution. Izumo1$_{22-254}$ is a monomer and adopts a distinct boomerang shape with dimensions of $\sim$85-Å x 25-Å x 22-Å. The overall structure consists of two domains: a rod-shaped N-terminal four-helix bundle (4HB; residues 22-134) and a C-terminal immunoglobulin-like (Ig-like; residues 167-254) domain (Figure 6.4 and Figure 6.5). Two anti-parallel $\beta$-strands ($\beta$1 and $\beta$2) function like a hinge between the 4HB and Ig-like domains.

The 4HB motif is a very common natural structural fold that occurs in proteins of diverse functions. The four helices in the Izumo1 4HB domain ($\alpha$1, $\alpha$2, $\alpha$3 and $\alpha$4) vary from 14 to 30 residues in length. The helices have amphipathic character with a polar surface exposed to solvent and hydrophobic residues packing into a core. The helices $\alpha$1-$\alpha$2 and $\alpha$3-$\alpha$4 are connected with short 5-residue loops (L1 and L3), while a longer 15-residue loop (L2) links $\alpha$2 to $\alpha$3. The 4HB and hinge regions are stabilized by an extensive network of disulphide linkages (C22-C149, C25-C152, C135-C159, C139-C165) and charge-charge interactions (H44-D101, E80-K154, and R96-E110) that are conserved in almost all Izumo1 orthologs and other Izumo family proteins (Figure 6.4 and 6.6). Commonly, helices in four-helix bundles are tilted $+25^\circ$ or $-35^\circ$ to maximize interhelical contacts and hydrophobic residues pack in a knobs-in-holes fashion to stabilize the core. Atypically, the Izumo1 4HB adopts a rare parallel/antiparallel arrangement of non-tilted helices with no knobs-in-holes packing, suggesting that the 4HB region may be more conformationally labile. Structural analysis using the DALI server reveals that the Izumo1 4HB domain shares similar structural architecture to the Plasmodium protein SPECT1 (rmsd 3.4-Å for 97 Ca atoms) (Figure 6.7). Interestingly, SPECT1 is required for pore
formation by highly motile sporozoites prior to invading hepatocytes; its mechanism is however poorly defined\textsuperscript{434,435}.

Figure 6.1. Alignment of Izumo1 protein sequences from various mammals. Izumo1 sequences from \textit{Homo sapiens} (human; GenBank: BAD91012.1), \textit{Macaca mulatta} (rhesus
macaque; GenBank: EHH30233.1), Gorilla gorilla (gorilla; Uniprot: G3QFY5), Pan paniscus (bonobo; NCBI: XP_003814124.1), Callithrix jacchus (marmoset; Uniprot: F7H859), Chlorocebus sabaeus (green monkey; Uniprot: A0A0D9S2Z4), Papio anubis (baboon; Uniprot: A0A0A0MU86), Nomascus leucogenys (gibbon; Uniprot: G1QXF7), Mus musculus (mouse; GenBank: BAD91011.1), Rattus norvegicus (rat; GenBank: BAD91013.1), Ictidomys tridecemlineatus (squirrel; Uniprot: I3N2L9), Cavia porcellus (guinea pig; Uniprot: H0UTJ7), Ochotona princeps (pika; NCBI: XP_004597241.1), Oryctolagus cuniculus (rabbit; Uniprot: G1TVX5), Felis catus (cat; NCBI: XP_006941089.1), Canis familiaris (dog; Uniprot: F6UM65), Ailuropoda melanoleuca (giant panda, Uniprot: G1M882), Equus caballus (horse; Uniprot: F6YE25), Bos taurus (cow; Uniprot: E1BDA8), Sus scrofa (pig; Uniprot: F1RIQ7), Capra hircus (goat; Uniprot: C6ZEA2), Ovis aries (sheep; Uniprot: W5PRD0), Sorex araneus (shrew; NCBI: XP_004619786.1), Pteropus vampyrus (megabat; NCBI: XP_011372928.1), Loxodonta africana (African elephant; NCBI: XP_003406572.1), and Dasypus novemcinctus (armadillo; NCBI: XP_004451154.1) are aligned. Red boxes indicate complete conservation of a given amino acid. N-linked glycosylation sequons (N-x-S/T) are indicated by red-coloured Y-shaped symbols. Secondary structural elements observed in the crystal structure of Izumo1 are shown as arrows for β-strands and coils for α-helices. Residues that interact with Juno are identified with asterisks, with those that form salt bridges and hydrogen bonds highlighted in blue and green boxes, respectively. Cysteine pairs involved in disulphide bond formation are numbered in red underneath each sequence.
Red boxes indicate complete conservation of a given amino acid. N-linked glycosylation sequons (N-x-S/T) are indicated by red-coloured Y-shaped symbols. Juno is anchored to the plasma membrane through a GPI anchor at S228 (shown as a green lollipop). Secondary structural elements observed in the crystal structure of Juno are shown as arrows for $\beta$-strands and coils for $\alpha$-helices. Residues that interact with Izumo1 are identified with asterisks underneath the sequence, with those that form salt bridges and hydrogen bonds highlighted in blue and green boxes, respectively. Cysteine pairs involved in disulphide bond formation are numbered in red underneath each sequence.
Figure 6.3. Purification and characterization of Izumo1 and Juno. (A) Superdex-75 10/300 GL size-exclusion chromatograms of Juno20-228, Izumo122-254, and the Izumo122-254-Juno20-228
complex. Eluted peak positions of protein standards are marked with triangles and dashed lines. (B) Coomassie-stained SDS-PAGE analysis of the purified Izumo122-258, Juno20-228 and Izumo122-258-Juno20-228 complex. (C) SEC-MALS profile of glycosylated human Izumo122-268. The detector response unit (mV) and molecular mass (kDa) are plotted against the elution volume from a Superdex-200 Increase 10/300 GL size exclusion column. SEC-MALS reveals an apparent molecular weight (MW) of 34.8 kDa (dashed blue line), which corresponds to a monomeric species. (D) A size distribution histogram from dynamic light scattering (DLS) measurements of Izumo122-254, Juno20-228 and Izumo122-254-Juno20-228 complex at 5 mg ml\(^{-1}\). Izumo122-254, Juno20-228, and Izumo122-254-Juno20-228 displays a hydrodynamic radii (R\(H\)) of ~3.0 nm, ~2.9 nm, ~3.9 nm, respectively. (E) Circular dichroism (CD) wavelength scan of human Izumo122-268 (blue) at 25°C shows mixed secondary structural characteristics. The crystal structure of Izumo122-268 aligns well with the secondary structural content calculated from the CD spectrum (35% \(\alpha\)-helical, 24% \(\beta\)-strand, and 41% random coil). A reconstructed CD wavelength scan (red) illustrates the agreement of the fit used in secondary structural content analysis. CD thermal denaturation profile of human Izumo122-268 at 222 nm is shown. CD signal was normalized between 0 (folded) and 1 (unfolded), and plotted as a function of temperature. The \(T_m\) value indicates the midpoint of the melting transition.

Figure 6.4. Overall structure of human Izumo1 and Juno. (A) Domain schematic of human Izumo1 and Juno. Human Izumo1 and Juno contain five and eight conserved disulphide linkages,
respectively (shown by red lines). Red Y-shaped and green lollipop symbols denote N-linked glycans and a glycosphatidylinositol (GPI)-anchor, respectively. Regions not observed in the crystal structure are shaded grey. Abbreviations are as follows: SP, signal peptide; 4HB, four-helix bundle; Hinge; Ig, immunoglobulin-like domain; Linker; TM, transmembrane region; CT, cytoplasmic tail. (B) Ribbon representation of unbound Izumo1_{22-254} and Juno_{20-228}. The Izumo1 4HB domain (orange), hinge region (green) and Ig-like domain (blue) are coloured according to the 2D schematic in panel (A). Loops that connect the four helices (α1, α2, α3, and α4) in the 4HB and β-strands (β1 and β2) in the hinge region are labelled as L1, L2, L3, L4 and L5. The β-strands in the Ig-like domain are labelled from A to G, according to convention for immunoglobulins. Ribbon diagram of Juno_{20-228} is coloured in purple. The helices and strands are labelled α1, α2, α3, α4, and α5, and β1, β2, β3, and β4, respectively. Cysteine residues that form conserved disulphide linkages are highlighted in red.
Figure 6.5. Overall structural architecture of Izumo1 and Juno. Ribbon and topology schematic of (A) Izumo122-254 and (B) Juno20-228. β-strands and α-helices are depicted in the topology diagram as arrows and cylinders, respectively. Disulfide bonds are shown as red sticks.
Figure 6.6. Electrostatic potential surface representation of Izumo1_{22-254}. (A) Electrostatic potential from -5 to +5 kT/e mapped onto the molecular surface of Izumo_{1,22-254} is shown in four orientations, 90° apart. (B) Multiple sequence alignment of human Izumo1, Izumo2, Izumo3 and Izumo4. The eight cysteines observed in Izumo1 are highly conserved in the human Izumo family. However, there is very poor overall sequence conservation in the family (<3% sequence identity). No residues involved in binding Juno are conserved (highlighted with asterisks).
Moreover, Izumo2, Izumo3 and Izumo4 do not appear to contain an intact Izumo1 Ig-like domain. Izumo4 also seems to lack a transmembrane anchor.

**Figure 6.7. Izumo1 4HB is structurally similar to SPECT1, a Plasmodium protein involved in host cell traversal.** A 3D structural database search using DaliLite\(^4\) (v.3) revealed similarity between Izumo1 4HB and SPECT1\(^4\). Overall, 97 positions between SPECT1 (PDB ID: 4U5A) and Izumo1 4HB (residues 22-134) superimpose with 3.4-Å rmsd.

The Izumo\(_{22-254}\) Ig-like domain resides at the membrane-proximal end of the molecule. It adopts a seven-stranded (A, B, C, C’, E, F, G) β-sandwich with the two β-sheets covalently linked with an Ig-superfamily (IgSF) conserved disulphide bond (C182-C233) between strands B and F (Figure 6.4). Seven-stranded Ig-like folds classically consist of a 3+4 arrangement with β-strands A, B and E forming β-sheet 1, and β-strands C, C’, F and G forming β-sheet 2\(^4\) (Figure 6.8). The Izumo\(_{22-254}\) Ig-like domain has a novel 2+5 organization representing a new IgSF subtype. In Izumo\(_{22-254}\), strand A interacts with β-sheet 2 rather than β-sheet 1. The disulphide bond preceding β-strand A (C139-C165) may constrain the movement of the strand toward β-sheet 1.
resulting in this strand switch (Figure 6.9). A comparative analysis of the Izumo1\textsubscript{22-254} Ig-like domain to structures deposited in the PDB using the secondary structure matching (SSM) server revealed that its closest structural neighbour is the murine fibroblast growth factor receptor-2 (FGFR2), which adopts a C2-type Ig-like fold (rmsd 1.3-Å for 69 C\textalpha{} atoms with a SSM Q-score=0.57) (Figure 6.9). The major difference between the C2-type and Izumo1 Ig-like folds is the 2+5 arrangement of the β-sandwich. We also determined the crystal structure of a slightly longer Izumo1\textsubscript{22-268} construct at 2.9-Å resolution to gain insights into the C-terminal linker region immediately following the Ig-like domain. The Izumo1\textsubscript{22-268} structure superimposes well with Izumo1\textsubscript{22-254} (rmsd 1.0-Å over all atoms) (Figure 6.10). However, no electron density was observed after residue 257, suggesting that the linker region following the Ig-like domain is flexible.

Juno (previously known as folate receptor-δ (FOLR-δ)) is a glycophosphatidylinositol (GPI)-anchored, cysteine-rich glycoprotein displayed on the egg surface that was demonstrated to be the egg receptor of Izumo\textsubscript{1}\textsuperscript{173}. We determined the crystal structure of the unbound Juno\textsubscript{20-228} at 1.8-Å resolution. Juno\textsubscript{20-228} has a globular architecture that is composed of five short α-helices (α1, α2, α3, α4 and α5), three 3\textsubscript{10} helices and two short two-stranded antiparallel β-sheets (Figure 6.4). Eight conserved disulphide bonds stabilize the core helices α1, α2, α3 and α4, and flexible loops. Juno shares sequence and structural similarity to the human FOLR-α and FOLR-β (~58% sequence identity and ~1-Å rmsd over 197 C\textalpha{} atoms)\textsuperscript{174,175} (Figure 6.11). Despite the close similarities, six key folate binding residues in FOLR-α and FOLR-β are not conserved in human Juno, thus supporting previous observations that Juno does not bind folate\textsuperscript{173–175} (Figure 6.11). While this manuscript was under review, a partial structure of murine Juno was determined\textsuperscript{439}. The human and murine Juno structures align well, however a number of loops are disordered and missing from the murine Juno structure (Figure 6.12).
Figure 6.8. Comparison of the Izumo1 Ig-like domain and various Ig-superfamily (IgSF) subtypes. The Izumo1 Ig-like domain adopts a 2+5 stranded topology; with sheet II (front sheet in this view) includes five strands \((a, g, f, c, c')\) and sheet I contains two strands \((b, e)\). The 2+2 stranded common structural core (strands coloured in red and labeled \(b, c, e, f\)) is conserved in all IgSF domains. Due to a disulphide bond preceding the Izumo1 Ig-like domain, strand A is forced to hydrogen bond with strand G in sheet I. The Izumo1 Ig-like domain is the only known member to adopt a 2+5 arrangement, thus suggesting a new IgSF subtype. Selected examples illustrating the various IgSF subtypes are as follows: C1-type, T-cell receptor constant (PDB ID: 1BEC); C2-type, CD4 constant (PDB ID: 3CD4); C3-type, Cu/Zn superoxide dismutase (PDB ID: 2SOD); C4-type, neocarzinostatin (PDB ID: 1NCO); V-type, Ig-variable domain heavy chain (PDB ID: 7FAB); Fn3-type, neuroglian (PDB ID: 1CFB); I-type, telokine (PDB ID: 1TLK); and H-type, galactose oxidase (PDB ID: 1GOG).
**Figure 6.9. Izumo1 Ig-like domain.** (A) The Izumo1 Ig-like domain adopts a novel 2+5 β-strand arrangement. A disulfide bond (shown as red sticks) between C139 and C165 stabilizes the interaction of the A-strand with β-sheet II. The Izumo1 Ig-like domain is coloured in blue and hinge region is shown in green. (B) Superimposition of Izumo1 Ig-like domain (shown in blue) and the C2-type Ig-like domain from murine FGFR2 (shown in red; PDB ID: 4HWU). The Izumo1 Ig-like domain adopts a C2-type Ig-like fold, similar to murine FGFR2 (rmsd 1.3 Å for 69 Cα atoms) with the exception of the 2+5 β-strand arrangement.
Figure 6.10. Crystal structure of Izumo1\textsubscript{22-268}. (A) A representative 2.9-Å resolution $\sigma_A$-weighted composite omit $2F_o-F_c$ electron density map of Izumo1\textsubscript{22-268} contoured at 1σ and superimposed with its final refined model. (B) Ribbon diagram of Izumo1\textsubscript{22-268}. The 4HB region, hinge region and Ig-like domain are shown in salmon, lemon and purple, respectively. Conserved disulphide linkages are depicted as red sticks. (C) Superimposition of the Izumo1\textsubscript{22-254} (shown in
salmon) and Izumo1_{22-268} (shown in green) crystal structures. Overall, the structures align with a rmsd of 1.0 Å over all C α atoms. Residues 258-268 in Izumo1_{22-268} are disordered suggesting that the linker region between the Ig-like domain and the transmembrane domain is highly flexible.

Figure 6.11. Structural comparison of Juno and the folate receptor family of proteins. (A) Structural superimposition of Juno_{20-228} with FOLR-α (PDB ID: 4LRH) and FOLR-β (PDB ID:
Experimentally bound folate (FOL), shown in white sticks, from the FOLR-α structure is positioned in the active site. (B) Superimposition of residues in the folate-binding site of human FOLR-α and FOLR-β, and equivalent residues in human Juno. Residue names shown in black are conserved between Juno, FOLR-α, and FOLR-β, and are numbered based on the FOLR-α sequence. Inset boxes highlight the residue differences between Juno, FOLR-α, and FOLR-β. Key hydrogen bond interactions are shown as dashed black lines. Mutagenesis studies showed that replacement of D103/D97, which forms strong interactions to the N1 and N2 nitrogen atoms of the pterin moiety, results in decreased affinity by more than one order of magnitude174. Six folate-binding residues observed in FOLR-α, and FOLR-β (FOLR-α/FOLR-β: D103/D97, W124/W118, R125/R119, V129/F123, H157/H151, and K158/R152) are not conserved in Juno. Four of these residues (FOLR-α/FOLR-β: D103/D97, W124/W118, R125/R119, and H157/H151) form key hydrogen bonds to anchor folate in the active site. In Juno, the substituted residues are not able to maintain the extensive hydrogen bond network seen in FOLR-α and FOLR-β to folate. (C) Homo sapiens FOLR-α (Uniprot: P15328), FOLR-β (Uniprot: P14207), FOLR-γ (Uniprot: P41439) and FOLR-δ (Uniprot: A6ND01) are aligned. Red boxes indicate complete conservation of a given amino acid. N-linked glycosylation sequons (N-x-S/T) are indicated by red-coloured Y-shaped symbols. Juno is anchored to the plasma membrane through a GPI anchor at Ser228 (shown as a green lollipop). Experimentally determined secondary structural elements are shown as arrows for β-strands and coils for α-helices. Key folate-binding residues, identified from the FOLR-α and FOLR-β crystal structures, are identified with an asterisk underneath the sequence. Key residues difference between Juno, FOLR-α and FOLR-β folate binding sites are highlighted in a blue box.
Figure 6.12. Structural comparison of human and murine Juno. During revisions for our manuscript, a murine Juno crystal structure (PDB code: 5EJN) was determined\textsuperscript{439}. The reported murine Juno model consists of only a partial structure, as a number of residues (residues 34-40, 67-78, 112-117 and 152-165 in chain A) in the loop regions are missing due to flexibility and disorder in the crystal lattice. Superimposition of human Juno (PDB code: 5F4Q) and murine Juno (PDB code: 5EJN) reveals no major structural differences in the core regions (rmsd=1.44 Å over 201 residues). However, two missing loops in murine Juno (residues 67-78 and 152-165) play an important role in Izumo1 binding.

In order to identify whether human Izumo1 and Juno interact, we used biolayer interferometry (BLI) and surface plasmon resonance (SPR) to measure the binding affinities of this receptor
pair. The interaction is stable with a dissociation constant ($K_d$) between ~48-60 nM (Figure 6.13 and Table 6.1). In addition, Izumo1$_{22-254}$ and Juno$_{20-228}$ co-purify as a 1:1 complex on size exclusion chromatography (Figure 6.3). Previously, the Izumo1-Juno interaction was characterized by SPR to be transient and low affinity ($K_d = 12 \, \mu\text{M}$). However, our results indicate that these two proteins instead engage in a stable complex during the gamete fusion process. To understand the precise molecular interactions of Izumo1-Juno, we crystallized the Izumo1$_{22-254}$-Juno$_{20-228}$ complex and determined the structure to 2.4-Å resolution. In the asymmetric unit, we observed one Izumo1$_{22-254}$ molecule binding to one Juno$_{20-228}$ molecule with an interface that spans over ~910 Å$^2$ surface area (Figure 6.13C). Formation of this interface results in a calculated free energy gain of -10.4 kcal/mol. Izumo1 has been shown to interact with Juno via its N-terminal domain. The crystal structure indicates that residues from all three Izumo1 regions (4HB, hinge and Ig-like) contact Juno$_{20-228}$ through extensive non-bonded van der Waals, hydrophobic and aromatic interactions (>60% of total interface interactions) (Figure 6.14). There are also two intermolecular salt bridges (K163$_{Juno}$-E71$_{Izumo1}$ and E45$_{Juno}$-R160$_{Izumo1}$) and eight hydrogen bond interactions at the interface (Figure 6.13 and 6.14). However, all of these interactions are >3.0-Å apart, suggesting these are weak in nature. No N-linked glycans on either Izumo1$_{22-254}$ or Juno$_{20-228}$ are involved in binding.

Structural comparison of Izumo1$_{22-254}$ and Juno$_{20-228}$ alone to their bound states revealed no major structural differences in Juno$_{20-228}$ or the Ig-like domain of Izumo1 upon complex formation (Figure 6.15). A major conformational change was observed in the 4HB and hinge regions. All four helices in the 4HB move ~20-Å towards Juno$_{20-228}$, whereas the L2 and hinge regions shift ~8-Å upon binding. As a result, the bound Izumo1$_{22-254}$ abandons its distinct boomerang shape and adopts an upright conformation. The structural constraints from the short loops between the $\alpha_1$-$\alpha_2$ and $\alpha_3$-$\alpha_4$ helices allow the 4HB domain to translate as a single unit. These conformational changes in the 4HB and hinge regions might be driven by the formation of new intermolecular interactions between Izumo1$_{22-254}$ and Juno$_{20-228}$ (Figure 6.13D and 6.15).

In order to understand the conformational dynamics of the Izumo1-Juno interaction in solution, we utilized a hybrid approach that combined small angle X-ray scattering (SAXS) and deuterium
exchange mass spectrometry (DXMS). SAXS provides medium-resolution visualization of
global structural conformational changes in solution, whereas DXMS, which measures kinetics
of backbone amide solvent exchange, is sensitive to local residue-level conformational
dynamics. Ab initio SAXS reconstructions of unbound Izumo1$_{22-254}$ revealed a distinct
boomerang shape, similar to its crystal structure (Figure 6.16 and 6.17). Upon Juno$_{20-228}$ binding,
Izumo1$_{22-254}$ adopts an upright conformation. Our DXMS studies revealed that the residues lining
the binding interface exhibited a reduced exchange profile in the complex compared to the
unbound states (Figure 6.16 and 6.18). Moreover, DXMS experiments performed on Izumo1
alone indicated a high level of exchange in the hinge region, thus suggesting dynamic flexible
motion within this region. Upon Juno binding, deuterium exchange of residues 127-140 in the
hinge region was reduced by $>50\%$, which is more than the reduction seen by residues at the
Izumo1-Juno interface (Figure 6.16). The strong level of H/D protection is due to the formation
of 10 additional main chain hydrogen bonds within residues 127-140 of the hinge region. This
suggests that the Izumo1 hinge region is stabilized in a “locked” upright position in the presence
of Juno.

Mutational studies at the Izumo1-Juno interface revealed the structural determinants required for
binding (Figure 6.13, 6.19, 6.20 and Table 6.1). Upon binding to Juno$_{20-228}$, a D72-Q130
hydrogen bond found between Izumo1 L2 and the 4HB $\alpha$4 helix is broken to form a new
intermolecular salt bridge (E71$_{Izumo1}$-K163$_{Juno}$) at the interface (Figure 6.15). Alanine mutations
to Izumo1 D72 and Q130 did not affect binding to Juno, as these residues are not involved at the
interface. Alanine and charge reversal mutations to intermolecular salt bridges (K163$_{Juno}$-
E71$_{Izumo1}$ and E45$_{Juno}$-R160$_{Izumo1}$) reduced the $K_d$ between $\sim$2- and 50-fold, suggesting these ion
pairs play a role in binding (Figure 6.20 and Table 6.1). Using SAXS, we characterized the
binding mode of Izumo1 and Juno salt bridge mutants that had $K_d$ values from nanomolar to
micromolar affinities. The SAXS scattering data and reconstructions show no major change
compare to the wild-type Izumo1-Juno complex, and suggest proper complex formation despite a
47-fold decrease in affinity (Figure 6.21 and 6.22). Significant reductions in binding ($>20$-fold)
were also observed with mutations to Izumo1 and Juno residues that are conserved in most
mammals (W148$_{Izumo1}$, H157$_{Izumo1}$, W62$_{Juno}$ and L81$_{Juno}$). In fact, an alanine mutation to
W148$_{Izumo1}$ completely abolished Izumo1-Juno binding. Taken together, our mutagenesis study
suggests that the interface is likely stabilized through the combined effects of multiple weak van der Waals, hydrophobic and aromatic interactions. This allows the Izumo1-Juno interface to be resilient to mutations.

Although we observed discernible sequence conservation at the complex interface in both proteins, comparative sequence analysis revealed significant variations among a number of interface residues. Approximately one-half of the residues (Juno: Y44, E45, L58, F77, M83, R87, M145, Y147, K163 and Izumo1: L69, V141, K150, N151, K153, E155, A158, Y163, N239, S241) vary across mammalian species (Figure 6.23). Similar to the species-specific recognition employed by glycoproteins in the zona pellucida that surrounds the egg, these residues may act to restrict a productive Izumo1 and Juno binding event to a specific pair of species. For example, in primates, these residues are mainly preserved (Figure 6.1, 6.2 and 6.23), suggesting that the species-specific diversification of Izumo1 and Juno may be restricted to non-primate mammals. The Izumo1-Juno axis may be an additional barrier for cross-species fertilization.

Cell-cell fusion is fundamental to the physiology of uni- and multi-cellular organisms ranging from yeast to humans. Malfunction in cellular fusion events in humans might be associated with diseases such as infertility, preeclampsia, and myopathies. Our structural characterizations indicate that the human Izumo1 does not have properties predictive of viral, intracellular and developmental fusogens, such as influenza A virus hemagglutinin (HA) and C. elegans epithelial fusion failure-1 (EFF-1) proteins. This suggests that Izumo1 does not seem to function as a direct fusion protein. At least three different fusion mechanisms are possible. Firstly, Izumo1 may act as a scaffold to recruit a protein complex that contains or regulates other fusion proteins. The requirement of a multiprotein complex for fusion is not unusual, as some viruses, such as herpes simplex virus-1 or Epstein Barr virus requires the formation of a multicomponent fusion complex. Alternatively, Inoue et al. proposed that monomeric Izumo1 on the sperm surface interacts with Juno. Subsequently, a protein disulfide isomerase facilitates the dimerization of Izumo1 to allow an interaction with another oocyte receptor to facilitate fusion. Finally, the tight heterotypic interaction between human Izumo1 and Juno
proteins may be sufficient to bring close apposition of the sperm and egg membranes leading to fusion. Regardless, the conformational changes within Izumo1 suggest that receptor adhesion triggers the progression of the 4HB domain to the vicinity of the egg membrane and the conformational switch may be part of the structural changes required for fusion. After fusion, the fertilized egg rapidly sheds Juno molecules into the perivitelline space\textsuperscript{173}. Given our measured tight nanomolar affinities between Izumo1 and Juno, shed Juno may essentially act as a rapid “sperm-sink” to neutralize incoming acrosome-reacted sperm as an additional polyspermy block. This process may be analogous to the shedding of viral glycoproteins (e.g. Ebola virus shed and soluble GP), which are able to act as a decoy to absorb antibody responses\textsuperscript{445,446}. This first glimpse into the molecular mechanisms of gamete recognition/fusion machinery provides a better understanding of human fertilization and reproductive medicine at the molecular level.
Figure 6.13. Izumo1-Juno heterotypic assembly. (A) Surface plasmon resonance (SPR) binding affinity and kinetic analysis of the human Izumo122-254 and Juno20-228 interaction. Human Juno20-228 was amine-coupled to the SPR sensor chip. The binding affinity of human Izumo122-254
was measured at 0.75, 0.50, 0.375, 0.25, 0.188 and 0.125 µM concentrations and kinetic parameters were derived from a Langmuir 1:1 binding model. (B) Biolayer interferometry (BLI) kinetic analysis of the human Izumo122-254 and Juno20-228 interaction. Human Juno20-228 was biotinylated and coupled to streptavidin-coated biosensors. The binding affinity of human Izumo122-254 was measured at 0.92, 0.46, 0.23, 0.12, and 0.06 µM concentrations and kinetic parameters were derived from a 1:1 binding model. The experimental curves are shown in colour superimposed with the fitted curves indicated as gray lines. (C) Crystal structure of the human Izumo122-254-Juno20-228 complex shown as a ribbon diagram. Izumo122-254 and Juno20-228 are coloured according to Figure 1. A disordered loop between the β1 and β2 strands of the Juno20-228 is shown with a black dashed line. (D) Binding site interactions of Izumo122-254 and Juno20-228. Side chains of key residues involved in hydrogen bond or salt bridge interactions are shown. (E) Electrostatic potential surface representation of the Izumo122-254-Juno20-228 binding interface. The footprints of the binding interface are shown by black dashed lines. R160 and E71 on Izumo1 form a salt bridge with E45 and K163 on Juno, respectively. Residues K150 and K153 form an additional positively charged patch on Izumo1. However, this positively charged patch does not likely play a major role in Juno binding, as there is no complementary negatively charged surface on Juno. In addition, neither K150 nor K153 are strictly conserved in other Izumo1 orthologues (Figure 6.1). The electrostatic charges from -5 to +5 kT/e are shown in a red-to-blue colour gradient. (F) BLI binding affinity analysis of Izumo122-254 and Juno20-228 interface mutants. Wild-type Izumo1-Juno interaction is normalized at 100% and the binding affinities (K_d) for each mutant is shown as a percent reduction to wild-type. All experiments in panel (A), (B) and (F) were performed with technical triplicates (n=3), and mean K_d values and standard deviations are presented in Table 6.1.
**Figure 6.14. Izumo1-Juno interface.** (A) 2D schematic of the interactions between Izumo1$_{22-254}$ and Juno$_{20-228}$. Residues from the Izumo1 4HB, hinge, and Ig-like regions, as well as Juno are coloured in orange, green, blue, and purple, respectively. Hydrogen-bond interactions are shown as dashed lines, and van der Waals forces are depicted as grey semi-circles. (B) Footprint of Juno on the surface of Izumo1 and Izumo1 on the surface of Juno. The molecular surfaces of Izumo1 and Juno are coloured white with residues forming interactions coloured similarly to panel (A).

**Figure 6.15. Conformational changes in Izumo1 upon binding to Juno.** Superimposition of structures of unbound Izumo1$_{22-254}$ and Juno$_{20-228}$ (shown in grey) on the Izumo1$_{22-254}$-Juno$_{20-228}$ complex (coloured according to Figure 1). Unbound Izumo1$_{22-254}$ abandons its distinct boomerang shape and adopts an upright conformation in the complex. The 4HB region of the unbound Izumo1$_{22-254}$ undergoes a ~20-Å upward displacement upon binding to Juno$_{20-228}$. The hinge region and the L2 loop also move ~8- to 12-Å towards Juno$_{20-228}$ upon complex formation. Black arrows highlight the positional changes in secondary structure with the corresponding distances shown in Ångstroms. The inset panel displays the conformational changes within the L2 region during formation of the complex. The L2 region residue D72 and α4 helix residue Q130 (both shown in grey) form a hydrogen bond in the unbound Izumo1$_{22-254}$ structure. Upon
binding to Juno$_{20-228}$, the L2 region residue E71 (orange) forms an electrostatic interaction with the Juno$_{20-228}$ K163 (purple). Electrostatic and hydrogen bond interactions are indicated by red dashed lines. A disordered loop between Juno$_{20-228}$ β1 and β2 strands is shown as a black dashed line.

Figure 6.16. Hybrid structural analyses of human Izumo1 and Juno in a solution state. (A) *Ab initio* small-angle X-ray scattering (SAXS) reconstruction of Izumo1$_{22-254}$, Juno$_{20-228}$, and the Izumo1$_{22-254}$-Juno$_{20-228}$ complex using the program DAMMIN$^{447}$. Twenty-three models from independent DAMMIN runs were aligned and averaged using DAMAVER$^{448}$ and shown as a
molecular surface using the program Chimera. The Izumo122-254, Juno20-228 and Izumo122-254-Juno20-228 complex crystal structures were docked into the SAXS reconstructed molecular envelopes. The boomerang shape and upright conformation seen in the crystal structures of unbound and bound Izumo122-254, respectively, were recapitulated by the SAXS reconstructions. Comparative deuterium exchange mass spectrometry (DXMS) profile of unbound and bound (B) Izumo122-254 and (C) Juno20-228. The plots reveal the change in individual deuterium exchange for all observable residues. The coloured lines above the residue numbers correspond to the observed regions in the crystal structures. Comparative DXMS profiles mapped onto the molecular surfaces of (D) Izumo122-254 and (E) Juno20-228. The residues identified from comparative DXMS to undergo exchange upon complex formation agree well with the residues observed at the interface in the crystal structure of the Izumo122-254-Juno20-228 complex. The change in deuterium exchange is shown using a colour gradient.
Figure 6.17. SAXS data and analysis. (A) Experimental scattering of deglycosylated Izumo122-254, Juno20-228 and Izumo122-254-Juno20-228 ranging from 1.0 to 5.0 mg ml\(^{-1}\). No concentration-dependent or radiation effects were observed. The inset box shows linearity in the Guinier plot at low \(q\) (\(qR_g < 1.3\)) for the Izumo122-254, Juno20-228 and Izumo122-254-Juno20-228 SAXS data at 5.0, 5.0 and 4.0 mg ml\(^{-1}\), respectively. (B) Normalized pair distance distribution function, \(P(r)\). (C) Kratky plot showing the degree of flexibility. (D) Summary of the experimentally derived SAXS parameters for Izumo122-254, Juno20-228 and Izumo122-254-Juno20-228. The program SCATTER\(^{450}\) was used to calculate the radius of gyration (Rg), maximum linear dimension (\(D_{max}\)), and to perform Porod-Debye analysis to obtain the Porod volume and P coefficient.
Figure 6.18. Comparative DXMS analysis. The DXMS exchange profiles of (A) unbound Izumo12-254, (B) Izumo12-254 bound to Juno, (C) unbound Juno20-228 and (D) Juno20-228 bound to

[Diagram showing DXMS exchange profiles for Izumo12-254 and Juno20-228]
Izumo1. Percentage deuteration exchange was determined over five different time points (10, 100, 1000, 10000 and 100000 seconds) for all observable Izumo122-254 and Juno20-228 residues in the bound and unbound states. The lines above the residue numbers are colour coded with the regions observed in the crystal structures (4HB, hinge and Ig-like domains, and Juno).

Figure 6.19. Biolayer interferometry binding affinity and kinetic analysis of Izumo1-Juno interface residues. Biotin-labeled wild-type or mutant Juno20-228 was coupled to streptavidin-
coated biosensors and monitored for binding to wild-type or mutant Izumo1\textsubscript{22-254} at various concentrations. The data was analyzed based on a 1:1 binding model using the BLIzt Pro software, with the fitted curves shown as grey lines.

Table 4. Izumo1-Juno binding interface mutations.

<table>
<thead>
<tr>
<th>Technique</th>
<th>Protein</th>
<th>Role of residue</th>
<th>$K_d$</th>
<th>$k_a \times 10^6$ (M$^{-1}$ s$^{-1}$)</th>
<th>$k_d \times 10^6$ (s$^{-1}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SPR</td>
<td>Wild-type</td>
<td>n/a</td>
<td>48 ±/− 4 nM</td>
<td>4.2 ±/− 0.2</td>
<td>19.4 ±/− 0.9</td>
</tr>
<tr>
<td>BLI</td>
<td>Wild-type</td>
<td>n/a</td>
<td>59 ±/− 1 nM</td>
<td>1.15 ±/− 0.02</td>
<td>6.7 ±/− 0.1</td>
</tr>
</tbody>
</table>

Izumo1 mutants:

<table>
<thead>
<tr>
<th>Technique</th>
<th>Protein</th>
<th>Role of residue</th>
<th>$K_d$</th>
<th>$k_a \times 10^6$ (M$^{-1}$ s$^{-1}$)</th>
<th>$k_d \times 10^6$ (s$^{-1}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>BLI</td>
<td>E71A Izumo1\textsubscript{22-254}</td>
<td>salt bridge with Juno K163, L2 leash</td>
<td>96 ±/− 13 nM</td>
<td>1.43 ±/− 0.04</td>
<td>13.7 ±/− 0.2</td>
</tr>
<tr>
<td>BLI</td>
<td>E71K Izumo1\textsubscript{22-254}</td>
<td>salt bridge with Juno K163, L2 leash</td>
<td>119 ±/− 19 nM</td>
<td>1.65 ±/− 0.04</td>
<td>19.5 ±/− 0.3</td>
</tr>
<tr>
<td>BLI</td>
<td>D772A Izumo1\textsubscript{22-254}</td>
<td>H-bond to Q138 in unbound state</td>
<td>71 ±/− 8 nM</td>
<td>1.78 ±/− 0.04</td>
<td>12.6 ±/− 0.2</td>
</tr>
<tr>
<td>BLI</td>
<td>Q130A Izumo1\textsubscript{22-254}</td>
<td>H-bond to D772 in unbound state</td>
<td>56 ±/− 19 nM</td>
<td>1.27 ±/− 0.03</td>
<td>6.8 ±/− 0.1</td>
</tr>
<tr>
<td>BLI</td>
<td>W148A Izumo1\textsubscript{22-254}</td>
<td>conserved interface residue</td>
<td>1.8 ±/− 1 μM</td>
<td>0.51 ±/− 0.05</td>
<td>79 ±/− 2</td>
</tr>
<tr>
<td>BLI</td>
<td>H157A Izumo1\textsubscript{22-254}</td>
<td>conserved interface residue</td>
<td>no binding detected</td>
<td>56.3 ±/− 0.9</td>
<td></td>
</tr>
<tr>
<td>BLI</td>
<td>R160A Izumo1\textsubscript{22-254}</td>
<td>salt bridge with Juno E45</td>
<td>730 ±/− 78 nM</td>
<td>0.78 ±/− 0.04</td>
<td>56.3 ±/− 0.9</td>
</tr>
<tr>
<td>BLI</td>
<td>R160E Izumo1\textsubscript{22-254}</td>
<td>salt bridge with Juno E45</td>
<td>2.2 ±/− 1.2 μM</td>
<td>0.32 ±/− 0.02</td>
<td>51.2 ±/− 1.1</td>
</tr>
</tbody>
</table>

Juno mutants:

<table>
<thead>
<tr>
<th>Technique</th>
<th>Protein</th>
<th>Role of residue</th>
<th>$K_d$</th>
<th>$k_a \times 10^6$ (M$^{-1}$ s$^{-1}$)</th>
<th>$k_d \times 10^6$ (s$^{-1}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>BLI</td>
<td>E45A Juno \textsubscript{22-254}</td>
<td>salt bridge with Izumo R160</td>
<td>681 ±/− 141 nM</td>
<td>0.37 ±/− 0.01</td>
<td>24.7 ±/− 0.5</td>
</tr>
<tr>
<td>BLI</td>
<td>E45K Juno \textsubscript{22-254}</td>
<td>salt bridge with Izumo R160</td>
<td>2.8 ±/− 0.2 μM</td>
<td>0.23 ±/− 0.02</td>
<td>62.4 ±/− 1.7</td>
</tr>
<tr>
<td>BLI</td>
<td>W62A Juno \textsubscript{22-254}</td>
<td>conserved interface residue</td>
<td>361 ±/− 34 nM</td>
<td>1.14 ±/− 0.04</td>
<td>41.2 ±/− 0.6</td>
</tr>
<tr>
<td>BLI</td>
<td>L81A Juno \textsubscript{22-254}</td>
<td>conserved interface residue</td>
<td>9.1 ±/− 0.4 μM</td>
<td>0.4 ±/− 0.3</td>
<td>349 ±/− 23</td>
</tr>
<tr>
<td>BLI</td>
<td>K165A Juno \textsubscript{22-254}</td>
<td>salt bridge with Izumo E71</td>
<td>97 ±/− 9 nM</td>
<td>1.38 ±/− 0.05</td>
<td>13.5 ±/− 0.2</td>
</tr>
<tr>
<td>BLI</td>
<td>K163E Juno \textsubscript{22-254}</td>
<td>salt bridge with Izumo E71</td>
<td>134 ±/− 4 nM</td>
<td>1.81 ±/− 0.05</td>
<td>24.2 ±/− 0.3</td>
</tr>
</tbody>
</table>

*Kinetic and binding affinity values are presented as the mean of technical triplicates (n=3) and errors as the standard deviation of the mean.*

$K_d$, dissociation constant

$k_a$, rate of association

$k_d$, rate of dissociation
Figure 6.20. Specific conserved interactions at the Izumo1-Juno binding interface. Residues involved in Izumo1-Juno binding are shown in stick representation. Key binding interface residues W62\textsubscript{Juno}, L81\textsubscript{Juno}, W148\textsubscript{Izumo1} and H157\textsubscript{Izumo1} are conserved in most mammals. Dashed lines indicate hydrogen bonds.
Figure 6.21. SAXS data and analysis of Izumo1 and Juno mutants. (A) Experimental scattering of wild-type (WT) and charge reversal mutant Izumo1_{22-254}-Juno_{20-228} complexes ranging from 1.0 to 5.0 mg ml^{-1}. No concentration-dependent or radiation effects were observed. The inset box shows linearity in the Guinier plot at low $q$ ($q R_g < 1.3$) for the E71K Izumo1_{22-254}-WT Juno_{20-228}, R160E Izumo1_{22-254}-WT Juno_{20-228}, WT Izumo1_{22-254}-E45K Juno_{20-228}, WT
Izumo$_{122-254}$-K163E Juno$_{20-228}$ complexed SAXS data at 4.0, 4.0, 3.0 and 4.0 mg ml$^{-1}$, respectively. (B) Normalized pair distance distribution function, $P(r)$. (C) Kratky plot showing the degree of flexibility.

**Figure 6.22. SAXS reconstruction of Izumo1 and Juno mutants.** (A) *Ab initio* small-angle X-ray scattering (SAXS) reconstruction of various wild-type (WT) and mutant Izumo$_{122-254}$-Juno$_{20-228}$ complexes using the program DAMMIN$^{447}$. Twenty-three models from independent DAMMIN runs were aligned and averaged using DAMAVER$^{448}$ and shown as a molecular surface using the program Chimera$^{449}$. The WT Izumo$_{122-254}$-WT Juno$_{20-228}$ complex crystal structure was docked into the SAXS reconstructed molecular envelopes. (B) Summary of the experimentally derived SAXS parameters for E71K Izumo$_{122-254}$-WT Juno$_{20-228}$, R160E Izumo$_{122-254}$-WT Juno$_{20-228}$, WT Izumo$_{122-254}$-E45K Juno$_{20-228}$, WT Izumo$_{122-254}$-K163E Juno$_{20-228}$ mutant complexes. The program SCATTER$^{450}$ was used to calculate the radius of gyration.
(Rg), maximum linear dimension (D\textsubscript{max}), and to perform Porod-Debye analysis to obtain the Porod volume and P coefficient.

Figure 6.23. Conservation of residues at the Izumo1-Juno interface. (A) Footprint of Juno on the molecular surface of Izumo1 (same as Extended Data Figure 6b). Representation of surface
residue conservation, calculated using ConSurf and the alignment of (B) all mammalian Izumo1 or (C) primate-only sequences from Extended Data Figure 1. Degree of residue conservation is coloured in a gradient from high (burgundy) to low (cyan) variability.

Figure 6.24. Comparison of Izumo1 with selected viral fusogens. A common feature of many viral fusogens is the presence of a hydrophobic fusion peptide or fusion loop. (A) A Kyte and Doolittle hydropathy plot was calculated for Izumo1, HIV-1 gp160, influenza A HA, Ebola virus

...
GP, Dengue virus type 2 E, and herpes simplex virus-1 gB to detect the presence of hydrophobic regions. Class I and class II viral fusion glycoproteins (GPs) contain three clear hydrophobic regions corresponding to the signal peptide (highlighted in grey), fusion peptide or loop (highlighted in red) and the transmembrane anchor (highlighted in blue). For class III viral GPs, the presence of a signal peptide and transmembrane anchor are clear, however the hydrophobic fusion loop is formed by two discontinuous regions. This results in a lower hydropathy scale that is more difficult to detect. Two regions of hydrophobic residues cluster at the tip of the GP (shown in red) and are thought to be the internal fusion loop. In all class I, II and III viral fusion GPs, a clustering of aromatic and hydrophobic residues in a loop or helical region are hallmark features of the proteins for fusion. In contrast, Izumo1 clearly does not have any hydrophobic regions or structural features similar to the viral fusogens that could insert into the egg membrane. (B) Molecular surface representation of the class I, II, and III viral GPs and Izumo1. The fusion peptide/loop is shown as red sticks and also coloured red on the GP surface. For the class I viral GPs, the metastable prefusion trimer is shown, with the receptor binding and fusion subunits shown in blue and green, respectively. For the class II and class III viral GPs, the postfusion trimer is shown with three hydrophobic fusion loops clustered at the tip of the molecule.
6.4 Materials and Methods

6.4.1 Protein expression and purification. The genes corresponding to full-length human Izumo1 (GenBank accession number: NM_182575, residues 1-350) and human Juno (GenBank accession number: NM_001199206, residues 1-250) were codon optimized for expression in *Drosophila melanogaster* and gene synthesized (Integrated DNA Technologies). The DNA sequences encoding the extracellular regions of Izumo1 (residues 22-254 and 22-268) and Juno (residues 20-228) with a BiP signal peptide were subcloned into a modified pMT-puromycin expression vector (Invitrogen). All protein constructs contain a thrombin cleavage site and 10X-His affinity tag at the C-terminus. Binding interface mutants of Izumo1 22-254 and Juno20-228 were generated using a modified QuikChange PCR-based site-directed mutagenesis protocol. The resulting wild-type and mutant Izumo1 22-254, Izumo1 22-268 and Juno 20-228 pMT expression plasmids were stably transfected in Drosophila S2 cells (Invitrogen) using Effectene transfection reagent (Qiagen), according to manufacturer’s protocol. Briefly, Drosophila S2 cells were cultured in Schneider’s medium (Lonza) supplemented with 10% (v/v) heat-inactivated fetal bovine serum (FBS) plus 1X antibiotic-antimycotic (Gibco), and propagated at 27°C. The day before transfection, 3x10^6 cells were seeded per well in a 6-well plate (Corning) with 3.0 ml complete growth medium and incubated overnight. On the day of transfection, 2 µg expression plasmid was mixed with the transfection reagents and the transfection complexes were added drop-wise onto the S2 cells. At 72 hours post-transfection, the cultured media were replaced with fresh S2 growth media supplemented with 6 µg ml^-1 puromycin (Bioshop). Subsequently, S2 cells were gradually adapted to FBS-free Insect-XPRESS growth media (Lonza) with 6 µg ml^-1 puromycin. Stably transfected cells were grown to 1x10^7 cells ml^-1 in Insect-XPRESS growth media using vented 2 L polycarbonate Erlenmeyer flasks (VWR) at 27°C. Protein expression was induced with 500 µM final concentration of sterile-filtered CuSO₄. Cultured media were harvested 6-days post-induction, clarified by centrifugation at 6750xg for 20 minutes, concentrated and buffer exchanged into Ni-NTA binding buffer (20 mM Tris-HCl [pH 8.0], 300 mM NaCl, 20 mM imidazole) using a Centramate tangential flow filtration system (Pall Corp.) All Izumo1 and Juno proteins were purified by Ni-NTA metal affinity chromatography. Eluted samples were buffer exchanged into TBS (10 mM Tris-HCl [pH 8.0], 150 mM NaCl) using a PD-10 desalting column (GE Life Sciences) and thrombin (EMD Millipore) digested at 22°C for 24 hours (1:2000 (w/w) enzyme to protein ratio). The cleaved protein samples were then buffer
exchanged to a low pH buffer (10 mM sodium acetate [pH 5.6], 150 mM NaCl) and enzymatically deglycosylated using 100 units Endoglycosidase H (New England Biolabs) per mg of Izumo1 or Juno at 22°C for 16 hours. To prepare the Izumo1-Juno protein complexes, deglycosylated Izumo1<sub>22-254</sub> and Juno<sub>20-228</sub> samples were mixed at a molar ratio of 1:1 and incubated at 22°C for 2 hours prior to size exclusion chromatography on a custom prep-grade Superdex-200 XK 16/70 column equilibrated with TBS. Peak fractions were pooled and protein concentrations were quantified by measuring A<sub>280</sub>.

6.4.2 Circular dichroism (CD) spectroscopy. CD spectra of human Izumo1<sub>22-268</sub> were acquired on a Jasco J-810 spectropolarimeter using a 1 mm quartz cuvette (Helma). CD measurements were conducted with 50 to 100 µM protein sample purified in 10 mM potassium phosphate [pH 7.5] and 150 mM NaCl buffer. Wavelength scans were recorded at 25°C between 190 nm and 250 nm and averaged over 5 accumulations. Data were converted to mean residue ellipticity and secondary structure content was estimated using the K2D algorithm within the DichroWeb analysis server<sup>412</sup>. Thermal denaturation assays were performed at a wavelength of 222 nm by increasing the temperature from 20°C to 99°C and monitoring the change in ellipticity as a function of temperature. The data were baseline corrected with buffer blank, normalized between 0 (folded) and 1 (unfolded) and fit to a nonlinear biphasic sigmoidal curve using GraphPad Prism (GraphPad Software, San Diego, CA USA).

6.4.3 Dynamic light scattering (DLS). Izumo1<sub>22-254</sub>, Juno<sub>20-228</sub>, and Izumo1<sub>22-254</sub>-Juno<sub>20-228</sub> complex samples were prepared in TBS with 2% (v/v) glycerol and concentrated to 0.5, 1.0, 1.5, 2.0, 2.5, 3.0, 4.0 and 5.0 mg ml<sup>-1</sup> prior to the DLS measurements. DLS experiments were performed at 25°C on a DynaPro Plate Reader II (Wyatt Technology). For each condition, 22 µl of sample was loaded in triplicate onto a black 384-well clear-bottom plate (Greiner). Data acquisition was recorded over 5 seconds with a total of 10 acquisitions for each concentration. The polydispersity and hydrodynamic radius (R<sub>H</sub>) of the molecules in solution was calculated using the Dynamics (v.7) software (Wyatt Technology).
6.4.4 Size Exclusion Chromatography-Multiangle Light Scattering (SEC–MALS). The oligomeric state of tag-removed, glycosylated Izumo122-268 was assessed by multiangle light scattering. Monomeric bovine serum albumin (BSA) standard (2 mg ml\(^{-1}\); \(M_w = 66,432 \text{ Da}\)) dissolved in PBS buffer (10 mM Phosphate [pH 7.4], 2.7 mM KCl, 137 mM NaCl) was used to calibrate the MALS detectors. Izumo122-268 was first purified on an analytical Superdex-75 10/300 GL size exclusion column equilibrated in PBS buffer to ensure proper monodispersity. Then, 600 µg of Izumo122-268 was applied onto a PBS-equilibrated Superdex-200 Increase 10/300 GL size exclusion column in-line with a Viscotek MALS detector (Malvern). The data were processed and weight-averaged molecular mass was calculated using the OMNISEC (v. 5.1) software package (Malvern).

6.4.5 Biolayer interferometry (BLI). The binding affinities of wild-type and mutant Izumo122-254 to Juno20-228 was measured by biolayer interferometry using a single-channel BLItz instrument (Pall FortéBio), based on protocols previously described\(^{451}\). Briefly, purified wild-type or mutant Juno20-228 in PBS buffer was biotinylated using the EZ link sulfo-NHS-LC-biotinylation kit (Thermo Pierce), according to the manufacturer’s instructions. Excess biotin reagent was removed by overnight dialysis in PBS. All streptavidin-coated (SA) biosensors were hydrated in BLI rehydration buffer (PBS, 0.5 mg ml\(^{-1}\) BSA and 0.01% (v/v) Tween-20) for 10 minutes. Biotinylated Juno20-228 (bait) was diluted in BLI kinetics buffer (PBS, 0.1 mg ml\(^{-1}\) BSA and 0.01% (v/v) Tween-20) to a final concentration of 20 µg/ml and immobilized onto a SA-biosensor for 90 seconds. Five different wild-type or mutant Izumo122-254 (analyte) concentrations were prepared in BLI kinetics buffer and association to Izumo122-254 was measured over 90 seconds at 20°C. Subsequently, the SA-biosensor was immersed into BLI kinetics buffer for 90 seconds to dissociate the analyte. All experiments were performed in triplicate. Two negative controls were performed: BSA and BLI kinetics buffer only against SA-biosensors loaded with biotinylated bait to detect non-specific binding. The data were analyzed and sensorgrams were step corrected, reference corrected and fit globally to a 1:1 binding model. The equilibrium dissociation constant (\(K_d\)), association (\(k_a\)) and dissociation (\(k_d\)) rate constants and its associated standard errors were calculated using the BLItz Pro data analysis (v. 1.1.0.16) software.
6.4.6 Surface plasmon resonance (SPR). Affinities and kinetics of wild-type Izumo1<sub>22-254</sub> binding to wild-type Juno<sub>20-228</sub> were assessed by SPR on a Biosensing Instruments BI-4000 system at 20°C using a CM-dextran sensor chip. Prior to immobilization, pH scouting between pH 4.5-6.5 was performed to identify the optimal pH for immobilization. Wild-type Juno<sub>20-228</sub> was immobilized using a coupling buffer containing 10 mM sodium acetate pH 5.0 onto one of two flow channels using the manufacturer’s standard amine-coupling protocol. Association of the wild-type Izumo1<sub>22-254</sub> analyte (0.75 µM, 0.5 µM, 0.375 µM, 0.25 µM, 0.188 µM, 0.125 µM, 0.0937 µM and 0 µM) was measured at a flow rate of 50 µl min<sup>−1</sup> for 90 seconds. The second flow cell, containing no bait, was injected with PBS buffer in a serial flow and used a reference. Subsequently, PBS buffer was injected at a flow rate of 50 µl min<sup>−1</sup> over 180 seconds to dissociate wild-type Izumo1<sub>22-254</sub>. The cells were regenerated between two analyte runs using the rapid injection protocol involving 8 cycles of 20 µl injections of 0.1 NaOH-acetate pH 9.0 followed by an equal volume of 1X PBS. Measurements were performed in triplicate. The resulting SPR sensorgrams were corrected with the reference and blank (0 µM analyte) curves, and fitted globally with a 1:1 Langmuir binding model using the BI-Data Analysis and BI-Kinetic Analysis SPR software.

6.4.7 Crystallization and X-ray data collection. Purified Izumo1<sub>22-254</sub>, Izumo1<sub>22-268</sub>, Juno<sub>20-228</sub>, and the Izumo1<sub>22-254</sub>-Juno<sub>20-228</sub> complex were concentrated to 10 mg ml<sup>−1</sup>. All crystallization trials were performed at 22°C by sitting drop vapour diffusion (300 nl protein and 300 nl mother liquor) in 96-well low profile Intelliplates (Art Robbins) using an Oryx8 protein crystallization robot (Douglas Instruments).

6.4.7.1 Izumo1<sub>22-254</sub> and Izumo1<sub>22-268</sub>. Initial sparse matrix screening of Izumo1<sub>22-254</sub> and Izumo1<sub>22-268</sub> constructs identified needle-shaped crystals in multiple conditions. Izumo1<sub>22-254</sub> and Izumo1<sub>22-268</sub> crystals were manually optimized in 48-well MRC Maxi crystallization plates using 2 µl sitting drops. Larger needle-shaped Izumo1<sub>22-254</sub> crystals appeared the next day and reached a maximum length of ~250 µm within 3-4 days in 0.07 M sodium acetate [pH 4.6], 5.6% (w/v) PEG 4000 and 30% (v/v) glycerol. Larger Izumo1<sub>22-268</sub> crystals were more difficult to obtain and required further optimization using random microseed matrix screening (rMMS) with the
Oryx\textsuperscript{452}. rMMS led to thicker needle crystals in 0.085 M HEPES sodium salt [pH 7.5], 8.5% (v/v) isopropanol, 17% (w/v) PEG 4000 and 15% (v/v) glycerol. These crystals reached the final length of ~200 µm within 4-5 days. All crystals were cryoprotected and flash-cooled in liquid nitrogen. Izumo\textsubscript{122-254} and Izumo\textsubscript{122-268} crystals diffracted to Bragg spacings of 3.1-Å and 2.9-Å, respectively, and datasets were remotely collected at the Canadian Light Source (CLS; Saskatoon, SK) 08ID-1 beamline (Figure 6.26).

6.4.7.2 Juno\textsubscript{20-228}. Rod-shaped Juno\textsubscript{20-228} crystals were grown in 0.02 M magnesium chloride, 0.1 M HEPES sodium salt [pH 7.5], 22% (w/v) Polyacrylic acid 5100. Crystals typically appeared after 3-4 days and reached full-size in 1 week. The mother liquor supplemented with increasing amounts of sucrose (up to 30% (w/v) in final solution) was used as a cryoprotectant before being rapidly cooled in liquid nitrogen. Juno crystals readily produced Bragg reflections better than 2.0-Å resolution on a Rigaku FR-E Superbright X-ray generator and Saturn A200 HD CCD detector (Rigaku Corp.), and a 1.8-Å resolution dataset was collected at the Structural Genomics Consortium X-ray diffraction facility (Toronto, ON, Canada) (Figure 6.27).

6.4.7.3 Izumo\textsubscript{122-254}-Juno\textsubscript{20-228} complex. Crystals of the protein complex were grown in sitting drops containing equal volumes (1 µl) of purified protein and crystallant (0.1 M MES [pH 6.5], 20% (w/v) PEG 4000 and 0.6 M NaCl). Crystals were observed after 3 days and matured to full-size within a week. Crystals were cryoprotected by sequential soaking in mother liquor with 5%, 10%, 20% and 30% (w/v) sucrose. Crystals were directly immersed into liquid nitrogen and screened at the CLS beamline 08ID-1. A dataset was collected from a single Izumo\textsubscript{122-254}-Juno\textsubscript{20-228} complex crystal diffracting to 2.4-Å resolution (Figure 6.26).

6.4.8 Structure determination and refinement. All diffraction data were integrated and reduced with the XDS program package\textsuperscript{425} and scaled using Aimless\textsuperscript{453} from the CCP4 program suite. Crystallographic data collection and final refinement statistics are presented in Table 6.2.
Figure 6.25. Izumo122-254 and Izumo122-254-Juno20-228. (A) Crystals of the unbound Izumo122-254 grown in 0.07 M sodium acetate (pH 4.6), 5.6% (w/v) PEG 4000 and 30% (v/v) glycerol. (B) A representative X-ray diffraction image of Izumo122-254 illustrates the anisotropic diffraction of the needle-shaped crystals. (C) Representative 3.1-Å resolution $\sigma_A$-weighted composite omit $2|F_o|-|F_c|$ electron density map of Izumo122-254 contoured at $1\sigma$ and superimposed with its final refined model. (D) Izumo122-254-Juno20-228 complex crystals were grown in 0.1 M MES (pH 6.5), 20% (w/v) PEG 4000 and 0.6 M NaCl. (E) X-ray diffraction image of the Izumo122-254-Juno20-228 complex. (F) Representative 2.4-Å resolution $\sigma_A$-weighted composite omit $2|F_o|-|F_c|$ electron density map of the Izumo122-254-Juno20-228 complex.
density map of Izumol$_{22-254}$-Juno$_{20-228}$ complex contoured at 1σ and superimposed with its final refined model.

**Figure 6.26. Juno$_{20-228}$ X-ray diffraction and twinning analysis.** (A) A representative example of an X-ray diffraction image of the merohedrally twinned Juno$_{20-228}$. The inset windows show
the medium-resolution and high-resolution parts of the image. (B) The self-rotation plot at $\kappa = 180^\circ$ reveals the presence of non-crystallographic symmetry. The plot was generated from the native data reduced to space group P4$_1$. The self-rotation function shows a clear two-fold axis at the centre (red arrow) and four equivalent peaks on the perimeter generated by non-crystallographic twin-related two-fold axes (blue arrow). The self-rotation function at $\kappa = 90^\circ$ shows the presence of the true crystallographic four-fold symmetry. (C) The Britton plot was generated after detwinning by plotting the number of negative intensities as a function of estimated twin fraction $\alpha^{454}$. An overestimation of $\alpha$ will result in increased number of negative intensities and the exact value of $\alpha$ can be extrapolated from this increase (black line). (D) Cumulative intensity distribution N(z) plots for Juno$_{20-228}$.$^{426}$ $I$ is the intensity of centric and acentric reflections and $Z = I/[I]$. (E) The cumulative distribution of $H$ determined using the twin operator $k, h, -l$ for merohedral twinning.$^{455}$ The Yeates $H$-plots indicate the twinning fraction of Juno$_{20-228}$ is larger than 0.4. Dashed lines indicate the theoretically determined twinning slopes for five different twinning fractions. (F) Twinning analysis using the $L$-test$^{426}$. The curve for the observed data (red) falls between the curves for theoretically perfectly twinned data (green) and theoretically untwinned data (blue). (G) Representative 1.8-Å resolution $\sigma_A$-weighted 2|Fo|-|Fc| electron density map of Juno$_{20-228}$ contoured at 1$\sigma$ and superimposed with its final refined model.

6.4.8.1 Juno$_{20-228}$. The structure of Juno$_{20-228}$ was determined by molecular replacement (MR) with Phaser$^{307}$ using the human folate receptor-\textalpha (PDB ID: 4LRH) as the search model. Initial characterization of the Juno$_{20-228}$ X-ray data using phenix.xtriage$^{428}$ and DETWIN$^{426}$ indicated translational pseudosymmetry$^{456}$ (TPS) and near-perfect merohedral twinning$^{457}$ with an estimated twin fraction of 0.45 (Figure 6.27). The twinning fraction was calculated from the cumulative distribution of $H$$^{455}$ and Britton plots$^{454}$, with the twin fractions related by the twin law $k, h, -l$. It was necessary to apply the twin law throughout the refinement to further refine the Juno$_{20-228}$ structure using phenix.refine$^{310}$.

6.4.8.2 Izumo1$_{22-254}$-Juno$_{20-228}$ complex. The initial phases for the Izumo1$_{22-254}$-Juno$_{20-228}$ complex were calculated via MR with Phaser$^{307}$, using the human folate receptor-\textalpha structure
(PDB ID: 4LRH) as an initial search model. One clear solution (Z=14.6) was identified. Strong electron density was observed for Juno20-228 and the 4HB domain of Izumo122-254. The polyalanine chain of the Izumo122-254 was initially traced by a combination of phenix.autobuild and manual building with Coot.

Validation of proper sequence registry was confirmed by locating the sulphur anomalous signals from methionine and cysteine residues. Multi-crystal sulphur anomalous data were collected on native Izumo122-254-Juno20-228 complex crystals. The X-ray beam was focused to 50 µm and the sulphur anomalous signal was measured at a wavelength of 1.7712-Å using a MarMosaic MX300 CCD detector (Rayonix). All crystals were rod-shaped and >400 µm in length, thus allowing us to translate along the rotation axis to expose a fresh undamaged part of the crystal. 360º of data with a rotation angle of 1.0º per frame were collected for each set before translating to a new part of the crystal. Each dataset was processed individually with anomalous signal using XDS. Twenty-four datasets with R_{meas}<10% were merged together using XSCALE and converted to CCP4 data format using XDSCONV, F2MTZ and CAD. The overall R_{merge} and anomalous multiplicity for the merged dataset to 2.8-Å resolution were 9.9% and 89.6, respectively. An anomalous difference Fourier electron density map was calculated using PHENIX and confirms the correct location for all 38 protein sulphur sites (Figure 6.28).

### 6.4.8.3 Izumo122-254 and Izumo122-268

The structures of Izumo122-254 and Izumo122-268 were determined by MR. An initial MR search using the refined Izumo122-254 structure from the Izumo122-254-Juno20-228 complex failed, likely due to conformational changes between the 4HB and Ig-like domains. A MR search was performed first using the Izumo1 Ig-like domain (residues 167-254) followed by a second MR search using the 4HB and hinge regions (residues 22-166). Clear solutions were identified for both sections in Izumo122-254 and Izumo122-268.

All structures were manually rebuilt using Coot and refined using phenix.refine. No non-crystallographic (NCS) symmetry restraints were employed except in the case of Juno20-228, where a 4-fold NCS was applied. All β-strands and α-helices were real-space refined with torsional secondary structural restraints using Coot. Torsion-angle simulated annealing
refinement, starting at 5000 K, with individual atomic displacement and Translation/Liberation/Screw (TLS) groups was carried out using Phenix. Due to the lower resolutions of the Izumo122-254 and Izumo122-268 data, these structures were refined with grouped B-factor refinement. Calculation of annealed $2|F_o|-|F_c|$ composite omit maps\textsuperscript{459} helped minimize model bias during rebuilding.

6.4.9 Validation and structure analysis. The stereochemical quality of the all refined models was validated using MolProbity\textsuperscript{429}, PROCHECK\textsuperscript{460} and Coot\textsuperscript{458}. No residues were identified in disallowed regions of the Ramachandran plot. Moreover, the R values, B-factors, and rmsd bond lengths and angles of all structural models are consistent with other deposited structures determined at similar resolutions, as validated by polygon.phenix\textsuperscript{428}. All structural representations were prepared using PyMOL (v. 1.7.4 Schrödinger, LLC.)
Figure 6.27. **Difference anomalous Fourier electron density map.** Overview of the anomalous difference Fourier electron density map for sulphur atoms in the asymmetric unit of the Izumo1-22-258-Juno20-228 complex. Twenty-four datasets collected at 1.7712 Å with $R_{\text{meas}} < 10\%$ were merged and scaled using XSCALE\textsuperscript{425}. The anomalous Fourier difference map was calculated using PHENIX\textsuperscript{428} and contoured at 3σ (blue mesh). Twenty-four strong electron density features were observed and this represent the entire S-atom substructure of Izumo1 and Juno. Note that the two sulphur atoms in a disulphide linkage display a single anomalous peak. Overall, the anomalous sulphur data confirm that the tracing and model building of Izumo1 and Juno were correct. The inset windows show magnified images of selected cysteine and methionine residues with corresponding anomalous signals.
6.4.10 Small-angle X-ray scattering (SAXS) data collection and reconstruction. SAXS experiments were performed by mail-in SAXS on beamline 12.3.1 (SIBYLS) at the Advanced Light Source (Berkeley, CA). Various protein concentrations of tag-cleaved, deglycosylated wild-type and mutant Izumo122-254, Juno20-228 and Izumo122-254-Juno20-228 complexes, along with matching buffer obtained from SEC (10 mM Tris-HCl [pH 8.0], 150 mM NaCl and 2% (v/v) glycerol), were loaded into a 96-well PCR plate (Axygen) and stored at 4°C prior to data collection. Samples were loaded into the SAXS sample cell using a Hamilton syringe robot. For the wild-type Izumo122-254, Juno20-228 and Izumo122-254-Juno20-228 complex, data were collected at a wavelength ($\lambda$) of 1.0-Å using a MarCCD 165 detector (Rayonix) positioned at a distance of 1.5 m, resulting in scattering vector $q$ of $0.01-\text{Å}^{-1} < q < 0.32-\text{Å}^{-1}$ (where $q = 4\pi\sin(\theta)/\lambda$ and $\theta$ is the scattering angle). Each data set was recorded at 283 K in a succession of three X-ray exposures of 0.5, 1, 2 and 5 seconds. For the mutant Izumo122-254-Juno20-228 complexes, data were collected at 1.0-Å wavelength using a Pilatus3X 2M detector (Dectris) (0.01-Å$^{-1} < q < 0.55$-Å$^{-1}$). Data for the mutant Izumo122-254-Juno20-228 complexes were recorded in a time slicing mode of 0.5 second exposures over 15 seconds (30 frames per sample). Data for buffer blanks were...
collected before each protein image and subsequently buffer subtracted. Sample radiation damage was assessed by overlaying short and long exposures and detecting for any shifts in the scattering curves using the program SCATTER\textsuperscript{450}. Concentration and aggregation effects were detected by comparing the lowest scattering angles for each of the protein samples. Fits to the Guinier region were made using autoRg. In order to maximize the signal-to-noise ratio, the SAXS scattering curve at the highest concentration that is free of interparticle interference was used for subsequent analysis. The characteristic real-space distance distribution function, $P(r)$, was determined from the scattering data using an indirect Fourier transformation and the maximum dimension, $D_{\text{max}}$\textsuperscript{461}. All \textit{ab initio} reconstructions of molecular envelopes from SAXS data were performed using the program DAMMIN\textsuperscript{447}. Multiple DAMMIN models were superimposed and averaged by the program DAMAVER\textsuperscript{448} to obtain a consensus averaged structure. Alignment of the SAXS reconstructions with the final refined crystal structures was performed using Chimera\textsuperscript{449}.

6.4.11 Deuterium exchange mass spectrometry (DXMS). Prior to performing deuterium exchange experiments, the optimal proteolysis conditions were established as previously described\textsuperscript{462,463} to maximize peptide sequence coverage of tag-cleaved, deglycosylated Izumo1\textsubscript{122-254} and Juno\textsubscript{20-228}. Briefly, 1 µl of diluted protein stock solution (2 mg ml\textsuperscript{-1} in 10mM Tris [pH 7.2], 150mM NaCl) was mixed with 5 µl of quench buffer (6.4 M GuHCl and 1.0 M TCEP in 0.8% (v/v) formic acid, 16.6% (v/v) glycerol). After incubating on ice for various times (2, 5, 10, 15 and 30 min), the quenched samples were mixed with 24 µl of dilution buffer (0.8% (v/v) formic acid, 16.6% (v/v) glycerol) and then subjected to proteolysis and LC-MS analysis. The Izumo\textsubscript{122-254}-Juno\textsubscript{20-228} complex was formed by mixing Izumo\textsubscript{122-254} and Juno\textsubscript{20-228} at 1:1.2 or 1.2:1 molar ratios and incubating the samples at 22°C for 2 hours. Deuterium exchange was initiated by mixing 3.5 µl of protein stock solution (Izumo\textsubscript{122-254}, Izumo\textsubscript{122-254}-Juno\textsubscript{20-228}, Juno\textsubscript{20-228} or Juno\textsubscript{20-228}-Izumo\textsubscript{122-254}) with 7 µl of D\textsubscript{2}O buffer (8.3 mM Tris [pH 7.2], 150 mM NaCl in D\textsubscript{2}O, pD\textsubscript{read} 7.2) and incubating at 0°C for 10, 100, 1000, 10,000 and 100,000 seconds. At indicated times, 2.1 µl of exchange samples were added to 10.5 µl quench solution to stop the D\textsubscript{2}O exchange reaction. After 5 minutes (Izumo\textsubscript{122-254} or Izumo\textsubscript{122-254}-Juno\textsubscript{20-228}) or 10 minutes (Juno\textsubscript{20-228} or Juno\textsubscript{20-228}-Izumo\textsubscript{122-254}) incubation on ice, quenched samples were diluted by addition of 48 µl of ice-cold dilution buffer, and then immediately frozen on dry ice and stored at
–80°C. The non-deuterated control samples and equilibrium-deuterated control samples were also prepared by mixing protein with H$_2$O buffer (8.3 mM Tris [pH 7.2], 150 mM NaCl in H$_2$O) and equilibrium-deuterated buffer (0.8% (v/v) formic acid in 99.9% D$_2$O)\(^{464}\). The frozen samples were then thawed at 5°C and passed over an immobilized pepsin column (16 µl bed volume) at a flow rate of 20 µl/min. The resulting peptides were collected on a C$_{18}$ trap for desalting and separated by a Magic AQ C$_{18}$ reverse phase column (Michrom BioResources) using a linear gradient of acetonitrile from 6.4% to 38.4% over 30 min. MS analysis was performed using the Orbitrap Elite Mass Spectrometer (ThermoFisher Scientific), with a capillary temperature of 200°C. Data were acquired in both data-dependent MS/MS mode and MS1 profile mode, and the data were analyzed by Proteome Discoverer software and DXMS Explorer\(^{465}\) (Sierra Analytics Inc., Modesto, CA).

### 6.4.12 Accession numbers

Atomic coordinates and structure factors have been deposited in the Protein Data Bank under accession codes: 5F4E (Izumo1$_{22-254}$-Jun0$_{20-228}$ complex); 5F4Q (Jun0$_{20-228}$); 5F4T (Izumo1$_{22-254}$); 5F4V (Izumo1$_{22-268}$).
6.5 Acknowledgements

This work was supported by a Canadian Institutes of Health Research (CIHR) Operating Grant (MOP-115066), NSERC Discovery Grant (RGPIN 435607-13), Ontario Early Researcher Award (ER-13-09-116), and a Canada Research Chair (Tier II) to J.E.L. S.L. is supported by grants from NIH (1U19AI117905, R01 GM020501 and R01 AI101436). Moreover, support for stipends were provided by University of Toronto and Ontario Graduate Scholarships to H.A., and a Natural Science and Engineering Council of Canada (NSERC) Undergraduate Summer Research Award to A.T.. The authors would like to thank Drs. Walid Houry and Trevor Moraes in the Department of Biochemistry (University of Toronto), and Christopher Spring at the Keenan Research Centre for Biomedical Science at St. Michael’s Hospital (Toronto, ON) for access to their CD, SEC-MALS and SPR systems, respectively. This work is based upon X-ray data collected at beamline 08ID-1 at the Canadian Light Source (CLS) and Structural Genomics Consortium (SGC), and SAXS data acquired at the Advanced Light Source (ALS) SIBYLS beamline 12.3.1. The CLS (Saskatoon, SK) is supported by NSERC, National Research Council of Canada, CIHR, the Province of Saskatchewan, Western Economic Diversification Canada, and the University of Saskatchewan. The SGC, headed by Drs. Aled Edwards and Cheryl Arrowsmith, is a not-for-profit organization funded by private-public partners. ALS (Berkeley, CA) is a national user facility operated by Lawrence Berkeley National Laboratory on behalf of the Department of Energy (Office of Basic Energy Sciences) through the Integrated Diffraction Analysis Technologies (IDAT) program (DE-AC02-05CH11231), supported by DOE Office of Biological and Environmental Research; additional support comes from the National Institute of Health project MINOS (R01GM105404). The authors would like to thank Farshad C. Azimi, Jonathan D. Cook, Aiping Dong (SGC) and Dr. Nguyen (Win) Ly (Biosensing Instruments) for technical support, and Drs. Erica Ollmann Saphire, Anthony S. Rocca and George Bikopoulos for critical reading of the manuscript.

6.6 Inter-Chapter Transition

Human surface proteins and receptors are responsible for a number of critical cellular functions, ranging from extracellular contact and recognition to transmission of intracellular signaling. Structural and functional analyses of these human proteins require the production of microgram
to milligram quantities of pure protein to pursue biochemical analysis. Atomic resolution structures of human sperm and egg surface proteins are quite rare given the fact that most of these surface molecules are cysteine-rich glycoproteins that are difficult to express from recombinant sources and are challenging targets for crystallization. In this chapter, we utilized Drosophila S2 insect cell expression platform to obtain milligram quantities of recombinant protein. Alternatively, mammalian cells such as adherent human embryonic kidney (HEK) 293T cell line can be used to produce recombinant cells in a native, mammalian environment. In chapter 7, we describe an alternative platform for the production of secreted proteins. Multiple protein constructs are transiently expressed in HEK 293T cell lines and screened on test-scale for high-level expression constructs. Protein production is easily scaled up for larger-scale optimization once initial constructs are identified. The test-scale protein expression platform was initially used to identify constructs for the human Izumo1-Juno project. We find this system provides several advantages over standard insect expression systems including incorporation of native mammalian glycans, facile switch from microscale screening to large-scale production, and importantly, more rapid turn-around of iterative construct design.
Chapter 7

A convenient and general expression platform for the production of secreted proteins from human cells

Halil Aydin1*, Farshad C. Azimi1*, Jonathan D. Cook1* and Jeffrey E. Lee1

1Department of Laboratory Medicine and Pathobiology, Faculty of Medicine, University ofToronto, Toronto, ON M5S 1A8, Canada

* These authors contributed equally to this work.

Published Manuscript
Journal of Visualized Experiments July 2012 (65): e4041
Copyright © JoVE 2012

- Author Contributions:

Halil Aydin: Optimized and performed small-scale protein expression experiments, assisted with cell propagation and performed immunoblotting.

Farshad C. Azimi: Performed large-scale protein expression experiments, purification and immunoblotting.

Jonathan D. Cook: Performed cloning, large-scale protein expression experiments, purification and immunoblotting.

Jeffrey E. Lee: Designed the project, performed cloning and supervised the research.
7.1 Summary

In the post-human genomics era, the availability of recombinant proteins in native conformations is crucial to structural, functional and therapeutic research and development. Here, we describe a test- and large-scale protein expression system in human embryonic kidney 293T cells that can be used to produce a variety of recombinant proteins.
7.2 Introduction

Recombinant protein expression in bacteria, typically *E. coli*, has been the most successful strategy for milligram quantity expression of proteins. However, prokaryotic hosts are often not as appropriate for expression of human, viral or eukaryotic proteins due to toxicity of the foreign macromolecule, differences in the protein folding machinery, or due to the lack of particular co- or post-translational modifications in bacteria. Expression systems based on yeast (*P. pastoris* or *S. cerevisiae*)\(^{466,467}\), baculovirus-infected insect (*S. frugiperda* or *T. ni*) cells\(^{468}\), and cell-free *in vitro* translation systems\(^{467,469}\) have been successfully used to produce mammalian proteins. Intuitively, the best match is to use a mammalian host to ensure the production of recombinant proteins that contain the proper post-translational modifications. A number of mammalian cell lines (*H*uman *E*mbryonic *K*idney (HEK) 293, *C*V-1 cells in *O*rigin carrying the *S*V40 large T-antigen (COS), *C*hinese *H*amster *O*vary (CHO), and others) have been successfully utilized to overexpress milligram quantities of a number of human proteins\(^{48,470–473}\). However, the advantages of using mammalian cells are often countered by higher costs, requirement of specialized laboratory equipment, lower protein yields, and lengthy times to develop stable expression cell lines. Increasing yield and producing proteins faster, while keeping costs low, are major factors for many academic and commercial laboratories.

Here, we describe a time- and cost-efficient, two-part procedure for the expression of secreted human proteins from adherent HEK 293T cells. This system is capable of producing microgram to milligram quantities of functional protein for structural, biophysical and biochemical studies. The first part, multiple constructs of the gene of interest are produced in parallel and transiently transfected into adherent HEK 293T cells in small scale. The detection and analysis of recombinant protein secreted into the cell culture medium is performed by western blot analysis using commercially available antibodies directed against a vector-encoded protein purification tag. Subsequently, suitable constructs for large-scale protein production are transiently transfected using polyethyleneimine (PEI) in 10-layer cell factories. Proteins secreted into litre-volumes of conditioned medium are concentrated into manageable amounts using tangential flow filtration, followed by purification by anti-HA affinity chromatography. The utility of this platform is proven by its ability to express milligram quantities of cytokines, cytokine receptors,
cell surface receptors, intrinsic restriction factors, and viral glycoproteins. This method was also successfully used in the structural determination of the trimeric ebolavirus glycoprotein\textsuperscript{48,474}.

In conclusion, this platform offers ease of use, speed and scalability while maximizing protein quality and functionality. Moreover, no additional equipment, other than a standard humidified CO\textsubscript{2} incubator, is required. This procedure may be rapidly expanded to systems of greater complexity, such as co-expression of protein complexes, antigens and antibodies, production of virus-like particles for vaccines, or production of adenoviruses or lentiviruses for transduction of difficult cell lines.
7.3 Protocol

7.3.1 Preparation work - constructs and cell cultures. Before starting the protocol, the gene of interest should be codon-optimized for expression in mammalian cells, and cloned into an appropriate expression vector using standard molecular biology techniques. In order to ensure the highest chance for successful expression, multiple variants of the gene of interest should be generated. Many mammalian expression vectors are available commercially and have various purification tags (polyhistidine, hemagglutinin, streptavidin, HALO-Tag, glutathione S-transferase, among others). We prefer to use the pDISPLAY vector, which encodes for a strong human cytomegalovirus promoter, an Igκ secretion signal, hemagglutinin purification tag, and has a C-terminal transmembrane anchor to target the protein through the secretory pathway for display on the plasma membrane. We usually insert a stop codon in front of the vector-encoded transmembrane anchor to allow the protein to be secreted into the conditioned medium.

Human embryonic kidney (HEK) 293T cells are widely available, and easily cultured and transfected. HEK 293T are routinely used for expression of mammalian proteins, but are considered biohazardous and should be handled at biosafety level 2. Please wear proper personal protective clothing; work should be performed in an approved biosafety cabinet using aseptic technique. All waste and surfaces should be disinfected according to institutional and governmental guidelines. It is recommended that cells be tested for mycoplasma contamination prior to use. Cells may be treated with ciprofloxacin (10 µg/ml) for ten days to eradicate any source of mycoplasma spp. contamination. General protocols to propagate HEK 293T cells are presented separately. Additional considerations for test- and large-scale protein expression are reviewed in\textsuperscript{475–479}.

7.3.2 Small-scale test expression. Once constructs have been designed and generated, small-scale test transfections may be performed using HEK 293T cells; a schematic summarizing the process is presented below (Figure 7.1)
7.3.2.1. Use T75 cm² or T225 cm² cell culture flasks (depending on the number of test expressions to be performed) to grow the HEK 293T cells and split the cells every 2-3 days when cells are 100% confluent.

7.3.2.2. Seed 2.5 x 10⁵ HEK 293T cells per well in a 6-well plate and add 2 ml DMEM with 1X pen/strep and 10% (v/v) FBS, swirl the plate gently to ensure even cell dispersal in each well, and incubate overnight at 37ºC in a 5% CO₂ humidified chamber.

7.3.2.3. When HEK 293T cells reach 40% confluency, discard the media and add fresh 2 ml DMEM with 1X pen/strep and 10% (v/v) FBS to the wells. Perform transfection assays.

7.3.2.4. Aliquot 90 µl serum-free DMEM into a sterile 1.5 ml microcentrifuge tube. Pipette 3 µl GeneJuice into the serum-free DMEM and gently mix the tube (finger vortex). Incubate for 5 minutes at room temperature.

7.3.2.5. Add 1 µg of MiniPrep purified plasmid DNA (DNA stock= 100 ng/µl) into DMEM-GeneJuice mixture, finger vortex, and incubate for 15 minutes at room temperature.

7.3.2.6. Pipette the transfection mixture dropwise onto HEK 293T cells, and swirl the 6-well plate gently to allow even distribution of the transfection mixture. Incubate the 6-well plate at 37ºC in a 5% CO₂ humidified chamber.

7.3.2.7. Add 1 ml fresh DMEM with 1X pen/strep and 10% (v/v) FBS to each well 24 hours post-transfection and incubate for another 48 hours (total 72 hours).

7.3.2.8. Harvest 1 ml of supernatant from each well at three days post-transfection and microcentrifuge the samples at 16,000g for 10 minutes at room temperature. Samples can be stored at 4ºC. The length of storage at 4ºC is protein dependent.
7.3.3 Large-scale test expression and purification. Once a construct has been identified, milligram quantity expression of recombinant protein is achieved by PEI transfection of adherent HEK 293T cells using 10-layer cell factories (6360 cm$^2$ surface area) (Figure 7.2). For more exploratory studies, smaller cell factories or T-flasks (Table 7.1) can be used.

7.3.3.1. Purify 1 mg of DNA for transfection using a MaxiPrep plasmid purification kit. A 500 ml overnight culture of XL-1 Blue cells should produce at least 1 mg of pure DNA. Check purity of DNA by measuring $A_{260}/A_{280}$ ratio; should be above 1.8.

7.3.3.2. Scale up HEK 293T cells to $2.0 \times 10^8$ cells. Each T225 cm$^2$ flask grown to 100% confluency contains and average of ~$2.25 \times 10^7$ cells.

7.3.3.3. Add 1.2 L DMEM with 5% (v/v) FBS to a 10-layer cell factory. Add $2.0 \times 10^8$ HEK 293T cells to the cell factory and distribute cells evenly to all layers of the vessel. It is very difficult to visualize the confluency of cells in the cell factory. As an alternative, set up a T75 cm$^2$ flask with an appropriate number of cells, using the same cell number to surface area ratio as performed with the 10-layer vessel. Monitor this flask for growth rates. See the associated video.
for instruction on handling the cell factory. Incubate overnight at 37°C with 5% CO₂ to allow for cell attachment and growth.

7.3.3.4. Perform large-scale transfection when adherent HEK 293T cells are 70% confluent. Prepare the PEI-DNA transfection mixture (3:1 w/w PEI to DNA ratio) in a biosafety cabinet using a sterile T75 cm² flask. Mix 0.84 mg of DNA with 84 ml of sterile 1X PBS, then add 2.5 ml of PEI (2.5 mg total PEI). Incubate at room temperature for 15 minutes. Solution should become cloudy.

7.3.3.5. Pour the PEI-DNA transfection mixture slowly into the cell factory and thoroughly distribute over all layers of the vessel. Optional: for increased expression yields, add valproic acid (4 mM final concentration). Incubate at 37°C with 5% CO₂ for four days.

7.3.3.6. Harvest supernatant four days post-transfection. Centrifuge the conditioned media at 6000g for 30 minutes at 4°C. Further filter the supernatant using a 0.22 µm Stericup vacuum filter apparatus. The 10-layer cell factory can be reused, see Box 3 for cleaning instructions. It is key that cleaning is initiated immediately after supernatant harvest; do not let the cells dry onto the vessel surface.

7.3.3.7. Concentrate the supernatant to 75 ml using the Centramate tangential flow filtration system.

7.3.3.8. Add 500 ml of PBS and re-concentrate to 75 ml. Repeat three additional times to completely buffer exchange the sample.

7.3.3.9. Equilibrate a 1 ml anti-HA affinity column with 1X PBS and apply concentrated sample by gravity flow at a rate <1 ml/min.

7.3.3.10. Wash the column with 30 ml of 1X PBS-Tween20.

7.3.3.11. Dissolve HA peptide in 1X PBS (1.0 mg/ml) and incubate at 37°C.

7.3.3.12. Apply 1 ml of the HA peptide to the anti-HA column and allow the peptide to flow into the resin. Collect the flow through. Stop the flow when the peptide solution reaches the bed height.
7.3.3.13. Incubate the entire anti-HA column at 37ºC for 15 minutes.

7.3.3.14. Repeat Step 12 two additional times.

7.3.3.15. Apply 1 ml of 1X PBS to the anti-HA column and flow into the resin until it reaches the bed height. Collect the flow through.

7.3.3.16. Regenerate the anti-HA column with 10 ml 0.1 M glycine pH 2.2. Wash with 10 ml PBS and store affinity column at 4ºC in PBS with 0.02% (w/v) NaN₃.

7.3.3.17. Perform SDS-PAGE analysis and pool the fractions accordingly. Note: The HA peptide will interfere with protein concentration measurements by A₂₈₀ or Bradford. To estimate the amount of protein present, load 5, 10, 15, 25 µg of BSA onto the SDS-PAGE gel as a standard and compare the band intensities. Steps 9-17 can be repeated to capture additional protein from the conditioned medium.

![10-layer CellSTACK](image)

**Figure 7.2. Corning 10-layer CellSTACK for large-scale protein expression.** Each layer contains 636 cm² surface area for cell attachment. A standard laboratory CO₂ incubator (6.0 cu. ft.) will comfortably hold four 10-layer cell factories.
Table 6. Comparison of cell culture vessels used for protein expression.

<table>
<thead>
<tr>
<th>Vessel</th>
<th>Surface area</th>
</tr>
</thead>
<tbody>
<tr>
<td>6-well plate</td>
<td>9.5 cm² (each well)</td>
</tr>
<tr>
<td>100 mm dishes</td>
<td>55 cm²</td>
</tr>
<tr>
<td>245 mm dishes</td>
<td>500 cm²</td>
</tr>
<tr>
<td>T75 cm² flask</td>
<td>75 cm²</td>
</tr>
<tr>
<td>T175 cm² flask</td>
<td>175 cm²</td>
</tr>
<tr>
<td>T225 cm² flask</td>
<td>225 cm²</td>
</tr>
<tr>
<td>Roller bottle- regular</td>
<td>850 cm²</td>
</tr>
<tr>
<td>Roller bottle- expanded surface</td>
<td>1700 cm²</td>
</tr>
<tr>
<td>1-layer CellSTACK</td>
<td>636 cm²</td>
</tr>
<tr>
<td>2-layer CellSTACK</td>
<td>1272 cm²</td>
</tr>
<tr>
<td>5-layer CellSTACK</td>
<td>3180 cm²</td>
</tr>
<tr>
<td><strong>10-layer CellSTACK</strong></td>
<td><strong>6360 cm²</strong></td>
</tr>
<tr>
<td>40-layer CellSTACK</td>
<td>25,440 cm²</td>
</tr>
</tbody>
</table>
7.4 Discussion

In this chapter, we describe and demonstrate a convenient expression platform for milligram-quantity production of human proteins that can subsequently be used for structural and functional studies. The screening of human protein constructs using HEK 293T cells in 6-well plates is efficient and effective in identifying constructs amenable to larger scale production. Commercial expression vectors can be transfected efficiently in HEK 293T cells using a variety of transfection reagents, such as GeneJuice, FuGene HD or PEI. We recommend the use of a commercial transfection reagent, such as GeneJuice or FuGene HD, for test expressions, as these reagents are more effective for poorer expressing proteins (Figure 7.3). Constructs selected for larger scale expression should be characterized by a single, strong intensity band, corresponding to the proper molecular weight on the western blot (Figure 7.3). Glycoproteins may migrate as a broader band due to heterogeneity in glycosylation. We have shown that a variety of macromolecules, ranging from viral glycoproteins, cytokines, cytokine receptors, and other surface proteins, can be expressed and purified to yield millgram quantities of protein using this general expression platform (Figure 7.4).

The 10-layer cell factories are an effective vessel for the production of milligram quantities of protein. A major advantage of using the cell factory over other traditional vessels, such as roller bottles, shake flasks or spinner flasks, is that they do not require the purchase of any additional laboratory equipment. A standard CO2 incubator (~6.0 cu. ft.) will easily accommodate four 10-layer cell factories (Figure 7.2). In addition, these vessels require less labor and space than dishes, flasks or roller bottles to produce cells and proteins; one 10-layer cell factory is equivalent to the use of 7.5 regular roller bottles (Table 7.1). More importantly, the HEK 293T cells adapt well in the vessels and are highly amenable to transient transfection by PEI, thus providing a low-cost and effective option to commercial transfection reagents. While cell factories are designed to be disposable cell culture plasticware, we have been able to efficiently clean and re-use these vessels multiple times without contamination issues, thus significantly reducing its cost per use. Transient protein expression using a 10-layer cell factory is capable of producing ~1-10 mg of purified protein (Figure 7.4), depending on the protein and its modifications. It is possible to go from cloning and small-scale expression to completion of
large-scale expression and purification in about 3-4 weeks. In summary, this is a rapid, general platform that is highly useful in the production of secreted human or viral surface proteins. Moreover, this platform can also be used to express antibodies, virus-like particles, and adenoviruses and lentiviruses for gene transfer. We have listed below common problems and potential solutions.

**Figure 7.3. Small-scale expression of various secreted proteins.** We performed a series of small-scale test expressions using common transfection reagents: GeneJuice, FuGene HD and PEI. (A) Western blot screening of selected human cellular proteins (tetherin), receptors (IL-2Rβ subunit) and cytokines (IL-2). Tetherin is a human membrane glycoprotein that restricts the release of nascent HIV-1 virions. The extracellular domain of tetherin exists as a glycosylated, disulfide-linked dimer of ~36 kDa. Under reducing conditions, as shown here, tetherin migrates as a monomer with an apparent molecular weight of ~22 kDa. Interleukin-2 (IL-2) is a cytokine (~17 kDa) involved in lymphocyte proliferation. It interacts with the IL-2 receptor complex, of which IL-2Rβ subunit (~26 kDa) is a component. CD21 is a membrane protein involved in the activation and maturation B-cells by the complement system, and is also a receptor for the Epstein-Barr virus. The glycosylated extracellular domain of CD21 migrates as a monomer with an apparent molecular weight of ~20 kDa. (B) Western blot screening of selected surface viral glycoproteins (XMRV Env and Ebola virus GP). XMRV and ebolavirus glycoproteins...
(transmembrane anchor deleted) exist at the viral membrane as trimeric spikes and are involved in host cell attachment and fusion. The ectodomain of XMRV Env and EBOV GP are heavily decorated with N-linked glycans and migrate at apparent molecular weights of 70 kDa and 75 kDa, respectively.

**Figure 7.4. Purified human cellular proteins from large-scale HEK293T cultures.** All proteins were expressed using a 10-layer cell factory, and concentrated and purified by anti-HA chromatography. As shown by Coomassie-stained SDS-PAGE analysis, the extracellular domains of the interleukin-2 receptor (IL-2R) α and γ subunits migrate at molecular weights of 40 kDa and 46 kDa, respectively. The extracellular domain of tetherin migrates as a dimer, under non-reducing conditions, with an apparent molecular weight of 36 kDa. Note that there is some BSA contamination that appears at an apparent molecular weight of 60 kDa. In addition, the heterogeneity of the N-linked glycans present on tetherin, IL-2Rα and IL-2Rγ causes band broadening on the SDS-PAGE gel. These complex-type N-linked glycans can be removed using peptide: N-glycosidase F.

7.4.1 No or very low protein expression.
7.4.1.1. Poor health of the cells- Stain the cells using trypan blue for viability and test for mycoplasma contamination. If cells have a high passage number (>25 passages) and are growing slowly, start fresh with new HEK 293T cells. Transfection efficiency and protein production are affected by the age of the cells in culture.

7.4.1.2. Poor transient transfection- Cells should be transfected at a confluency level of 40%. The ratios of DNA to transfection reagent should be optimized.

7.4.1.3. Protein is unstable or not folded- Check the cell pellet for insoluble protein. Test the expression of other constructs or homologous protein.

7.4.2 No protein is eluted from column.
7.4.2.1. Protein retained on the column- Elute the protein of interest with 0.1 M glycine pH 2.2 and analyze by western blot. If protein is retained, use a higher concentration of synthetic HA peptide for elution or incubate the column at 37°C for longer periods.

7.4.2.2. anti-HA column is damaged- While the column, if cleaned and stored properly, can be reused multiple times, the efficiency of the column will decrease over time. Use a new anti-HA column.

7.4.2.3. Protein degradation- Add protease inhibitors and keep protein on ice during the entire purification process.

7.4.2.4. Protein has precipitated or aggregated on column- If the protein is unstable, try different constructs or buffer conditions (ie. pH, salts, additives, etc.) to improve protein stability.

7.4.2.5. Protein failed to bind to column- The HA-tag may be buried or inaccessible. Check to see that the protein is not in the flow through or wash.

7.4.3 Serum albumin contamination.
Bovine serum albumin (BSA) tends to bind non-specifically to the anti-HA resin and may contaminate the HA peptide elutions. In Fig. 4, BSA migrates as a compact band at an apparent molecular weight of ~60-65 kDa. In this case, increase the Tween-20 concentration in the 1X PBS-Tween20 wash to 0.2% (v/v) and increase the volume of the wash. Alternatively, BSA contamination may be removed using size exclusion or ion exchange chromatography.
7.5 Materials and Methods

7.5.1 General protocols for cell propagation.

7.5.1.1. Grow HEK 293T cells in Dulbecco’s Modified Eagle’s Medium (DMEM) supplemented with 10% (v/v) fetal bovine serum (FBS), 1X pen/strep at 37°C in a 5% CO₂-humidified atmosphere.

7.5.1.2. Observe cells under an inverted microscope. When cells are at 100% confluency, remove and discard culture medium.

7.5.1.3. Rinse cells with 5 ml of sterile 1X PBS to remove traces of serum. Discard the PBS wash.

7.5.1.4. Add 2 ml of 0.05% (w/v) trypsin-EDTA solution to a T225 cm² flask (or 1 ml of 0.05% (w/v) trypsin-EDTA for T75 cm² flask) and incubate at room temperature until cells detach from surface. It is also possible to use sterile 1X PBS-EDTA to detach HEK 293T cells.

7.5.1.5. Add 13 ml of DMEM with 10% (v/v) FBS to a T225 cm² flask (or 9 ml of DMEM with 10% (v/v) FBS for T75 cm² flask) to inhibit the trypsin reaction.

7.5.1.6. Split cells 1:5. For a T225 cm² flask, add 3 ml of cell suspension to 27 ml of fresh DMEM with 1X pen/strep and 10% (v/v) FBS in a new culture vessel. For a T75 cm² flask, add 2 ml of cell suspension to 8 ml of fresh growth media. Incubate cultures at 37°C in a 5% CO₂-humidified atmosphere. Cells should grow to 100% confluency within two days.

7.5.2 Western immunoblotting.

7.5.2.1. Add 10 µl SDS-PAGE reducing sample buffer to 30 µl cell culture supernatant. Load samples and prestained molecular weight markers onto polyacrylamide gels, and electrophorese using 1X SDS-PAGE running buffer at 175 V for 1 hour or until molecular weight markers are well resolved.

7.5.2.2. Soak the PVDF Immobilon-P membrane in 100% methanol for 1 minute to activate.
7.5.2.3. Assemble western blot apparatus. Ensure the PVDF membrane faces the positive electrode and keep all components wet with 1X transfer buffer. Avoid bubbles between the polyacrylamide gel and the membrane.

7.5.2.4. Completely fill the electrophoresis chamber with 1X transfer buffer and transfer for 1 hour at 100 V.

7.5.2.5. Block the membrane with 5% (w/v) skim milk in 1X PBS-Tween20 for 1 hour at room temperature, or overnight at 4°C.

7.5.2.6. Incubate with monoclonal primary antibody (ie. 1:1000 dilution anti-HA mAb or another appropriate antibody) dissolved in 5% (w/v) skim milk in 1X PBS-Tween20 for 1 hour at room temperature or overnight at 4°C.

7.5.2.7. Wash the membranes in 1X PBS-Tween20 for 10 minutes. Repeat two extra times.

7.5.2.8. Incubate with a monoclonal secondary antibody conjugated with alkaline phosphatase (1:1000 dilution in 5% (w/v) skim milk in 1X PBS-Tween20) for 1 hour at room temperature, or overnight at 4°C.

7.5.2.9. Wash the membranes in 1X PBS-Tween20 for 10 minutes. Repeat two extra times.

7.5.2.10. Place the membrane into a small container, add 5 ml alkaline phosphatase substrate (BCIP/NBT) solution. Colour development should occur within 1-5 minutes. Once desired band intensity is reached, wash with deionized water and air dry. Colour may fade over time; electronically scan the membranes once dry.

7.5.3 Cleaning and recycling of cell culture vessels. While cell factories are designed to be single use, these vessels can be recycled for additional large-scale transfections using the following cleaning protocol:

7.5.3.1. Immediately after decanting the supernatant from the 10-layer cell factory, add 20% (v/v) bleach and shake vigorously to detach cells. Incubate at room temperature for three hours.
7.5.3.2. Empty the vessel and add fresh 20% (v/v) bleach and incubate at room temperature overnight.

7.5.3.3. Empty the vessel and wash with 1.5 L of deionized water. Repeat three times.

7.5.3.4. Empty the 10-layer cell factory and add 0.5 L of sterile 1X PBS supplemented with 10X Antibiotic/Antimycotic solution. Store the vessel at room temperature and replace venting caps for filling ports (standard 33 mm threaded caps) prior to next usage. Note: all layers should be completely clear after cleaning, if not, do not use and dispose the cell factory according to institutional guidelines.

7.5.4 List of Reagent Recipes.

100X Ciprofloxacine For 10 ml solution, add 10 ml deionized water to 10 mg ciprofloxacine. Add 10 ml 6N HCl to completely dissolve the ciprofloxacine.

PEI (1 mg/mL) For a 100 ml solution, dissolve 100 mg of 25 kDa linear PEI in deionized water and heat to 80°C. Cool solution to room temperature, adjust pH to 7.2, 0.22 µm filter sterilize, aliquot and freeze at -20°C for long-term storage.

1X PBS For 1 L aqueous solution: 8.0 g NaCl, 0.2 g KCl, 1.4 g Na₂HPO₄ (anhydrous), 0.24 g KH₂PO₄. Adjust pH of solution to 7.4 and fill to 1.0 L

1X PBS-Tween-20 For 1 L aqueous solution: 8.0 g NaCl, 0.2 g KCl, 1.4 g Na₂HPO₄ (anhydrous), 0.24 g KH₂PO₄, 1 ml Tween-20. Adjust pH of solution to 7.4 and fill to 1.0 L

1X transfer buffer For 1 L aqueous solution: 3.0 g Tris base, 14.4 g glycine, 150 ml methanol

1X SDS-PAGE running buffer For 1 L aqueous solution: 3.0 g Tris base, 14.4 g glycine, 1.0 g SDS

SDS-PAGE reducing sample buffer For 10 ml solution: 0.6 g SDS, 3 ml glycerol, 1.8 ml 1.0 Tris-HCl pH 6.8, 1 mg bromophenol blue, 5% (v/v) 2-mercaptoethanol
### 7.5.5 List of Specific Reagents and Equipment.

<table>
<thead>
<tr>
<th>Name of reagent</th>
<th>Company</th>
<th>Catalogue number</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alkaline phosphatase (BCIP/NBT) liquid substrate solution</td>
<td>Sigma</td>
<td>B6404</td>
</tr>
<tr>
<td>Antibiotic/Antimycotic, 100X</td>
<td>Invitrogen</td>
<td>15240062</td>
</tr>
<tr>
<td>Anti-HA affinity matrix, clone 3F10</td>
<td>Roche</td>
<td>1815016</td>
</tr>
<tr>
<td>Anti-HA murine mAb, clone 16B12</td>
<td>Covance</td>
<td>MMS-101P</td>
</tr>
<tr>
<td>Cell culture flask, T75 cm$^2$ tissue culture treated</td>
<td>Corning</td>
<td>430641</td>
</tr>
<tr>
<td>Cell culture flask, T225 cm$^2$ tissue culture treated</td>
<td>Corning</td>
<td>431082</td>
</tr>
<tr>
<td>Cell culture plates, 6-well tissue culture treated</td>
<td>Corning</td>
<td>3516</td>
</tr>
<tr>
<td>Cell factory, 10-layer CellSTACK</td>
<td>Corning</td>
<td>3312</td>
</tr>
<tr>
<td>Centramate-Omega 5K Cassette</td>
<td>Pall</td>
<td>OS005C12</td>
</tr>
<tr>
<td>Centramate-Omega 30K Cassette</td>
<td>Pall</td>
<td>OS030C12</td>
</tr>
<tr>
<td>Chromatography glass column, 1.0x10 cm</td>
<td>Kontes</td>
<td>4204001010</td>
</tr>
<tr>
<td>Ciprofloxacin</td>
<td>Sigma</td>
<td>17850</td>
</tr>
<tr>
<td>CO$_2$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dulbecco’s modified Eagle’s media (DMEM)</td>
<td>Sigma</td>
<td>D5796</td>
</tr>
<tr>
<td>Fetal bovine serum (FBS), heat inactivated</td>
<td>Invitrogen</td>
<td>12484-028</td>
</tr>
<tr>
<td>FuGENE HD transfection reagent</td>
<td>Promega</td>
<td>4709691001</td>
</tr>
<tr>
<td>GeneJuice transfection reagent</td>
<td>EMD/Merck</td>
<td>70967-6</td>
</tr>
<tr>
<td>Glycine</td>
<td>Sigma</td>
<td>G8898</td>
</tr>
<tr>
<td>Goat anti-mouse IgG F(ab’)$_2$ alkaline phosphatase-conjugated antibody</td>
<td>Thermo Scientific</td>
<td>31324</td>
</tr>
<tr>
<td>Hemagglutinin (HA) peptide, 100 mg (sequence: YPYDVPDYA; 95% purity)</td>
<td>Genscript</td>
<td>custom synthesis</td>
</tr>
<tr>
<td>HEK 293T cells</td>
<td>ATCC</td>
<td>CRL-11268</td>
</tr>
<tr>
<td>Item</td>
<td>Company</td>
<td>Code</td>
</tr>
<tr>
<td>----------------------------------------------------------------------</td>
<td>---------------</td>
<td>-----------</td>
</tr>
<tr>
<td>Household bleach</td>
<td>various brands are available</td>
<td></td>
</tr>
<tr>
<td>(4% w/v sodium hypochlorite)</td>
<td></td>
<td>IPVH07850</td>
</tr>
<tr>
<td>Immobilon-P PVDF membrane</td>
<td>Millipore</td>
<td>IPVH07850</td>
</tr>
<tr>
<td>MiniPrep plasmid purification kit, PureLink Quick</td>
<td>Invitrogen</td>
<td>K2100-11</td>
</tr>
<tr>
<td>MaxiPrep plasmid purification kit, PureLink</td>
<td>Invitrogen</td>
<td>K2100-07</td>
</tr>
<tr>
<td>HiPure NaN₃</td>
<td>Sigma</td>
<td>S8032</td>
</tr>
<tr>
<td>pDISPLAY expression vector</td>
<td>Invitrogen</td>
<td>V660-20</td>
</tr>
<tr>
<td>Penicillin/streptomycin (pen/strep), 100X</td>
<td>Invitrogen</td>
<td>15140-122</td>
</tr>
<tr>
<td>Phosphate-buffered saline (PBS), sterile 1X</td>
<td>Sigma</td>
<td>D8537</td>
</tr>
<tr>
<td>Polyethyleneimine (PEI), linear 25 kDa</td>
<td>Polyscience</td>
<td>23966</td>
</tr>
<tr>
<td>Skim milk dry powder</td>
<td>Carnation</td>
<td></td>
</tr>
<tr>
<td>Stericup-GP PES vacuum filtration unit, 0.22 mm, 500 ml capacity</td>
<td>Millipore</td>
<td>SCGPU05RE</td>
</tr>
<tr>
<td>Trypan blue</td>
<td>Invitrogen</td>
<td>15250061</td>
</tr>
<tr>
<td>Trypsin-EDTA, 0.05% (w/v)</td>
<td>Invitrogen</td>
<td>25300-054</td>
</tr>
<tr>
<td>Tween-20</td>
<td>Sigma</td>
<td>P7949</td>
</tr>
<tr>
<td>Valproic acid</td>
<td>Sigma</td>
<td>P4543</td>
</tr>
</tbody>
</table>

**Name of equipment**

<table>
<thead>
<tr>
<th>Name of equipment</th>
<th>Company</th>
</tr>
</thead>
<tbody>
<tr>
<td>Centramate tangential flow system</td>
<td>Pall</td>
</tr>
<tr>
<td>CO₂ humidified incubator, standard 6.0 cu. ft.</td>
<td>various brands are available</td>
</tr>
<tr>
<td>Electrophoresis and transfer unit</td>
<td>various brands are available</td>
</tr>
<tr>
<td>Incubator, 37°C</td>
<td>various brands are available</td>
</tr>
</tbody>
</table>
7.6 Acknowledgements

This work was supported by an Ontario HIV Treatment Network Research Operating Grant (ROG-G645) and Canadian Institutes of Health Research New Investigator Award (MSH-113554) to JEL, and University of Toronto Fellowships to HA, FCA, and JDC. The authors would like to thank Marnie Fusco, Dafna Abelson and Dr. Erica Ollmann Saphire at The Scripps Research Institute (La Jolla, CA) for providing cells, Ebola virus GP expression vector and general advice.
Chapter 8
General Discussion and Conclusions

Portions of this chapter were modified from the following publications:


8.1 Insights into the general mechanisms of viral fusion

While there is significant diversity among viral fusion proteins in terms of architecture, size and physiological properties, the mechanisms of viral fusion are generally conserved. All class I viral glycoproteins utilize a similar mechanism of fusion, in which structural rearrangements of two highly conserved heptad repeats juxtapose the viral and host cell membranes to form the fusion pore. The conformational changes necessary for membrane fusion requires the formation of an energetically stable six-helix bundle structure in the postfusion state. The transition from the higher energy metastable prefusion to the lower energy six-helix bundle postfusion conformation provides the energetics for fusion. In chapters 2, 3, and 4 the crystal structures of class I viral fusion proteins allowed us to evaluate and compare the pH-independent, two-step, and pH-dependent nature of viral fusion processes to identify general trends and features required for catalyzing the merger of viral and host cell membranes (Figure 8.1).

8.1.1 Fusion proteins are most stable at the pH of the site of fusion. In general, the stability of the fusion proteins tends to mimic the environment where they fuse. For example, HTLV-1 fuses at the host plasma membrane at neutral pH. The HTLV-1 gp21 fusion subunit is most stable at pH values above 7.0 (61.0°C at pH 5.0 vs. >99.0°C at pH 7.5)418. EBOV fusion occurs at the endolysosome, and its GP2 subunit is most stable at pH values below 5.5 (86.8°C at pH 5.3 vs. 49.8°C at pH 6.1)290. The observation that the stability of the fusion subunit mimics the environment where they fuse also holds true for avian sarcoma leukosis virus (ASLV). ASLV is a retrovirus that undergoes a unique two-step entry mechanism that involves first receptor-binding at the plasma membrane, followed by low pH activation in the endosome317. The ASLV TM fusion protein is stable over a broad range of pH values (67.0°C to 73.8°C at pH values between 5.0 and 8.5)418. Although SARS-CoV does not undergo a two-step mechanism of entry, SARS-CoV entry may be promiscuous as it is able to enter target cells through either a low pH endosomal route or direct fusion at the plasma membrane at neutral pH379,380,382,409,483. Consistent with ASLV TM, the linked SARS-CoV S2 is stable between pH 4.0 and 8.5. This provides further support that the SARS-CoV S2 fusion subunit is able to maintain the core stability regardless of its route of entry and environments encountered.
8.1.2 Electrostatic charges are important to the stability of pH-independent viral TMs. MPMV, HTLV-1, and XMRV belong to the β-, δ-, and γ-retrovirus genus, respectively, and are well known to undergo a fusion process at the plasma membrane at neutral pH\(^3\). Structures of the β-, δ-, and γ-retroviral fusion subunits reveal a clustering of electrostatic interactions at the heptad repeat and chain reversal regions\(^2\). The presence of salt bridges on pH-independent viral TMs aligns with the life cycle of retroviruses, as the electrostatic interactions will provide the utmost stability at neutral pH. Biophysical and mutagenic studies revealed that the HR1-HR2 interface and CR regions of these retroviral TMs are strongly stabilized by its electrostatic interactions. Moreover, previous reports showed that the alanine mutations to these residues resulted in melting temperature changes of >30°C and rendered the virus incapable of entry\(^2\). For instance, Maerz et al. identified a CR region salt bridge (R380 and E398) on HTLV-1 gp21 and an asparagine substitution on Glu398 residue reduced the fusion activity by more than 5-fold compared to wild-type\(^2\). Similar results were obtained from a previously identified HIV-1 salt bridge (K574 and D632) and inhibitors binding to a hydrophobic pocket that interrupts this salt bridge were shown to hinder the cellular entry of infectious HIV-1 particles\(^2\). Electrostatic interactions may be a common strategy used by viruses to stabilize the fusion protein at the plasma membrane (Figure 8.1).

8.1.3 Hydrophobic residues play a greater role in stabilizing the fusion subunits of pH-dependent viruses. In ASLV, which has been used extensively as a model virus for both pH-dependent and pH-independent entry, the fusion domain crystal structure also revealed a lining of electrostatic salt bridges. However, these electrostatic interactions do not play a major role in stabilizing the six-helix bundle, as they do in fusion proteins of retroviruses with pH-independent entry. Instead, hydrophobic residues in the ASLV fusion protein play a large role in stabilizing the postfusion state\(^4\). The use of hydrophobic residues for stabilization is consistent with ASLV’s mode of entry requiring low pH\(^4\), as the strength of hydrophobic interactions are not affected by pH changes in the environment. Other class I viruses such as EBOV, influenza A virus (IAV), and lymphocytic choriomeningitis virus (LCMV), which solely enter host cells through low pH endosomes, also contain both ionic and hydrophobic interactions within their fusion core in the postfusion state\(^5\). It was demonstrated that EBOV also utilizes hydrophobic residues to stabilize the postfusion state\(^7\). On the other hand, SARS-CoV is able to enter through
direct fusion at the plasma membrane and the low pH environment of the endosome; therefore, we expected that the fusion subunit would have structural features typical of both types of viral fusion proteins. Our biophysical study on the linked SARS-CoV S2 now provides additional evidence to support this hypothesis. Salt bridge interactions in SARS-CoV S2 play a minor role in stabilizing the six-helix bundle. Moreover, the lack of a role for salt bridges is supported by the poor conservation of some salt bridge residues (i.e. E918-R1167) in all CoVs. Like ASLV, other pH-dependent viruses may exploit hydrophobic interactions as their primary mechanism for maintaining postfusion stability within the endosomes (Figure 8.1). In summary, viral fusion proteins that mediate entry at the plasma membrane tend to display salt bridges in its heptad repeat and chain reversal regions, whereas viruses entering through an endosomal pathway tend to utilize non-ionizable hydrophobic residues for stability.

8.1.4 A positive charge clusters in the chain reversal region of pH-dependent viral TMs.
Structural analysis of class I fusion proteins from viruses that require low pH for entry, such as influenza A virus HA2, EBOV GP2, MARV GP2 and LCMV GP2, reveals a layering of outer and inner positive charges at the HR1 base or chain reversal region (Figure 3.5A). In ASLV TM, IAV HA2, EBOV GP2, MARV GP2 and LCMV GP2, a histidine residue is present in an outer layer on the fusion protein. In EBOV GP2 and MARV GP2, the outer histidine is structurally conserved with ASLV TM H492, suggesting an important role for the outer histidine. In IAV HA2, the outer histidine (H142) is suggested to act as a pH sensor to trigger conformational changes at low pH. Based on the prefusion EBOV GP structure, H602 is at the GP1-GP2 interface, thus supporting a role in also activating the glycoprotein for fusion. Histidine residues are ideal low pH sensors as the pK<sub>a</sub> of its imidazole side chain is 6.0. Thus, at the pH of an early or late endosome, histidine residues become protonated to trigger conformational changes. However, in addition to its potential role in conformational changes, the outer layer positive charge may be involved in stabilizing the negative helix dipole moment. Mutation of ASLV TM H492A and the addition of a negative charge (H492E) decreased the apparent T<sub>m</sub> by >20°C at low pH (Figure 3.5B), suggesting that the presence of a positive charge on the histidine residue is important for stabilization of the fusion subunit. In contrast, the incorporation of a histidine residue on the outer layer of HTLV-1 gp21 (K394H), as expected, destabilized the fusion protein at all pH values (Figure 3.5B). Due to the weak dipole moment of HTLV-1 gp21, a protonated histidine residue did not increase the stability at low pH. At pH values >7.0, a neutral histidine
residue destabilized the fusion subunit, suggesting that a positive charge in the outer layer is important for postfusion HTLV-1 gp21 stability. In summary, the outer layer positive charge may play two roles: a) triggering the conformational changes for fusion, and b) to stabilize the negatively charged helix dipole in the postfusion six-helix bundle.

8.1.5 An inner layer of positive charge is important in stabilizing the negative dipole moment. A positive charge is also commonly found on an inner layer of the fusion subunit. In ASLV TM, IAV HA2, EBOV GP2, MARV GP2 and LCMV GP2, a positive charge is found on an inner layer of the HR1 core. In IAV HA2 and ASLV TM, H106 and H490 reside at the base of the HR1 core, respectively. Histidine residues located at the C-terminus of a helix have been commonly shown to interact with a helix dipole to increase protein stability through charge-dipole interactions350,352,353. In EBOV GP2, MARV GP2 and LCMV GP2, an arginine residue is located at the base of HR1 in a position similar to H490 in ASLV TM. The substitution of H490R in ASLV TM resulted in maintained thermal stability ($T_m > 70^\circ$C) over a wide pH range (Figure 3.5B). Our results show that a positively charged arginine residue can also stabilize the postfusion conformation of viral fusion proteins at low pH. Regardless of the virus and type of amino acid, the positive charge at the base of the helical core will be able to counterbalance the negative helix dipole moment. For the fusion proteins of pH-independent viruses, such as HTLV-1 and MoMLV, there are no inner layer histidine or positively charged residues. Given that the pH-independent viral fusion subunits have a smaller negative helix dipole moment, end capping of the helix dipole moment may not be necessary for this class of fusion proteins. In summary, a basic residue is likely required on an inner layer of the HR1 core to counterbalance the negative helix dipole moment and stabilize the postfusion six-helix bundle for viral fusion proteins that encounter a low pH environment.
Figure 8.1. Summary of pH-dependent and pH-independent fusion protein properties. A cartoon of the viral fusion subunit in the postfusion conformation. Viruses that require low pH for entry, such as ASLV, IAV, and LCMV, fuse at the endosomal membrane, and have strong positive and negative charges at their ends due to its helix dipole moment. The negatively charged helix dipole is stabilized by two layers of positively charged histidine or arginine capping residues (only histidine is shown in the figure). In addition, the outer layer of histidine residues may be involved in triggering conformational changes in the prefusion viral glycoprotein upon entering a low pH environment. The HR1-HR2 interface is in general mediated by largely hydrophobic interactions. Viruses that undergo a pH-independent entry process, such as retroviruses, fuse at the plasma membrane. Its helix dipole moment is decreased, and no histidine or arginine residues are located in the chain reversal region to stabilize the negative helix dipole. Moreover, the HR1-HR2 interface is stabilized with a larger number of electrostatic interactions rather than hydrophobic forces.
Targeting viral membrane fusion and its inhibition. Viruses are obligate intracellular parasites and they must enter the cells to perpetuate their life cycle. In recent years, many class I viral fusion protein structures were determined in both pre- and post-fusion states to understand the mechanisms of viral and host cell membrane fusion \cite{21,48,53–58}. These structures provide valuable information about the molecular architecture of many viral fusion proteins. However, the molecular details of structural rearrangements between pre- and post-fusion states remain poorly understood. Recently, two separate groups obtained the first structural images of IAV HA intermediate conformations by using cryo-electron tomography (cryo-ET) \cite{488,489}. The intermediate states, in which the protein is associated with two membranes simultaneously, are unstable and rapidly converted to post-fusion conformation leading to efficient enveloped virus fusion. Inhibitors that bind to the intermediate conformations may disrupt fusion by preventing formation of the post-fusion trimer-of-hairpins (Figure 8.2). Such inhibitors include, C-peptides, small molecules, viral antigens, and neutralizing antibodies that bind to critical fusion protein epitopes. Designing C-peptides is a proven strategy as the FDA-approved HIV-1 drug enfuvirtide (T20 and also marketed as Fuzeon™) is a C-peptide derived from the C-terminal HR2 region of gp41 \cite{399,400,490}. Other inhibitors such as neutralizing antibodies, viral antigens, and D-peptides were also shown to be potent and broadly neutralizing inhibitors targeting viral fusion machinery \cite{299,491–494}. Ultimately, it is important to develop low molecular weight fusion inhibitors having a similar potent activity as C-peptides or neutralizing antibodies. A small molecule bound in to a pocket on the fusion protein surface should interfere with viral entry through the disruption of protein-protein interactions or critical salt bridges important in the fusion mechanism (Figure 8.2). Interestingly, a previously described cavity in HIV-1 gp41 is located in the middle of its HR1 region contains a salt bridge (K574 and D632) critical for HIV-1 entry \cite{297,298,495}. Moreover, small molecule inhibitors and D-peptide analogues have been developed against the hydrophobic pocket in HIV-1 gp41 and are able to disrupt viral fusion \cite{299,484–486}. Despite the evidence suggesting that the gp41 pocket is an attractive drug target, it has been challenging to identify small molecule inhibitors that bind with high affinity and high specificity. This is a novel strategy that opens new lines of inquiries for effective inhibition of these viruses. The general presence of these pockets in retroviruses, coronaviruses, and filoviruses, and previously designed antiviral inhibitors against the HIV-1 gp41 pocket suggest that a common strategy targeting the fusion protein core might be used for the development of drug-like inhibitors against these viruses. Future studies will further elucidate structural and
functional properties of intermediate states of the viral fusion machinery, which will likely contribute to the development of orally bioavailable antiviral drugs.

Figure 8.2. Schematic representation of the class I viral fusion protein prehairpin intermediate and its inhibition. Activation by receptor binding, low pH or proteolytic cleavage triggers the release of the fusion peptide, followed by its insertion into the target cell membrane. The structural rearrangements yield the transient intermediate state in which the fusion protein associated with two membranes simultaneously. The prehairpin intermediate exposes critical sites on fusion protein that may be disrupted by inhibitors in order to prevent the formation of trimer-of-hairpins\textsuperscript{496}. Previously, neutralizing antibodies that bind to the fusion peptide\textsuperscript{494}, helical core\textsuperscript{491,492}, and membrane proximal external region (MPER)\textsuperscript{497} of gp41 were identified and shown to be effective inhibitors targeting HIV-1 fusion\textsuperscript{491,492,494,497}. Other molecules such as C-peptides\textsuperscript{399,400,490} and D-peptides\textsuperscript{299}, small molecules\textsuperscript{484–486}, and 5-Helix\textsuperscript{498} also bind to the transient intermediate structure of the fusion protein and irreversibly inactivate membrane fusion.
8.2 Structural basis of cell-cell fusion mechanisms

While the protein nanomachines that mediate virus-cell fusion events are now fairly well characterized, cellular fusogens and their mechanisms of action are largely unknown for many cell fusion events. Cell-cell fusion is an emerging field and it is crucial at many intersections, from early stages of development, such as sperm and egg fusion, to the fusion of myoblast cells that is essential for muscle development\(^{41,42}\). Despite their seemingly different biological functions, the placental syncytins and the \textit{C. elegans} epithelial cell fusion protein, EFF-1, share strong structural and functional similarities with class I and class II viral fusion proteins, respectively\(^{22,31,308}\). These proteins may have diverged from a common ancestral gene and may use similar mechanisms to promote fusion. However, other cellular fusogens, such as those governing macrophage fusion or stem cell fusion, may adopt a different mechanism to catalyze the merger of two cell membranes. It is a challenging task to define whether a protein is a true fusogen or it is part of a multi-protein fusion complex. Another major obstacle is to differentiate a true fusogen from a recognition/adhesion molecule. Recently, the gene deletion experiments on animal models have been proven to be an unbiased approach to identify candidate proteins involved in particular cell fusion events. Cell-cell fusion is still poorly understood, and identifying the right candidate fusogen for various cell fusion event is the first step to elucidate the molecular mechanisms of these processes.

8.2.1 From pathology to physiology: syncytins in placental cell fusion and beyond. For millions of years, infection of a germline cell by a retrovirus has lead to integration and inheritance of provirus that is known as endogenous retrovirus (ERV). In humans, ERVs constitute about 7\% to 8\% of the genome\(^{13}\) and expression of particular human ERVs (HERVs) is associated with physiological functions, such as syncytin-mediated trophoblast fusion in placentation\(^{205}\), as well as certain pathologies, such as HERV-K expression in carcinogenesis\(^{499–501}\). Captured syncytins are \textit{bona fide} cellular fusogens integrated to mammalian genomes as early as >80 Million years ago, each providing its host a positive selective advantage\(^{189,190}\). So far, there have been ten different \textit{syncytin} genes identified among various clades of eutherian mammals and marsupials\(^{189,190}\). However, these \textit{syncytin} genes only share <2\% sequence identity; thus, introducing a major challenge for comparative sequence analysis and functional
Increasing numbers of whole genome sequences from a wide variety of living organisms will allow us to identify ‘new’ genes encoding endogenous retroviral proteins that have been captured and domesticated on multiple occasions, through convergent evolution.

In chapter 5, we determined the crystal structure of the human syncytin-1 TM and demonstrated that electrostatic interactions within the fusion subunit are critical for post-fusion stability and syncytium formation. Our results strongly indicate that the syncytin-1 TM share similar structural and functional features with the pH-independent class I viral fusion proteins; thus, strengthening the notion that syncytin-1 may exert a similar mechanism compared to pH-independent class I viral fusion proteins in order to promote trophoblast fusion in placentation. However, there are no existing structures of any full-length syncytin Env in its pre-fusion and/or intermediate states, and we are still lacking a significant amount of functional information to conclusively address the mechanism of the syncytin-mediated cell-cell fusion. In humans, the discovery of syncytin-2 \(^{183}\) also added another layer on the complexity of trophoblast fusion. Syncytin-1 and syncytin-2 are encoded by two different ERV loci \(^{182,183}\) and they display differences regarding their expression, localization patterns, and mechanism of action \(^{193,195-197,207}\). For instance, syncytin-1 utilizes neutral amino acid transporter ASCT2 as its cell surface receptor \(^{181,502}\), while syncytin-2 has been associated with essential fatty acid carrier MFSD2 \(^{503}\). In addition, the syncytin-2 ISD is endowed with an IS function, whereas the syncytin-1 ISD is non-IS \(^{207}\). In chapter 5, we compared the structures of the syncytin-1 and syncytin-2 ISDs and demonstrated that the dissimilarities in syncytin-induced immuneosuppression is not due to the gross structural differences but rather depend on a single amino acid at position 14 of the ISD. An important question is still pending on whether the fusogenic activity or the IS function of the syncytins is more critical for their physiological role in placental morphogenesis. Placental expression of syncytins also provides an additional layer of regulation to the fusogenic activity of these proteins. Syncytin-1 is predominantly expressed in all villous cytotrophoblast, syncytiotrophoblast and extravillous cytotrophoblast cell lineages, whereas syncytin-2 expression is restricted to villous cytotrophoblast cells \(^{195-197}\). Hence, strict regulation of syncytin gene expression suggests that syncytin-mediated trophoblast fusion might be controlled at the transcriptional level. This is in contrast with the tissue specific activity of viral glycoproteins where the cell specific expression and distribution of surface receptors posttranscriptionally
determine the site of fusogen activation and membrane fusion. Further structural and functional studies are needed to mechanistically explain the exact function of syncytins in placentation.

Recently, placental exosomes isolated from the culture supernatant of primary villous cytotrophoblast cells and from the sera of pregnant women were shown to bear syncytin-1 and syncytin-2 proteins at the surface. Incorporated syncytin proteins could potentially migrate more distally from the placenta and interact with different cell types in an uncontrolled fashion. These findings now open up an entirely fresh line of inquiry on how these placental fusogens may be involved in cancer progression. So far, syncytin expression was determined in endometrial and breast cancers. These might not be the only types of cancers that exhibit syncytin expression. Deregulation of syncytins in cancer cells may contribute to uncontrolled cancer cell fusion and induce tumor development and metastasis. However, the possible involvement of syncytins in cancer progression is still poorly understood and requires further investigations to elucidate the mechanisms of syncytin-mediated disorders.

8.2.2 Sperm meets egg: Izumo1 and Juno are essential for fertilization. In sexually reproducing species, the fusion of sperm and egg plasma membranes results in the creation of a new genetically distinct diploid organism termed zygote. During fertilization, mature sperm undergoes an acrosome reaction and penetrates through the egg zona pellucida to reach the perivitelline space. The acrosome reaction also causes relocalization of Izumo1 to the sperm equatorial segment. In chapter 6, we demonstrated that Izumo1 adopts a monomeric boomerang conformation on the surface of the sperm membrane. Upon binding to the Juno egg receptor, Izumo1 undergoes a conformational change. The 4HB region migrates towards the egg membrane. Moreover, the hinge region of Izumo1 becomes more rigid and "locks" the molecule into an upright position. The formation of the Izumo1 and Juno complex provides a direct physical link between the egg and sperm membranes. Despite the importance of sperm and egg binding for fertilization, the transition from Izumo1-Juno binding to membrane fusion remains uncertain. It is currently not clear whether Izumo1 requires a post-Juno binding event to trigger the fusion process. Nonetheless, at least three potential mechanisms are possible (Figure 8.3). The heterotypic assembly of Izumo1 and Juno, or a secondary conformational
change in Izumo1, may bring the egg and sperm membranes into close proximity for fusion to take place. Inoue et al. proposed that subsequent to Izumo1-Juno binding, a protein disulphide isomerase (PDI) catalyzes a thio-disulphide exchange reaction that leads to structural conformation change and dimerization of Izumo1\textsuperscript{444}. The Izumo1 dimer releases Juno and contacts a yet to be discovered oocyte receptor that facilitates membrane fusion\textsuperscript{444}. In addition, Ohto et al. demonstrated that the disulfide bonds in Izumo1 are easily broken; hence it was suggested that conformational changes upon Juno binding could expose critical disulfide bonds for exchange\textsuperscript{508}. Alternatively, Izumo1 may act as a scaffold to recruit other sperm or egg protein partners to form a multiprotein fusion complex in a manner similar to some viral fusogens\textsuperscript{442,443}. Further studies will shed light on the precise molecular events required for the merger of sperm and egg. Following fusion, Juno is rapidly shed into extracellular vesicles from the fertilized oocyte. Within 30-40 minutes, Juno is weakly or barely detectable on the membrane surface of zona-intact or anaphase II-stage zona-free fertilized oocytes, and undetectable at the pronuclear stage\textsuperscript{173}. Izumo1 binds Juno tightly and rapidly and once shed, Juno is able to bind to exposed Izumo1 on incoming acrosomal-reacted sperm in the perivitelline space to act as a “sperm-sink” to block polyspermy (Figure 8.3). Over the years, a large numbers of sperm and egg genes have been identified as candidates in sperm-egg recognition and fusion during mammalian fertilization\textsuperscript{143}. Until the discovery of Izumo1\textsuperscript{168} and Juno\textsuperscript{173}, almost all the participants reported to be important in sperm-egg interaction were shown to be not essential for fertilization. Nevertheless, it seems likely that Izumo1 and Juno are not the only proteins essential for fertilization and there might be other unknown sperm or egg proteins that may take a part in sperm-egg fusion. So far, no mammalian proteins have been found to function as a direct fusogen that may catalyze the merger of two plasma membranes. Our complex structure of human Izumo1 and Juno lays the foundation for further studies that could resolve the intricate mechanisms of sperm-egg fusion and lead to the development of non-hormonal contraceptives and fertility treatments for humans and other mammals.
8.3 Conclusions

Different biological systems utilize distinct mechanisms to mediate the process of membrane fusion. Although no universal mechanism may exist, identification of novel fusogens unifies the concepts and provides a better understanding of precise structural and functional signatures essential for viral-cell and cell-cell fusion\textsuperscript{141}. At the same time, we are exploring the specific functions of existing fusogens in diverse biological processes to elucidate their roles in...
physiological events as well as pathologies regarding their malfunction. We now have a good understanding about the mechanisms of viral and target cell membrane fusion. The molecular mechanisms of viral-cell fusion are highly conserved within the structural classes and appear to be shared by most of the studied viruses. However, with the exception of syncytins and EFF-1, many developmental and cellular fusogens remain undiscovered. It is tempting to speculate that the structures of the cellular fusogens, activation mechanisms, and requirements of in one or both membranes might be completely different and not conserved between cellular fusogens. Divergent mechanisms of cell membrane fusion raise new challenges for finding the missing fusogens, and characterizing their regulation and complex triggering mechanisms. The diversity and complexity of membrane fusion draws significant attention to this field of biology and explains why even the most crucial events for human development, such as the fusion of sperm and egg, remain poorly understood after so many years of active investigation.
References


346. Delos, S. E., La, B., Gilmartin, A. & White, J. M. Studies of the ‘chain reversal regions’ of the avian sarcoma/leukosis virus (ASLV) and ebolavirus fusion proteins: analogous residues are important, and a His residue unique to EnvA affects the pH dependence of ASLV entry. *J. Virol.* 84, 5687–5694 (2010).


263


