Experimental Determination of the Topology of the HIV-1 gp41 C-Terminal Tail

by

Jonathan Damien Steckbeck

B.S. in Microbiology, the Pennsylvania State University, University Park, 1999

M.B.A. in Business Strategy, Joseph M. Katz Graduate School of Business, University of Pittsburgh, 2003

Submitted to the Graduate Faculty of

The School of Medicine in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy in Biochemistry and Molecular Genetics

University of Pittsburgh

2011

UNIVERSITY OF PITTSBURGH

School of Medicine

This dissertation was presented

by

Jonathan Steckbeck

It was defended on

December 15, 2011

and approved by

Michael Cascio, PhD, Associate Professor, Duquesne University

Billy W. Day, PhD, Professor, Department of Pharmaceutical Sciences

Judith Klein-Seetharaman, PhD, Associate Professor, Department of Structural Biology

Thomas E. Smithgall, PhD, Professor and Chair, Department of Microbiology and Molecular Genetics

Dissertation Advisor: **Ronald C. Montelaro, PhD**, Professor, Department of Microbiology and Molecular Genetics

Experimental Determination of the Topology of the HIV-1 gp41 C-Terminal Tail

Jonathan D. Steckbeck, PhD

University of Pittsburgh, 2011

The C-terminal tail (CTT) of HIV gp41 has been traditionally viewed as a cytoplasmic domain. Genetic studies demonstrating functional interactions between the CTT and various intracellular partners have implicitly reinforced this view. However, antibody neutralization data and biochemical studies have suggested that the CTT is, or can be, externally localized under certain condition. Additionally, other studies have demonstrated that the CTT is dispensable for *in vitro* virus replication. After nearly three decades of HIV research, the function and structure of the CTT remain elusive.

Our goals, then, were twofold: (i) to determine the overall conservation of the CTT in an attempt to provide an understanding of the functional and structural relevance of the CTT; and, (ii) to provide an experimental topological map of the CTT in an attempt to understand and align observed CTT topology(ies) with the functional necessity of a cytoplasmic CTT. We believe that we made significant contributions to the understanding of CTT topology and its relationship to current published functional studies.

The initial studies demonstrated that the CTT sequence is conserved at a level that is intermediate between the highly variable gp120 region and the relatively conserved gp41 ectodomain. Additionally, physicochemical and structural properties of CTT sequences were found to be conserved in spite of the relatively high sequence

variability. These studies demonstrated for the first time that the CTT sequence, while highly variable, contains highly conserved structural and chemical properties that suggest a functional requirement for the CTT.

Topology studies of the CTT indicated that the topology of the CTT can be distinct between the surface of Env-expressing cells and viral particles. Additionally, dynamic rearrangement of the CTT was observed as a function of antibody neutralization. These findings prompted a theoretical study of gp41 CTT predicted topology and the proposal of a topological model that we believe is consistent with all published studies regarding the localization of the CTT.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS XVI

1.0 INTRODUCTION TO THE HUMAN IMMUNODEFICIENCY VIRUS AND PRINCIPLES REGARDING THE GENERATION OF MEMBRANE PROTEIN TOPOLOGY 1

1.1 HUMAN IMMUNODEFICIENCY VIRUS 1

1.1.1 Viral genome 3

1.1.2 Virion morphology and viral protein functions 5

1.1.3 Gag proteins 6

1.1.4 Pol proteins 7

1.1.5 Env proteins 8

1.1.5.1 gp120 structure and function 9

1.1.5.2 gp41 structure and function 11

1.2 MEMBRANE PROTEIN TOPOLOGY 36

1.2.1 Membrane protein classifications 37

1.2.2 Membrane protein assembly and topogenic signals 40

1.2.3 Transmembrane domain and topology prediction 43

1.2.3.1 A biological hydrophobicity scale 44

1.2.4 Alternative protein topologies 46

1.2.4.1 Dynamic topology 47
1.2.4.2 In-plane membrane anchoring sequences 48
1.2.5 Effects of lipids on membrane protein topology 49
2.0 HYPOTHESIS AND SPECIFIC AIMS 52
2.1 HYPOTHESIS 52
2.2 SPECIFIC AIMS 53
2.2.1 Aim 1. To define the conservation of gp41 CTT sequence and structural elements from diverse HIV strains. 53
2.2.2 Aim 2. To determine the native topology of the Env CTT on the surface of intact cells and viral particles and examine the effect of the fusion process on the topology of the CTT. 54
3.0 HIGHLY CONSERVED STRUCTURAL PROPERTIES OF THE C-TERMINAL TAIL OF HIV-1 GP41 DESPITE SUBSTANTIAL SEQUENCE VARIATION AMONG DIVERSE CLADES: IMPLICATIONS FOR FUNCTIONS IN VIRAL REPLICATION ... 56
3.1 INTRODUCTION 57
3.2 MATERIALS AND METHODS 61
3.2.1 Sequence acquisition and processing 61
3.2.2 Sequence alignments and evolutionary analyses 61
3.2.3 Characterization of physicochemical properties of consensus LLP peptide sequences 63
3.2.4 PEP-FOLD structure prediction for LLP peptides 63
3.2.5 Circular dichroism (CD) spectroscopy to determine peptide secondary structure 64

3.3 RESULTS 65
3.3.1 CTT sequence variation is intermediate between gp120 and the gp41 ectodomain 65
3.3.2 Physicochemical characteristics of CTT domains are conserved 67
3.3.3 Secondary structure of the LLP regions of the CTT is predicted to be highly conserved 74
3.3.4 LLP peptide analogs adopt helical structure in membrane-mimetic and hydrophobic environments 81
3.4 DISCUSSION 85
4.0 TOPOLOGY OF THE C-TERMINAL TAIL OF HIV-1 GP41: DIFFERENTIAL EXPOSURE OF THE KENNEDY EPITOPE ON CELL AND VIRAL MEMBRANES 92
4.1 INTRODUCTION 93
4.2 MATERIALS AND METHODS 99
4.2.1 Cells and virus preparations 99
4.2.2 VSV-G epitope tag substitutions 99
4.2.3 Cell culture and transfections 100
4.2.4 Cellular FACS analysis 100
4.2.5 Isolation of detergent resistant membranes 101
4.2.6 Viral immunoprecipitation and western blotting 102
4.2.7 Western blot quantitation 103
4.2.8 SPR spectroscopy analysis of antibody-virus interactions 103
4.3 RESULTS 104

4.3.1 Exposure of CTT sequences on the surface of Env-transfected cells 104
4.3.2 KE-specific antibodies do not bind to intact viral particles 107
4.4 DISCUSSION 113
5.0 EXISTENCE OF A CELL SURFACE TOPOLOGY FOR THE HIV GP41 C-TERMINAL TAIL THAT IS DISTINCT FROM THE VIRAL TOPOLOGY 119
5.1 INTRODUCTION 119
5.2 MATERIALS AND METHODS 123
5.2.1 Cell and virus construct 123
5.2.2 VSV-G epitope tag substitutions 123
5.2.3 Cell culture and transfections 125
5.2.4 Cellular FACS analysis 125
5.2.5 Production of pseudoviral particles with VSV-G tagged Env 126
5.2.6 Pseudoviral immunoprecipitation and Western blotting 126
5.2.7 Western blot quantitation 128
5.3 RESULTS 128
5.3.1 The majority of the CTT sequence is exposed on the cell surface 128
5.3.2 CTT-exposed Env constitutes a large proportion of cell-surface Env 131
5.3.3 CTT exposure is not apparent on the surface of VSV-G-tagged pseudovirus particles 133
5.3.4 Dynamic exposure of CTT sequences during viral infection 135

5.3.5 Prediction of HIV gp41 sequences to act as an in-plane membrane anchor 138
5.4 DISCUSSION 139
6.0 GENERAL DISCUSSION OF RESULTS AND FUTURE RESEARCH DIRECTIONS 148
6.1 SUMMARY OF MAJOR FINDINGS 148
6.1.1 Physicochemical and structural characteristics of the CTT are widely conserved across genetically distinct viral sequences 149
6.1.1.1 Summary of CTT conservation 149
6.1.1.2 Implications of CTT property and structural conservation 150
6.1.2 Distinct CTT topologies on the cell and virion surface 154
6.1.2.1 Summary of CTT cellular and viral topology studies 154
6.1.2.2 Implications of different cellular and virion CTT topologies 155
6.2 A COMPREHENSIVE WORKING MODEL FOR THE CTT 160
6.2.1 Revisiting the Membrane Association of gp41 - A Theoretical Study 160
6.2.2 A Proposed Topological Model for the HIV gp41 CTT 163
6.3 FUTURE DIRECTIONS 169
6.3.1 Did the epitope mapping technique provide a non-native topology for the CTT? 170
6.3.2 Why is arginine so highly conserved in the CTT? 171

6.3.3 Can the CTT be a target for small molecule drugs as HIV inhibitors? 172
6.3.4 Do CTT sequences act as immunological decoys? 174
6.3.5 What is the measured potential for predicted transmembrane region insertion? 174
6.4 CLOSING REMARKS 175
BIBLIOGRAPHY 177

LIST OF TABLES

Table 1. Average cRMSD of PEP-FOLD-predicted structures 80

Table 2. Sequences of peptides synthesized for CD spectra determination 82

LIST OF FIGURES

Figure 1. World-wide HIV prevalence and AIDS deaths 2

Figure 2. HIV genome organization 4

Figure 3. HIV virion structure 6

Figure 4. Env protein primary structure 9

Figure 5. Schematic of gp41 ectodomain 12

Figure 6. gp41 ectodomain six helix bundle in fusion-competent state 14

Figure 7. Proposed models for the membrane-spanning sequences of HIV gp41 17

Figure 8. Sites of functional consequence in genetic studies of the HIV gp41 CTT 26

Figure 9. Proposed topology models for HIV gp41 CTT 28

Figure 10. CTT-dependence of Env conformation 35

Figure 11. Hydrogen bonding between peptide main chain groups 39

Figure 12. Sequence properties influence final topology 41

Figure 13. Phylogenetic analysis of Env domains 66

Figure 14. Comparison of average diversity in Env indicates intermediate CTT variation relative to other domains 68

Figure 15. Comparison of the amino acid sequence of the CTT domain 69

Figure 16. LLP Sequences have different physicochemical properties than MPER or MSD sequences 71
Figure 17. Helical wheel diagrams demonstrate the conservation of the amphipathic nature of LLP regions despite sequence variation 73
Figure 18. Specific arginine residues are highly conserved in LLP1 and LLP2 75
Figure 19. LLP sequences are predicted to be structurally homogeneous 77
Figure 20. PEP-FOLD simulations return multiple clusters for some LLP peptides 78
Figure 21. LLP peptide structural alignments reveal high level of structural similarity across clades/groups 79
Figure 22. SDS and TFE induce increased helicity in synthetic LLP peptides 83
Figure 23. Relative helicity of LLP peptides by mean residue ellipticity 84
Figure 24. Schematic models of the HIV-1 CTT 95
Figure 25. Sequence alignment of variant Env CTT 98
Figure 26. FACS analysis of intact Env-expressing cells demonstrates extracellular exposure of the KE sequence 106
Figure 27. VSV-G epitope insertions do not disrupt Env association with detergent-resistant membranes 108
Figure 28. Anti-KE MAbs do not bind to intact virions 110
Figure 29. Comparison of MPER and KE MAb binding using SPR spectroscopy 112
Figure 30. Replacement of HIV gp41 sequences with VSV-G sequences 124
Figure 31. HIV Env staining of live intact cells 130
Figure 32. Inhibition of HA-specific antibody binding by VSV-G-specific antibody 132
Figure 33. Relative infectivity of pseudoviruses containing VSV-G-tagged Env 134

Figure 34. Immunoprecipitation of VSV-G-tagged pseudoviral particles 136

Figure 35. Post-attachment neutralization of HIV-1 89.6 by anti-CTT antibody..... 137

Figure 36. In-plane anchor prediction for HIV gp41..... 140

Figure 37. Schematic model for exclusion of CTT-exposed Env from viral assembly and budding sites 146

Figure 38. Proposed topological model for the HIV gp41 CTT 165

Figure 39. Proposed MSD sequence rearrangement upon trimerization..... 167

ACKNOWLEDGEMENTS

This dissertation is dedicated to my wife,

Jodi,

without the support of whom

none of this would have been possible,

and our greatest collaborative experiment,

our beautiful daughter

Zooey Camille,

who is a constant inspiration to see

wonder in the world.

I love you.

1.0 INTRODUCTION TO THE HUMAN IMMUNODEFICIENCY VIRUS AND PRINCIPLES REGARDING THE GENERATION OF MEMBRANE PROTEIN TOPOLOGY

1.1 HUMAN IMMUNODEFICIENCY VIRUS

The human immunodeficiency virus (HIV) is the causative agent of AIDS. HIV infection of the human population has led to a world-wide AIDS epidemic that has recently passed 30 years since its recognition as a distinct disease, and 27 years since the identification of the responsible virus (Barré-Sinoussi et al., 1983; Gallo et al., 1984). In the three decades of the epidemic, more than 25 million people have died as a result of HIV infection, and an estimated 40 million people are currently infected (UNAIDS Global Report 2010). In 2009, the last year for which data has been reported, 2.6 million (2.3 - 2.8 million) new infections were reported, with 1.8 million (1.6 - 2.1 million) deaths. These numbers represent 8% and 21% decreases, respectively, from the incidence and mortality data reported in 2004. While the epidemic is truly global in nature, its weight is borne predominantly by the developing world, in particular sub-Saharan Africa, with 69% of new infections and 72% of deaths attributable to AIDS or AIDS related diseases (UNAIDS Global Report 2010). A comparison of land area maps adjusted for population

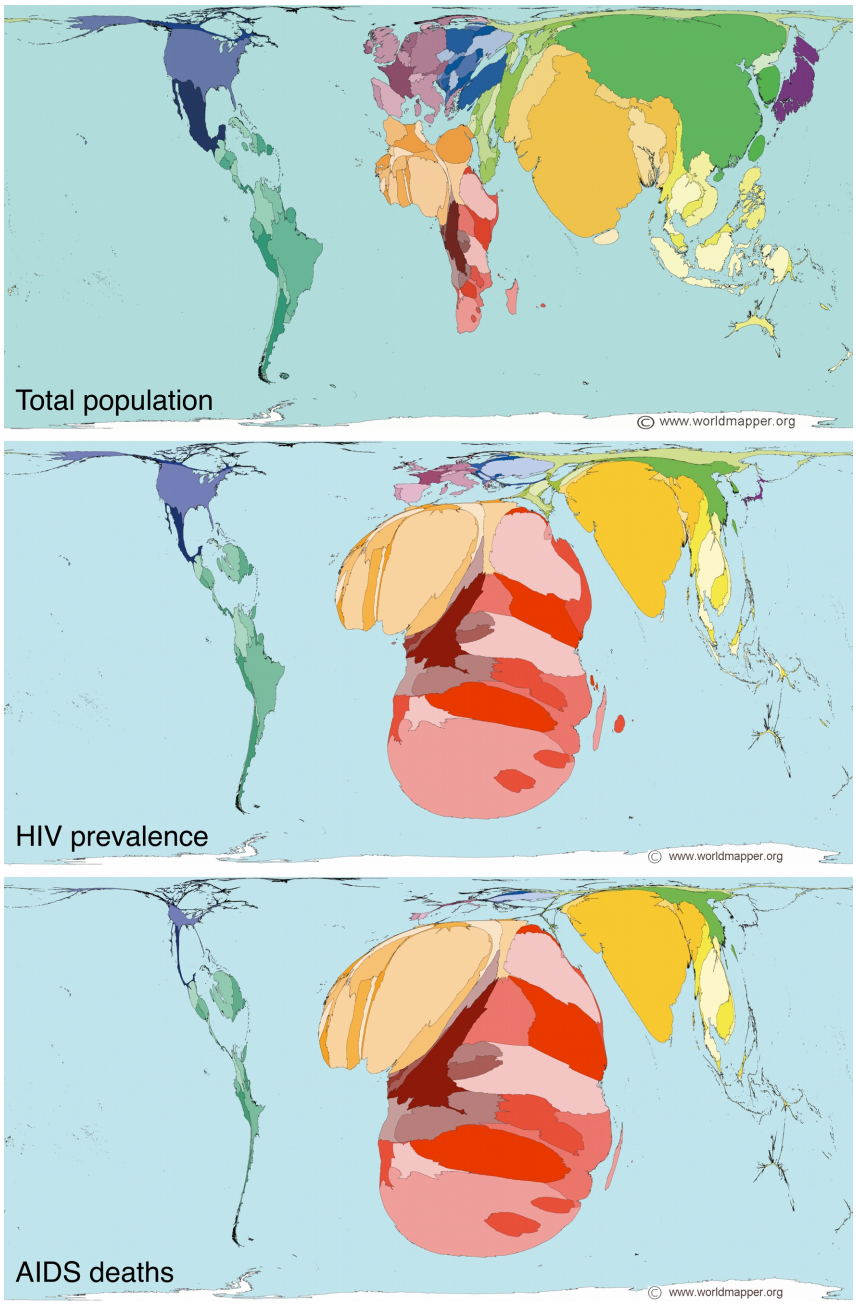

Figure 1. World-wide HIV prevalence and AIDS deaths

World maps with land area adjusted for total population (top), HIV prevalence (middle), and AIDS deaths (bottom). Maps used by permission of WorldMapper.org under a Creative Commons License.

and HIV prevalence and deaths demonstrate this point more clearly than any words (Figure 1). Of concern epidemiologically is that the potential impact of the epidemic has only recently been considered in China, where poor reporting and governmental obfuscation likely have led to underestimation of the number and extent of infections (Beck, 2004). However, notwithstanding the geographical and cultural borders separating the most-affected from the least-affected countries, HIV infection and AIDS must be considered among the most important, and devastating, infectious epidemics humanity has encountered.

1.1.1 Viral genome

HIV is a member of the Retroviridae family, and belongs to the Lentiviridae genus (lenti- : derived from Latin, meaning slow). As a retrovirus, RNA is the genetic material for HIV, and each virion contains two positive strand copies of the viral genome containing a 5' cap and a 3' poly-A tail (Luciw et al., 2002). As such, the viral genome is not infectious. The genome organization of HIV is shown in Figure 2. As with all retroviruses, the HIV genome contains the three primary structural and enzyme-coding genes, *gag* (structure), *pol* (enzyme), and *env* (structure), but also encodes for six accessory genes (*vif*, *vpr*, *vpu*, *tat*, *rev*, and *nef*) that determine the replicative and pathogenic potential, which tend to vary by the genetic subtypes (described below) (Levy, 2009). *gag* encodes the major structural proteins matrix/p17, capsid/p24, nucleocapsid/p7, and p6, in addition to minor proteins p1 and p2. The *pol* gene encodes the viral enzymes protease (PR), reverse transcriptase (RT), and integrase (IN). *env* predominantly codes

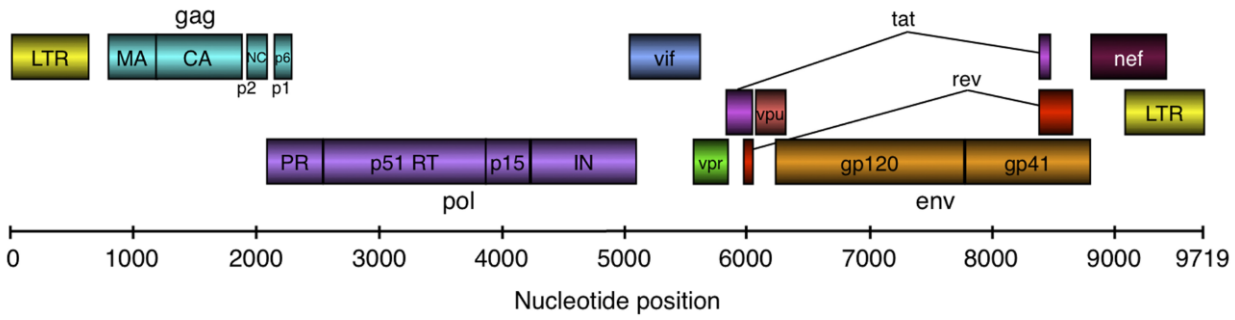

Figure 2. HIV genome organization

HIV genes and gene products are organized relative to their genomic nucleotide position.

for the viral envelope (Env) glycoprotein, but also contains coding sequences for the accessory proteins Vpu, Tat, and Rev. The coding sequence for Vif overlaps with the 3' end of *pol*. Of the accessory proteins, only Vpr and Nef utilize coding regions solely outside the major *gag*, *pol*, and *env* genes.

HIV is divided into phylogenetically distinct groups based on sequence differences in the *env* gene (Korber et al., 2001). The three main groups are M (for “main”), N (for “non-M and non-O”), and O (for “other”). Most circulating viruses and the vast majority of clinical isolates derive from group M. Group M is further divided into additional phylogenetically distinct groups, called clades (designated A, B, C, D, F, G, H, J, and K), that display distinct geographical distributions. Viruses with recombinant Env, likely resulting from superinfection, are also among the circulating strains and further increase the genetic diversity of Env.

1.1.2 Virion morphology and viral protein functions

Morphologically, HIV is an irregular, roughly spherical virus with a characteristic electron-dense conical core surrounded by a lipid envelope that is derived from the host cell during the budding process (Luciw et al., 2002). The virion lipid content is distinct from the overall lipid content of the parental cell, however, and is suggestive of budding from cholesterol- and sphingomyelin-enriched membrane microdomains (Brügger et al., 2006). A schematic of the virion detailing the location and identity of viral proteins in the mature virion is shown in Figure 3.

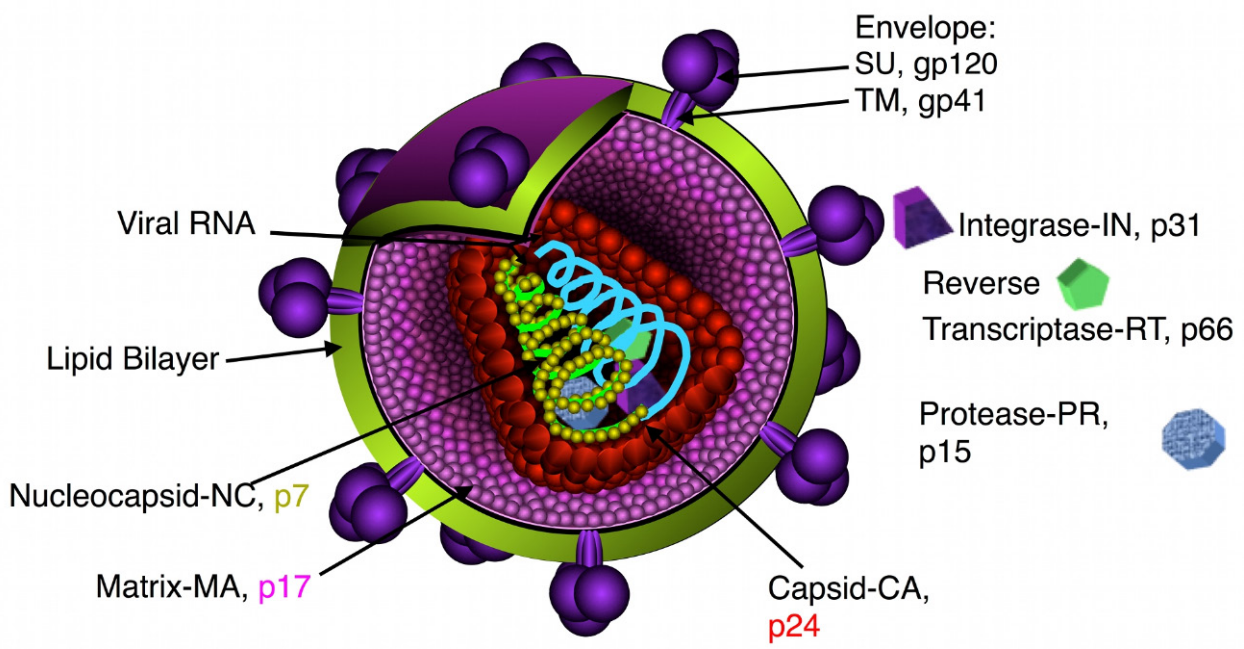

Figure 3. HIV virion structure

Schematic representation of a mature HIV virion detailing the localization of viral proteins and the approximate virion structure. The representation is not to scale. Virion schematic courtesy of Jodi Craigo, Ph.D.

1.1.3 Gag proteins

gag is translated as the Gag polyprotein, $Pr55^{Gag}$, that is subsequently cleaved to yield the individual Gag domains (Luciw et al., 2002). $Pr55^{Gag}$ is trafficked to the cellular plasma membrane by interactions of the matrix (MA) domain with the lipid bilayer. MA contains an N-terminal myristic acid as well as a basic patch near the membrane-proximal side of the folded domain that keeps the protein localized to the membrane (Hill et al., 1996; Massiah et al., 1994; Parent et al., 1994). MA also functions to direct the incorporation of Env into the budding virion, which will be considered in more detail below. The capsid domain (CA) participates in Gag-Gag interactions during virion

assembly (Byeon et al., 2009; Pornillos et al., 2009) and is the structural component of the conical viral capsid in the mature virion (Luciw et al., 2002). Nucleocapsid (NC) facilitates the binding, and thus incorporation, of the viral genomic RNA into the budding virion, and also contributes to Gag-Gag interactions (Bieniasz, 2006; Demirov and Freed, 2004; Hurley et al., 2010; Morita and Sundquist, 2004). Finally, p6, through its late domain, recruits (i.e. "hijacks") the cellular ESCRT assembly machinery normally utilized by the cell for membrane budding and scission events, to effect virion budding, predominantly from the plasma membrane (Balasubramaniam and Freed, 2011; Finzi et al., 2007; Jouvenet et al., 2008; Welsch et al., 2007). Following budding, $Pr55^{Gag}$ is cleaved by the viral protease, after which CA undergoes condensation from the viral membrane to form the characteristic conical core structure that encapsulates the NC-associated viral genome (Ganser-Pornillos et al., 2008).

1.1.4 Pol proteins

The Pol proteins are translated and incorporated into the virion as the gag-pol polyprotein $Pr160^{gag\text{-}pol}$ (Luciw et al., 2002). Viral protease cleaves $Pr160^{gag\text{-}pol}$ after viral budding from the cell as part of the process of virion maturation (Luciw et al., 2002). These virally-encoded enzymes carry out functions specific to the virus, and for which no cellular functionality is known (Luciw et al., 2002). Following infection of a target cell, reverse transcriptase (RT) converts the genomic viral single-stranded RNA into double stranded DNA through coordinated action of its RNAse H and DNA polymerase functionalities (Luciw et al., 2002). Importantly, RT lacks a genomic proofreading capability, making it the major contributing factor in the observed genetic diversity

among HIV strains (Korber et al., 2001). Finally, integrase contains both DNA cleavage and strand-transfer (joining) functions to incorporate the newly reverse-transcribed viral DNA into the host cell genome, resulting in a cell capable of producing progeny virions (Luciw et al., 2002).

1.1.5 Env proteins

Env is the only viral protein exposed on the virion surface, and it is the main target of the host humoral immune system (McElrath and Haynes, 2010). The targeting of Env, through direction of the immune response by vaccination, or by traditional small-molecule pharmaceuticals provides the best hope for controlling the epidemic (McElrath and Haynes, 2010). As such, the study of Env structure and function has yielded important insights into the properties and features of these viral proteins. Functional virion-associated Env is a trimer of heterodimers derived from extensive processing of the *env* gene product, presented schematically in Figure 4. Env is translated and cotranslationally N-glycosylated as a 160 kDa polyprotein (gp160) in the endoplasmic reticulum. Env gp160 is thought to multimerize in the ER into predominantly trimers, before trafficking to the Golgi (Earl et al., 1990; Earl et al., 1991; Pinter et al., 1989). In the Golgi, Env gp160 is cleaved by a cellular furin-like protease at a conserved K/R(X)K/RR motif into gp120 (or surface unit (SU)) and gp41 (or transmembrane (TM)) (Freed et al., 1989; McCune et al., 1988). The cleaved gp120 and gp41 products non-covalently associate to form the active trimeric Env unit.

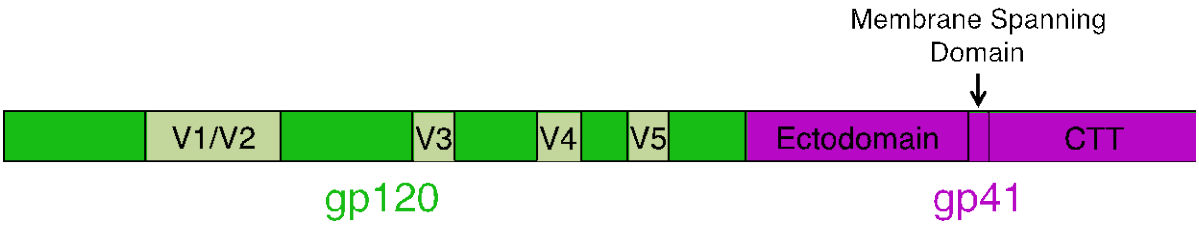

Figure 4. Env protein primary structure

Schematic representation of the primary structural layout of HIV-1 Env. The variable loops of gp120 are labeled as V1, V2, V3, V4, and V5. The three primary domains of gp41 are indicated.

1.1.5.1 gp120 structure and function

gp120 is a structurally complex molecule. Sequence analyses of gp120 from diverse HIV-1 isolates identified a high degree of sequence variation, determining the presence of five variable regions (V1 - V5) sequentially interspersed in the primary sequence with five relatively conserved regions (C1 - C5) (Starcich et al., 1986; Willey et al., 1986). Sequence analyses also identified 18 highly conserved cysteine residues that form nine disulfide bonds that play an important role in the formation of gp120 tertiary structure (Leonard et al., 1990). The variable regions determined by sequence analysis coincide with loop structures predicted to be formed by the disulfide bonds (Leonard et al., 1990). These loop regions vary considerably in size, with the V1/V2 loop ranging from 50 - 90 amino acids in length, while V4 ranges from 19 - 44 amino acids, and V5 from 14 - 36 amino acids (Chohan et al., 2005; Kitrinos et al., 2003; Masciotra et al., 2002; Palmer et al., 1996; Sagar et al., 2006; Shioda et al., 1997). V3, on the other hand, does not vary appreciably in size. The presence of these loops has been confirmed by x-ray crystal structures of gp120 proteins (Huang et al., 2005; Kwong et al., 1998). In addition, gp120 contains 20-30 canonical N-linked glycosylation sites and is thus extensively glycosylated with oligosaccharides accounting for approximately half of its molecular

mass. The location and number of glycosylation sites also varies by strain and change over time during the course of infection (Allan et al., 1985; Montefiori et al., 1988).

HIV-1 gp120 has been studied extensively, and a number of atomic resolution structures have been elucidated, all in conjunction with one or more binding partners, including CD4 and coreceptor binding site MAbs 17B (Kwong et al., 1998) and X5 (Huang et al., 2005), or CD4 binding site MAbs b12 (Zhou et al., 2007) and VRC01 and VRC03 (Wu et al., 2010). Due to what is likely extensive flexibility, the V1/V2 loops were deleted from gp120 proteins used for structure determination (Kwong et al., 1998). The relative location of the V1/V2 loops on the virion-associated Env trimer has been recently determined, however, by fitting (V1/V2 deleted) gp120-CD4 crystal structures into cryo-electron microscopy (cryo-EM) density maps of CD4-bound Env (Liu et al., 2008). The inclusion of CD4 in the cryo-EM Env structure determination allowed for unambiguous orientation and fitting of preexisting structures that led to the identification of V1/V2 density in the cryo-EM Env structure (Liu et al., 2008; White et al., 2010).

The primary function of gp120 is to initiate and modulate virus-cell interactions through binding to the primary receptor, CD4, and the coreceptor, predominantly CCR5 or CXCR4. gp120 initiates virus-cell contact by binding to CD4 and subsequently undergoes structural rearrangements that result in the formation of the coreceptor binding site. These conformational changes allow for high affinity binding of gp120 to coreceptor, that in turn leads to conformational changes in gp41 to begin the process of virus-cell membrane fusion (Myszka et al., 2000). Because of its receptor binding functionality and the identification of antibodies that exhibit cross-clade neutralizing activity (broadly neutralizing antibodies), gp120 is a major focus for experimental

vaccines (McElrath and Haynes, 2010). In addition, one commercially-licensed small molecule therapeutic, maraviroc, targets the gp120-coreceptor binding interaction (Dragic et al., 2000). However, maraviroc acts by binding CCR5, and does not specifically interact with gp120 and is inactive against viruses that utilize CXCR4 as coreceptor (Dragic et al., 2000).

1.1.5.2 gp41 structure and function

gp41 can be structurally subdivided into three major domains (see Figure 4): the extracellular domain (or ectodomain); the membrane-spanning domain (MSD); and the C-terminal tail (CTT), also referred to classically as the intracytoplasmic tail. Of these, experimental protein structures for only the gp41 ectodomain have been determined at atomic detail, and then only for the six-helix bundle of what is generally considered the fusion-competent state (Caffrey et al., 1998; Chan et al., 1997; Shi et al., 2010). Little to no information exists on the native structure of gp41. Even recent cryo-EM structures of Env on the virion surface provide little insight into the structure of gp41 beyond that it appears to present as a thin stalk upon which gp120 rests (Liu et al., 2008; White et al., 2010).

gp41 ectodomain The location and function of the gp41 ectodomain is well studied and generally unquestioned. gp41 is also N-glycosylated, although to a much lesser extent than gp120, at three to five potential sites (Montefiori et al., 1988). The ectodomain contains the major determinants of membrane fusion (Figure 5): a hydrophobic N-terminal domain termed the fusion peptide (FP) (Bosch et al., 1989; Freed et al., 1992; Freed et al., 1990); two heptad repeat regions, termed HR1 and HR2 (also referred to

Figure 5. Schematic of gp41 ectodomain

Schematic representation of the gp41 ectodomain, with the N-terminus located to the left and the C-terminus located to the right of the MPER. FP = fusion peptide; HR = heptad repeat; MPER = membrane proximal external region

as the N- and C-helix, respectively) that form alpha-helical coiled-coil structures and are linked by a disulfide-bridged loop (Caffrey et al., 1998; Chan et al., 1997; Dubay et al., 1992; Lu et al., 1995; Weissenhorn et al., 1997); and a tryptophan-rich region referred to as the membrane-proximal external region (MPER) (Muñoz-Barroso et al., 1999; Salzwedel et al., 1999). The MPER is an important target for broadly neutralizing MAbs.

The primary function of the gp41 ectodomain is to drive virus-cell membrane fusion. The gp41 fusion peptide (FP) is normally buried in the gp120/gp41 complex, but following conformational rearrangements induced by receptor and coreceptor binding, FP is exposed and inserts into the target cell membrane. The three HR1 domains then fold into a stable six-helix bundle in an anti-parallel fashion with the three HR2 domains (Figure 6) to bring the virus and cell into close proximity to allow fusion to occur (Chan et al., 1997; Weissenhorn et al., 1997). The mechanism by which the actual fusion step occurs is not well elucidated. However, peptide analogs of both the fusion peptide and the MPER have been demonstrated to both disorder and reduce the bending modulus for lipid bilayers, and thus may lower the free energy barrier to membrane fusion, a decidedly non-spontaneous process (Greenwood et al., 2008; Tristram-Nagle et al., 2010; Tristram-Nagle and Nagle, 2007).

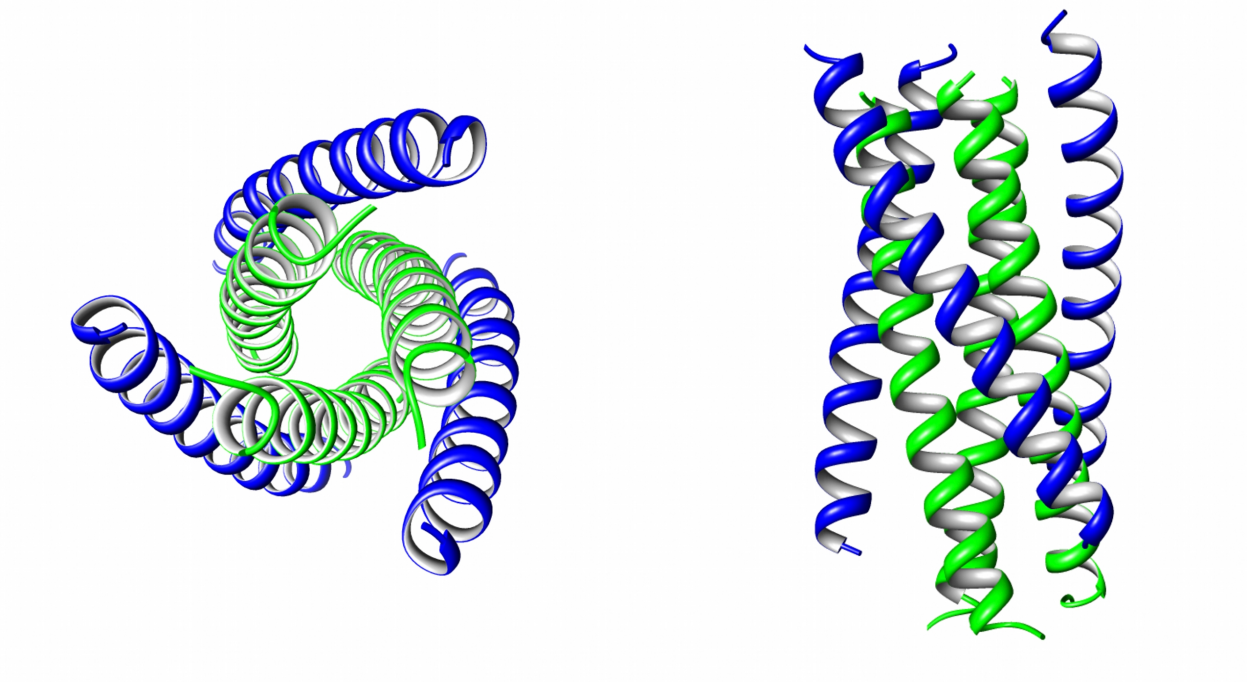

Figure 6. gp41 ectodomain six helix bundle in fusion-competent state

The gp41 six helix bundle of gp41 is shown in a bottom-up view (from the cell membrane perspective; left) and a side perspective (right), where the N-terminus of HR1 is at the top, and the N-terminus of HR2 is toward the bottom. HR1 and HR2 are colored as in Figure 5. Figure made in Chimera using PDB coordinates 1AIK.

gp41 Membrane-Spanning Domain (MSD) In contrast to the gp41 ectodomain, the gp41 MSD remains, despite much study, structurally enigmatic. Early studies identified a hydrophobic 25 amino acid sequence that acted as a ribosomal stop-transfer signal, leading to the classical identification of the gp41 ectodomain, MSD, and intracytoplasmic tail (Haffar et al., 1988). The MSD amino acid sequence is highly conserved among HIV isolates, and contains a conserved GXXXG helix-helix interaction motif and an absolutely conserved midspan arginine residue. The classical structural view of the MSD is that of a single membrane spanning alpha helix. The presence of the midspan arginine, as well as results from both previous and more recent studies, however, lead to conflicting views of the sequence and structural characteristics of the MSD in contrast to the classical definition. These studies are briefly discussed below.

Prior studies attempting to generate antibodies (in rabbits) to peptides spanning the full Env protein (gp120 and gp41) identified a strongly reactive peptide derived from positions 728 to 745 that is referred to as the Kennedy epitope (Chanh et al., 1986; Kennedy et al., 1986). This sequence is located in what is now known as the CTT, but is classically thought of as the intracytoplasmic tail. Intriguingly, the sera directed at the Kennedy epitope strongly neutralized virus (Chanh et al., 1986), suggesting an external localization of the epitope on the virion, as antibody cannot cross intact lipid membranes. These results were largely dismissed in light of the identification of the stop-transfer signal and the division of gp41 into the classical external, membrane-spanning, and intracytoplasmic domains (Haffar et al., 1988). More recently, however, accumulating evidence also points to possibly different MSD sequence and/or structure. Studies examining the exposure of the Kennedy epitope on the surface of Env-

expressing cells suggest that under some circumstances the Kennedy epitope is extracellularly exposed, indicating a need for either additional MSD sequences, or an alternative structure for the MSD other than the traditionally-held alpha-helix (Cheung et al., 2005; Cleveland et al., 2003; Heap et al., 2005; Reading et al., 2003). The two proposed MSD models are shown in Figure 7.

One of the most successful techniques for examining the structure of the MSD has been the use of molecular dynamics (MD) simulations. Two recent studies use MD simulations to examine the stability of monomeric and trimeric forms of the MSD in lipid membranes (Kim et al., 2009) and conformational variation in the MSD (Gangupomu and Abrams, 2010). A general caveat of these techniques, however, is that the starting assumptions have a major impact on the outcome. As such, both studies begin with the assumption that the MSD sequence adopts a helical conformation in the membrane. In general, this is a reasonable assumption given that the most energetically-favorable conformation of a peptide in a membrane is an alpha helix due to the internal stabilization of all potential hydrogen-bonding partners (White and Wimley, 1999). It does not, however, account for the experimental data that suggests alternative gp41 topologies, and thus alternative sequences or structures for the MSD. With that understanding, both studies provide interesting insights into the potential structure of the MSD. Engelman and colleagues demonstrated a remarkable stability of the HIV MSD sequence (using FIMIVGGLVGLRIVFAVLSI) as a monomeric helix in the lipid membrane, but rapid unfolding of the helix in water (Kim et al., 2009). They also

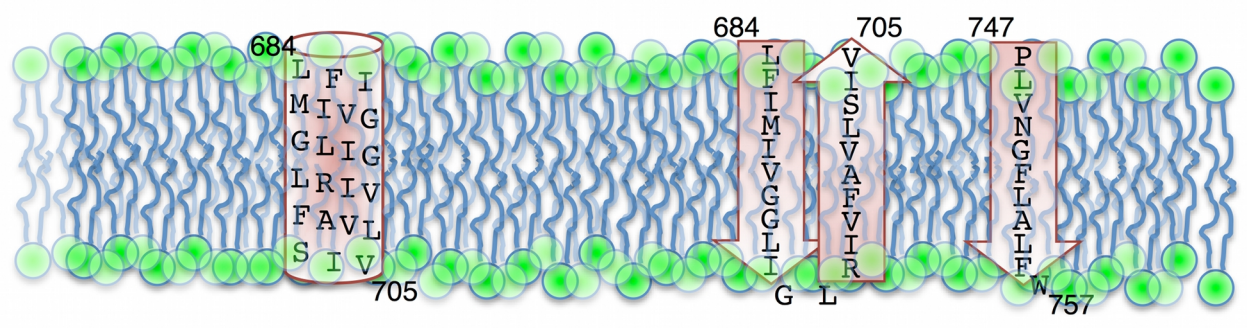

Figure 7. Proposed models for the membrane-spanning sequences of HIV gp41

The two models for the HIV gp41 MSD are shown schematically using the sequence from HIV-1 89.6. (Left) The traditional model, with the MSD as a single alpha helix. (Right) The alternative model proposed by Hollier and Dimmock (Hollier and Dimmock, 2005) with three membrane-spanning beta sheets. Residues are numbered according to the standard HXB2 sequence, and residues 706 - 746 are left out for simplicity.

demonstrated conformational stability of a right-handed helical bundle that increased further when the central arginine residues were deprotonated (Kim et al., 2009). These results suggest that a right-handed helical bundle with deprotonated arginine residues is the most stable trimeric conformation for the gp41 MSD sequence. In contrast, Gangupomu and Abrams examined the conformational flexibility of a monomeric MSD sequence (using KLFIMIVGGLVGLRIVFAVLSIVNRVR). Their findings indicate that the MSD monomer forms a stable tilted helix in the membrane, but that metastable states are formed when the midspan arginine "snorkels" to either side of the membrane (Gangupomu and Abrams, 2010). When the arginine snorkels towards the outer leaflet, the N-terminal end of the MSD is unfolded. However, when the midspan arginine snorkels towards the inner leaflet the MSD kinks at phenylalanine 697, and the remaining 10 amino acids form a helix along the inner leaflet at the water-membrane interface (Gangupomu and Abrams, 2010). Collectively, and importantly, these MD simulation results suggest multiple (meta)stable conformations of the MSD, even in an alpha-helical state.

Functionally, the primary role of the MSD has been thought to be to anchor the Env protein to the cellular and viral membrane. Additional studies have identified a role for the MSD in overall Env function. In particular, mutations in the MSD have been shown to have an effect on Env-mediate fusion such that mutations of conserved residues in the MSD led to decreased cell-cell fusion and decreased viral infection (Kondo et al., 2010; Shang and Hunter, 2010; Shang et al., 2008; Yue et al., 2009). The mechanism(s) by which these mutations act to decrease fusion are not well understood.

gp41 C-terminal tail

Structure Of the three gp41 domains, the structure of the gp41 CTT is the least understood. Results from early topogenesis studies led to the view of gp41 (and thus Env as a whole) as a type I membrane protein, with an extracellular N-terminus, a single MSD, and an approximately 150 amino acid long cytoplasmic CTT (Haffar et al., 1988). Comparisons of this model to other retroviral Env proteins supported this view, with a major difference in the length of the proposed cytoplasmic CTT. For most retroviral Env proteins, the cytoplasmic CTT sequences are generally between 20 - 40 amino acids long (Checkley et al., 2011). The presence of a very long CTT is not unique to HIV, as other lentiviruses such as the simian immunodeficiency virus (SIV) and the equine infectious anemia virus (EIAV) have similarly long CTTs, at 150 and 200 amino acids, respectively (Rushlow et al., 1986). The presence of a long CTT in most lentiviruses suggests an important functional role, as viruses do not generally replicate non-functional sequences. This functional importance is supported by the finding that truncation of the CTT leads to *in vivo* suppression of viral replication in animal models (Shacklett et al., 2000). A decade earlier, however, studies demonstrated that the CTT was dispensable for *in vitro* viral replication (Chakrabarti et al., 1989; Hirsch et al., 1989; Kodama et al., 1989), leading to a long-held view that the CTT was not functionally important. This view was subsequently popularized by the finding that truncation of the CTT led to increased Env incorporation into the virion, which was an important consideration at the time as a means by which to boost the anti-Env immune response in vaccine studies. It is now generally accepted that the CTT plays multiple important functional roles in the virus life cycle. Structurally, very little is still known, and no full-

length atomic level structures exist for any CTT sequences.

Some CTT sequences have been studied structurally. The most well-studied domains, the lentivirus lytic peptides (LLPs), are sequences that were initially identified by sequence scanning as having extraordinarily high hydrophobic moments (Eisenberg and Wesson, 1990). Subsequent studies on peptide analogs of these domains demonstrated high levels of structural similarity with naturally occurring cytolytic peptides (Miller et al., 1991) as well as the ability of the peptides to alter cell membrane permeability (Miller et al., 1993). Peptide analogs of these domains have been demonstrated to be generally unstructured in aqueous buffer, but to rapidly adopt amphipathic alpha-helical structure in membrane or membrane-mimetic environments as determined by circular dichroism spectroscopy (Chernomordik et al., 1994; Kliger and Shai, 1997; Srinivas et al., 1992). There are no atomic level structures of the peptides interacting with membranes or membrane mimetics.

Topology and function

Contradicting data and competing topological models As the intracytoplasmic localization of the CTT has been predominantly reinforced implicitly with functional data, the topology and function of the CTT will be presented together. As alluded to above, the topology of the CTT has long been considered to have an entirely cytoplasmic localization. Accumulating recent evidence suggests that this may not be the case, or is at least not always the case. Initial studies suggestive of a non-cytoplasmic localization for the CTT were not explicitly performed to determine the CTT topology, and thus appear to have been largely ignored. These studies found that serum from a rabbit

immunized with a gp41 peptide (residues 728 - 745, the Kennedy epitope) bound to HIV Env, and that serum from HIV-1 infected patients bound the synthetic peptide (Kennedy et al., 1986). More importantly, subsequent studies demonstrated that anti-Kennedy epitope serum could specifically neutralize HIV *in vitro* (Chanh et al., 1986). Since antibody cannot cross intact lipid membranes, these results implied the localization of the Kennedy epitope on the outer surface of the virion, in direct contrast to the presumed intravirion localization. A few years later, topogenesis studies suggested the organization of Env as a type I membrane protein (Haffar et al., 1988), a view that was predominant and was implicitly reinforced by a number of experimental functional studies.

Nearly all functional studies to date reinforce the traditional cytoplasmically-localized CTT model. This is largely due to genetic evidence demonstrating functional, but not physical interactions of the CTT with intracellular or intravirion partners. These functional interactions have been found to regulate: (i) Env incorporation into viral particles (Freed and Martin, 1995, 1996; Jiang and Aiken, 2007; Murakami, 2008; Murakami and Freed, 2000a, b); (ii) virion maturation (Jiang and Aiken, 2007; Joyner et al., 2011; Kol et al., 2007; Wyma et al., 2004); and, (iii) endocytosis of Env from the cell surface to late endosomes (Byland et al., 2007; Ohno et al., 1997).

Virion Env incorporation In the mid-1990s, studies were published that demonstrated a functionally important role for the CTT. These studies are crucial to the current CTT field in that they established conclusively that the CTT was of functional importance in contrast to the view that arose from studies published in the late 1980s which

demonstrated that the CTT was dispensable for virus replication (Chakrabarti et al., 1989; Hirsch et al., 1989; Kodama et al., 1989). In addition, these studies demonstrated increased Env content in virions containing a CTT-deleted Env and proposed that deletion of the CTT would be a means to boost the anti-Env immune response through the increase in Env per virion. This was the prevailing attitude until seminal studies by Freed and Martin demonstrated that the CTT was implicated in the incorporation of Env into virions through interactions with the MA domain of Gag. Their first study used site-directed mutagenesis to demonstrate that mutations in specific amino acids in Gag MA resulted in deficiencies in incorporation of Env with a full-length CTT (Freed and Martin, 1995). The MA mutations did not, however, lead to reduced incorporation of heterologous retroviral Env proteins with naturally short CTT sequences. Further, specific truncation of CTT sequences to seven or 47 residues (from the original 150) reversed the Env-incorporation block imposed by the MA mutations. This paper demonstrated for the first time conclusive evidence of a functional role for the CTT through direct or indirect interactions with an intracellular partner, Gag MA.

The role and localization of sequences important in the MA-CTT interaction has been further elucidated. A subsequent paper by the same authors demonstrated that truncating up to 56 residues from the CTT (leaving a 94 residue CTT) resulted in no Env incorporation in MA viruses containing mutations leading to deficiencies in Env incorporation, but that truncations of 93 amino acids or greater (leaving only 57 amino acid CTT) relieved the block caused by MA mutations, resulting in efficient Env incorporation (Freed and Martin, 1996). Later papers from this group localized the MA-CTT interaction to LLP2 in the CTT (Murakami and Freed, 2000a) and further

demonstrated that the interaction was crucial for Env virion incorporation in a cell-type specific manner (Murakami and Freed, 2000b).

The above studies, while important in establishing a functional role for the CTT, do not provide evidence for a direct interaction of the CTT with MA; in fact, there is to date no experimental evidence for a direct MA-CTT interaction. The closest evidence for a stable interaction between Gag and the Env CTT also suggested a role of the CTT in HIV virion maturation, and as such, will be described below.

Virion maturation Aiken and colleagues have come the closest to demonstrating a direct interaction of the Env CTT with MA by examining the effect of pelleting immature HIV virus particles (in which the $Pr55^{Gag}$ has not been cleaved) through detergent (Wyma et al., 2000). Pelleting particles through detergent strips away the viral lipid membrane in which the CTT is embedded, leaving viral protein cores. Their reasoning was if the CTT and $Pr55^{Gag}$ interact, then gp41 should be associated with the pelleted viral cores. They found that the pelleted cores retained the major fraction of the gp41 found in untreated virions (Wyma et al., 2000). Virions treated similarly but with a truncated CTT did not remain associated with the pelleted cores (Wyma et al., 2000). This study does not preclude, however, the presence of a mediating protein that facilitates the interaction between gp41 CTT and $PR55^{Gag}$. Interestingly, a previous study using mature viral particles for a similar analysis demonstrated no association of the CTT with MA (Kotov et al., 1999). Thus, taken together these results suggested that the association of the CTT with $Pr55^{Gag}$ was dependent on the maturation state of the virion.

This maturation-dependent association of the CTT with Gag has been extended to

examine the role of the CTT in viral fusion. It is a well-established fact that HIV particles are not infectious until proteolytic cleavage of the $Pr55^{Gag}$ into its constituent domains (MA, CA, and NC, predominantly). Using a reporter assay to measure virus-cell fusion, Wyma et al. demonstrate that immature virions are less fusogenic than mature virions in a manner that is dependent on the CTT, as truncations of the CTT resulted in identical fusogenicity of immature and mature viral particles (Wyma et al., 2004). More recent results suggest that the extreme C-terminus of the CTT, LLP1, modulates the maturation dependence of infectivity, while deletion of this region does not affect the incorporation of Env into immature viral particles (Jiang and Aiken, 2007). The CTT has also been shown to modulate the mechanical stability of immature virus particles relative to mature virus particles, presumably through its interactions with $Pr55^{Gag}$ (Kol et al., 2007). These results collectively demonstrate that the interaction of the CTT and Gag have specific effects on the infectivity and structural stability of HIV particles.

Env endocytosis Another functional role of the CTT has implicitly reinforced the traditional topological model. The gp41 CTT of both SIV and HIV have long been known to contain endocytic signals. The first was demonstrated in SIV, where a consensus YXXΦ motif in the CTT was shown to interact with members of the adaptor protein medium chain family (Ohno et al., 1997). It has also been demonstrated that CTT sequences interact specifically with the AP-2 clathrin adaptor (Boge et al., 1998). Most recently, Byland et al. demonstrated that the CTT contains two functional endocytic signals: a ^{711}GYXXΦ motif located near the N-terminus of the CTT, and a dileucine motif at the extreme C-terminus (Byland et al., 2007). Their results indicate that in order to

completely abolish endocytosis of Env from the cell surface to late endosomes, both motifs must be mutated, suggesting that both the N- and C-termini of the CTT are cytoplasmically-localized.

Overall, known CTT functions in Env virion incorporation, virion maturation, and endocytosis of Env from the cell surface all provide support for the traditional intracytoplasmic (intravirion) model for the CTT. Figure 8 presents an overview of the known functionalities associated with the CTT sequences.

Evidence for alternative CTT topologies

Surface exposure of the Kennedy epitope The derivation of an alternative topological model for the CTT comes from a mixture of indirect (inferred from neutralization of HIV infection by anti-CTT antibodies) and direct biochemical evidence, as opposed to the indirect evidence supporting the tradition model. The earliest studies by Kennedy and colleagues, discussed above, demonstrated that antiserum directed against a CTT peptide both bound gp160, and most importantly, neutralized HIV *in vitro* (Chanh et al., 1986; Kennedy et al., 1986). These are the earliest studies where the

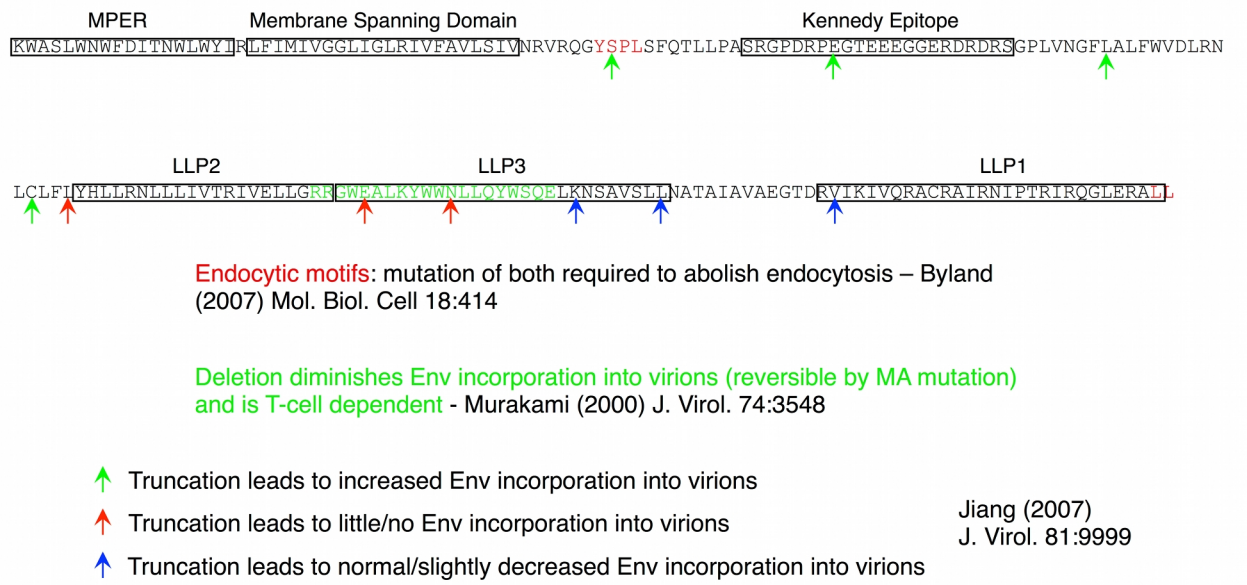

Figure 8. Sites of functional consequence in genetic studies of the HIV gp41 CTT

A visual overview of a subset of the sites that demonstrate functional alterations when mutated in genetic studies of the gp41 CTT. The wide distribution of sites over the entirety of the CTT sequence suggests the internal localization of the CTT, consistent with the traditional model. The sequence is from HIV-1 89.6.

results suggest an external localization of CTT sequences. However, for reasons that have also been detailed above, the traditional model was, and still remains, the predominating topological model.

In 2003, the mantle for the exposure of CTT sequences was assumed by a professor at the University of Warwick in the United Kingdom, Dr. Nigel Dimmock. Over three years and five manuscripts, his group provided the most extensive, although at times contradictory, evidence for the extravirion and/or extracellular localization of CTT sequences, and it is his group that proposed the alternative topological model presented in Figure 9.

The initial evidence for external localization of Kennedy epitope sequences were the result of antibody neutralization and binding studies of anti-CTT antibodies to viral particles (Cleveland et al., 2003). This study used epitope-purified antiserum (against the epitope ^{746}ERDRD^{750}, and termed EPES) to demonstrate high levels of neutralization (Concentration necessary to achieve 50% neutralization (NC_{50}) = 0.3 µg/ml) of HIV *in vitro*. Importantly, EPES did not have any neutralizing activity against viruses with a CTT-deleted Env. They further demonstrated that EPES could bind to intact viral particles (as evidenced by lack of binding by anti-MA and anti-CA antibodies), and that this binding was abolished by protease treatment of the particles (Cleveland et al., 2003). This report provided the first direct evidence for exposure of CTT sequences on viral particles.

A second study from Dimmock's group described a monoclonal antibody, termed SAR1 that is directed against the Kennedy epitope, as well as its neutralization activity

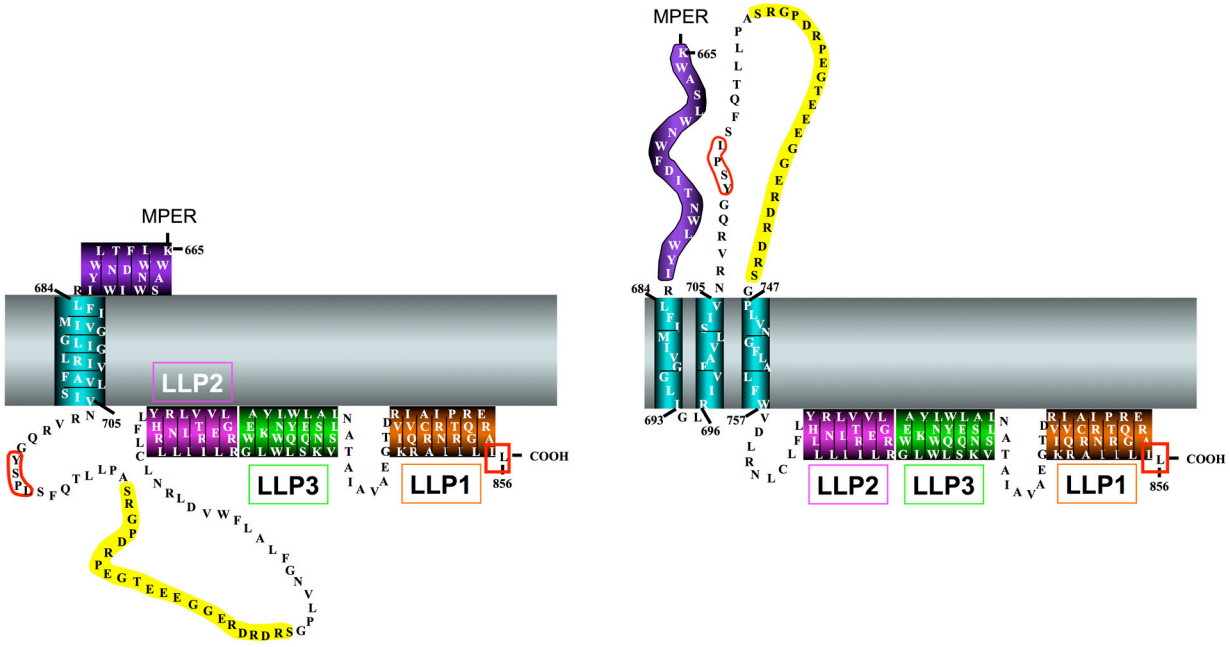

Figure 9. Proposed topology models for HIV gp41 CTT

The two predominant CTT topology models are presented. (Left) Traditional CTT topology model, with completely intracytoplasmic CTT. (Right) Model proposed by Hollier and Dimmock (Hollier and Dimmock, 2005), with the Kennedy epitope (yellow) extracellularly localized. Transmembrane domains are as presented in detail in Figure 7. Red boxes indicate functional endocytic sequences as determined in (Byland et al., 2007).

(Reading et al., 2003). SAR1 was derived by immunizing a mouse with a plant virus engineered to express the ^{745}GERDRDR^{751} sequence from the Kennedy epitope in the CTT. The resulting antibody was found to bind to both soluble peptides and proteins containing the GERDRDR sequence. More interesting, however, was that SAR1 bound to cells infected with all different HIV strains tested, but only to some virions representing the same strains. Moreover, SAR1 did not exhibit neutralization activity in a standard neutralization assay, but was able to neutralize effectively in a post-attachment neutralization (PAN) assay. In a standard neutralization assay, antibody and virus are incubated together prior to adding to the cells to allow the interaction to reach equilibrium before addition to target cells; neutralizing antibodies reduce viral infection in a concentration-dependent manner. In a PAN assay, however, antibody, virus, and target cells are incubated together, but at temperatures that do not support virus-cell fusion (usually ≤ 31 °C) for a defined period of time before washing free virus from the cells and incubating at 37 °C to allow infection to occur (or not, if the antibody neutralizes) (Reading et al., 2003). The inability of SAR1 to neutralize in a traditional assay, but exhibit PAN activity, suggested a transient exposure of the Kennedy epitope during the fusion process. The mechanism of the PAN activity of SAR1 was confirmed in a subsequent paper to act through the blocking of virus-cell fusion (Heap et al., 2005). Reading, et al. also demonstrated that SAR1 acted to block the production of infectious progeny virus from infected cells (Reading et al., 2003). These results provided increased evidence for an alternative topology for the CTT, where the Kennedy epitope is exposed to antibody binding, although not universally as SAR1 did not bind to virions from all HIV strains tested.

Finally, in an effort to provide a condensation of results from these studies and to provide a working topological model explaining their results, Hollier and Dimmock used sequence analyses to provide theoretical support to their observed exposure of CTT sequences. Using a combination of secondary structure predictions with hydrophobicity analyses, Hollier and Dimmock proposed that the CTT can form a tail loop structure supported by three membrane-spanning beta sheets that result in the extracellular localization of the Kennedy epitope (Hollier and Dimmock, 2005). This is the alternative topological model presented in Figure 9. Hollier and Dimmock acknowledge that their proposed topological model positions the N-terminal GYXXΦ endocytic signal in an extracellular position, implying that it cannot be functional. As discussed above, this model is in direct contrast with subsequent results demonstrating the GYXXΦ signal to be active in AP-2 mediated endocytosis (Byland et al., 2007). Hollier and Dimmock propose that the majority of cell surface Env is in a state similar to the traditional topological model, with a single MSD, but that a minor population exists in their newly proposed topological state, and that it is this Kennedy-epitope exposed Env that is preferentially incorporated into progeny virions (Hollier and Dimmock, 2005). It is important to note that at the time of its proposal the alternative model was consistent with a majority of published studies, both functional studies discussed in the previous section as well as their own biochemical data, excluding the inconsistent HIV virion binding. A major argument against this model is the improbable existence of a MSD composed entirely of beta-sheet with an odd number of membrane-interacting sequences. One of the "construction rules" of beta sheet membrane proteins is that they must have an even number of MSDs as beta sheets hydrogen bond most efficiently in

an anti-parallel fashion (Schulz, 2000). An odd number of beta sheet MSD sequences would leave a lone beta strand and its hydrogen-bonding backbone amide and carboxyl groups exposed to the hydrocarbon interior of the membrane, an energetically unfavorable situation (Schulz, 2000). This untenable arrangement would be further compounded upon Env trimerization, where there would now be three naked beta sheets in the interior of the membrane, all arranged in a parallel fashion and unable to relieve the unfavorable interactions.

Dynamic exposure of LLP2 sequences In addition to the evidence for possible Kennedy epitope exposure, recent studies present data supporting the exposure of additional CTT sequences, in particular the LLP2 domain. A study by Lu, et al. determined the exposure of CTT sequences in Env-expressing cells under native conditions as well as during the fusion process (Lu et al., 2008). Using antibodies directed at the C-terminal 90 amino acids of the CTT, and to the LLP2 region only, Lu et al. determined that no exposure of the 90 C-terminal amino acids were detected under normal conditions, nor did they observe exposure during cell-cell fusion carried out at 37 °C as measured by flow cytometry. They did, however, observe LLP2 exposure during cell-cell fusion by slowing the fusion reaction by incubation at 31.5 °C (Lu et al., 2008). These same antibodies were also demonstrated to inhibit cell-cell fusion in a concentration-dependent manner at 31.5 °C, but not at 37 °C. Their results suggest a transient exposure of LLP2 sequences during membrane fusion, and are consistent with results demonstrating membrane association of LLP2 both pre- and post-fusion (Viard et al., 2008).

CTT influences overall Env structure The final known function of the CTT is its role in modulating overall Env structure. Data from these studies provide no direct topological implications and are included and discussed as further evidence supporting the functional importance of the CTT.

One of the earliest observations that the CTT could influence overall Env structure was provided by the insight that viruses with CTT-deleted Env proteins could infect target cells in a CD4-independent manner (Edinger et al., 1997). Until that time, the paradigm for HIV infection of target cells was that gp120 binding to CD4 was necessary to induce conformational changes that allowed binding to coreceptor. Edinger et al. demonstrated that viruses with a CTT-deleted Env were able to infect CD4 negative, coreceptor positive cells, suggesting a different conformation for the Env that was dependent on the CTT.

Initially, direct evidence for CTT-dependent alterations in Env structure was provided by differential reactivity between CTT-deleted and wild-type Env with conformationally-dependent antibodies. Edwards, et al. demonstrated that truncation of the CTT to 27 amino acids resulted in increased binding by monoclonal antibodies directed to both the CD4 binding site and the CD4-induced coreceptor binding site (Edwards et al., 2002). In Env with a full-length CTT, monoclonal antibodies directed to the coreceptor binding site only bound if the Env was preincubated with soluble CD4. The study also demonstrated differential reactivity of conformational monoclonal antibodies directed at the ectodomain of gp41 between CTT-truncated and full-length CTT (Edwards et al., 2002). This study was the first to demonstrate that the CTT

modulates the conformation of overall Env structure, even in the non-covalently attached gp120 that is located (presumably) on the opposite side of the membrane.

Additional, and arguably more interesting, data on this was published a few years later from our lab. Kalia et al. demonstrated that, like the Edwards study, alterations in the CTT could modulate the antigenic conformation of both gp120 and the gp41 ectodomain. Instead of utilizing large deletions in the CTT, however, Kalia demonstrated that mutation of two conserved arginine residues in LLP2 to glutamate was sufficient to alter the conformation of both gp120 and the gp41 ectodomain on the surface of Env-expressing cells (Kalia et al., 2005). The mutations were shown in a previous manuscript to have no effect on the levels of virion Env incorporation or viral replication, but did decrease the efficiency of cell-cell fusion (Kalia et al., 2003). In addition to demonstrating antigenic distinctions between wild-type and mutant Env on the cell surface, differences were seen in the viral sensitivity to antibody-mediated neutralization by antibodies directed at the CD4 binding site, with the mutant virus demonstrating an approximately 40-fold decrease in neutralization sensitivity (Kalia et al., 2005). This study importantly demonstrated that point mutations in the CTT were sufficient to alter overall Env conformation, and equally importantly, provided the first glint of information regarding the critical nature of the conserved arginine residues in the LLP regions.

A very recent paper has extended the study of CTT-dependent alterations in overall Env antigenicity from the cell surface to the surface of viral particles (Joyner et al., 2011). Under the guise of maturation-induced cloaking of neutralization epitopes, Joyner et al. demonstrated that Env on immature virions (protease deleted) reacted differently to a number of conformationally-dependent monoclonal antibodies in a

manner that was dependent on the presence or absence of the CTT (Joyner et al., 2011). This study provided the first direct evidence that the CTT plays a major role in modulating the conformation of the CTT in the virion in addition to on the cell surface.

There is also recent structural evidence for the role of the CTT in the modulation of overall Env conformation. Subramaniam and colleagues recently published the cryo-EM structures of CTT-deleted Env on the surface of SIV viral particles (White et al., 2010). In comparison to wild-type virus containing the full-length CTT, CTT-deleted Env existed in a naturally "open" state, displaying large differences in the localization of electron density that are consistent with the conformational changes normally associated with CD4 binding (Figure 10). This study, finally, provides the first conclusive structural evidence that the CTT sequences, aside from any functional roles, also plays a critical role in the modulation of overall Env structure and function.

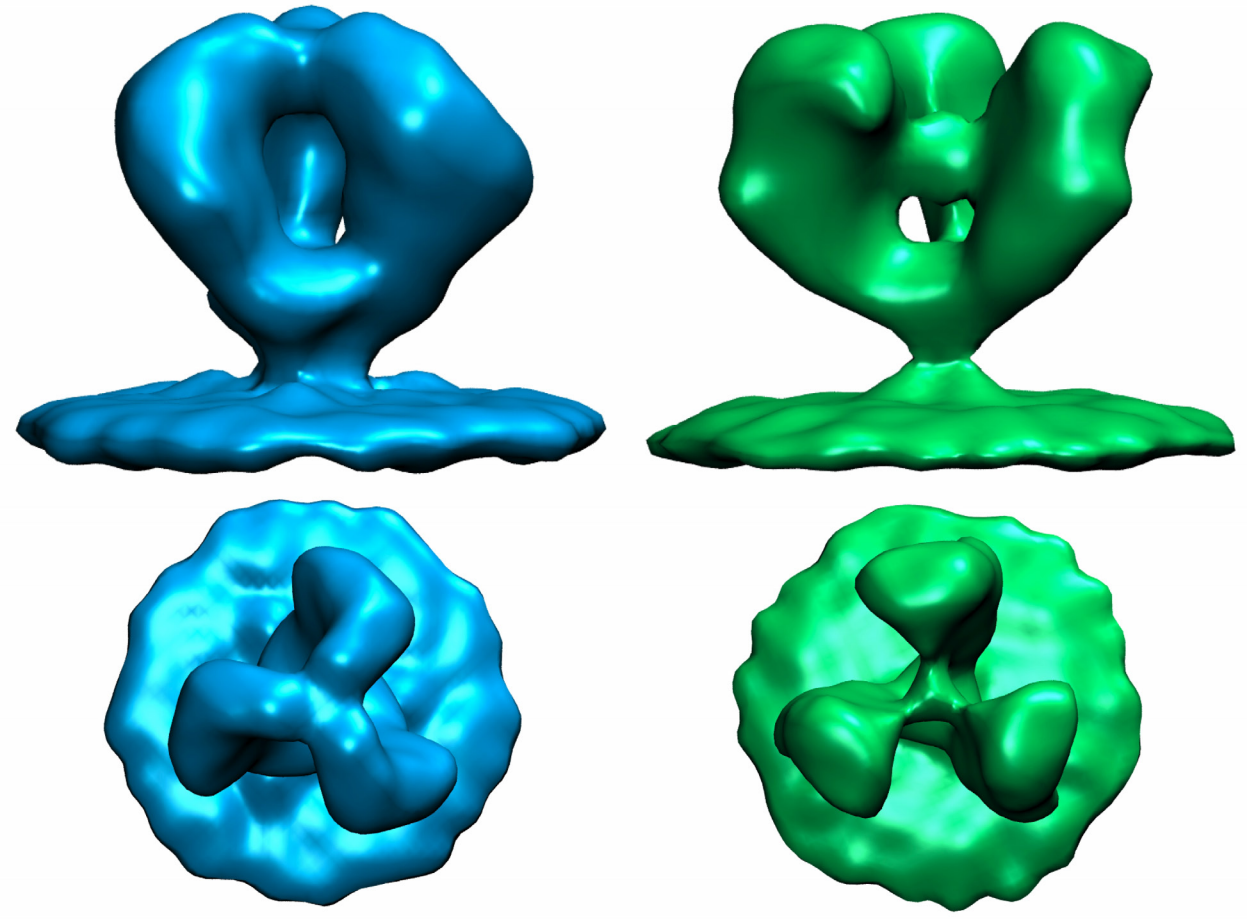

Figure 10. CTT-dependence of Env conformation

Cryo-EM structures of SIVmacE11S (full-length Env; left) and SIV CP-MAC (CTT-truncated Env; right). The CTT-truncated Env displays large alterations in the positioning of the gp120 units, as a function of both the monomeric unit as well as in the overall quaternary arrangement. Figure made using VMD 1.8.7 and EM Database coordinates EMD-5272 (E11S) and EMD-5274 (CP-MAC). Contour levels were set as recommended.

CTT conclusions I hope that it is clear to the reader at this point that the questions and data surrounding the topology of the gp41 CTT present a complicated, and often conflicting, picture. While the majority of functional data support the traditional idea that the CTT sequences must be cytoplasmically localized, accumulating biochemical data suggest that some of the CTT sequences are localized to the exterior of the cellular or viral membrane, if not always, then at least some of the time. The conflicting data provided what we believed to be an excellent space to attempt to find answers explaining the topological confusion regarding the gp41 CTT. Before discussing our hypotheses regarding CTT topology, however, it is necessary to provide some background on the theoretical considerations of membrane protein topogenesis and folding.

1.2 MEMBRANE PROTEIN TOPOLOGY

In vivo, membrane protein topology is derived from the cotranslational insertion of the protein (regions) into the lipid membrane by the cellular translation and translocation machinery (White and von Heijne, 2008). While this sounds like a relatively straightforward process, insomuch as it can be described in a single sentence, resulting membrane protein topologies can be anything but simple. The study of membrane protein topology has evolved from basic descriptions of "on which side of the membrane do sequences reside" to concepts including topological distributions of single protein sequences with differing numbers of transmembrane domains, dual-topology proteins that arrange in opposite topologies in equal stoichiometry, and dynamic reorientation of

protein sequences across a membrane in response to changes in membrane lipid composition (reviewed in (Dowhan and Bogdanov, 2009; von Heijne, 2006; White and von Heijne, 2008)). In the years since the proposal of the traditional and alternative CTT topological models, the understanding of membrane protein topology has undergone a "quiet revolution" that has not yet reached mainstream biology. Thus, to be considered valid, any proposed model for CTT topology must account for all experimental data, both functional and biochemical, as well as conform to the principles of membrane protein folding and topology. As has been discussed above, the alternative model with its three membrane-spanning beta strands does not conform to the known principles for beta sheet membrane proteins and cannot be considered a valid model. As such, to provide proper context for the data surrounding the topology of the gp41 CTT, the following section will provide an overview of the current understanding of alpha-helical membrane protein topology.

1.2.1 Membrane protein classifications

It is perhaps most useful to start with a brief description of the basic classifications of membrane proteins and their topology. Membrane proteins belong to two broad classes: alpha-helical (bundle) membrane proteins and beta barrel membrane proteins (Elofsson and von Heijne, 2007; White and Wimley, 1999). As their names suggest, helical membrane proteins span the lipid bilayer as a single (or bundle of) alpha helix(es), while the beta barrel proteins are constructed of an even number (8 - 22) of membrane-spanning beta sheets arranged in an anti-parallel fashion (Elofsson and von Heijne, 2007). It is likely that helices and anti-parallel beta sheets are the major secondary

structural elements of membrane proteins as these arrangements minimize the energy of the protein-lipid interaction by satisfying the hydrogen bonding potential of the peptide bond (Figure 11) (White and Wimley, 1999). Additionally, depending on the number of membrane-spanning segments, membrane proteins can be classified as monotopic (having an in-plane anchor or a re-entrant loop), bitopic (one membrane spanning domain), or polytopic (two or more membrane spanning segments) (Blobel, 1980).

Membrane proteins can also be classified into three distinct groups based on the localization of their termini: (i) Type I membrane proteins have a cleavable membrane-localization signal peptide and an exoplasmic N-terminus and cytoplasmic C-terminus; (ii) Type II membrane proteins contain a signal anchoring domain and a cytoplasmic N-terminus and exoplasmic C-terminus: and, (iii) Type III membrane proteins have a reverse signal anchor and an exoplasmic N-terminus and a cytoplasmic C-terminus, similar to Type I proteins (White and von Heijne, 2004). These topological classifications are generally held to describe both single-membrane spanning proteins, as well as more complex multi-domain proteins, where ultimate topology depends on the number of transmembrane domains (Goder and Spiess, 2001). HIV Env is generally considered to be a bitopic type I membrane protein with an extreme cleavable N-terminal signal before the soluble gp120, and a single membrane spanning domain in gp41.

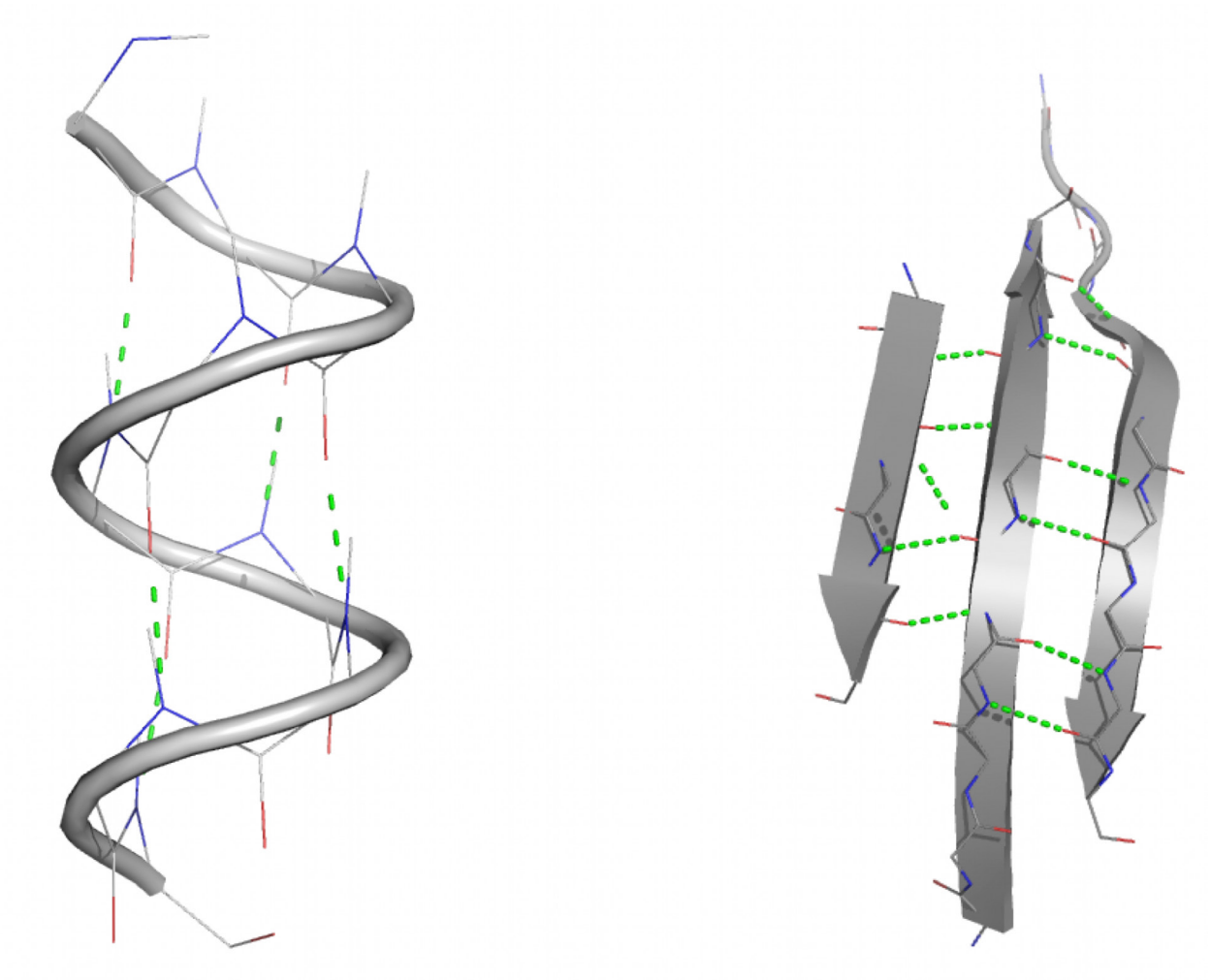

Figure 11. Hydrogen bonding between peptide main chain groups

Hydrogen bonding (green dashed lines) is demonstrated between backbone amide nitrogen (blue) and carbonyl oxygen (red). (Left) Hydrogen bonding between i and i+4 residues in an alpha helix. (Right) Hydrogen bonding between residues in anti-parallel beta sheets. Image created using MacPymol.

1.2.2 Membrane protein assembly and topogenic signals

Membrane protein assembly and topogenesis requires the coordinated activity of several events (reviewed in (von Heijne, 2006; White and von Heijne, 2004, 2008). Very briefly, the major steps that apply to all transmembrane proteins are the recognition and targeting of the nascent translating chain to the membrane-localized translocation machinery, including the translocon (Sec61 in eukaryotes), which coordinates the selection, orientation, and membrane insertion of transmembrane helices into the bilayer milieu. Bitopic proteins are generally considered "finished" at this step, while polytopic membrane proteins undergo further helical packing interactions to attain their properly folded tertiary structure. Details of the process as they potentially impact topology will be added in following sections.

For membrane proteins, topology is ultimately governed by a set of topogenic signals that are encoded in the amino acid sequence. Additional information is captured by rules that have been determined by statistical analyses of known transmembrane proteins. The main topogenic signals are detailed below and presented in Figure 12. The major signal for membrane integration of a protein segment is its overall hydrophobicity (von Heijne, 2006; White and von Heijne, 2008). The orientation of a transmembrane domain, however, is influenced by a complex confluence of a number of signals. These signals include: (i) the presence and orientation of charged residues flanking the helix, with positive charges localizing internally (Goder et al., 2004; Goder

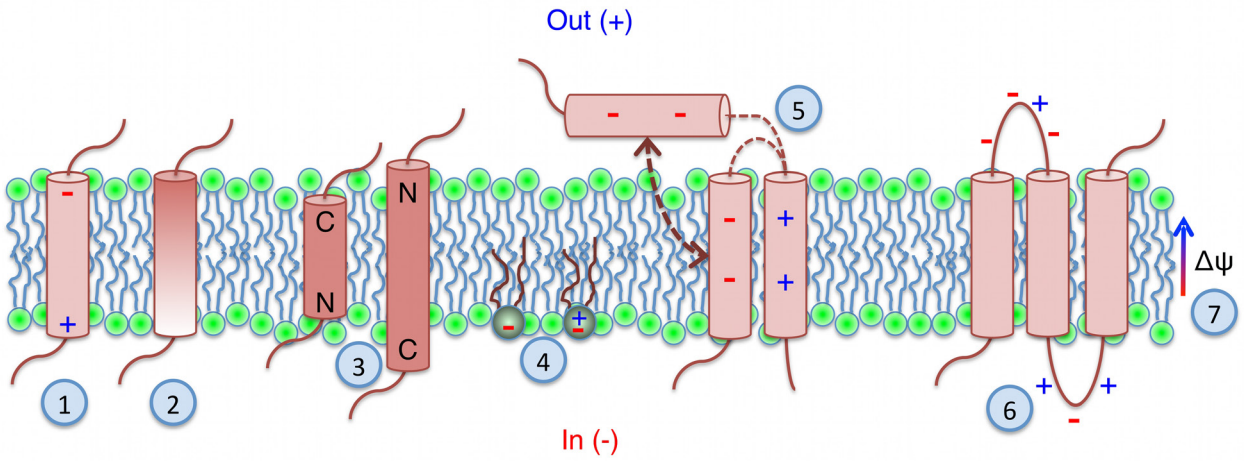

Figure 12. Sequence properties influence final topology

The main determinant of transmembrane insertion is the overall hydrophobicity of a sequence (depicted as a gradient of red, with darker hues indicating greater hydrophobicity). Additional factors that exert an influence on membrane protein topology are: (1) the presence and orientation of charges flanking the transmembrane region; (2) a hydrophobic gradient that leads to the more hydrophobic end being oriented out; (3) transmembrane region length, where for very hydrophobic sequences, short transmembrane regions orient C-terminal end out, while long sequences orient N-terminal end out; (4) the presence and relative abundance of charged lipids; (5) long-range interactions such as salt-bridging can lead to the insertion of a charged helix; (6) the positive inside rule; and (7) the positive out membrane potential leads to translocation of negative sequences.

and Spiess, 2003); (ii) the hydrophobic gradient of the domain, with the more hydrophobic end oriented outward (Harley et al., 1998); and, (iii) preferential orientation of highly hydrophobic transmembrane domains, with the C-terminus oriented out if the domain is short, and the N-terminus out if it is long (Cao et al., 1994). Long-range interactions can also influence the topology during polytopic protein folding, where salt-bridge formation can stabilize the transmembrane orientation of an otherwise poorly-inserting charged domain (Bogdanov et al., 2008). Finally, the generally positive outward membrane potential favors the translocation of net negative domains and the retention of net positive domains (Dowhan and Bogdanov, 2009).

Additionally, topological rules for membrane protein topology have been gleaned from statistical analyses of membrane protein structures. The most well known is the positive-inside rule. This rule was identified as a four-fold overrepresentation of positively charged residues in the cytoplasmic, as opposed to non-cytoplasmic domains, of proteins (von Heijne, 1986). Subsequent studies have determined that the molecular mechanism(s) of action for the positive inside rule include charged residues flanking the transmembrane domain interacting directly with the translocon (Goder et al., 2004; Goder and Spiess, 2003), charged residues in the protein interacting with ionizable head groups on membrane phospholipids (Dowhan and Bogdanov, 2009), and the overall positive out cellular membrane potential leading to translocation of negative domains and retention of positively charged protein domains (Dowhan and Bogdanov, 2009). In addition to the skewed cytoplasmic:non-cytoplasmic distribution of positive charges, statistical analyses have identified distinct distribution profiles for all the amino acids relative to their position on transmembrane helices (Ulmschneider et al., 2005).

This study demonstrated that the frequency of the hydrophobic residues alanine, isoleucine, leucine, phenylalanine, and valine peaks at the center of the membrane, while the aromatic residues tyrosine and tryptophan (but not phenylalanine) peak at the membrane-water interface. In contrast, the polar and charged residues are largely absent from the membrane center. These rules have held up remarkably well as new membrane protein structures have been identified, and contribute positively to the prediction of transmembrane domains and topology in protein sequences of undetermined structure.

1.2.3 Transmembrane domain and topology prediction

One of the first modern techniques used to identify helical membrane proteins was to computationally scan protein sequences for (predominantly) hydrophobic stretches of amino acids long enough to span the membrane lipid bilayer (Henderson and Unwin, 1975; Khorana et al., 1979; Ovchinnikov et al., 1977, 1979; Tomita and Marchesi, 1975). Hydrophobicity scanning, and thus topological prediction, was performed by scanning the amino acid sequence of a protein using a moving average of a specified length, usually 10 - 20 residues (Engelman and Steitz, 1981; Kyte and Doolittle, 1982; von Heijne and Blomberg, 1979). Early hydrophobicity studies were performed using hydropathy scales derived from experimental studies of amino acid side chain analogs partitioning between water and hydrophobic solvents or calculations from accessible surface areas of amino acids in helical peptides (Engelman and Steitz, 1981; Kyte and Doolittle, 1982; von Heijne and Blomberg, 1979). Topological predictions from these data relied on the fact that transmembrane helices are on average more hydrophobic

than extramembranous loops. Results from these studies provided good information on what sequences were likely to be transmembrane helices, but the identification of sidedness was less accurate. The inclusion of the positive inside rule into topological prediction algorithms greatly increased the accuracy of sequence assignment to the appropriate side of the membrane (von Heijne, 1986, 1992).

An important recent development in the prediction of transmembrane helices is the determination of more relevant hydrophobicity scales. As stated above, the original hydrophobicity scales were derived from the partitioning free energy of amino acid side chain analogs into hydrophobic solvents (Kyte and Doolittle, 1982). These scales did not take into account the partitioning free energy for the entire amino acid, particularly in a peptide environment, including the polar peptide bond, as it would be present in a protein sequence. To this end, Wimley and White derived whole-residue hydrophobicity scales based on partitioning free energies for amino acids in host pentapeptides, at both a membrane interface as well as into *n*-octanol as a measure of partitioning into the apolar environment of the membrane center (White and Wimley, 1998, 1999; Wimley et al., 1996; Wimley and White, 1996). These scales generally predicted lower free energy barriers to partitioning of charged and polar amino acids than the older scales, while still maintaining the ability to reliably identify transmembrane segments (White and Wimley, 1999).

1.2.3.1 A biological hydrophobicity scale

All the predictive scales suffered from the same criticism, namely that they are based on the partitioning of elements (side chain analogs, amino acids, or peptides) solely between water and a hydrophobic (or interfacial) environment, a situation that is not

necessarily similar to what occurs *in vivo*. This criticism was elegantly addressed by the generation of a biological hydrophobicity scale by von Heijne and colleagues (Hessa et al., 2005a). Using a glycosylation reporter system to directly measure the efficiency of transmembrane insertion, they determined the apparent free energy, ΔG^{app}, for each amino acid in the middle of a transmembrane helix, as well as for all positions in a transmembrane helix (Hessa et al., 2005a). The results from these studies correlated very well with the Wimley-White predictive scale (R = 0.85), suggesting that the predictive scale based on peptide partitioning provides a good approximation of the *in vivo* situation (Hessa et al., 2005a). The major implication of this correlation is that the insertion of membrane proteins into the lipid bilayer by the protein translocation machinery takes place as a partitioning event governed by the thermodynamics of the protein-lipid interaction (Hessa et al., 2005a; von Heijne, 2006; White and von Heijne, 2008). Importantly, this suggests that transmembrane segments with $\Delta G^{app} \approx 0$ exist as a thermodynamic equilibrium between inserted and non-inserted states, leading to the potential for proteins with a distribution of topological states.

This biological hydrophobicity scale is currently considered to be the most accurate of the predictive scales. Importantly, it has been used recently to demonstrate that transmembrane helices can reposition during membrane folding (Kauko et al., 2010). Here, the authors noted that the transmembrane helices predicted using the position-dependent biological hydrophobicity scale did not match the position of the helices relative to the membrane in the crystal structure of the glutamate transporter homolog GltPh from *Pyrococcus horikoshii*. The authors demonstrated that three of the GltPh transmembrane helices are inefficiently recognized and inserted by the

translocon, but that they overlap with more hydrophobic segments with higher insertion efficiencies. They then postulated that tertiary and quaternary rearrangements (i.e. long-range interactions) must occur during folding-induced shifts in the positions of these helices relative to the membrane. This induced shift resulted in the insertion of polar and charged residues present on two of the helices into the membrane, presumably in such a manner that they do not directly contact the hydrocarbon interior of the membrane (Kauko et al., 2010). These results provoke an interesting concept in the structure of membrane proteins, i.e. that the structure inserted by the translocation machinery may not exactly match the protein's "final" functional structure.

1.2.4 Alternative protein topologies

An implicit assumption underlying the discussion of membrane protein topology, as presented above, is that translation and translocation of a protein sequence results in a single protein structure with a defined, stable topology. The goal of this section is to address the potential for protein sequences to adopt "non-standard" topologies, or topologies not taught in introductory biochemistry. As many comprehensive, full reviews of membrane protein insertion and topologies exist (Elofsson and von Heijne, 2007; Goder and Spiess, 2001; von Heijne, 2006; White and von Heijne, 2004, 2008; White and Wimley, 1999), in an attempt to keep the discussion more relevant, only examples demonstrated to exist in viral proteins will be discussed.

1.2.4.1 Dynamic topology

As indicated above, the use of the biological hydrophobicity scale has led to a larger-scale identification of protein sequences that reorient post-insertion to arrive at their final topology with respect to the membrane (Kauko et al., 2010). These relatively minor rearrangements are interesting, but can be considered mild when considering the topological changes that have been observed for some viral proteins. A dramatic example is presented by the hepatitis B virus large envelope glycoprotein. In this case, the glycoprotein is initially synthesized with a cytoplasmically-localized N-terminus. During a post-translational, chaperone-dependent maturation process, the N-terminus is translocated across the membrane into the lumen in about 50% of the molecules, resulting in a protein of mixed topology in the cell (Lambert et al., 2004; Lambert and Prange, 2001; Lambert and Prange, 2003). There is also evidence for a similar functionally-necessary cytoplasmic to periplasmic translocation of a highly-charged N-terminal domain in the phage lambda S107 holin protein that is apparently driven by a decrease in the membrane electrochemical potential, leading to hole formation and lysis of the infected cell (Graschopf and Bläsi, 1999).

The envelope virus protein from the hepatitis C virus displays a different, but no less complicated, dynamic topology (Cocquerel et al., 2002). In this case, the envelope glycoprotein is synthesized as a polyprotein consisting of the individual proteins E1, E2, and p7. The E1 protein has a cleavable N-terminal signal peptide and a hydrophobic C-terminal anchoring sequence. This E1 anchoring sequence additionally acts as the signal peptide for E2 that also has a C-terminal anchor that is the signal peptide for p7. The anchor/signal peptide segments are thought to initially insert as helical hairpins with

sharp cytoplasmic loops. Following cleavage of the signal peptides by the signal peptidase, located in the lumen, the C-termini of the signal peptide reorients across the membrane, finally localizing to the cytoplasm (Cocquerel et al., 2002).

1.2.4.2 In-plane membrane anchoring sequences

In-plane membrane anchors have also been described for viral proteins, with the majority of examples coming from the Flaviviridae family in the Pestivirus genus. In the Pestivirus viruses, both bovine viral diarrheal virus (BVDV) and the hepatitis C virus have been shown to produce proteins that associate with the membrane through the use of in-plane membrane anchors, resulting in proteins with a monotopic topology (Brass et al., 2007; Sapay et al., 2006b; Tews and Meyers, 2007). For BVDV and hepatitis C, the non-structural protein NS5A has been demonstrated to associate with the membrane by means of a 28 residue N-terminal amphipathic helix that is required for the assembly of a functional replication complex (Brass et al., 2007). NMR structure determinations and MD simulations indicate that the N-terminal amphipathic helix mediates membrane association through stable localization to the membrane-water interface (Sapay et al., 2006b).

An additional pestivirus protein has been demonstrated to associate with membranes via an in-plane helix. The envelope glycoprotein E^{rns} from BVDV is anchored to the membrane by a C-terminal amphipathic helix, albeit on the lumenal side of the membrane, as opposed to the cytosolic localization of the other pestivirus proteins with in-plane anchors (Tews and Meyers, 2007). Importantly, mutations that resulted in destabilization of the amphiphilicity of the anchor sequence resulted in decreased membrane association of the protein concomitant with decreased viral

infectivity. This was the first report of a glycoprotein anchored to the membrane via an in-plane helix. The existence of in-plane anchoring sequences has important implications for topology prediction because current predictive methods do not distinguish well between monotopic and bitopic proteins (Bernsel and von Heijne, 2005), and the very recent finding that minor flanking sequence differences can result in adoption of monotopic (in-plane or re-entrant loop) or bitopic (transmembrane) topology for anchoring segments (Norholm et al., 2011).

1.2.5 Effects of lipids on membrane protein topology

Probably the most recent, and controversial, topic regarding membrane protein topology is the concept that the lipid environment in which the protein exists can, by itself, play a determining role in the protein's topological arrangement. Key studies demonstrating this lipid-dependence of protein topology were carried out on the LacY lactose transporter for *Escherichia coli* by Dowhan and colleagues (Bogdanov et al., 2002; Bogdanov et al., 2008; Dowhan and Bogdanov, 2009, 2011; Wang et al., 2002; Xie et al., 2006; Zhang et al., 2003; Zhang et al., 2005). Crucial to these studies was the use of *E. coli* genetically engineered to inducibly synthesize phosphatidylethanolamine (PE), a zwitterionic lipid. Wild-type *E. coli* membranes are composed of ~70% PE, 20% phosphatidylglycerol (PG, an anionic lipid), and 5% cardiolipin (CL, also an anionic lipid) (Dowhan and Bogdanov, 2009). LacY folding in wild-type *E. coli*, then, takes place in an abundance of PE, which serves to mitigate the level of negative charge at the membrane imposed by PG and CL (Dowhan and Bogdanov, 2009). In the PE-inducible *E. coli*, however, protein synthesis, and thus membrane protein folding, can occur in the

absence of PE, and in the presence of a large negative charge on the cytosolic leaflet owing to the overwhelming presence of PG and CL (Dowhan and Bogdanov, 2011). This system was used to demonstrate that LacY synthesized in the absence of PE contained a topologically-inverted N-terminal six-helix bundle that resulted in protein inactivity (Bogdanov et al., 2008). Following the blockage of new protein synthesis and induction of PE synthesis, the authors demonstrated that the N-terminally-inverted LacY reoriented the N-terminal bundle to assume a native state that restored uphill lactose transport activity (Bogdanov et al., 2008). This result suggests that alterations in lipid content alone can result in topological reorientation of membrane proteins outside the environment of the translocation machinery. The authors further demonstrated that the addition of negative charge to loops normally located in the cytoplasm was sufficient to lead to inverted N-terminal topology, even in wild-type *E. coli*, suggesting that increasing the negative charge of the protein can override the positive inside rule (Bogdanov et al., 2008). Overall, results from Dowhan and colleagues collectively suggest that the interaction of the lipid and protein play a role on the final topology of the protein (Bogdanov et al., 2002; Bogdanov et al., 2008; Dowhan and Bogdanov, 2009, 2011; Wang et al., 2002; Xie et al., 2006; Zhang et al., 2003; Zhang et al., 2005). The application of these concepts to eukaryotic systems is uncertain, however, as membrane content cannot be controlled to the same extent as in the prokaryotic system, nor do there exist such extensive natural variations in membrane content. Of potential interest, however, is the fact that the eukaryotic endoplasmic reticulum (ER) membrane displays a symmetrical distribution of lipids between the lumenal and cytosolic leaflets. Further, asymmetry with respect to the lipid content of the bilayer

leaflets is only introduced to the biosynthetic pathway in the Golgi network, with PE and negatively charged phospholipid phosphatidylserine (PS) being transported to and maintained on the cytosolic leaflet in an energy-dependent fashion (Dowhan, 1997; van Meer et al., 2008). This presents the hypothetical possibility for post-translational lipid-dependent protein topological changes based on membrane leaflet segregation, where the negatively charged PS becomes solely cytoplasmically-localized in the Golgi after the protein is translocated and membrane inserted in the ER. It is then possible, if lipid-protein interactions can determine topology, that as the protein traffics from the ER to the Golgi, topological rearrangements may occur due to the interaction between protein charges and the now cytoplasmically-oriented PS.

2.0 HYPOTHESIS AND SPECIFIC AIMS

2.1 HYPOTHESIS

Retroviral envelope proteins serve as major determinants for replication, immunologic, and pathogenic properties, and variations in envelope (Env) sequences can markedly alter viral phenotypes. Studies of the HIV-1 Env have focused predominantly on the surface unit (SU) gp120 and the ectodomain of the transmembrane (TM) Env gp41, reflecting an early bias toward a negligible functional role for the gp41 CTT. Concurrently, Env immunology has generally ignored the CTT since most Env-based vaccines are engineered without CTT sequences to increase Env expression or to make a soluble protein. More recent studies indicate (i) that HIV-1 and SIV CTT sequences mediate specific interactions with cellular protein components of endocytic and signal transduction pathways, (ii) that deletions and truncations in HIV-1 and SIV CTT sequences can alter viral Env conformation, function, and antigenicity, and (iii) that CTT mutations can strongly attenuate SIV replication and pathogenesis in experimentally infected monkeys. Recent results from our lab demonstrated that select point mutations in the CTT can markedly alter the antigenic and functional properties of gp120 and gp41, indicating that even minor variations in CTT sequences affect Env structure and function. In addition, a series of experimental and modeling studies of CTT structure

suggest that portions of the CTT may be exposed on the membrane surface, contrasting the generally-held view of gp41 as a type I integral membrane protein. This proposed topology is supported by studies identifying CTT sequences as targets for neutralizing monoclonal and polyclonal antibodies *in vitro*. Finally, recent studies suggest a transient exposure of a discrete CTT region during the fusion process, providing further evidence for complex CTT topology. Thus, a definitive topological analysis of the Env CTT is warranted.

Based on these observations, I hypothesize that the gp41 CTT serves as a critical determinant of Env structure and function by modulation of its topological arrangement with and within the membrane lipid bilayer. To address this hypothesis, I propose to: (i) define the conservation of gp41 CTT sequence and structural elements from diverse HIV strains; (ii) determine the native topology of the Env CTT on the surface of intact cells and viral particles; and (iii) examine the influence of the fusion process on the topology of the CTT.

2.2 SPECIFIC AIMS

2.2.1 Aim 1. To define the conservation of gp41 CTT sequence and structural elements from diverse HIV strains.

The sequence conservation of the gp41 CTT was compared to that of gp120 and the gp41 ectodomain. Phylogenetic and diversity analyses of approximately 2000 sequences from HIV clades A, B, C, D, F, G and groups N and O demonstrated that the

CTT was intermediately variable between the highly variable gp120 and the more conserved gp41 ectodomain. In spite of relatively high variability, the gp41 CTT displayed remarkable conservation of the physicochemical characteristics of the lentivirus lytic peptide (LLP) regions, with an extraordinary conservation of arginine residues, especially as compared to lysine. Computational folding of consensus LLP peptide sequences from all clades/groups demonstrated overall conservation of helical secondary structure. In addition, circular dichroism studies of synthetic LLP peptides representing consensus as well as naturally-occurring sequences from clades B, C, and group O confirmed a high level of helicity in membrane mimetic environments including sodium dodecyl sulfate (SDS) micelles and trifluoroethanol. Results from these studies indicate that the CTT contains conserved structural elements in spite of high levels of sequence variation.

2.2.2 Aim 2. To determine the native topology of the Env CTT on the surface of intact cells and viral particles and examine the effect of the fusion process on the topology of the CTT.

The topology of the CTT was determined by a combination of epitope tagging of the CTT sequence and the use of monoclonal antibodies targeting the wild-type sequence. For epitope tagging studies, the VSV-G epitope tag (amino acid sequence YTDIEMNRLK) was used to serially replace HIV sequences along the entirety of the CTT. To map the topology of the CTT, intact Env-expressing cells and pseudoviral particles were analyzed for reactivity with anti-CTT antibodies by flow cytometry, immunoprecipitation, and surface plasmon resonance (SPR) spectroscopy assays.

Results from these studies demonstrated topological differences in Env expressed on the cell surface compared to virion-associated Env.

The potential for the CTT to undergo dynamic topological changes was addressed through the use of a post-attachment neutralization assay. When the fusion process was slowed by incubation at 31 °C, a CTT-specific monoclonal antibody neutralized virus infection, but did not neutralize under standard neutralization at 37 °C. These results indicate that the CTT, which is not exposed under native conditions on the virion, undergoes dynamic, transient rearrangement during the fusion process.

3.0 HIGHLY CONSERVED STRUCTURAL PROPERTIES OF THE C-TERMINAL TAIL OF HIV-1 GP41 DESPITE SUBSTANTIAL SEQUENCE VARIATION AMONG DIVERSE CLADES: IMPLICATIONS FOR FUNCTIONS IN VIRAL REPLICATION

This research was originally published in the Journal of Biological Chemistry. Jonathan D. Steckbeck, Jodi K. Craigo, Christopher O. Barnes, Ronald C. Montelaro. Highly Conserved Structural Properties Of The C-Terminal Tail Of HIV-1 gp41 Despite Substantial Sequence Variation Among Diverse Clades: Implications For Functions In Viral Replication. Journal of Biological Chemistry. 2011; 286: 27156-27166.

3.1 INTRODUCTION

The envelope (Env) proteins of retroviruses in general, and HIV-1 in particular, have been shown to serve as a major determinant of viral replication, antigenic, and pathogenic properties, such that natural and experimental variations in envelope sequence can markedly alter these viral phenotypes (Evans and Desrosiers, 2001; Levy, 2009; Roux and Taylor, 2007). Fundamentally, HIV-1 Env is the mediator of viral infection. Env is translated as a full-length gp160 polyprotein that is post-translationally cleaved into non-covalently associated gp120 and gp41 subunits that yield the functional Env monomer that then forms the virion associated Env trimer complex. The gp120 and gp41 subunits of Env each mediate specific functions during the infection process. The function of gp120 is to bind the primary receptor, CD4, and subsequently the co-receptor, primarily CCR5 or CXCR4. Receptor binding leads to conformational changes in gp41 such that the previously sequestered fusion peptide at the N-terminus of gp41 inserts into the cellular membrane. This is followed by refolding events that function to bring the cellular and viral membranes into proximity finally leading to fusion and infection.

The gp41 protein is organized into three distinct domains: (i) the ectodomain, including the fusion peptide, the N- and C-terminal heptad repeat regions, and the membrane proximal external region (MPER), a target for broadly-neutralizing monoclonal antibodies (MAbs); (ii) the membrane spanning domain (MSD), the sequence that anchors Env into the membrane; and (iii) the C-terminal tail (CTT),

demonstrated to function in cellular Env trafficking (Byland et al., 2007; Lopez-Vergès et al., 2006; Ohno et al., 1997) and virion incorporation (Kalia et al., 2003; Murakami and Freed, 2000b; Wyma et al., 2000). The gp41 ectodomain has historically been the focus of many studies examining its role in the fusion process and as a target for neutralizing antibodies, while the MSD has been the subject of detailed studies to determine its precise length and structure (Gangupomu and Abrams, 2010; Kim et al., 2009; Shang et al., 2008; West et al., 2001; Yue et al., 2009). In contrast to these gp41 domains, the CTT has been relatively less studied, having only recently been demonstrated as playing a major role in Env trafficking and viral Env incorporation (Byland et al., 2007; Kalia et al., 2003; Lopez-Vergès et al., 2006; Murakami and Freed, 2000b; Ohno et al., 1997; Wyma et al., 2000).

The CTT first came to prominence in the 1980s when it was reported that serum targeting a peptide from the CTT could bind gp160 (Kennedy et al., 1986) and neutralize virus infectivity (Chanh et al., 1986; Ho et al., 1987). This apparent exposure of the CTT to neutralizing antibodies implied a surface exposure not predicted from the intracytoplasmic location predicted for the CTT sequences. In the following years, however, the CTT was thought to be dispensable for virus replication based on studies indicating the replication competence of proviral constructs in which the CTT was deleted from the Env gene (Chakrabarti et al., 1989; Hirsch et al., 1989; Kodama et al., 1989). More recently the CTT has been demonstrated to contain a number of trafficking signals that control the level of Env expression at the cellular membrane. The CTT contains two functional endocytic motifs that lead to the rapid removal of Env from the surface of Env-expressing cells into late endosomes (Byland et al., 2007; Ohno et al.,

1997). The CTT has also been shown to contain sequences involved in virion incorporation (Kalia et al., 2003; Lopez-Vergès et al., 2006; Murakami and Freed, 2000a, b; Wyma et al., 2000), particularly through a proposed interaction with the matrix domain of HIV Gag (Murakami, 2008; Murakami and Freed, 2000a). Additionally, various studies have demonstrated that the CTT is a critical determinant of overall Env conformational properties, as variations in CTT sequences can alter Env fusogenicity and antigenicity mediated by the gp120 protein and the gp41 ectodomain (Kalia et al., 2005). Finally, the CTT has been demonstrated to contain sequences required for association with detergent-resistant membranes, including a crucial role for the alpha-helical nature of the LLP regions in maintaining lipid raft association (Yang et al., 2010).

While high-resolution structures have been determined for sequences in the gp41 ectodomain, little is known about the structure of the CTT. There are no high-resolution structures for any sequences of the CTT, although various studies have indicated that discrete domains in the CTT, known as the lentivirus lytic peptides (LLP), are helical in membrane mimetic environments (Chernomordik et al., 1994; Kliger and Shai, 1997; Srinivas et al., 1992). Topologically, gp41 is generally considered to be a type I membrane protein with an extracellular N-terminus and a cytoplasmic C-terminal tail following the predicted membrane spanning domain. The model of the CTT as an exclusively intracytoplasmic domain is consistent with the location of functional endocytic motifs that affect Env trafficking. However, this generalization is challenged by conflicting data indicating that the Kennedy epitope is the target for neutralizing antibodies, suggesting an external localization of CTT sequences, perhaps most evident after receptor binding (Chanh et al., 1986; Heap et al., 2005; Ho et al., 1987). Indeed,

recent studies have also demonstrated differential exposure of CTT sequences on viral and cellular membranes (Steckbeck et al., 2010), and biochemical studies have revealed transient exposure of CTT sequences during the Env fusion to cellular membranes (Lu et al., 2008).

Sequence analyses of Env have focused almost exclusively on the gp120 subunit and the gp41 ectodomain (Korber et al., 2001; Pieniazek et al., 1998) and the effects of sequence variation on viral phenotypes, e.g. tropism (Scarlatti et al., 1997) and neutralization (Karlsson Hedestam et al., 2008). To date, there have been no comprehensive analyses of the sequence variability of the CTT domain. In light of recent observations that even minor variations in CTT sequence can markedly alter Env structural and functional properties, the current study was designed to determine the extent of amino acid sequence variation in the CTT of diverse HIV-1 strains and to analyze the effects of this sequence variation on the physicochemical and structural properties of variant CTT species. The results of these studies for the first time provide evidence for evolutionary pressures on CTT sequences that are balanced by a remarkable conservation of uniquely characteristic physicochemical and structural properties, apparently to maintain critical functional roles in virus replication. These data indicate the need for further characterization of these CTT-mediated functions.

3.2 MATERIALS AND METHODS

3.2.1 Sequence acquisition and processing

Sequences were obtained from the Los Alamos National Labs (LANL) HIV sequence database. Full-length (gp160) HIV-1 Env amino acid alignments were collected from years including 1997 through 2008 and were separated into clades A, B, C, D, F, and G, and Groups N and O for eight total sequence groupings. The sequences in each clade/group were further processed as full-length Env sequences to remove duplicate sequences using the ElimDupes tool from Los Alamos National Labs (http://www.hiv.lanl.gov/content/sequence/ELIMDUPES/elimdupes.html). Sequences were then separated into the three main domains of Env (gp120, the gp41 ectodomain, and the gp41 CTT) and aligned as described in the following section. The collection of sequences and subsequent processing yielded a total of approximately 2,000 unique sequences. Total sequences in each clade/group were as follows: 150 in clade A; 888 in clade B; 626 in clade C; 118 in clade D; 34 in clade F; 48 in clade G; 6 in group N; and 126 in group O.

3.2.2 Sequence alignments and evolutionary analyses

Sequences from each clade/group were individually aligned within the respective domain (gp120, gp41 ectodomain, and gp41 CTT) for a total of 24 alignments (8 clades/groups and 3 protein domains). Global alignments were performed using Geneious Pro 5.0.4 (Biomatters Ltd., NZ) (Drummond et al.) using the Pam250 cost

matrix. The gap opening/extension penalties were set to 16/8, with 30 refinement iterations performed for each alignment. Sequences were hand-edited where necessary, particularly around the variable regions in gp120. Evolutionary analyses included amino acid diversity computations and phylogenetic tree construction. The diversity of each domain (gp120, gp41 ectodomain, gp41 CTT) was measured by comparing the calculated distances of all sequences within a group in the alignment to one another. Genetic distances of aligned sequences were determined utilizing Kimura population distance calculations for amino acid sequences with no gap weighting, as executed in the program distmat from the EMBOSS package of software. Statistical significance values of the mean diversity of the domains were calculated in GraphPad Prism 5.0 (GraphPad Software, San Diego, CA). The statistical significance of diversity was calculated using two-tailed unpaired t-tests, as implemented in GraphPad InStat version 3.00 (GraphPad Software, San Diego California USA). Phylogenetic analyses were performed on the individual domain alignments of the hand edited, aligned sequences. The phylogenetic trees were constructed by the neighbor-joining method of Kimura corrected measurements with the optimality criterion set to distance as measured in PAUP (Swofford) and implemented in Geneious Pro version 5.0.4 (Drummond et al.). Statistical significance of branchings and clustering were assessed by bootstrap re-sampling of 1000 pseudoreplicates on the complete data set and the confidence level set to 75. The trees were edited for publication using FigTree version 1.1.2 (Morariu et al., 2008).

3.2.3 Characterization of physicochemical properties of consensus LLP peptide sequences

MPER, MSD, and LLP sequences from all clades/groups were extracted from consensus sequences determined from alignments generated with Geneious Pro 5.0.4 (Drummond et al.). The physicochemical characteristics (hydrophobic moment, net charge, and hydrophobicity) and helical wheel diagrams of the extracted regions were determined using HeliQuest (http://heliquest.ipmc.cnrs.fr/). One-way ANOVA using Tukey's multiple comparison test and Bartlett's test for equal variances was performed using GraphPad Prism 5.0 (GraphPad Software, San Diego, CA).

3.2.4 PEP-FOLD structure prediction for LLP peptides

LLP and Kennedy epitope (KE) sequences for all clades/groups were extracted from the consensus sequences generated by the global multiple sequence alignments. Sequences were submitted to the PEP-FOLD server (Maupetit et al., 2009; Maupetit et al., 2010; Maupetit et al., 2007) for peptide folding including PSIPRED predictions (Jones, 1999). The number of clusters and the percentage of total structures occupying each cluster were graphed to demonstrate conformational diversity for each peptide sequence. Conservation of secondary structure for the peptides was determined by finding the root mean square deviation (RMSD) for the peptides as follows. For KE, LLP1, and LLP3, the models were aligned as described (Kabsch, 1976, 1978) using the RMSD align function of VMD (Humphrey, 1996). Due to the predicted random coil nature of the C-terminal half of the group O LLP2 model, the LLP2 models were aligned

using the STAMP structural alignment method (Russell, 1992), instead of aligning to the minimum RMSD. To take into consideration the differences in the predicted differences in LLP2 lengths, the LLP2 analysis was done using the first 14 residues of clades/groups A, C, G, N, and O and residues 3-16 of clades B, D, and F. Average RMSD was then determined for all residues on the STAMP-aligned models.

3.2.5 Circular dichroism (CD) spectroscopy to determine peptide secondary structure

LLP sequences for clades B and C and group O were extracted from the consensus sequences generated by the global multiple sequence alignments. In addition to the domains derived from the consensus sequences, we also determined the sequence with the lowest average divergence (LAD) from all sequences within each clade/group to select a representative natural isolate for secondary structure characterization. LLP sequences were also extracted from these LAD sequences. All 17 peptides (LLP2 from clade B consensus and LAD were identical) were synthesized using standard FMOC chemistry with a carboxylated N-terminus and amidated C-terminus to mimic their native chemistry within a longer polypeptide sequence. Peptides were resuspended in 10 mM HEPES, 10 mM SDS in 10 mM HEPES, or 60% TFE in 10 mM HEPES to a concentration of 50 μM. CD spectra were determined at 37 °C on a Jasco 810 CD spectrometer from 195 - 250 nm. Spectra are presented as the average of 10 acquisitions. Peptide mean residue ellipticity was determined using CDPro (Sreerama and Woody, 2000; Sreerama and Woody, 2004).

3.3 RESULTS

3.3.1 CTT sequence variation is intermediate between gp120 and the gp41 ectodomain

Initial analyses of the variation present in the individual domains of the Env protein involved a comparative alignment of the amino acid sequences of the approximately 2,000 isolates available in the Los Alamos database (see Materials and Methods for inclusion/exclusion determinations). Alignments of the respective gp120, gp41 ectodomain, and the CTT domains were performed within and between all sequences of all clades and groups (data not shown). Subsequently, phylogenetic analyses by construction of neighbor-joining trees were performed to demonstrate the fundamental ancestral associations between the clades/groups at these three domains (Figure 13). All phylogenetic inferences were made by comparing the resultant ancestral relationships of the three domains at a similar scale. The results demonstrated the least genetic relatedness for the gp120 sequences within a clade/group, as evidenced by the greatest average distance between sequences. In contrast, the gp41 ectodomain sequences were most closely related to each other within a clade/group, as exhibited by the least average distance between sequences. Finally, the gp41 CTT sequences were

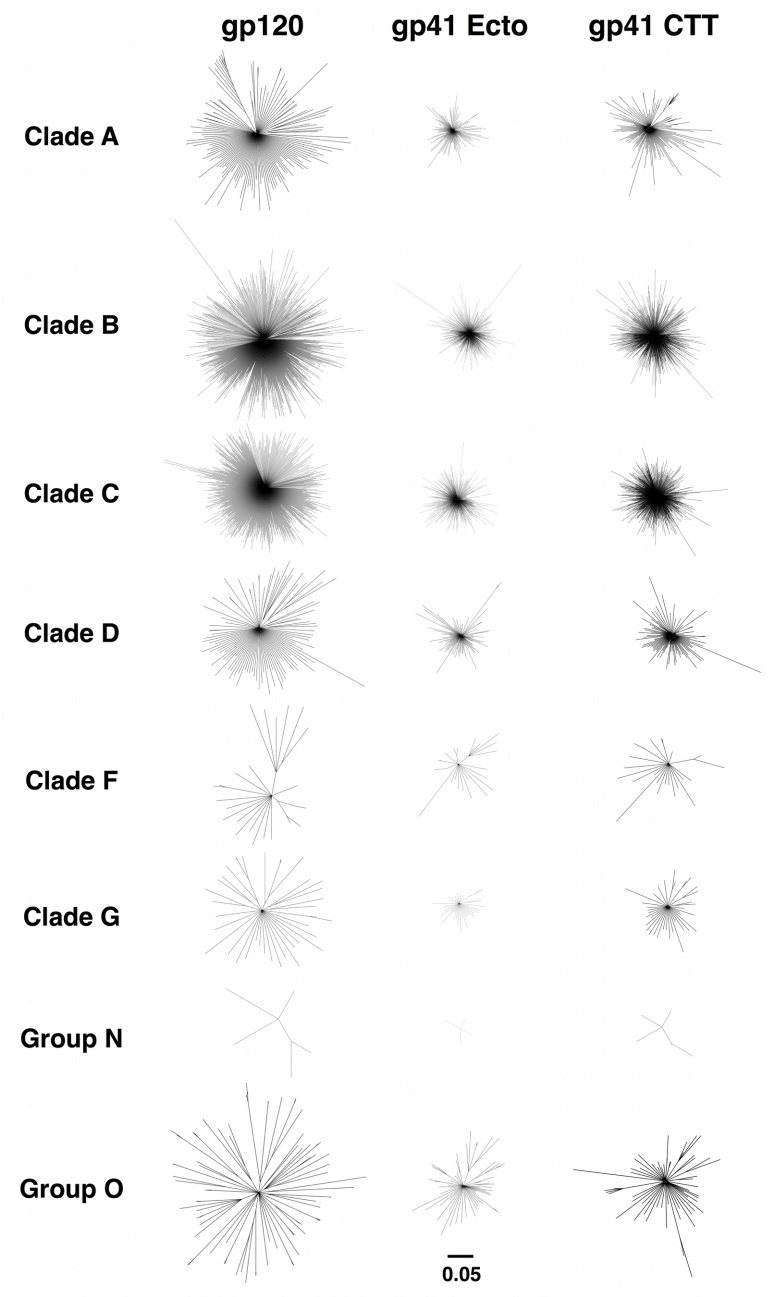

Figure 13. Phylogenetic analysis of Env domains

The phylogenetic relationship between the Env domains of clades A, B, C, D, F, G and groups N and O were examined. A neighbor-joining tree of aligned amino acid sequences was constructed from Kimura corrected distances. Bootstrap values were determined over 1000 iterations and the confidence level set to 75. Branch lengths are proportional to the phylogenetic distance existing between the sequences.

more related to each other than gp120 sequences, but less related to each other than the gp41 ectodomain sequences.

Diversity analyses were performed to provide a more quantitative distance measure between the sequences of each clade/group (Figure 14). Similar to the results from the phylogenetic analyses, gp120 was observed to have the highest overall diversity within each clade/group, with the mean diversity ranging from 20.32 - 37.75%. The gp41 ectodomain sequences displayed the lowest mean diversity, ranging from 6.63 - 18.96%. Finally, CTT sequences exhibited intermediate diversity between gp120 and gp41 ectodomain, with a mean diversity range of 12.36 - 26.78%.

Together these results demonstrate an intermediate level of amino acid sequence relatedness and diversity for the CTT in comparison to gp120 and gp41 ectodomain domains, suggesting a previously unrecognized evolutionary pressure on the CTT domain.

3.3.2 Physicochemical characteristics of CTT domains are conserved

In light of the relatively high level of diversity calculated for the CTT domain, we next sought to analyze the effects of CTT sequence variation on the physicochemical properties of defined segments of the CTT domain. To achieve a practical means of comparing the large number of CTT sequences in the database, we chose to additionally perform an alignment of the consensus sequences of the CTT domain for all of the clades (Figure 15). Consensus sequences of the entire CTT region of each clade

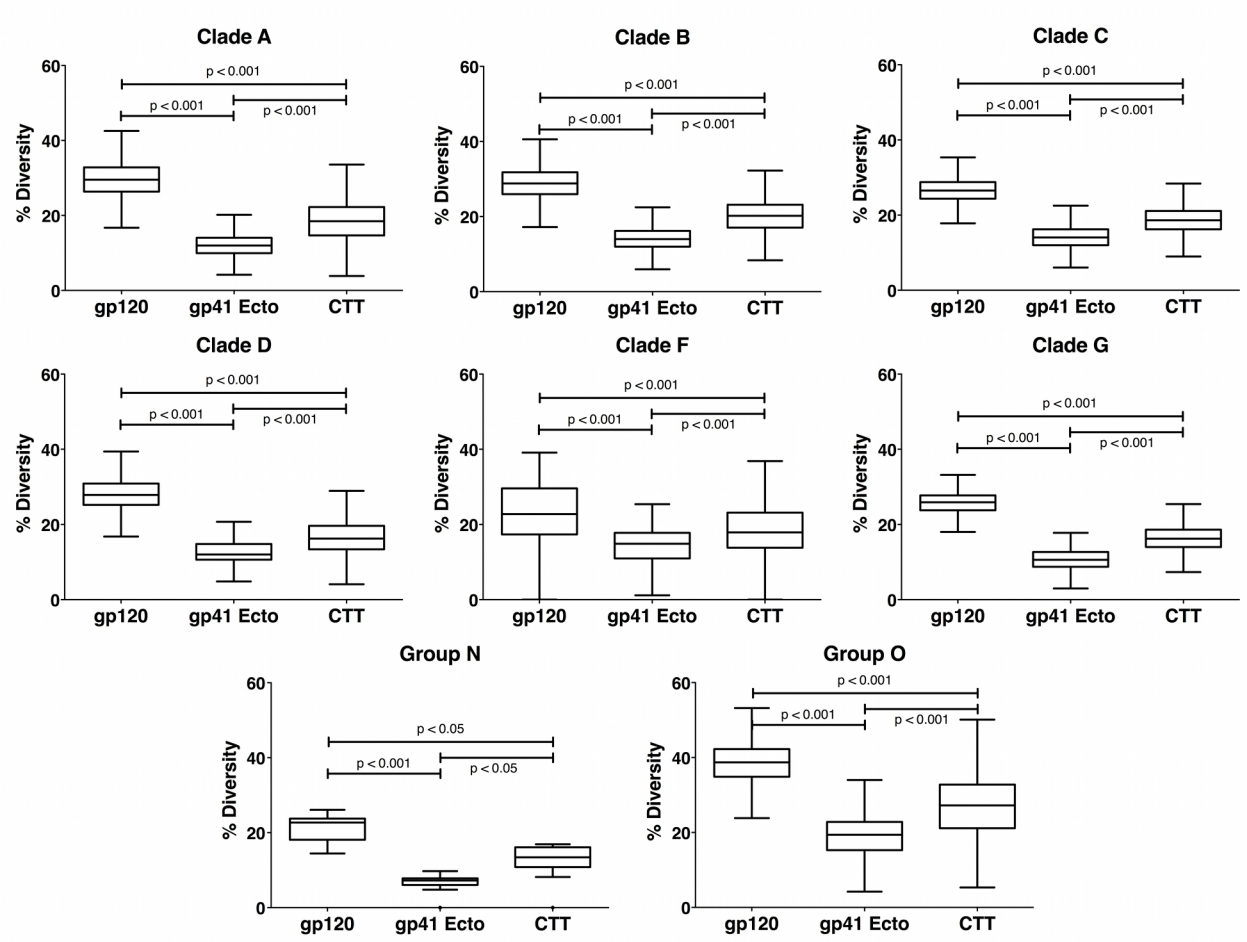

Figure 14. Comparison of average diversity in Env indicates intermediate CTT variation relative to other domains

Tukey box and whisker plots (GraphPad Prism v. 5.0) of Env gp120, gp41 Ecto, and the CTT domain amino acid diversity are plotted for all clades/groups. Statistical significance of the differences between all populations (box and whisker plots) was performed using an unpaired two-tailed t-test with Welch's corrections. The differences in all the medians are statistically significant.

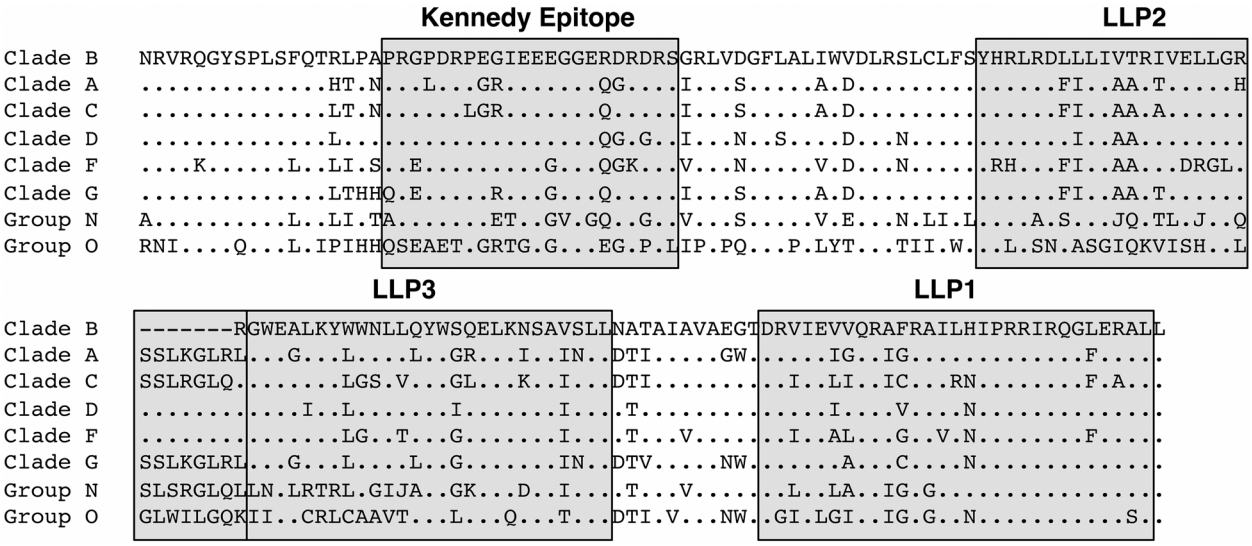

Figure 15. Comparison of the amino acid sequence of the CTT domain

Consensus sequences for the CTT region of the Env of each clade or group were derived from hand-edited alignments of all the isolates downloaded from the Los Alamos database (see Materials and Methods). Derived CTT region sequences were aligned and compared to the clade B consensus sequence for visualization purposes. Only the amino acid residues different from clade B are reported on the alignment. Dots (·) indicate residues identical to the clade B consensus sequence; dashes (-) indicate alignment gaps. Previously described LLP regions and the Kennedy epitope are boxed.

were derived from hand-edited alignments and aligned to one another as described in Materials and Methods. To more thoroughly understand the potential impact of amino acid variability, we investigated the conservation of the physicochemical characteristics of discrete, putative membrane-associated domains of the CTT compared to two other defined membrane-associated domains of gp41, the MPER and the MSD. We calculated the hydrophobic moment (μH), net charge, and hydrophobicity (H) of the peptide domains derived from the consensus sequences for the MPER, MSD, and the LLP regions of the CTT for all clades/groups (Figure 16). The results of these analyses reveal that the LLP domains have significantly higher hydrophobic moments (mean μH = 0.6003, 0.3640, and 0.3945 for LLPs 1, 2, and 3, respectively) than the MPER and MSD (mean μH = 0.1520 and 0.1153, respectively), and that the calculated LLP1 hydrophobic moment is significantly higher than LLP2 and LLP3 (Figure 16A). Both LLP1 and LLP2 (+5 and +3 mean charges, respectively) are significantly more cationic than the MPER, MSD, or LLP3 (-0.25, +1, and +0.25 mean charges, respectively) (Figure 16B). The LLP regions were also found to be significantly less hydrophobic (0.3643, 0.4974, and 0.6650 mean H for LLPs 1, 2, and 3, respectively) than either the MPER (0.8268 mean H) or the MSD (1.133 mean H) (Figure 16C). Comparison of the variances of the calculated hydrophobic moments and net charge between the clades/groups indicate in general a similar level of variation in the sequences across clades/groups. However, the variances between clades/groups did differ for hydrophobicity, with the MPER and MSD exhibiting lower variance than the LLP regions

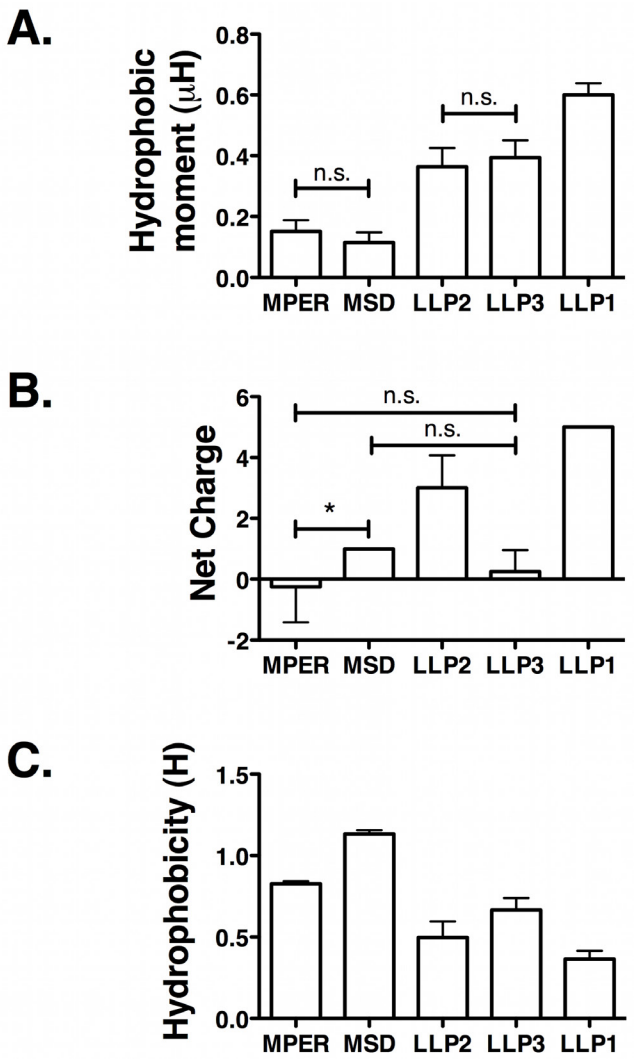

Figure 16. LLP Sequences have different physicochemical properties than MPER or MSD sequences

Physicochemical properties of lipophilic domains from gp41 were determined from consensus sequences from each clade/group. Data presented as the mean and standard deviation of clades/groups, and differences between columns are significant at $p < 0.001$ unless otherwise noted. A.) Hydrophobic moments were calculated to determine amphipathic potential, and LLPs were found to be significantly more amphipathic than MPER or MSD sequences. B.) Average net charges demonstrate that LLPs 1 and 2 are highly cationic. C.) MPER and MSD are significantly more hydrophobic than LLP regions. n.s. - not significant; * $p < 0.05$

across all clades/groups, indicating that the hydrophobicity of the MPER and MSD peptide sequences is more conserved than for the LLP peptide sequences.

As the consensus LLP sequences were found to have high hydrophobic moments, and thus high amphipathic potential, we modeled the LLP peptide sequences using helical wheel diagrams (Figure 17). Inspection of the helical wheel diagrams confirms the amphipathic nature of the peptide sequences, with the charged and polar residues concentrated on one helical face and the hydrophobic and nonpolar residues concentrated on the opposite helical face. This amphipathic nature is highly conserved across all groups and clades, indicated by the similar incidence and position of the charged residues on the polar helical face.

While overall charge for the LLP sequences was conserved (Figure 16B), the consensus sequences suggest a preference for arginine relative to lysine in the cationic LLP1 and LLP2 sequences. To test this apparent preference for arginine in LLP sequences, we calculated the frequency of arginine and lysine in the CTT of HIV-1 Env using 1,675 sequences from clades A, B, C, D, F, and G, and groups N and O from the Los Alamos National Labs 2009 HIV-1 Env sequence alignment (http:// www.hiv.lanl.gov/) and compared the calculated frequencies to the arginine and lysine frequencies for all proteins represented in the UniProtKB/Swiss-Prot protein knowledgebase release 2010_09 (http://ca.expasy.org/sprot/relnotes/relstat.html). Overall, arginine was found to occur at a frequency of 12.7% in the CTT (5.9% for all of Env) compared to a frequency of 5.5% observed for all proteins in the UniProtKB/Swiss-Prot protein knowledgebase. This increased frequency of arginine was accompanied by

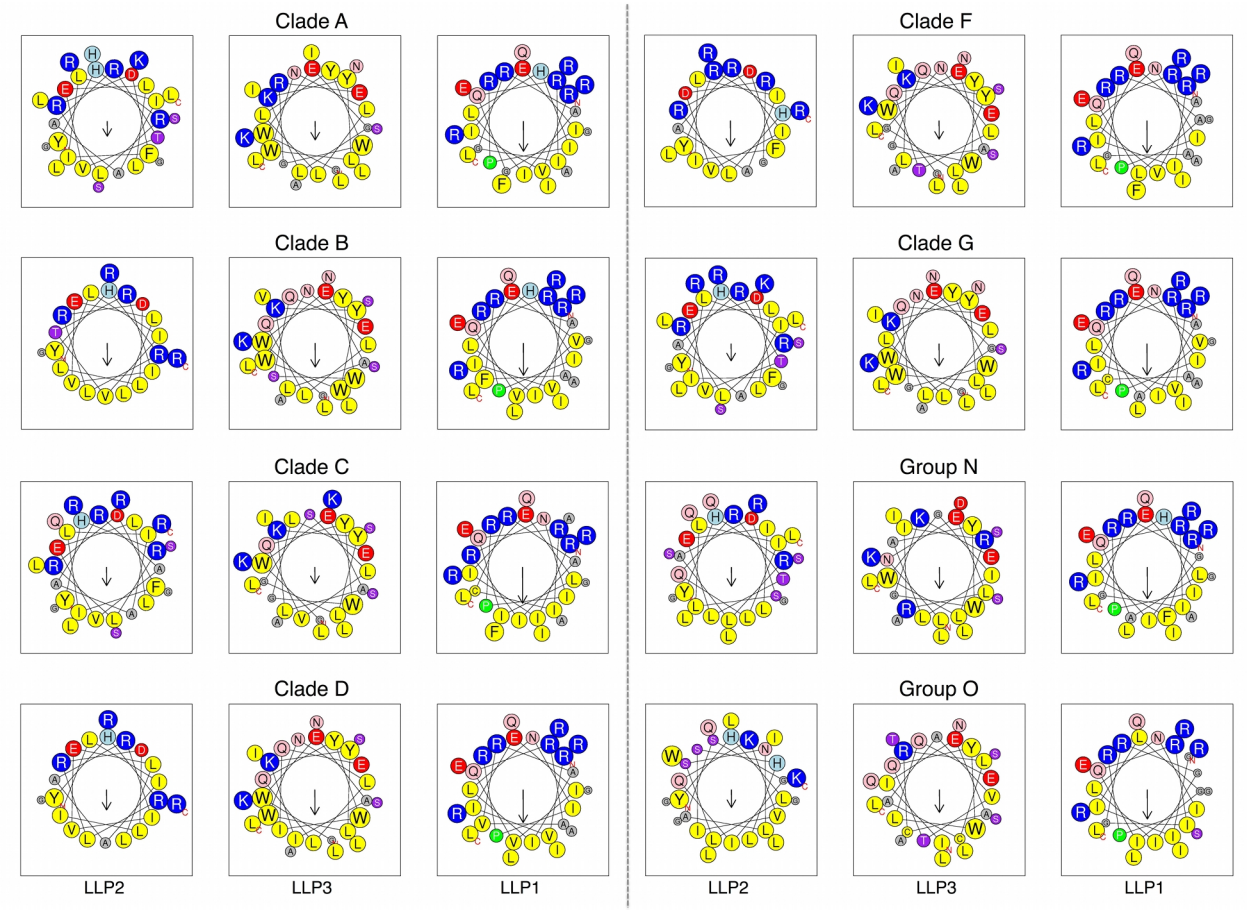

Figure 17. Helical wheel diagrams demonstrate the conservation of the amphipathic nature of LLP regions despite sequence variation

Consensus LLP sequences were plotted as helical wheels to observe the relative conservation of amphipathicity predicted by the hydrophobic moment analysis. The direction and relative magnitude of the hydrophobic moment is shown by the arrow. Amphipathicity of the sequences is clearly observed as a clustering of charged and polar residues on one helical face opposite the direction of the hydrophobic moment with concomitant clustering of hydrophobic residues on the opposite helical face.

a marked decrease in the frequency of lysine (1.9% CTT vs. 5.9% UniProtKB/Swiss-Prot; 5.0% for all of HIV Env), implying an important functional role for the arginine residue relative to lysine in the CTT. To ascertain position-specific arginine/lysine substitutions, we prepared logo diagrams for the LLP regions to determine the most conserved positively charged positions (Figure 18). Positions where the probability of arginine/lysine was >0.9 across all clades/groups were considered to be highly conserved. For the most highly conserved arginine residues in LLP1 and LLP2, lysine was substituted for arginine 0.40% of the time (40 out of 10,044 possible instances), while in LLP3 the conserved lysines were replaced by arginine in 4.84% of instances (162 of 3,348). Collectively these results indicate that the physicochemical properties of the LLP regions are conserved in spite of high sequence variation and suggest a preference for arginine over lysine, particularly in LLP1 and LLP2.

3.3.3 Secondary structure of the LLP regions of the CTT is predicted to be highly conserved

After determining that the physicochemical characteristics of the LLP regions were conserved, we examined the conservation of LLP secondary structure using PEP-FOLD computational peptide folding (Maupetit et al., 2009; Maupetit et al., 2010; Maupetit et al., 2007). Since peptide analogs of the LLP regions are proposed to be helical in conformation (Chernomordik et al., 1994; Kliger and Shai, 1997; Srinivas et al., 1992), we included as a control another reference CTT sequence, the Kennedy epitope (KE), that is predicted to exist in a non-helical conformation (Hollier and Dimmock, 2005).

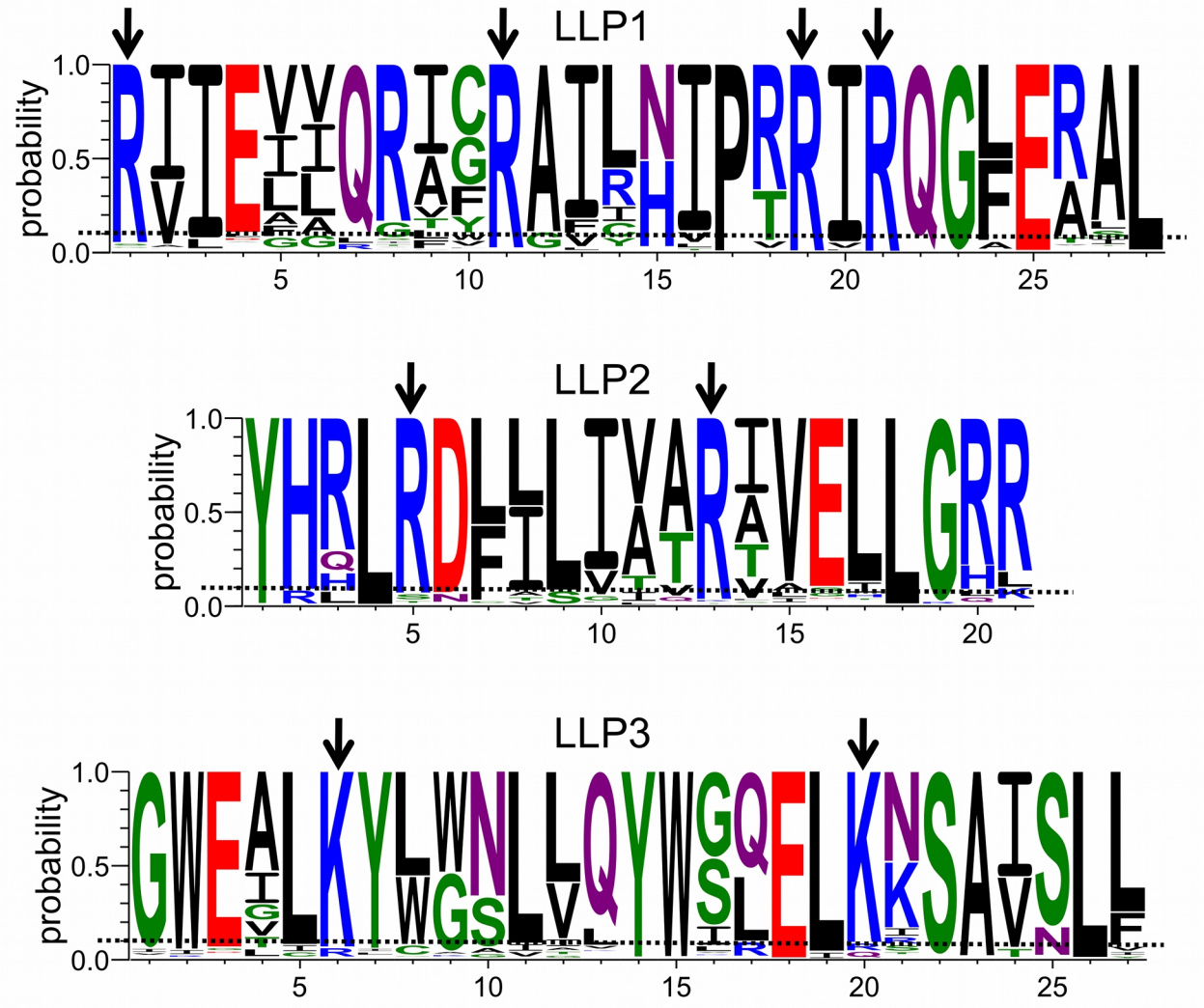

Figure 18. Specific arginine residues are highly conserved in LLP1 and LLP2

Logo diagrams were created using the probability scale to determine the most highly conserved residues in the LLP regions. In the cationic LLP1 and LLP2, there are six arginine residues that have ≥ 0.9 probability of occurrence across all clades/groups. In LLP3, arginine is not preferred, while there are two highly conserved lysine residues. Arrows represent positions with ≥ 0.9 probability of occurrence of the indicated amino acid.

Each PEP-FOLD run included 50 simulations followed by comparison of the returned models by cluster analysis to assess conformational diversity, with models < 3 Å Cα-RMSD (cRMSD; for peptides > 20 residues) grouped together in a cluster (Maupetit et al., 2009). Figure 19 shows the number of clusters returned by PEP-FOLD for each CTT domain and the relative percentage of the total population comprised by each cluster. The conformational diversity of the KE was high for all the clades/groups, with between 13 and 33 clusters identified, consistent with a random coil conformation. The LLP regions, on the other hand, were more conformationally homogeneous, with 21 of the 24 LLP regions tested returning only one cluster. Models from the three LLP domains that returned more than one cluster are shown in Figure 20. Clade A LLP2 (Figure 20A) and clade C LLP1 (Figure 20B) both returned two clusters of high helical content, with a kinked second, higher-energy structure. Group O LLP2 returned seven clusters. However, the 16 N-terminal amino acids (LLSNLASGIQKVISHL) were consistently helical while the remaining nine C-terminal amino acids (GLGLWILGQ) adopted a random coil conformation (Figure 20C). Finally, to assess overall conservation of the predicted secondary structure of the LLP domain, we compared the lowest energy PEP-FOLD models for each consensus domain sequence from all of the clades/groups (Figure 21) and determined the average cRMSD (Table 1). In general for the LLP peptides the average cRMSD was <1 indicating a high level of structural conservation. In contrast to the other consensus LLP sequences, a cRMSD value of approximately 6 was calculated for the group O LLP2 consensus sequence due to the predicted random coil nature of the C-terminal nine amino acids. As expected from the

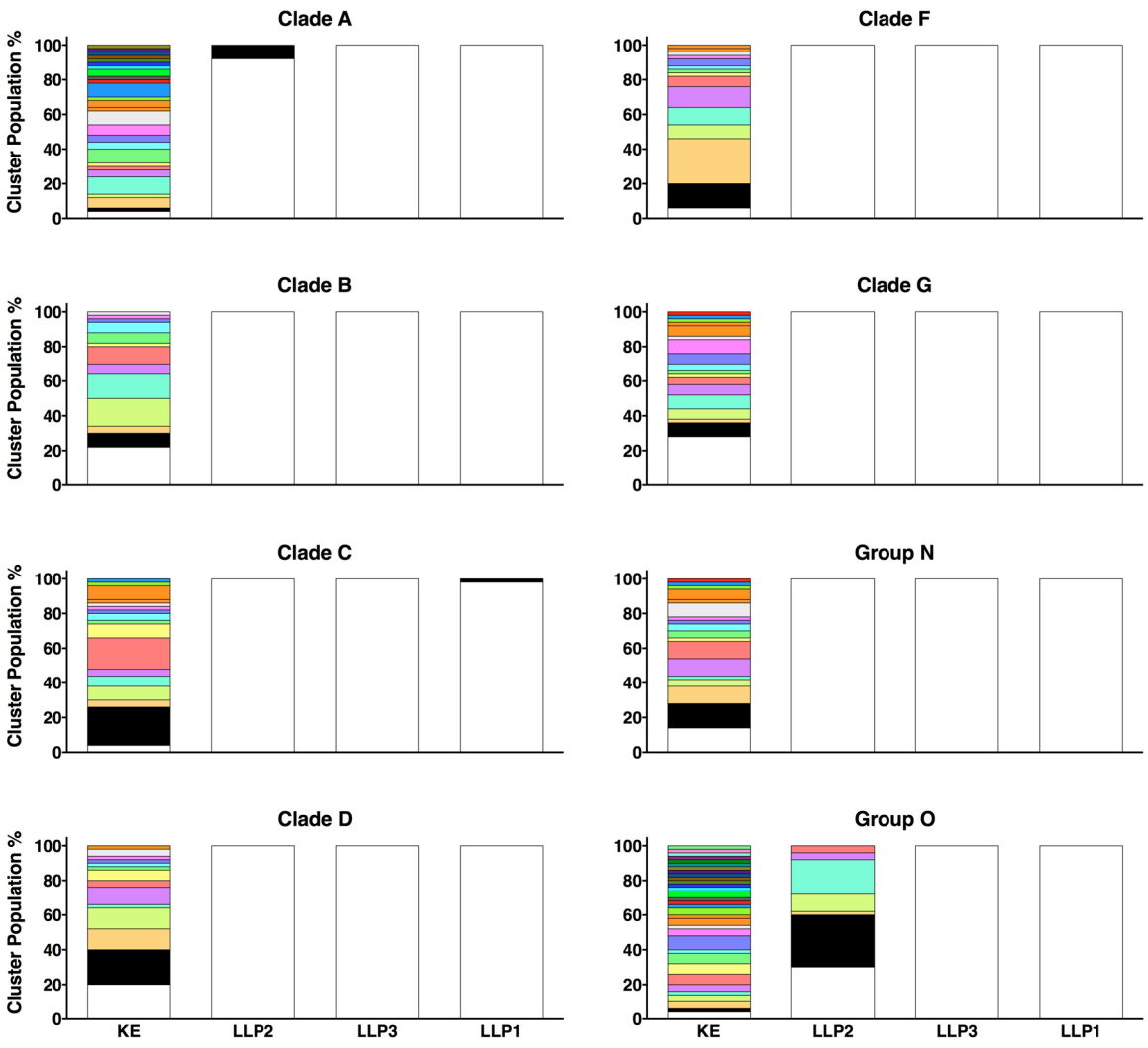

Figure 19. LLP sequences are predicted to be structurally homogeneous

Results from PEP-FOLD-based folding simulations were analyzed to determine structural homogeneity of CTT sequences. The number of clusters returned by PEP-FOLD for each peptide sequence was plotted with each structurally similar cluster represented by a different hue. For each cluster (hue), the height of each bar represents the relative prevalence of the cluster as a percentage of the total number of simulations. Fewer clusters indicate greater predicted structural homogeneity.

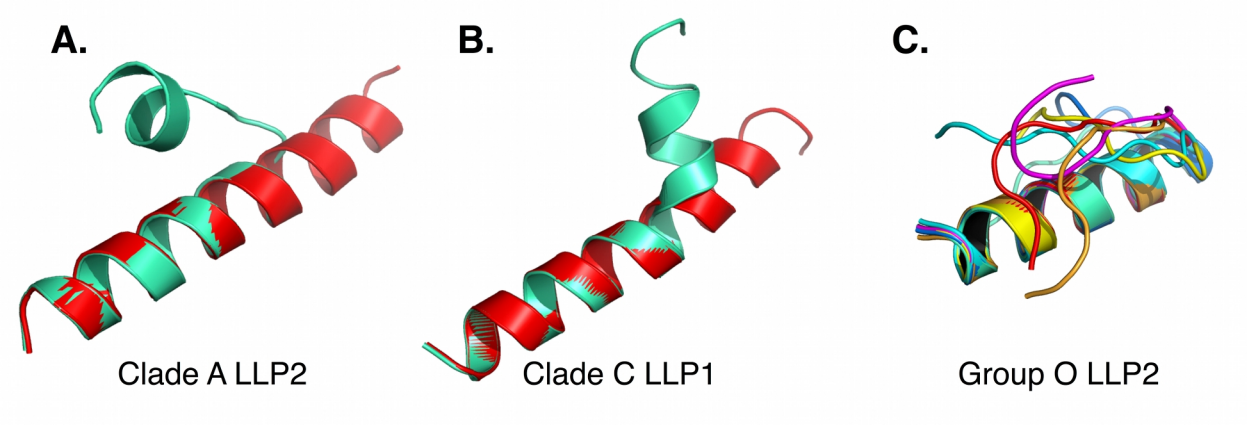

Figure 20. PEP-FOLD simulations return multiple clusters for some LLP peptides

The lowest energy structures from each returned cluster were aligned using the STAMP structural alignment method for (A) clade A LLP2, (B) clade C LLP1, and (C) group O LLP2. Each cluster is represented as a different hue. (A) and (B) contain two clusters of predominantly helical content, while (C) demonstrates greater heterogeneity due to predicted random conformations for the nine C-terminal amino acids.

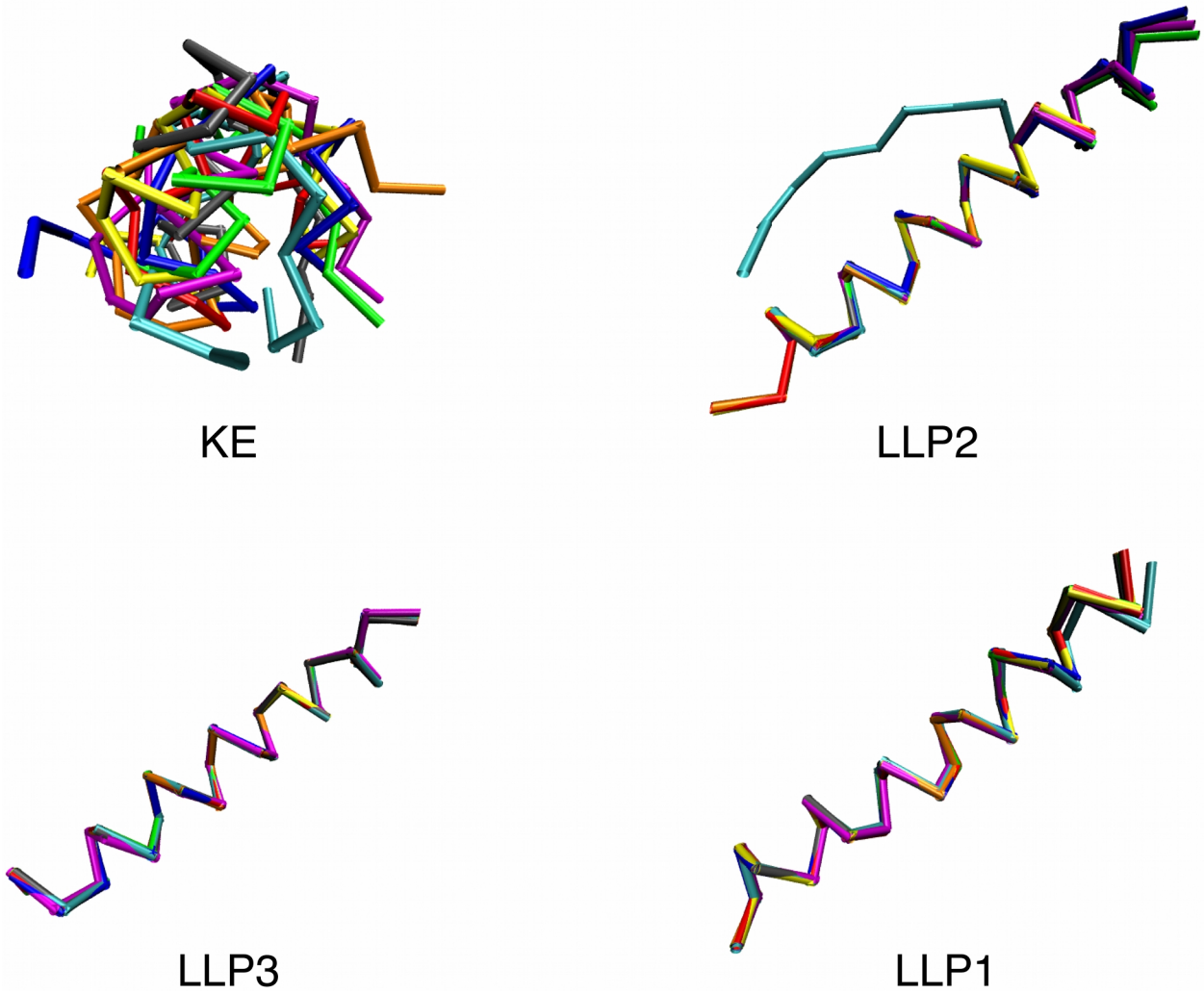

Figure 21. LLP peptide structural alignments reveal high level of structural similarity across clades/groups

The Cα traces from the lowest energy PEP-FOLD models for each peptide sequence from each clade/ group were aligned as described in Materials and Methods. Peptides are presented with the N-terminus on the left and the C-terminus on the right. KE demonstrated little structural similarity, consistent with a random coil conformation. Cα traces for the LLP peptides demonstrate high structure conservation across clades/groups with the exception of Group O LLP2. Clade A – blue; clade B – red; clade C – gray; clade D – orange; clade F – yellow; clade G – green; group N – magenta; group O – cyan

Table 1. Average cRMSD of PEP-FOLD-predicted structures

The lowest energy models from each clade/group were aligned and the average cRMSD was determined and compared between all clades/groups for each peptide region (KE, LLP2, LLP3, and LLP1). KE – Kennedy epitope; LLP – lentivirus lytic peptide

Clade/Group	KE	LLP2	LLP3	LLP1
Clade A	3.59427	0.83105	0.14509	0.12035
Clade B	3.57269	0.88990	0.07386	0.14126
Clade C	3.64798	0.91100	0.08960	0.11856
Clade D	5.87318	0.95789	0.05237	0.14982
Clade F	3.21947	0.98239	0.06398	0.12540
Clade G	2.92300	0.98992	0.06398	0.10418
Group N	3.17061	0.90760	0.26369	0.17850
Group O	6.47101	6.04013	0.12948	0.33790

predicted random coil conformation of the KE consensus sequence, the average cRMSD calculated for the KE ranged from ~2.9 - 6.5. These predictive experiments and analyses suggest that the alpha helical structure of the LLP regions is conserved across all the clades/groups for the consensus sequences generated from multiple sequence alignments.

3.3.4 LLP peptide analogs adopt helical structure in membrane-mimetic and hydrophobic environments

As the PEP-FOLD experiments predicted that consensus sequences of the LLP regions adopt stable helical conformations, we endeavored to explicitly measure the secondary structure of LLP sequence peptide analogs. Towards this objective, peptides were synthesized corresponding to the LLP regions from the consensus sequences from clades B and C, and group O and from the lowest average divergence (LAD) sequences determined from the CTT alignments for the same clades/groups (Table 2). These clades/group were chosen as they represent the most widely studied HIV-1 sequences (clades B and C) and the most structurally divergent as predicted by the PEP-FOLD analyses (group O, particularly for LLP2). CD spectroscopy measurements of the LLP peptides were performed in three solutions: 10 mM HEPES; 10 mM SDS in 10 mM HEPES; and 60% TFE in 10 mM HEPES. For peptides in aqueous solution, spectra were consistent with predominantly random coil conformation with the exception of LLP1 from the clade C LAD sequence and LLP2 from clade B (Figures 22A and 23). Addition of 10 mM SDS to 10 mM HEPES resulted in an increase (or a maintenance, in

Table 2. Sequences of peptides synthesized for CD spectra determination

Underlined residues are different between the Consensus and LAD sequences.

		LLP1	LLP2	LLP3
Clade B	Consensus	RVIEVVQRAFRAILHIPRRIRQGLERAL	YHRLRDLLLIVTRIVELLGRR	GWEALKYWWNLLQYWSQELKNSAVSLL
	LAD	RVIEVVQRACRAVLHIPVRIRQGLERAL	YHRLRDLLLIVTRIVELLGRR	GWEVLKYWWNLLQYWSQELKNSAVSLL
Clade C	Consensus	RIIELIQRICRAIRNIPRRIRQGFEAAL	YHRLRDFILIAARAVELLGRSSLRGLQR	GWEALKYLGSLVQYWGLELKKSAISLL
	LAD	RIIELIQGICRAIRNIPRRIRQGFEAAL	YHQLRDFILIAARAAELLGRSSLRGLQR	GWETLKYLGSLVQYWGLELKKSAISLL
Group O	Consensus	GIILGIQRIGRGILNIPRRIRQGLERSL	YHLLSNLASGIQKVISHLGLGLWILGQK	IIEACRLCAAVTQYWLQELQNSATSLL
	LAD	SIILGIQRIGRGIYNIPRRIRQGLERSL	YHLLSNLASGIQTVISHLGLGLRILGQK	IIDACRICTAVTQYWLQELQNSATSLL

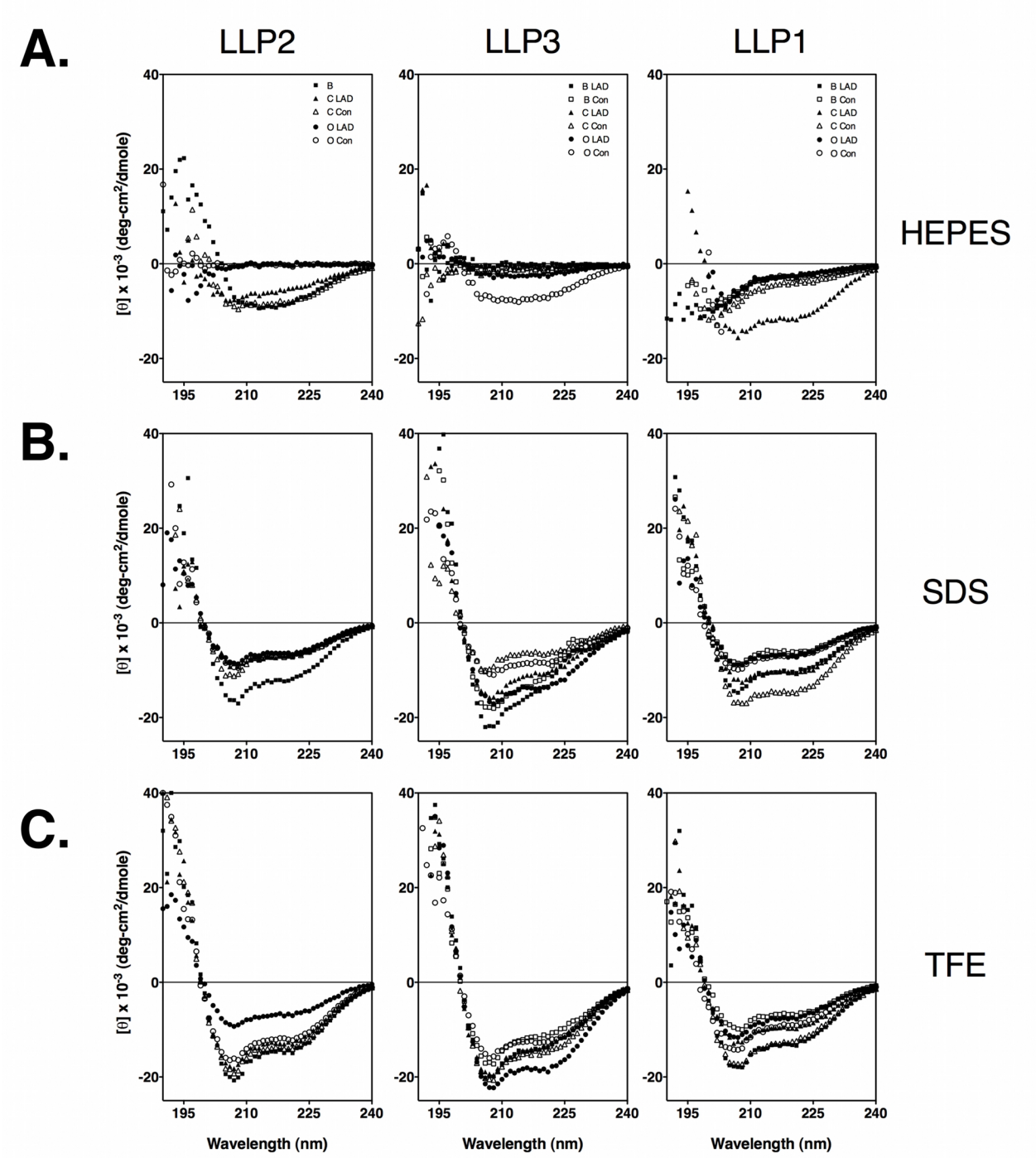

Figure 22. SDS and TFE induce increased helicity in synthetic LLP peptides

CD spectra from synthetic LLP peptides were determined in (A) 10 mM HEPES; (B) 10 mM HEPES + 10 mM SDS; and (C) 10 mM HEPES + 60% TFE.

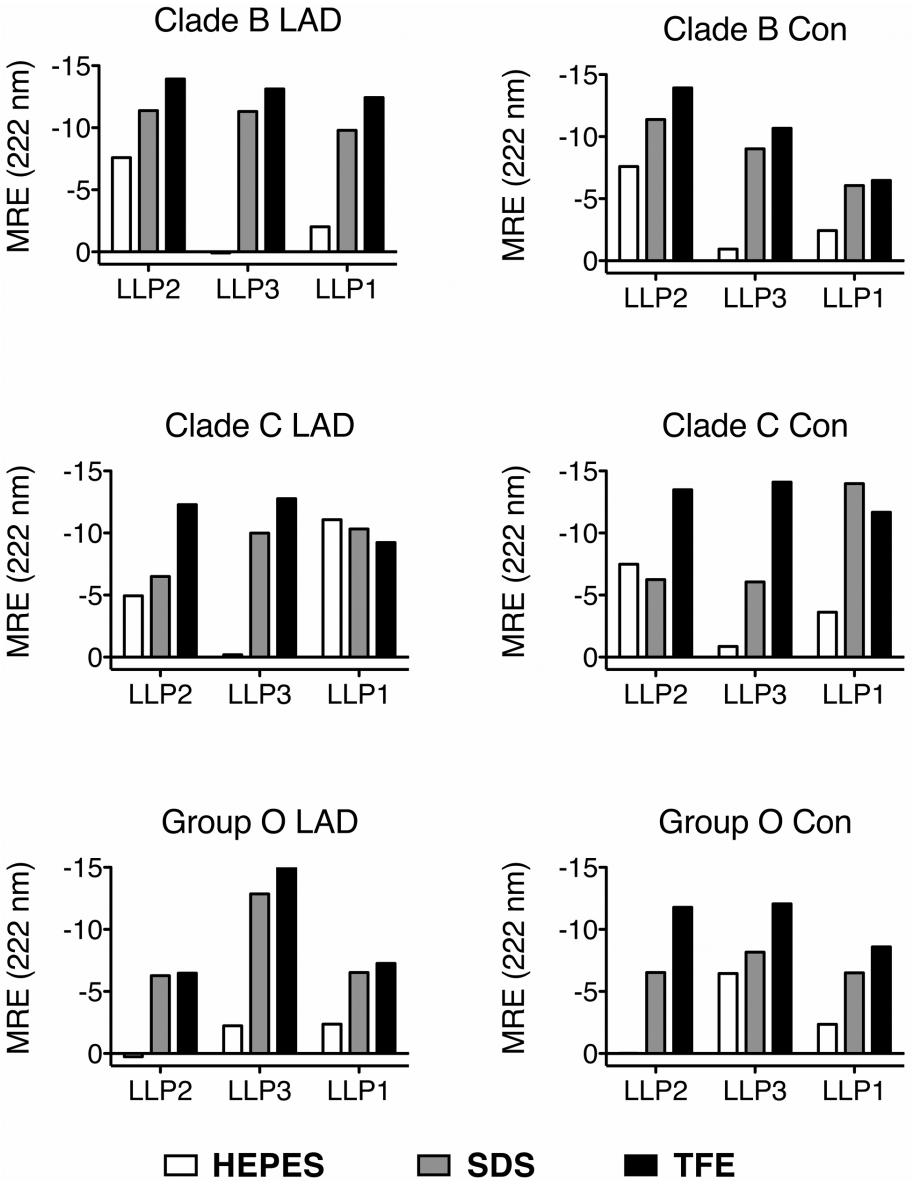

Figure 23. Relative helicity of LLP peptides by mean residue ellipticity

Mean residue ellipticity (MRE) at 222 nm was calculated for each peptide from the data presented in Figure 22. Results were plotted for each peptide-solute combination to facilitate a simple visual comparison of relative helicity of each peptide in the different solutes.

the case of clade B LLP2 and clade C LAD LLP1) in helical content relative to 10 mM HEPES (Figures 22B and 23). Likewise, addition of 60% TFE to 10 mM HEPES maintained, or moderately increased, helical content compared to that observed with 10 mM SDS (Figures 22C and 23). Results from CD spectroscopy measurements of LLP peptide analogs confirm predictions obtained by PEP-FOLD analysis that the alpha helical nature of the LLP peptides in membrane mimetic environments (i.e. SDS micelles and/or TFE) is broadly conserved among clades/groups in spite of extensive amino acid variation.

3.4 DISCUSSION

Contemporary research has placed Env structure and function as a high priority research focus for HIV-1. While the majority of published studies focus on the gp120 protein or gp41 ectodomain, recent studies have implicated the CTT in the modulation of overall Env structure and function (Kalia et al., 2003, 2005; Newman et al., 2007). These recent observations increase the need for a more detailed characterization of CTT sequences among diverse isolates of HIV-1 and the effects of sequence variation on CTT structure. Thus, the current study was designed to provide a comprehensive comparative analysis of CTT species among different groups and clades of HIV-1.

In the present study, we examined the conservation of the CTT of HIV-1 gp41 in clades A, B, C, D, F, and G and groups N and O as compared both to gp120 and to the gp41 ectodomain. In addition, we characterized the conservation of physicochemical and structural properties among variant CTT sequences. Our results demonstrate that

the CTT is intermediately conserved relative to the highly variable gp120 and the more conserved gp41 ectodomain. In spite of the relatively high level of amino acid diversity present in the CTT, the physicochemical properties of LLP domains are remarkably conserved across the clades/groups examined. Secondary structure predictions using PEP-FOLD indicate a high level of structure conservation in the LLP regions, results that are confirmed by CD spectroscopy of peptide analogs of the LLP regions from consensus and isolate sequences from clades B and C and group O.

The finding that the CTT exhibits intermediate sequence variation between the highly variable gp120 and the least variable gp41 ectodomain provokes thought as to the source of the evolutionary pressures and significance of the substantial sequence evolution on putative CTT functions. The relatively high level of sequence variation observed in HIV-1 gp120 is due to a combination of random genomic alterations associated with reverse transcription and selective pressure by host immune responses leading to evolution of Env gp120 populations during persistent infection. While the gp41 gene is also similarly subject to mutations during reverse transcription, the relatively low variation in the gp41 ectodomain is likely due to more strict requirements for structure and function in light of large conformational rearrangements that occur during the fusion process. Intermediate variation observed in the CTT suggests two possibilities: (i) CTT sequences are exposed to the immune system and therefore variation is observed as a result of immune selection; or (ii) the CTT is not an important functional domain in Env and as such can vary randomly without detrimental effect. In our opinion, the observed conservation of LLP structure and the selective incorporation of arginine residues indicate a functional importance that is incompatible with random

variation. Additionally, when we examined whether the presence of accessory protein coding sequences (Tat and Rev) affected CTT sequence variation, we found no significant differences in sequence variability between regions utilizing one (CTT-only), two (CTT and Rev), or all three (CTT, Rev, and Tat) open reading frames (data not shown). As LLP2 and LLP3 occur in the region including the Rev coding sequence, and LLP1 occurs in the CTT-only region, the conservation of the LLP regions is not likely due to extraneous influence (e.g. accessory proteins) and is further indicative of a functional relevance for the sequences. This perspective is further supported by the observations that minor point mutations in LLP sequences can result in major alterations in overall Env structural and functional properties (Kalia et al., 2005).

In stark contrast to the overall observed sequence variability in the CTT is the conservation, and indeed prevalence, of arginine in the CTT, especially relative to the other basic amino acid, lysine (c.f., Figure 18). While the two highly basic amino acids lysine and arginine are generally considered to be interchangeable and able to readily substitute for one another, the data presented here directly controvert that assumption for the CTT. As indicated by the current analyses of CTT basic residues, the CTT arginine residues are highly conserved and only very rarely replaced by lysine. In fact, arginine replacement by lysine in the CTT occurs at a much lower incidence as compared to its general prevalence in proteins. This selective preference for arginine in the CTT is likely due to functional requirements for the unique properties of the arginine side chain that terminates in a guanidinium group. Arginine demonstrates properties that make it particularly well suited to occur in membrane-interactive regions of proteins, as exemplified by the highly conserved arginine in the MSD of HIV-1 gp41. Primary is

the potential for the arginine side chain to “snorkel” from a membrane-embedded main chain position into the extramembranous space, providing an energetically favorable interaction with aqueous substrates (Killian and von Heijne, 2000). Additional recent controversy surrounds the potential for arginine to reside in or near the membrane center, although the mechanism by which this may occur is unclear (Dorairaj and Allen, 2007; Krishnakumar and London, 2007; MacCallum et al., 2007, 2008). It is thought that the guanidinium group of arginine may be hydrated by membrane-resident waters (Krepkiy et al., 2009). More clearly demonstrated and well-established is the cation-pi interaction of the arginine side chain with proximal phenylalanine, tryptophan, and tyrosine side chains (Dougherty, 1996; Ma and Dougherty, 1997). In this interaction the dispersed positive charge of the guanidinium group interacts with the quadrupole moment of planar hydrocarbon rings to form an energetically stable “bond” that can subsequently pass into the membrane hydrocarbon core (Dougherty, 1996; Jing et al., 2003). Any and all of these arginine-specific properties may play a role in the well-documented ability of arginine-rich sequences to pass through biological membranes (Futaki, 2005; Mitchell et al., 2000; Tung and Weissleder, 2003), and it is this metaproperty that likely is of the greatest significance to the prevalence of arginine in the CTT.

The sequence variation that is characteristic of the CTT appears to indicate an exposure to immune selection perhaps indicative of surface accessibility of the target sequences. Recent findings have demonstrated transient exposure of LLP2 sequences on the surface of Env-expressing cells (Lu et al., 2008). In particular, Chen and colleagues determined that LLP2 is exposed to antibody binding during cell-cell fusion

under conditions that slow the fusion process (i.e. at 31 °C), but not under physiological conditions (37 °C), suggesting highly dynamic LLP2 exposure. This is consistent with published data demonstrating lipid-association of LLP2 both pre- and post-fusion (Viard et al., 2008). The arginine-rich nature of LLP2 may contribute to the demonstrated transient exposure of LLP2 requiring membrane penetration and traversion, as arginine-rich peptides have been demonstrated to freely cross biological membranes, and indeed have been utilized to transport soluble proteins into live cells (Futaki, 2005; Inomata et al., 2009; Mitchell et al., 2000; Tung and Weissleder, 2003). Additionally, this ability of arginine-rich peptides to carry soluble proteins across membranes may be the mechanism by which the KE appears to be transiently exposed on the virion surface during the fusion process (Heap et al., 2005; Reading et al., 2003). As LLP2 is moving through the membrane, it is plausible that it can transport the highly charged KE through to the extravirion space where it becomes a neutralizing epitope under proper conditions.

In addition to conservation of specific arginine residues in the CTT in an environment of generally high sequence variation, we also found that the physicochemical and structural properties of the CTT are well conserved across clades/ groups. LLP regions of the CTT display significantly higher hydrophobic moments than the MPER and MSD regions, the other membrane-interactive regions of gp41, indicating a preference for the membrane-water interfacial region (Eisenberg et al., 1984). LLP1 and LLP2 also display conserved cationic properties, due to the enrichment in arginine described above, and the amphipathic nature of the LLP sequences is also conserved across clades/groups. Conservation of these biochemical properties suggests that there

is functionality to the CTT beyond current characterizations. Conservation of amphipathicity and high hydrophobic moments suggest that the LLP sequences help to localize and anchor the majority of the CTT sequence to the inner leaflet of the cell membrane. This is consistent with findings indicating membrane association of LLP1 and LLP2 in virus and cells (Viard et al., 2008).

The finding that the secondary structure of the LLP peptide sequences is conserved is also suggestive of critical but undefined functionality for the CTT. That LLP2 has been demonstrated to alternate accessibility to antibody binding during the fusion process suggests that the LLP2 sequence must traverse the membrane during the fusion process. The adoption of helical secondary structure would facilitate this process as helical structure internally satisfies all hydrogen bonding requirements and minimizes the free energy of the peptide-lipid interaction (White and Wimley, 1999). In conjunction with the demonstrated preference for arginine in the CTT, conserved helicity may be a means by which LLP2 is able to pass through the biological membrane during the fusion process and become accessible to antibody binding. Indeed, when conserved arginines in LLP2 were replaced by glutamate, Env-mediated fusogenicity and virion infectivity was abolished, even though Env virion incorporation was similar to wild-type virus (Kalia et al., 2003).

Recent studies have highlighted the ability of the CTT to modulate HIV-1 Env structure and functional properties, but the mechanisms for this control remain to be defined. In addition, the current CTT sequence and structure studies indicate highly conserved physicochemical and structural properties that may mediate critical functions in the life cycle of HIV-1. In light of the genetic economy intrinsic to viruses, there

appears to be increasing data for a reevaluation of CTT functional relevance in HIV replication and Env structural and antigenic properties.

4.0 TOPOLOGY OF THE C-TERMINAL TAIL OF HIV-1 GP41: DIFFERENTIAL EXPOSURE OF THE KENNEDY EPITOPE ON CELL AND VIRAL MEMBRANES

 Jonathan D. Steckbeck, Chengqun Sun, Timothy J. Sturgeon, Ronald C. Montelaro. Topology of the C-terminal tail of HIV-1 gp41: Differential exposure of the Kennedy epitope on cell and viral membranes. PLoS One. 2010; 5:e15261.

4.1 INTRODUCTION

Human immunodeficiency virus (HIV) infects humans predominantly through interaction of the viral envelope glycoprotein (Env) with the primary receptor CD4 and coreceptors CCR5 or CXCR4 on the surface of target cells. Env is produced as a 160 kDa polyprotein that is subsequently processed by extensive glycosylation, multimerization, and proteolytic cleavage to yield the virion-associated trimeric complexes of non-covalently associated gp120-gp41 dimers (Evans and Desrosiers, 2001; Luciw et al., 2002). Numerous studies have identified Env as a primary determinant of viral phenotypes; variations in Env sequence can affect cellular tropism, viral replication levels, immune recognition, and pathogenesis (Evans and Desrosiers, 2001; Luciw et al., 2002). Additionally, Env sequence variation has recently been experimentally demonstrated to be a primary determinant of lentivirus vaccine efficacy (Craigo et al., 2007). The majority of Env structural studies have focused on gp120 and the ectodomain of gp41; there is to date no definitive structural information on the approximately 150 amino acid long C-terminal tail that follows the proposed membrane-spanning domain (MSD) of gp41.

Studies addressing the CTT have traditionally examined: (i) the role of the CTT in viral Env incorporation (Freed and Martin, 1995, 1996; Jiang and Aiken, 2007; Murakami and Freed, 2000a, b); (ii) the influence of the CTT on virion maturation (Jiang and Aiken, 2007; Kol et al., 2007; Wyma et al., 2004); and (iii) the function of predicted endocytic signals present in the CTT (Byland et al., 2007; Ohno et al., 1997). Various

studies of the interactions of both cellular and other viral proteins as intracellular partners with Env have implicitly reinforced the traditional model of CTT topology as being localized completely within the inner surface of the cell or viral lipid membrane (Figure 24, left).

Early evidence for an alternative topological model for the CTT was provided by Kennedy and colleagues (Chanh et al., 1986; Kennedy et al., 1986) who first reported that antiserum produced against a synthetic peptide from gp41 (residues 728-745) bound to HIV-1 Env, and that serum from HIV-1-infected humans also recognized this synthetic peptide (Kennedy et al., 1986). Importantly, this group subsequently reported that antiserum raised against this synthetic peptide could specifically neutralize HIV *in vitro* (Chanh et al., 1986). These observations indicated exposure of the Kennedy epitope (KE) on the virion surface to allow antibody binding and neutralization, in direct contrast to the presumed intracytoplasmic location of the entire C-terminal sequences of gp41 following the MSD.

More recently, Dimmock and colleagues have attempted to address this apparent discrepancy between the traditional model of an exclusively intracytoplasmic CTT and an alternative model where the KE is exposed (Cleveland et al., 2003; Hollier and Dimmock, 2005). Using antibodies directed to the ^{739}ERDRD^{743} sequence and MAbs directed to the upstream ^{727}PDRPEG^{732} and ^{733}IEEE^{736} sequences in the KE, Dimmock and colleagues demonstrated virion binding and viral neutralization that was abrogated after pre-exposing virions to proteases (Cleveland et al., 2003). Subsequent

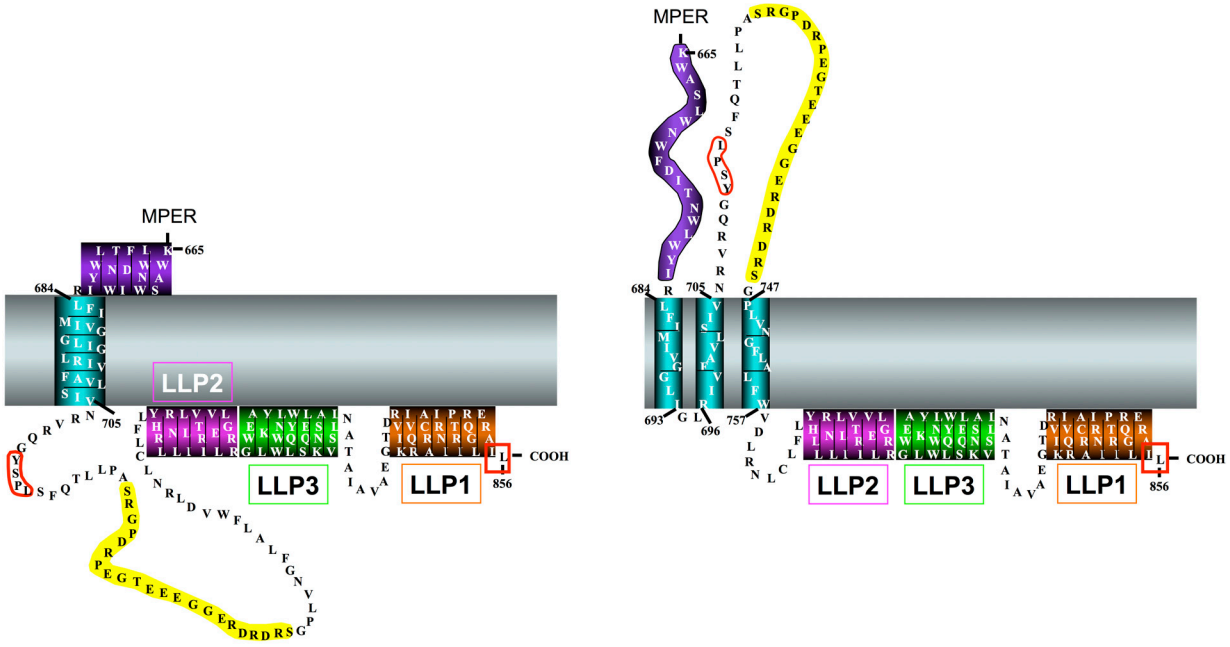

Figure 24. Schematic models of the HIV-1 CTT

(Left) Traditional CTT model with one membrane-spanning α-helix and a completely intracytoplasmic localization of the remaining CTT sequence. LLP domains have been placed at their presumed membrane-localized position. (Right) Alternative CTT model with multiple MSD segments as proposed by Hollier and Dimmock (Hollier and Dimmock, 2005). This model proposes three membrane-spanning β-sheets and an extracellular localization of the KE.

experiments using a MAb, SAR1, directed to ^{739}ERDRD^{743} demonstrated that SAR1 effectively neutralized HIV-1 by apparently post-attachment binding at lower temperatures; negligible neutralization was observed when virus and cells were pre-incubated at 37 °C (Heap et al., 2005; Reading et al., 2003). SAR1 was also shown to effectively inhibit virus replication in HIV-1 infected cells incubated in the presence of SAR1 antibody (Reading et al., 2003), likely through a decrease in cell-cell spread mediated by SAR1 inhibition of HIV-1 gp41-induced cell-cell fusion (Cheung et al., 2005). On the basis of these antibody accessibility studies, Dimmock and colleagues have proposed an alternative model for the topology of the HIV-1 gp41 CTT that predicts multiple MSDs and surface loops (Figure 24, right) (Cleveland et al., 2003; Hollier and Dimmock, 2005).

This multiple MSD topology model is compatible with various studies that have identified intracellular partners proposed to interact with CTT sequences (Freed and Martin, 1996; Lopez-Vergès et al., 2006; Murakami and Freed, 2000a; Ohno et al., 1997; Wyma et al., 2000), but is incompatible with recent data indicating a requirement for both the tyrosine-based sorting signal ^{711}GYSPL^{715} and the extreme C-terminal dileucine sorting signal (boxed in red in Figure 24) in the CTT to facilitate Env endocytosis from the cell surface (Byland et al., 2007; Ohno et al., 1997; Rowell et al., 1995).

Finally, more recent studies have examined the accessibility of the lentivirus lytic peptide (LLP) regions of the CTT (c.f. Figure 24) to antibody on Env-expressing cells (Lu et al., 2008). Under native conditions, cell surface Env was not reactive with LLP1 and LLP2-specific antibodies, consistent with a sequestered localization of the LLP1

and LLP2 sequences (Viard et al., 2008). Interestingly, however, when Env-expressing cells were incubated with CD4- and CCR5-expressing target cells at 31.5 °C, binding of LLP2-specific antibodies was observed (Lu et al., 2008). LLP2-specific MAbs did not bind when effector and target cells were incubated at 37 °C, indicating a highly transient exposure of the LLP2 sequence (Lu et al., 2008), consistent with prior reports of LLP2 membrane association post-fusion (Viard et al., 2008). These results indicate the need to reconsider the current static model of CTT topology and suggest that the CTT may be a more dynamic structure than previously thought.

In the current study, we sought to test the proposed topology models by experimentally determining the surface exposure of the KE sequences of gp41 for comparison to the external and internal locations predicted by the respective CTT models. Towards this goal, we have examined the accessibility of the KE to antibody binding on the surface of Env-expressing cells and on the surface of intact mature virions using a combination of flow cytometry, immunoprecipitation, and surface plasmon resonance (SPR) spectroscopy techniques that minimize the potential for disruption of the lipid membrane, the integrity of which is crucial to topology mapping studies (see Figure 25 for CTT sequences used). Our results indicate that the KE is topologically distinct on the cell surface as compared to the virion and that the topology of the CTT may be more complex and dynamic than indicated by current Env structural models.

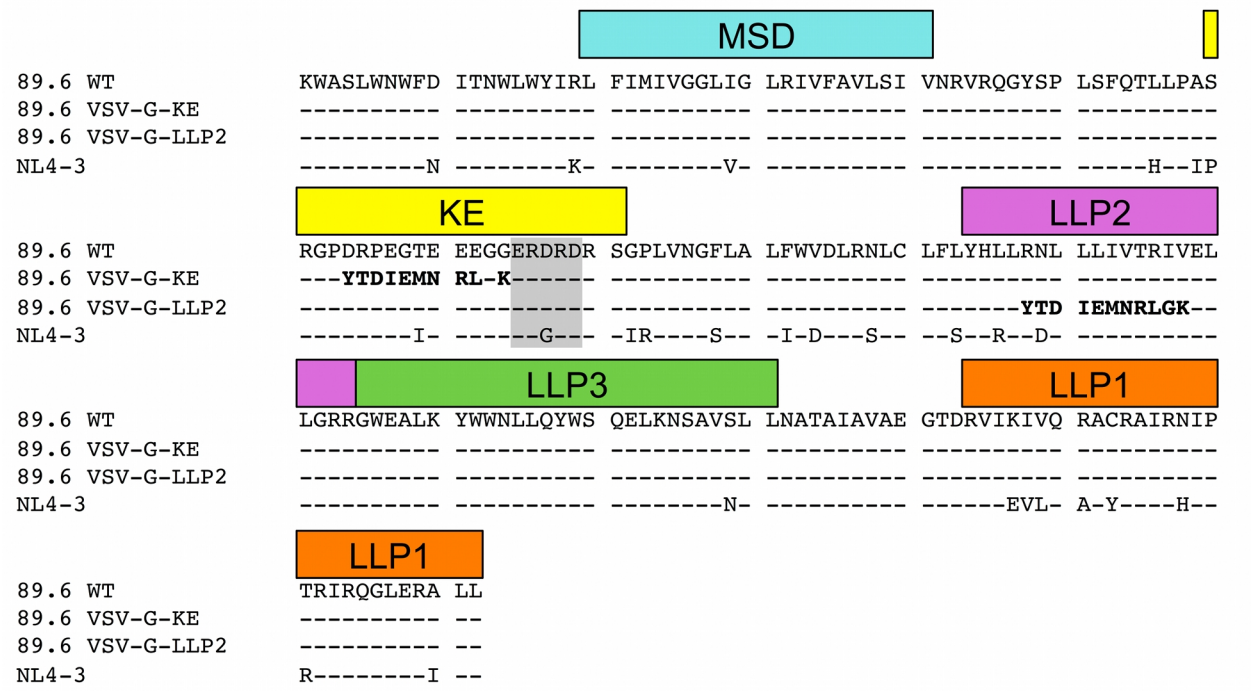

Figure 25. Sequence alignment of variant Env CTT

Sequence alignment of the membrane-spanning domain (MSD) and CTT of the Env proteins used in this study. All Env proteins were full-length gp160, however, only the sequences from K665 to the C-terminus are shown for simplicity. Structural and sequence domains are indicated in boxes above the corresponding sequence. VSV-G epitope tag indicated in bold; SAR1 and 1577 epitope indicated in gray box. KE – Kennedy epitope; LLP – lentivirus lytic peptide, MSD – membrane spanning domain

4.2 MATERIALS AND METHODS

4.2.1 Cells and virus preparations

HEK293T cells (ATCC) were maintained in DMEM (Invitrogen) supplemented with 10% (v/v) FBS. Cells were used at passage numbers less than 30. Gradient purified aldrithiol-2-inactivated HIV-1 NL4-3 virus (generously provided by Dr. Jeff Lifson, SAIC Frederick, Inc) used in this study has been described previously (Rossio et al., 1998) and recently used for cryo-electron microscopy studies of virion-associated Env (Liu et al., 2008; Zanetti et al., 2006; Zhu et al., 2006; Zhu et al., 2008). The SAR1 hybridoma was a kind gift from Dr. Nigel Dimmock.

4.2.2 VSV-G epitope tag substitutions

As described for topology mapping of other membrane proteins (Anand, 2000; Das et al., 2008; Obermeyer et al., 2007), we determined the antibody reactivity of selected Env constructs containing selected VSV-G epitope substitutions in the CTT. For these studies, the codon-optimized Env gene of HIV-1 (89.6) was cloned into a p2CI vector derived from PCR2.1 by insertion of the CMV promoter and polyA signal sequence from pcDNA3.1(hygro) (Invitrogen) and PCR-amplified IRES-Neomycin resistant sequence from pFB-Neo-LacZ vector (Stratagene). Overlapping PCR was used for the construction of the VSV-G substitution mutants of gp160. The two hybrid primers were

constructed containing VSV-G tag sequences at the 5'-ends and HIV gp160 specific sequences at the 3'-ends of both the primers used for the substitutions. Final overlapping PCR products were then subcloned into the HIV-1 gp160 expression vector using the appropriate restriction enzymes and the Rapid DNA Ligation Kit (Roche Applied Science). All mutants were verified by DNA sequencing.

4.2.3 Cell culture and transfections

HEK293T cells were plated into six-well plates 24 hours prior to transfection. 80% confluent cells were transfected with 2.5 µg of the selected HIV-1 Env mutant DNA using Lipofectamine™ LTX reagent and PLUS™ reagent, as recommended by the manufacturer (Invitrogen).

4.2.4 Cellular FACS analysis

Cells were harvested 24 hours post-transfection by treatment with 2 mM EDTA. Cells were resuspended and washed twice with FACS wash buffer (1X PBS with 5% FBS) at 4 °C prior to antibody staining. Reference MAbs to VSV-G (Roche Diagnostics), gp41 (SAR1), and actin (Sigma-Aldrich) were labeled immediately prior to staining using Zenon labeling kits (Invitrogen) following manufacturer's instructions. 10^6 cells were dual-stained by incubating with 5 µg fluorophore-labeled antibody (VSV-G or gp41) and 7.5 µg fluorophore-labeled actin antibody for 30 minutes on ice. Following staining, cells were washed thrice with FACS wash buffer at 4 °C. Washed cells were resuspended and stained with 7-amino-actinomycin D (7-AAD). Cells were not fixed prior to analysis.

Fluorescently-labeled cells were analyzed on a FACSAria (BD Biosciences). Live, intact cells (7-AAD negative) were selected for scatter characteristics, including selection of single-cell populations by doublet-discrimination analysis. PMT settings were adjusted on identically-stained mock transfected cells prior to analysis of Env-transfected cells. Data was collected for $5x10^4$ 7-AAD negative cells and analyzed for reactivity with fluorescently-labeled VSV-G or gp41 and actin antibodies.

4.2.5 Isolation of detergent resistant membranes

HEK293T cells were transfected with 89.6 WT, VSV-G-KE, or VSV-G-LLP2 plasmids. 48 hours post-transfection, cells were subjected to sucrose gradient floatation as previously described (Chan et al., 2005). Briefly, cells were collected and washed in ice-cold 1X phosphate buffered saline (PBS). Washed cells were extracted on ice for 30 minutes with 0.5 ml ice-cold 1% Triton X-100 in NTE buffer (25 mM Tris-HCl; 0.15 M NaCL; 5 mM EDTA; pH 7.5) with phenylmethylsulfonyl fluoride and complete protease inhibitor (Roche). Cell lysates were centrifuged at 8,000 x g for 10 minutes at 4 °C. Clarified extracts were mixed with an equal volume of 85% sucrose in NTE and loaded into a prechilled centrifuge tube. Extracts were overlaid with 6 ml 30% sucrose in NTE, followed by 5 ml 5% sucrose in NTE. Tubes were centrifuged at 100,000 x g for 18 hours at 4 °C. The gradients were fractionated from the top manually by pipetting into ~1 ml fractions. Aliquots of the samples were precipitated with trichloroacetic acid, resuspended in SDS-PAGE running buffer, and subjected to SDS-PAGE followed by western blot with anti-gp41 MAb Chessie 8. Blots were developed with PicoWest substrate (Pierce) and visualized on X-ray film.

4.2.6 Viral immunoprecipitation and western blotting

Protein G Dynabeads (Invitrogen) were prepared according to the manufacturer's directions. Briefly, anti-Env or anti-Gag antibodies (4 µg) were incubated with 20 µl protein G Dynabeads in 35 µl citrate-phosphate buffer, pH 5.0, with gentle shaking for 45 minutes at room temperature. Isotype-matched IgG controls were used for each species (murine, human, etc.) from which a MAb was derived. Beads were washed thrice with 0.5 ml citrate-phosphate buffer followed by resuspension in either 26 µl PBS (for intact virus) or 26 µl PBS with 1% Triton X-100 (for lysed virus) and 4 µl NL4-3 virus (equivalent to 1.1 µg p24). Virus-bead suspensions were incubated at 4 °C for one hour with gentle shaking and subsequently washed thrice with 1X PBS. Following the final wash, beads were resuspended in NuPAGE SDS-PAGE buffer, heated at 70 °C for 10 minutes, and the supernatant loaded onto 4-12% Bis-Tris NuPAGE gels. Gels were electrophoresed followed by transfer to polyvinylidene fluoride (PVDF) membranes using the Invitrogen iBlot system. Blots were blocked for one hour in 5% blotto (1X PBS with 5% dry milk). After blocking, blots were cut to allow separate staining of gp120 (>60 kDa), gp41 (30 – 60 kDa), and p24 (<30 kDa). gp120 was stained with rabbit anti-gp120 (Advanced Biotechnologies, Inc.), gp41 stained with Chessie 8, and p24 stained with Ag3.0 for 1.5 hours at room temperature. Blots were washed thrice with 1X PBS and 0.025% Tween 20 (PBS-T), followed by incubation with appropriate secondary antibody (anti-rabbit IgG or anti-mouse IgG conjugated to horseradish peroxidase) for one hour at room temperature. Blots were washed thrice in PBS-T with the gp120 blot receiving an additional wash in 1X PBS with 0.1% Triton X-100. Finally, blots were incubated with

PicoWest substrate (Pierce) for one minute and reassembled for visualization on X-ray film.

4.2.7 Western blot quantitation

Antibody IPs were quantified by densitometry analysis. For each IP, three independent X-ray exposures were scanned and analyzed using ImageJ (NIH). For intact virus, p24 bands, and for solubilized virus bands from each MAb's target antigen, were selected for each protein, and the integrated area under the densitometry curve was compared to that of the viral input band to yield percent of the input protein in the immunoprecipitate. Percent inputs for each protein for each of the three exposures were averaged to yield the overall percent input immunoprecipitated per experiment. This was repeated for three independent experiments, and the results were averaged to yield the final percent input immunoprecipitated per antibody. One-way ANOVA analysis was performed in Prism 5.0b with the Dunnett Multiple Comparison test and IgG as the control column. The P value of the test was <0.0001. Individual P values for antibodies compared to the IgG control are indicated in each data set.

4.2.8 SPR spectroscopy analysis of antibody-virus interactions

Antibody-virus interactions were analyzed using SPR spectroscopy using a Biacore 3000 instrument. Protein A was immobilized on CM5 sensor chips as described (Steckbeck et al., 2006; Steckbeck et al., 2005). For analysis of the binding interaction, MAbs were captured on the protein A surface to ~1500 RU with a protein A-only surface

serving as a reference surface to account for virus-protein A non-specific interactions. Virus (2.5 mg/ml total protein; 172 µg/ml p24) diluted in 10 mM sodium phosphate buffer, pH 7.2 was injected over parallel reference and MAb-captured flowcells at 10 µl/ min for 2 minutes. A buffer-only injection served to account for MAb dissociation from the protein A surface and was subtracted from each MAb binding experiment. Thus, presented sensorgrams are double referenced (Myszka, 1999).

4.3 RESULTS

4.3.1 Exposure of CTT sequences on the surface of Env-transfected cells

To experimentally determine the exposure of the KE on the surface of Env-expressing cells, the antibody reactivity of transfected cells expressing HIV-1 89.6 Env with the VSV-G tag substituted in the KE (VSV-G-KE) or LLP2 (VSV-G-LLP2) segments was analyzed. The selected epitope allowed the use of a high affinity commercial anti-VSV-G MAb to compare the accessibility of a common epitope in the context of the KE segment and in a putative internal LLP2 segment of the CTT (Lu et al., 2008; Viard et al., 2008). Cells transfected with VSV-G-KE or VSV-G-LLP2 were compared to cells transfected with 89.6 WT following staining with anti-VSV-G and anti-actin antibodies. Anti-actin was used as a control for cellular membrane integrity, providing the ability to distinguish VSV-G positive intact cells from cells with compromised membranes.

As shown in Figure 26A, VSV-G-KE-expressing cells (middle) exhibit increased VSV-G-specific staining with no increase in actin staining compared to 89.6 WT (left)

transfected cells, while VSV-G-LLP2-expressing cells (right) stain for neither VSV-G nor actin, similar to the control cells. Since the insertion of the VSV-G tag into the KE strategically conserved a key epitope for an existing KE-specific MAb (SAR1), cells expressing the various Env constructs were also tested for reactivity with this native sequence-specific MAb. All cells expressing Env show similarly increased SAR1 reactivity compared to mock-transfected cells without an increase in actin staining (Figure 26B). In addition to confirming the KE antibody accessibility results obtained with VSV-G-KE, this result also suggests that the observed lack of staining of 89.6 WT and VSV-G-LLP2 in Figure 26A is not due to decreased cell surface gp160 expression, as SAR1 reacted similarly with all Envs. Finally, to demonstrate the sensitivity of the actin stain to detect permeable membranes, SAR1 staining was repeated on permeabilized cells (Figure 26C). When the cell membrane integrity was compromised, anti-actin fluorescence intensity increased at least 10-fold, and 75-85% of permeabilized cells stained positive for actin. This result indicates that the lack of actin staining

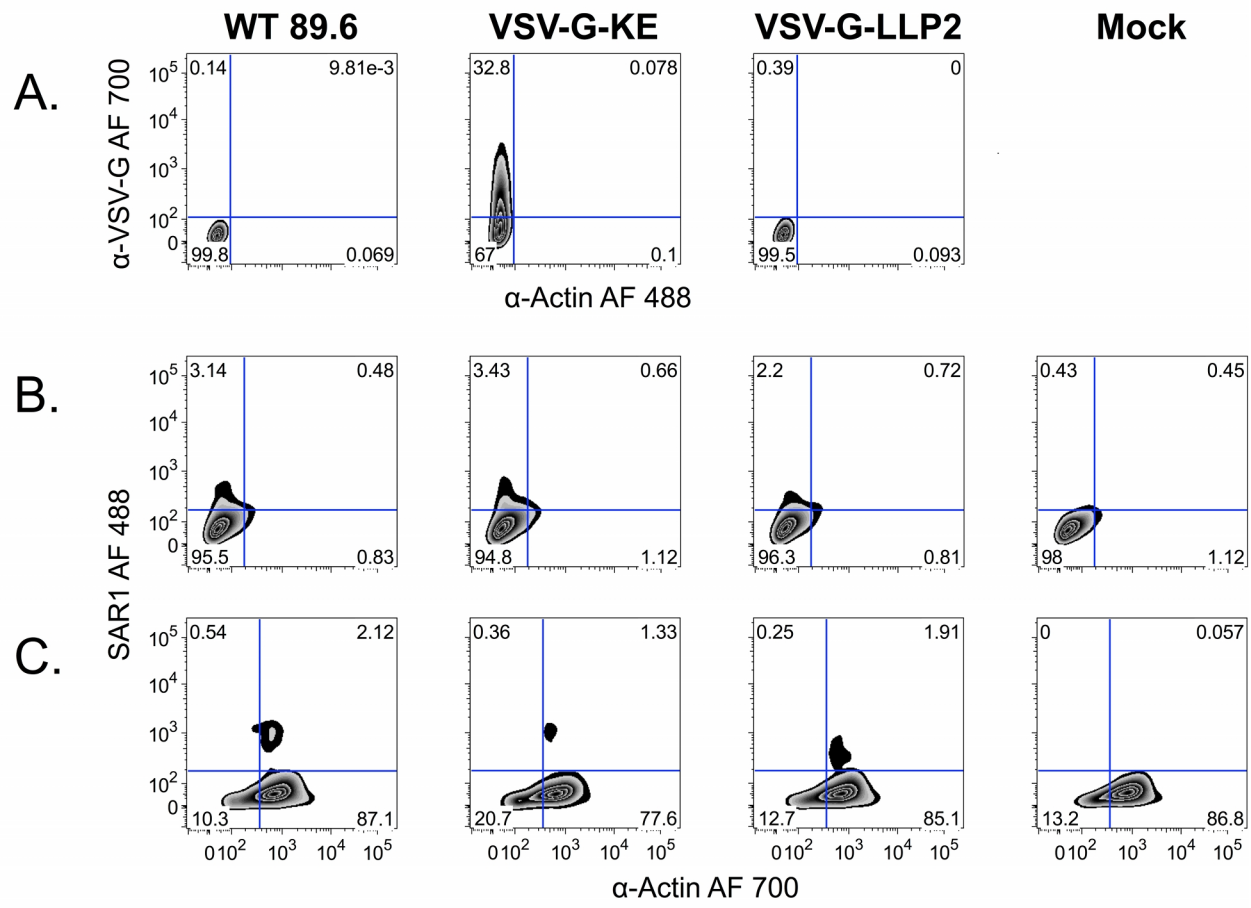

Figure 26. FACS analysis of intact Env-expressing cells demonstrates extracellular exposure of the KE sequence

Cells transfected with HIV-1 89.6 WT, VSV-G-KE, or VSV-G-LLP2 were analyzed by FACS to determine VSV-G epitope accessibility to anti-VSV-G MAb. A.) Intact, non-permeabilized Env-expressing cells were stained with α-VSV-G (AlexaFluor (AF) 700) and α-actin (AF 488). B.) Intact cells expressing Env were stained for surface exposure of the KE using a native KE antibody, SAR1. C.) Same as (B) except cells were permeabilized prior to staining.

observed in Figure 26A and 26B reflects intact cell membranes and not poor actin reactivity.

Finally, as the CTT has been shown to contain signals important for association with detergent-resistant membranes (DRMs) (Yang et al., 2010), association of the Env constructs with detergent-resistant membranes (DRMs) was determined. Cells expressing 89.6 WT, VSV-G-KE, and VSV-G-LLP2 were subjected to membrane sucrose floatation gradients. As seen in Figure 27, all three Env constructs exhibited similar association with DRMs, suggesting that the VSV-G tag replacement strategy did not result in disruption of cell surface Env distribution. Taken together, these results indicate that the KE can be exposed on the surface of the plasma membrane, as the KE sequence is accessible to antibody binding in intact cells expressing Env, both in the context of the VSV-G epitope tag substitution and the native sequence.

4.3.2 KE-specific antibodies do not bind to intact viral particles

Next, immunoprecipitation studies were performed to assess the ability of MAbs directed at the KE to pull down intact viral particles using protein G-coated paramagnetic beads. Both a gp120-specific antibody (termed C11) and a gp41 ectodomain specific antibody (7B2) were used to provide positive controls for Env binding and virion pull down. A p24-specific MAb (183-H12-5C) was used to detect disrupted virions in the pull down assays as a control for virion integrity. Of particular importance, several reference MAbs (2F5, 4E10, Z13e1) directed to the membrane proximal external region (MPER) of gp41 were used as controls, as this epitope is

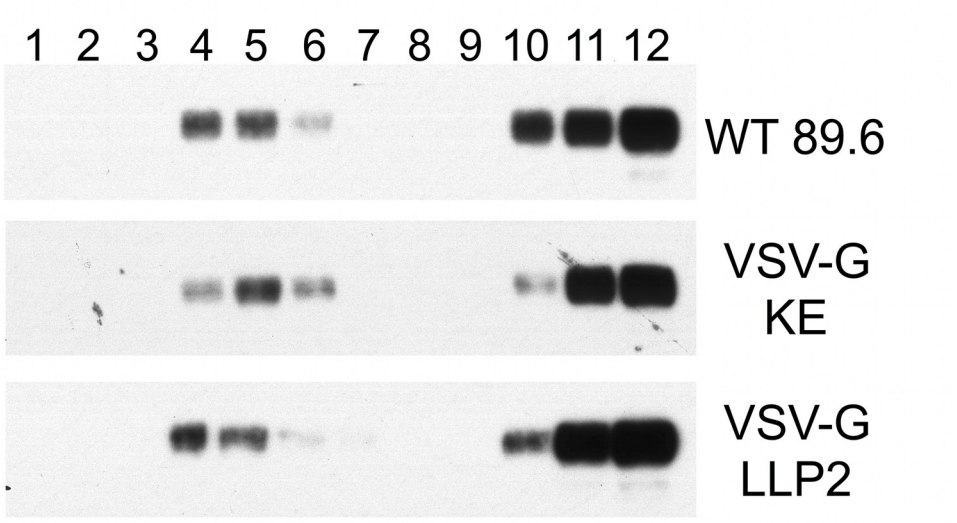

Figure 27. VSV-G epitope insertions do not disrupt Env association with detergent-resistant membranes

Env association with detergent resistant membranes was determined by sucrose gradient floatation followed by western blot analysis. Lane 1 represents the top of the gradient and lane 12 the bottom of the gradient. Blots were stained with anti-gp41 MAb Chessie 8 to determine the localization of gp41 in the bands shown.

known to be closely associated with the membrane surface (Haynes and Alam, 2008; Schibli et al., 2001; Sun et al., 2008; Zwick, 2005). Together these MAbs provided highly specific controls for antibody accessibility to viral membrane-associated epitopes, as predicted for the KE sequences in the Dimmock topology model (see Figure 24, right). For intact virus immune precipitations, the level of antibody reactivity to its target antigen was measured by the relative level of viral p24 precipitated by the anti-Env MAbs (without lytic agents) compared to the amount of p24 (virus) input into the reaction.

To ensure that all the reference MAbs were able to bind their target protein, immunoprecipitations were initially performed with MAbs incubated with purified virions solubilized with Triton X-100. The product immunoprecipitates were then characterized by SDS-PAGE and western blot for the respective target protein (e.g. C11 immunoblotted for gp120, 7B2 immunoblotted for gp41, etc.). The open bars in Figure 28 summarize the average results obtained for each MAb from three independent reactions with detergent-solubilized viral preparations. These data demonstrate that all of the reference MAbs effectively pulled down ≥50% of the total amount of target protein from solubilized virions; in contrast, the nonspecific IgG control precipitated less than 3% of the virion p24 indicating minimal nonspecific precipitation under these experimental conditions.

Having established that the MAbs were able to bind their target antigen in this assay format, we next tested the ability of the reference MAbs to precipitate intact virions under conditions similar to those used above without detergent solubilization.

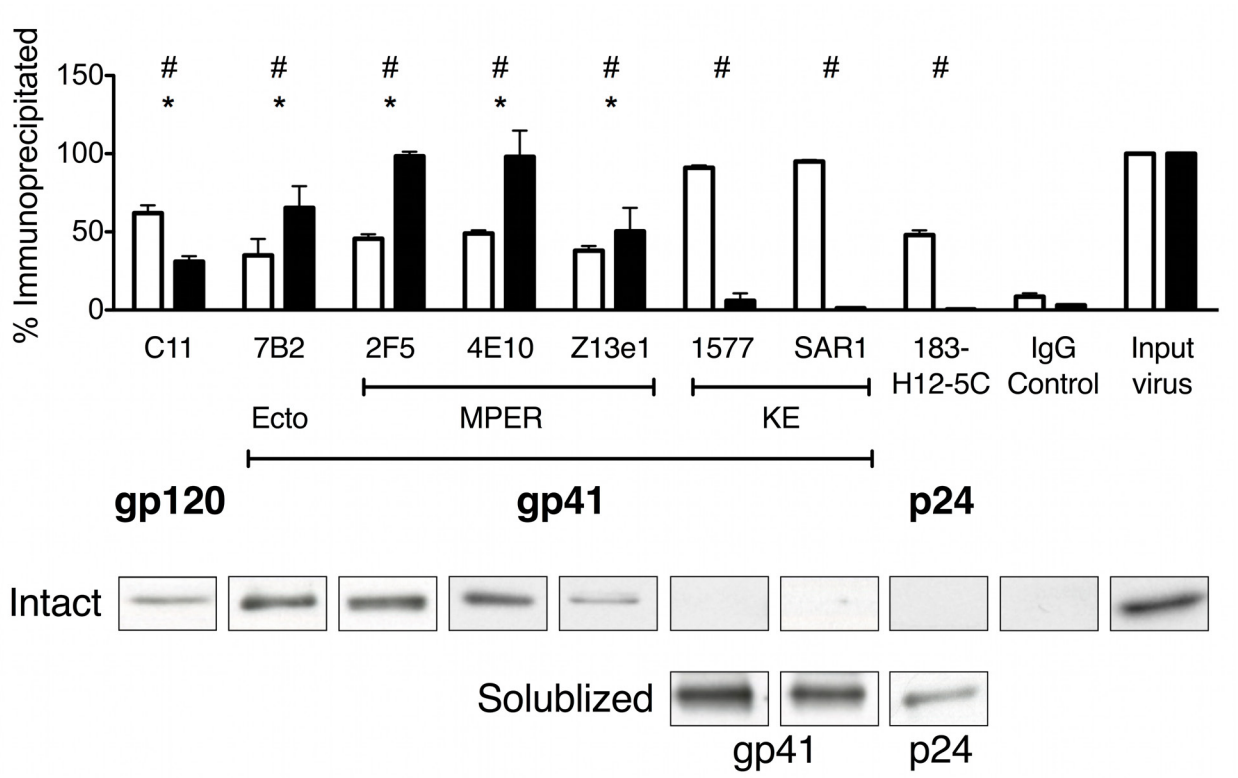

Figure 28. Anti-KE MAbs do not bind to intact virions

The indicated proteins and viral particles were immunoprecipitated using reference MAbs coupled to protein G-coated paramagnetic beads. (Top) Open bars represent % of target antigen precipitated when incubated with solubilized virus, while closed bars represent the % of input p24 precipitated by the corresponding MAb under native (intact virus) conditions. # = $p < 0.05$ for MAbs compared to IgG control with solubilized virus; * = $p < 0.05$ for MAbs with intact virus compared to IgG control. (Bottom) Representative p24 bands immunoprecipitated using the MAbs indicated in top panel with intact virus (Intact) or the bands of the target antigen from each MAb in detergent-disrupted virus (Solubilized).

Results presented in Figure 28 (solid bars) demonstrate that only MAbs specific for gp120 (C11) or the known gp41 ectodomain epitopes (7B2, 2F5, 4E10, and z13e1) effectively pulled down the intact virus particles, as measured by p24 levels in the immunoprecipitate. In marked contrast, both KE-specific MAbs (1577 and SAR1) failed to immunoprecipitate virus (p24) to significant levels above the IgG control, indicating a lack of reactivity of these reference MAbs with the KE in the context of intact virions. Additionally, the p24-specific MAb (183-H12-5C) also failed to pull down p24 in intact virions, confirming the integrity of the virion preparation. Taken together, the results from these immunoprecipitation experiments indicate that the KE is not accessible to antibody binding in the context of intact viral particles.

To complement these immunoprecipitation studies, binding analyses of gp41-specific MAbs to purified virions in solution were performed using SPR spectroscopy. This technique allowed the assessment of qualitative differences in relative antibody affinity in real time during the course of the interaction. Specifically, SPR assays were used to address the possibility that both KE and MPER-specific MAbs bound viral particles equally well, but that KE MAb binding to virions was unstable during the washing procedures used in the immunoprecipitation assay.

As shown in Figure 29, the SPR results demonstrate stable binding of MAb 2F5 to the MPER of intact HIV-1 virions, but no detectable binding of SAR1 to the KE of gp41 in intact virions, despite the highly sensitive nature of this assay. These results are consistent with our viral immunoprecipitation observations and confirm that the KE is apparently inaccessible to antibody binding on the surface of viral particles and that the

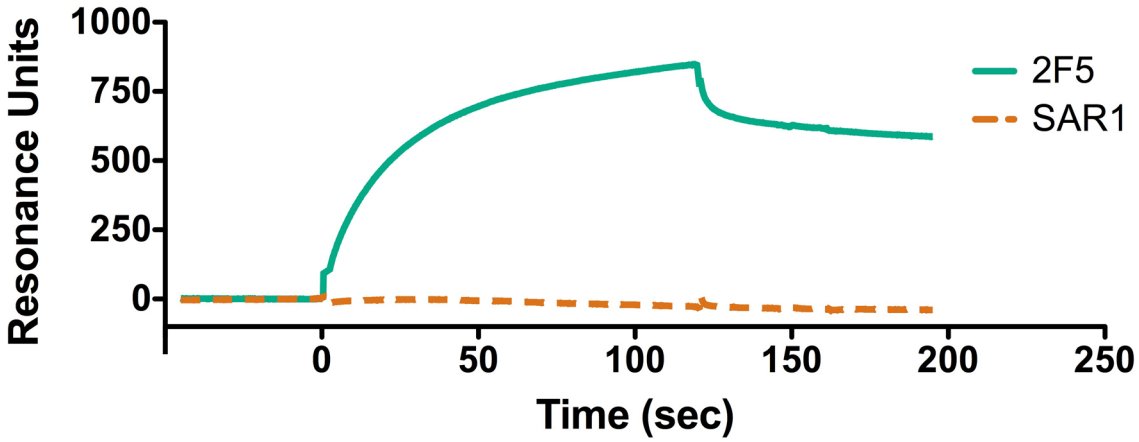

Figure 29. Comparison of MPER and KE MAb binding using SPR spectroscopy

MAbs 2F5 (MPER) and SAR1 (KE) were compared for relative binding rates and affinity using SPR spectroscopy to monitor antibody binding to purified intact HIV-1 virions. RU – resonance units.

observed difference in KE accessibility in cells and viruses cannot be attributed to differences in the stability of antibody binding.

4.4 DISCUSSION

The current studies were designed to distinguish between the two distinct topological models currently proposed for HIV-1 gp41 by determining the accessibility of key CTT sequences in the context of virions and Env-expressing cells to binding by reference MAbs. The results of these studies for the first time reveal in parallel assays a marked difference in the accessibility to antibody binding of the KE in a primary-isolate Env expressed on the surface of cells compared to intact virions. Thus, these observations indicate that the CTT may exist in different topologies depending on the membrane environment, suggesting that the CTT may actually be dynamic in its association with different lipid membranes or other cofactors.

While the goal of this study was to distinguish between the alternative models of CTT topology, the current data is only partially compatible with both models. In the case of virion-associated Env, the lack of KE exposure to antibody binding is consistent with the classical model of Env topology (Figure 24, left) in which the entire CTT sequence is sequestered within the lipid bilayer. In the case of Env expressed on the surface of cells, the accessibility of KE to antibody binding (and the inaccessibility of the LLP2) is consistent with an alternative Env topology (Figure 24, right) in which parts of the CTT are exposed to the exterior of the lipid bilayer. To reconcile these two different models, it seems necessary to discard the static model of Env topology and to assume a new

perspective in which CTT association with membranes is dynamic, perhaps controlled by interactions with other membrane binding proteins or lipid composition.

A dynamic model for CTT topology is consistent with the published studies by the Dimmock lab (Heap et al., 2005; Reading et al., 2003) indicating that the KE is inaccessible to inactivation by SAR1 in free virions, but that the KE becomes exposed to SAR1 inactivation only after virion binding to target cells. The current data are also consistent with initial studies by Kennedy et al (Chanh et al., 1986; Kennedy et al., 1986) demonstrating that rabbit antiserum prepared against a KE synthetic peptide could inactivate HIV-1, as these early studies used an assay format in which the serum antibody was incubated with the virus and the target cells. Further supporting the concept of a dynamic CTT topology is the recent report indicating that the LLP2 domain of the CTT is only accessible to antibody during cell-cell membrane fusion, suggesting major changes in virion Env topology post receptor binding (Lu et al., 2008). Interestingly, this would not be the first case of a viral envelope protein exhibiting a dynamic topology. The large envelope glycoprotein of hepatitis B virus is initially inserted into the endoplasmic reticulum membrane with its N-terminus located in the cytoplasm (von Heijne, 2006). After a post-translational maturation process, the N-terminus is translocated across the membrane by insertion of another transmembrane helix in about 50% of the molecules (Lambert and Prange, 2003). It will be interesting and important to determine the extent and precise localization of KE-exposed Env on the cell surface.

Recent structural studies of diverse membrane proteins have resulted in an increased appreciation of the complexity and dynamics of membrane protein topology

(von Heijne, 2006). Earlier topology models were based on the simple concept that membrane proteins of the helix-bundle type were assembled from long hydrophobic alpha helices that were oriented perpendicular to the membrane plane (von Heijne, 1999). However, based on high resolution structures of membrane proteins obtained during the past decade, it is now recognized that membrane helices can adopt a remarkable range of orientations relative to a membrane lipid bilayer and that membrane helices can dynamically reorient across the membrane in response to changes in the environment (von Heijne, 2006). Specifically, a recent study has demonstrated that the six helix N-terminal domain of LacY can undergo topological inversion due solely to the introduction of phosphatidylethanolamine into the lipid membrane, illustrating that environmental changes can cause specific alterations in membrane protein topology (Bogdanov et al., 2008). Additionally, recent advances in understanding the properties of the lipid membrane have led to new insights into membrane protein-lipid interactions (Engelman, 2005; Phillips et al., 2009). In particular, hydrophobic matching of protein transmembrane segment length to lipid domains of similar thickness has been shown to affect membrane protein segregation, structure, and function (Andersen and Koeppe, 2007; Epand, 2008; Lee, 2004; Phillips et al., 2009). It is intriguing to consider, in the context of data presented here, the possibility that the gp41 CTT may undergo a dynamic change in topology due to alterations in the lipid environment. Alternatively, in order to minimize hydrophobic mismatch, Env with the CTT in the traditional topology could be selectively incorporated into the lipid rafts that serve as the sites of virion assembly.

A second aspect of current membrane protein structure studies that appears to be relevant to the case for a reconsideration of HIV-1 CTT membrane topology is the unique role of arginine (Arg) in membrane-associated proteins. While the early paradigm predicted Arg residues to be excluded from hydrophobic environments like a membrane lipid bilayer, more recent structural data indicate unique biochemical properties of Arg that produce unexpected functional properties for membrane proteins. For example, Arg residues can be incorporated as critical components of the membrane spanning domains of proteins, as observed with highly conserved Arg residue in the MSD of HIV-1 (c.f. Figure 24). The ability of the positively charged Arg to be located in a lipid environment is evidently due to the capacity of the Arg guanidinium group to form an energetically favorable cation-π bond with tryptophan, tyrosine, or phenylalanine residues in a stacked helix (Dougherty, 1996; Ma and Dougherty, 1997). Interestingly, a highly conserved phenylalanine is located three residues C-terminal to the Arg in the HIV-1 MSD (Figure 25).

A second unique property of Arg is its ability in a repeat sequence (or homopeptide) to mediate transport of proteins and DNA or RNA across intact biological membranes, in contrast to homopeptides of other charged amino acids such as lysine (Lys) and histidine (His) that fail to efficiently cross lipid bilayers (Mitchell et al., 2000). Thus membrane protein segments rich in Arg residues have a propensity to cross biological membranes under certain environmental conditions, offering the potential for a dynamic membrane topology. For example, the Arg-rich helical S4 segment of the voltage-gated potassium channel KvAP is proposed to move across the membrane as the channel opens and closes in response to changes in the membrane electric field

(Jiang et al., 2003). While this was initially controversial due to the number of Arg residues (four) in the S4 segment, more recent studies have demonstrated that the S4 segment can insert into the membrane of eukaryotic cells as a transmembrane helix in spite of the four Arg residues (Hessa et al., 2005b). These findings, when viewed in light of observations by Lu et al (Lu et al., 2008) demonstrating transient LLP2 exposure on Env-expressing cells undergoing fusion, may provide an explanation for the predominance of Arg residues in the HIV gp41 CTT relative to other cationic residues.

A unique characteristic of the gp41 CTT of HIV-1 is that it contains two cationic amphipathic helical segments (LLP1 and LLP2) that are rich in Arg. In addition, the CTT contains a number of other highly conserved Arg residues. We and others have previously shown that synthetic LLP1 and LLP2 peptides assume a random coil conformation in aqueous environments, but rapidly convert to a helical conformation in a hydrophobic or membrane mimetic environment (Fujii et al., 1992; Srinivas et al., 1992; Yuan et al., 1995). In addition, we have shown that synthetic LLP1 and LLP2 peptides are able to readily traverse biological membranes of intact mammalian cells. These observations are consistent with the concept of the Arg-rich CTT being intimately associated with the membrane lipids and being able to cross the lipid bilayer of an intact viral or cell membrane.

Finally, we and others have previously reported that modifications in CTT length or amino acid composition can have marked affects on global Env structure and function, including virion incorporation, fusogenicity, and antibody reactivity. For example, we found that the mutation of two Arg residues to glutamate (Glu) residues in LLP2 of an HIV-1 provirus substantially reduced Env fusogenicity and altered the

antibody reactivity and neutralization sensitivity of epitopes located in both the gp120 protein and the ectodomain of gp41 (Kalia et al., 2005). These observations have been interpreted to indicate that changes in CTT result in allosteric alterations in global Env conformation that are reflected in altered functional properties, including reduced fusogenicity. It is interesting to speculate that altering the Arg content may affect the association of the LLP segments with the lipid bilayer and perhaps the ability of the CTT to traverse the membrane during the fusion process.

As often happens in HIV-1 research, our efforts to test two relatively simple models for CTT topology have indicated the need to consider more complicated and dynamic models of CTT membrane topology. This is perhaps a timely pursuit in light of the new paradigms of membrane structure and the importance of understanding the mechanisms by which the CTT serves as a determinant of HIV-1 Env structure and function on the surface of virions and cells.

5.0 EXISTENCE OF A CELL SURFACE TOPOLOGY FOR THE HIV GP41 C-TERMINAL TAIL THAT IS DISTINCT FROM THE VIRAL TOPOLOGY

5.1 INTRODUCTION

The envelope (Env) protein of HIV, as the only virally-encoded protein present on the surface of the virion, is the primary target of the humoral immune response (McElrath and Haynes, 2010). Env is composed of two subunits that are translated as a 160 kDa polyprotein that is post-translationally cleaved to yield the highly glycosylated gp120 (or surface unit - SU) and the transmembrane (TM) protein gp41 (Luciw et al., 2002). In addition, gp41 is composed of three distinct domains: the ectodomain; the (putative) membrane spanning domain (MSD) that is thought to anchor Env in the membrane; and the C-terminal tail (CTT). gp120 functions to mediate binding to the primary receptor, CD4, and the coreceptor, primarily CXCR4 or CCR5, while gp41 mediates fusion of the viral and cellular membranes, resulting in infection (Luciw et al., 2002).

The gp120 protein and the gp41 ectodomain have both been extensively studied, both structurally and functionally, as they appear to be the important targets of the antibody response in infected individuals (McElrath and Haynes, 2010). Likewise, the gp41 MSD is the focus of intensive study to determine the exact sequences involved in spanning the cellular and viral lipid bilayers (Haffar et al., 1988; Miyauchi et al., 2005;

Shang and Hunter, 2010; Shang et al., 2008; West et al., 2001; Yue et al., 2009). The CTT, on the other hand, has largely been studied at a functional level, and has been demonstrated to play a role in viral Env incorporation (Freed and Martin, 1995, 1996; Murakami and Freed, 2000a, b), virion maturation (Jiang and Aiken, 2007; Kol et al., 2007; Wyma et al., 2004; Wyma et al., 2000), cellular Env trafficking (Byland et al., 2007; Ohno et al., 1997), and more recently, as a modulator of Env gp120 conformation on both the cell and virion surfaces (Joyner et al., 2011; Kalia et al., 2005). However, relatively little is known about the structure of the CTT aside from characterizations of peptide analogs of CTT subdomains, known as the lentivirus lytic peptides (LLPs), that have been demonstrated to be predominantly helical in membrane and membrane-like environments (Fujii et al., 1992; Kliger and Shai, 1997; Srinivas et al., 1992; Steckbeck et al., 2011).

The topology of the CTT has been largely ignored as a topic of study in the otherwise vastly explored field of Env structure. The prevailing model, which has never been experimentally demonstrated, is that Env gp41 exists as a type 1 membrane protein, with an extracellular (or extravirion) N-terminal domain (the ectodomain), a single (helical) transmembrane domain, and a cytoplasmically-localized, approximately 150 amino acid long C-terminal domain. Originally, this model was described based on sequence and structural comparisons with other retroviral Env proteins, particularly the oncogenic retroviruses that have a single transmembrane domain followed by a short cytoplasmic tail (Miller et al., 1991). More recently, an alternative topology for the CTT has been proposed based on reactivity of viral particles (Cheung et al., 2005; Heap et al., 2005; Hollier and Dimmock, 2005) and Env-expressing cells with a monoclonal

antibody (MAb) directed at the CTT (Cheung et al., 2005; Heap et al., 2005; Hollier and Dimmock, 2005; Steckbeck et al., 2010). Interestingly, this alternative topology was suggested in the 1980s when Kennedy and colleagues discovered that serum from rabbits immunized with a peptide from the gp41 CTT could neutralize virus (Kennedy et al., 1986), and that antibodies reactive to that particular peptide were found in infected humans (Chanh et al., 1986; Ho et al., 1987). These results suggested exposure of the peptide epitope, known as the Kennedy epitope (KE), on the virion surface, as antibody cannot cross intact lipid membranes.

More recent studies have yielded interesting, if sometimes conflicting, results regarding CTT topology. Initial studies of CTT topology by Dimmock and colleagues suggested that the KE was exposed on both the cellular and viral surface (Cleveland et al., 2003; Reading et al., 2003). Later experiments, however, demonstrated that the KE appeared to be exposed on the cell, but not the virion (Cheung et al., 2005; Heap et al., 2005). Our lab has recently published similar results demonstrating the exposure of the Kennedy epitope on the surface of intact Env-expressing cells, but no apparent exposure on the virion (Steckbeck et al., 2010). These cellular results are in contrast to another recent study using GFP-fused gp41 truncation mutants (Liu et al., 2010). In that study, the gp41 CTT was only observed in a cytoplasmic orientation in Env-expressing cells (Liu et al., 2010).

Complicating CTT topology studies even further are results suggesting that the CTT undergoes dynamic rearrangements during the fusion process. This was first observed using the post-attachment neutralization (PAN) assay and finding that an anti-KE MAb, SAR1, could neutralize HIV when preincubated with virus and cells at 25 °C,

but did not neutralize either at 37 °C (in the PAN format) or in a traditional neutralization assay (Cheung et al., 2005; Heap et al., 2005). A more recent study has also indicated that topological rearrangements of the CTT take place during the membrane fusion process. Using an antibody directed at LLP2, Chen and colleagues found that they could detect LLP2 if cell-cell fusion was carried out at 31.5 °C, but not at 37 °C (Lu et al., 2008). These results are similar to those obtained with the PAN assay in that CTT sequence exposure is only observed when the fusion process is delayed using a lower temperature (Lu et al., 2008) suggesting that the exposure during fusion is likely to be highly transient. This model is consistent with data demonstrating that the LLP2 sequence is membrane associated both during and after membrane fusion (Viard et al., 2008). Together, the current published results present an uncertain picture of CTT membrane topology.

In an effort to provide a comprehensive experimental CTT topological map, we have extended previously published experiments (Steckbeck et al., 2010) using the VSV-G epitope tag insertion technique to determine the reactivity of the entirety of the CTT sequence, both on the surface of Env-expressing cells and on viral particles. Results indicate that a majority of the CTT sequence is accessible to antibody binding on the surface of intact, Env-expressing cells. On the contrary, no apparent exposure of CTT sequences was observed on pseudoviral particles containing VSV-G-tagged Env. These results provide further evidence for a complex CTT topology beyond the widely accepted intracytoplasmic model.

5.2 MATERIALS AND METHODS

5.2.1 Cell and virus construct

HEK293T cells (ATCC) and TZM-bl cells (from the AIDS Reference and Reagent Program) were maintained in DMEM (Invitrogen) supplemented with 10% (v/v) FBS. Cells were used at passage numbers less than 30. The pSG3ΔEnv used for pseudovirus production was provided by the NIH Aids Reference and Reagent Program.

5.2.2 VSV-G epitope tag substitutions

The VSV-G epitope tag was cloned serially into the CTT of HIV-1 89.6 Env as previously described (Steckbeck et al., 2010), and shown schematically in Figure 30. Briefly, the codon-optimized Env gene was cloned into a p2CI vector derived from PCR2.1 by insertion of the CMV promoter and polyA signal sequence from pcDNA3.1(hygro) (Invitrogen) and PCR-amplified IRES-Neomycin resistant sequence from pFB-Neo-LacZ vector (Stratagene). Overlapping PCR was used for the construction of the VSV-G substitution mutants of gp160. The two hybrid primers were

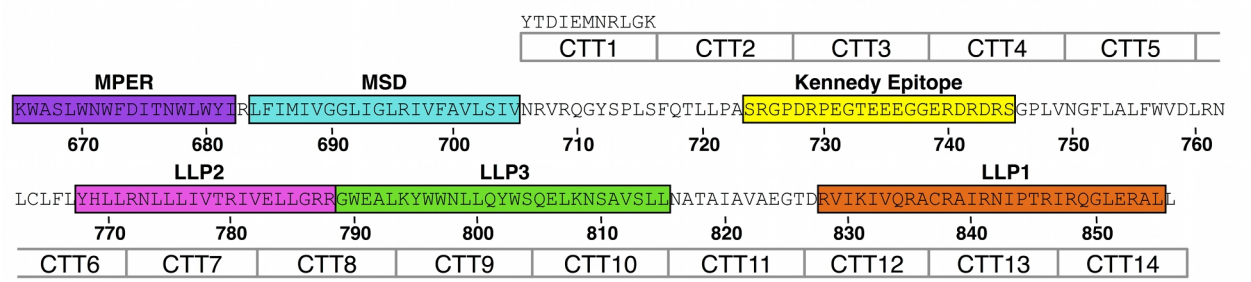

Figure 30. Replacement of HIV gp41 sequences with VSV-G sequences

The VSV-G epitope tag was used to replace gp41 CTT sequences serially across the length of the CTT as indicated, from CTT1 to CTT14. The VSV-G epitope tag sequence is listed above CTT1.

constructed containing VSV-G tag sequences at the 5'-ends and HIV gp160 specific sequences at the 3'-ends of both primers used for the substitutions. Final overlapping PCR products were then subcloned into the HIV-1 gp160 expression vector using the appropriate restriction enzymes and the Rapid DNA Ligation Kit (Roche Applied Science). All VSV-G substitutions were verified by DNA sequencing.

5.2.3 Cell culture and transfections

HEK293T cells were plated into six-well plates 24 hours prior to transfection. For evaluation of VSV-G epitope exposure on Env-expressing cells, 80% confluent cells were transfected with 2.5 µg of the selected HIV-1 Env mutant DNA using Lipofectamine™ LTX reagent and PLUS™ reagent, as recommended by the manufacturer (Invitrogen).

5.2.4 Cellular FACS analysis

Cells were harvested 24 hours post-transfection by treatment with 2 mM EDTA. Cells were resuspended and washed twice with FACS wash buffer (1X PBS with 5% FBS) at 4 °C prior to antibody staining. Reference MAbs specific for VSV-G (Roche Diagnostics), gp41 (SAR1), and actin (Sigma-Aldrich) were labeled immediately prior to staining using Zenon labeling kits (Invitrogen) following manufacturer instructions. A sample of 10^6 cells was dual-stained by incubating with 5 µg fluorophore-labeled antibody (VSV-G- or gp41-specific) and 7.5 µg fluorophore-labeled actin antibody for 30 minutes on ice. Following staining, cells were washed thrice with FACS wash buffer at 4

°C. Washed cells were resuspended and stained with 7-amino-actinomycin D (7-AAD). Cells were not fixed prior to analysis. Fluorescently-labeled cells were analyzed on a FACSAria (BD Biosciences). Live intact cells (7-AAD negative) were selected for scatter characteristics, including selection of single-cell populations by doublet-discrimination analysis. PMT settings were adjusted on identically-stained mock transfected cells prior to analysis of Env-transfected cells. Data was collected for $5x10^4$ 7-AAD negative cells and analyzed for reactivity with fluorescently-labeled VSV-G or gp41 and actin antibodies.

5.2.5 Production of pseudoviral particles with VSV-G tagged Env

To produce VSV-G tagged pseudovirions, HEK293T cells were plated and transfected as above, with 1.25 µg VSV-G tagged Env DNA and 1.25 µg pSG3ΔEnv. Cellular supernatants were harvested 48 hours post-transfection and clarified at 1,500 x g for 10 minutes. The clarified supernatant was pelleted for 1.5 hours at 21,000 x g over a 20% glycerol cushion. The pelleted virus was resuspended in 1X PBS and further purified on a 30%/45% sucrose step gradient. The viral band was collected, pelleted through a 20% glycerol cushion, and resuspended at a 1000X concentration in 1X PBS.

5.2.6 Pseudoviral immunoprecipitation and Western blotting

Protein G Dynabeads (Invitrogen) were prepared according to the manufacturer's directions. Briefly, anti-Env or anti-Gag antibodies (4 µg) were incubated with 20 µl protein G Dynabeads in 35 µl citrate-phosphate buffer, pH 5.0, with gentle shaking for

45 minutes at room temperature. Isotype-matched IgG controls were used for each species (murine, human, etc.) from which a MAb was derived. Beads were washed thrice with 0.5 ml citrate-phosphate buffer followed by resuspension in either 26 µl PBS (for intact virus) or 26 µl PBS with 1% Triton X-100 (for lysed virus) and 4 µl NL4-3 virus (equivalent to 1.1 µg p24). Virus-bead suspensions were incubated at 4 °C for one hour with gentle shaking and subsequently washed thrice with 1X PBS. Following the final wash, beads were resuspended in NuPAGE SDS-PAGE buffer, heated at 70 °C for 10 minutes, and the supernatant loaded onto 4-12% Bis-Tris NuPAGE gels. Gels were electrophoresed followed by transfer to polyvinylidene fluoride (PVDF) membranes using the Invitrogen iBlot system. Blots were blocked for one hour in 5% blotto (1X PBS with 5% dry milk). After blocking, blots were cut to allow separate staining of gp120 (>60 kDa), gp41 (30 - 60 kDa), and p24 (<30 kDa). gp120 was stained with rabbit anti-gp120 (Advanced Biotechnologies, Inc.), gp41 stained with Chessie 8, and p24 stained with Ag3.0 for 1.5 hours at room temperature. Blots were washed thrice with 1X PBS and 0.025% Tween 20 (PBS-T), followed by incubation with appropriate secondary antibody (anti-rabbit IgG or anti-mouse IgG conjugated to horseradish peroxidase) for one hour at room temperature. Blots were washed thrice in PBS-T with the gp120 blot receiving an additional wash in 1X PBS with 0.1% Triton X-100. Finally, blots were incubated with PicoWest substrate (Pierce) for one minute and reassembled for visualization on X-ray film.

5.2.7 Western blot quantitation

Antibody IPs were quantified by densitometry analysis. For each IP, three independent X-ray exposures were scanned and analyzed using ImageJ (NIH). When applicable, p24 bands were selected for each protein, and the integrated area under the densitometry curve was compared to that of the viral input band to yield percent of the input protein in the immunoprecipitate. Percent inputs for each protein for each of the three exposures were averaged to yield the overall percent input immunoprecipitated per experiment. This was repeated for three independent experiments, and the results were averaged to yield the final percent input immunoprecipitated per antibody.

5.3 RESULTS

5.3.1 The majority of the CTT sequence is exposed on the cell surface

To map the topology of the CTT, a VSV-G epitope tag was used to serially replace CTT sequences over the entirety of the CTT (Anand, 2000; Das et al., 2008; Obermeyer et al., 2007; Steckbeck et al., 2010) (Figure 30). Additionally, the HA epitope tag was inserted into the V3 region of gp120 to allow for detection of the relative levels of Env in cells transfected with the various VSV-G substituted Envs. The resulting constructs were then used in assays to determine reactivity with Env on the surface of intact live cells, or on the surface of pseudovirus particles (see below).

VSV-G-labeled Env constructs were transfected into 293T cells and analyzed for epitope tag reactivity to either anti-HA or anti-VSV-G MAbs as described previously (Steckbeck et al., 2010). Initially, cells were dual stained with anti-HA and anti-VSV-G, however, no HA (gp120) reactivity was observed in spite of positive VSV-G staining. It was determined by separate staining for HA and VSV-G reactivity in aliquots from the same transfection that the anti-VSV-G (CTT) MAb was apparently inhibiting the binding of the anti-HA MAb (data not shown). Therefore, to determine overall levels of Env expression on the cell surface, transfected cells were first stained for reactivity with anti-HA MAb (Figure 31A). In general, staining of all CTT constructs resulted in between 20% to 40% of cells labeling positively for gp120 on the surface of intact, non-permeabilized cells, with the exceptions of CTT11 and CTT13. To determine VSV-G, and thus CTT, exposure on the cell surface, aliquots of the respective transfected cells stained with anti-HA (gp120) were stained in parallel with anti-VSV-G (Figure 31B). In general, a similar overall level of staining was observed with the anti-VSV-G MAb, with the majority of CTT constructs showing approximately 20% positive labeling, with the exception of CTT 6, 7, 8, 12, and 13. Results with CTT3 and CTT7 are consistent with results obtained previously with these constructs (Steckbeck et al., 2010). These results suggest that the majority of the CTT sequence of gp41 is exposed and reactive with antibody on the surface of intact, non-permeabilized Env-expressing cells.

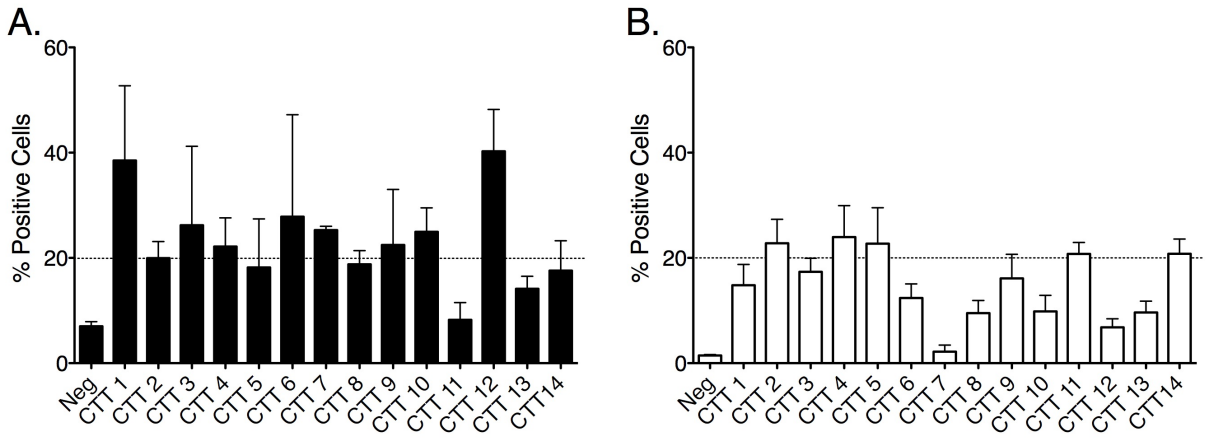

Figure 31. HIV Env staining of live intact cells

Live intact 293T cells transfected with the indicated VSV-G containing Env were stained with (A) anti-HA, or (B) anti-VSV-G monoclonal antibodies. (A) Total cell surface Env expression was determined by staining cells with anti-HA, targeting HA epitope tag inserted into gp120. (B) CTT exposure was determined by staining cells with anti-VSV-G, targeting the VSV-G epitope tag in the CTT sequence as demonstrated in Figure 30. The line at 20% staining is for visual reference only.

5.3.2 CTT-exposed Env constitutes a large proportion of cell-surface Env

While the above experiments demonstrated that a majority of the CTT sequence appears to be exposed on the surface of intact Env-expressing cells, they provide no information on the proportion of CTT-exposed Env relative to the total Env on the cell surface. In order to determine the relative extent of CTT-exposed Env, we took advantage of our previous observation that anti-VSV-G (anti-CTT) antibody prevented anti-HA (anti-gp120) binding during attempts at dual staining for FACS analysis. The rationale for this experiment is that blocking of anti-gp120 antibody binding by anti-CTT antibody allowed a relative quantitation of CTT-exposed Env as a function of the total surface Env. As total Env cell surface expression levels differed between each construct (Figure 31A), results were normalized to total HA binding in the absence of VSV-G antibody to allow a comparison between constructs. Results are presented as the percent HA binding inhibited by VSV-G antibody (Figure 32). The majority of constructs resulted in >20% inhibition of gp120 binding, with CTT3, CTT12, and CTT13 approaching 50% inhibition. CTT2, CTT5, CTT7, and CTT11 all demonstrated <15% average inhibition. The apparent lack of antibody competition with the CTT7 construct is consistent with our previous data indicating that this segment is not exposed to antibody binding on the surface of transfected cells (this work and (Steckbeck et al., 2010)). CTT1 costaining with anti-120, on the other hand, apparently led to an increase in

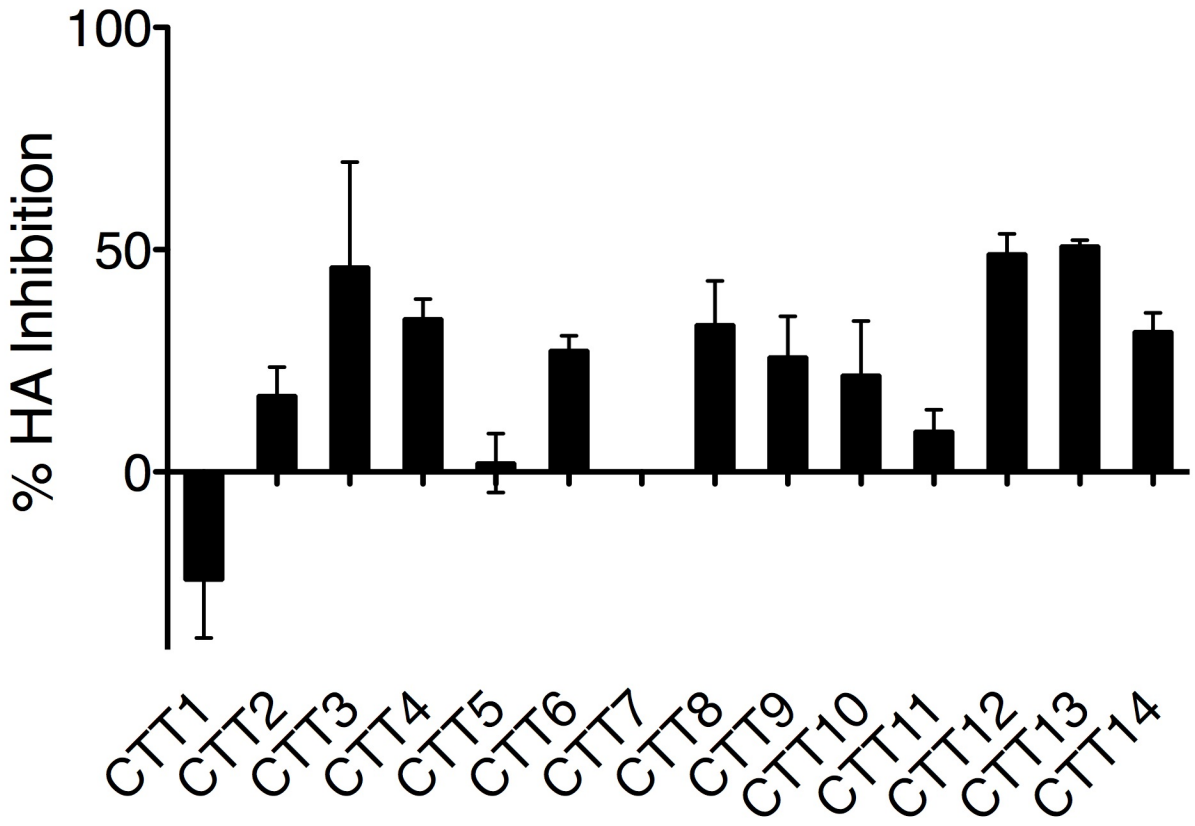

Figure 32. Inhibition of HA-specific antibody binding by VSV-G-specific antibody

Env-expressing cells were surface stained first with unlabeled anti-VSV-G (CTT) antibody followed by staining with fluorescently-labeled anti-HA (gp120). The percent reduction in HA-labeled cells compared to HA-stained cells without VSV-G staining is presented as percent HA-inhibition.

antibody reactivity, as evidenced by negative inhibition. These results indicate that CTT-exposed Env constitutes at least 20-50% of Env on the cell surface.

5.3.3 CTT exposure is not apparent on the surface of VSV-G-tagged pseudovirus particles

To provide a direct comparison of the VSV-G epitope tag based topology mapping studies on the cell surface, pseudoviruses were made by cotransfecting the VSV-G-tagged Env with pSG3ΔEnv. This approach has been widely used to generate infectious pseudoviruses for use in neutralization assays (Platt et al., 2009; Platt et al., 1998). Infectivity of the resulting transfection supernatants was determined using the TZM-bl assay and results are presented in Figure 33. Importantly, all VSV-G-tagged Env clones yielded infectious pseudovirus when transfected with pSG3ΔEnv. Luciferase activity from all the clones was >200-fold higher than that observed for the mock transfection and of the same order of magnitude as the wild-type Env. These results suggest that the VSV-G tag substitutions did not evidently interfere with the functional activity of Env as the resulting pseudovirions were infectious at a similar level to the wild-type Env from which they were derived. As such, the resulting pseudovirus particles were used to map the topology of the CTT by immunoprecipitation using the VSV-G epitope tag as the potentially exposed target.

Immunoprecipitations were performed as described previously (Steckbeck et al., 2010). As a control for antibody accessibility to an exposed membrane-localized epitope, immunoprecipitations with the MPER-directed antibody 2F5 were performed in

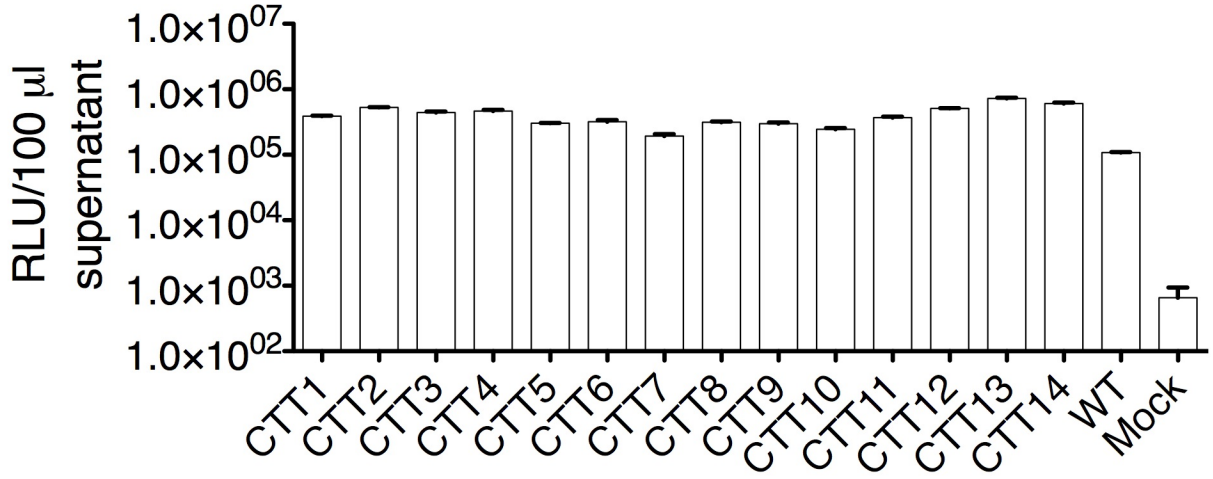

Figure 33. Relative infectivity of pseudoviruses containing VSV-G-tagged Env

Supernatants from 293T cells cotransfected with pSG3ΔEnv and the indicated Env construct were used to infect TZM-bl cells to determine infectivity of the resulting pseudovirus particles. Results are presented as relative light units (RLU) per 100 µl of transfection supernatant.

parallel with the anti-VSV-G pull-downs. As these MAbs are Env-directed, precipitation of intact virions was determined by measurement of the amount of p24 bound to MAb-coated beads relative to the input p24. The rationale is that p24 could only be bound to the beads if the particular MAb of interest bound to an intact virion. Results from the immunoprecipitation assays are presented in Figure 34. The MPER-directed MAb 2F5 was observed to immunoprecipitate intact particles from all VSV-G-tagged pseudoviral clones to varying degrees (open bars), ranging from 30-80% of the input p24 bound to the beads. In contrast, however, the anti-VSV-G MAb did not pull down p24 from any of the pseudoviral clones (closed bars). These results suggest that the CTT sequences are not apparently exposed on the surface of intact virions, in distinct contrast to the observed exposure on the cell surface.

5.3.4 Dynamic exposure of CTT sequences during viral infection

The current results demonstrate that a majority of the CTT is exposed to antibody binding on the surface of Env-expressing cells, but is apparently not natively exposed on the virion. To address the potential for CTT sequences to demonstrate transient exposure, post-attachment neutralization (PAN) assays were performed using a monoclonal antibody (SAR1) directed at the Kennedy epitope in the CTT (Figure 35). The assay was performed at both 37 °C and 31 °C, as prior studies demonstrated PAN activity at temperatures that do not support membrane fusion (Heap et al., 2005). Results from this assay demonstrate that SAR1 can mediate PAN when antibody and virus are incubated with cells at 31 °C, but not 37 °C. 50% neutralization was reached at

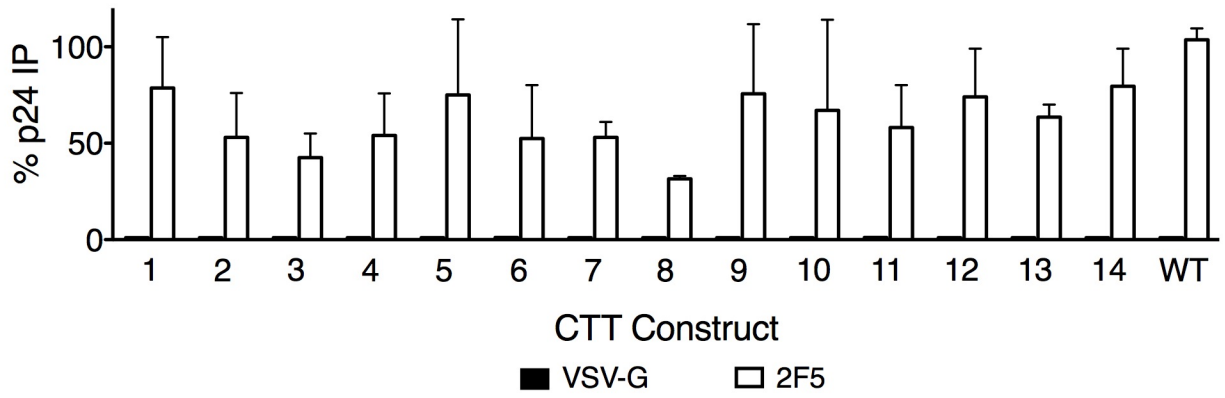

Figure 34. Immunoprecipitation of VSV-G-tagged pseudoviral particles

Pseudovirus particles were immunoprecipitated using monoclonal antibodies specific for the VSV-G epitope tag (closed bars) or the gp41 MPER (2F5, open bars). Results are presented as the percent of input p24 that was immunoprecipitated (IP).

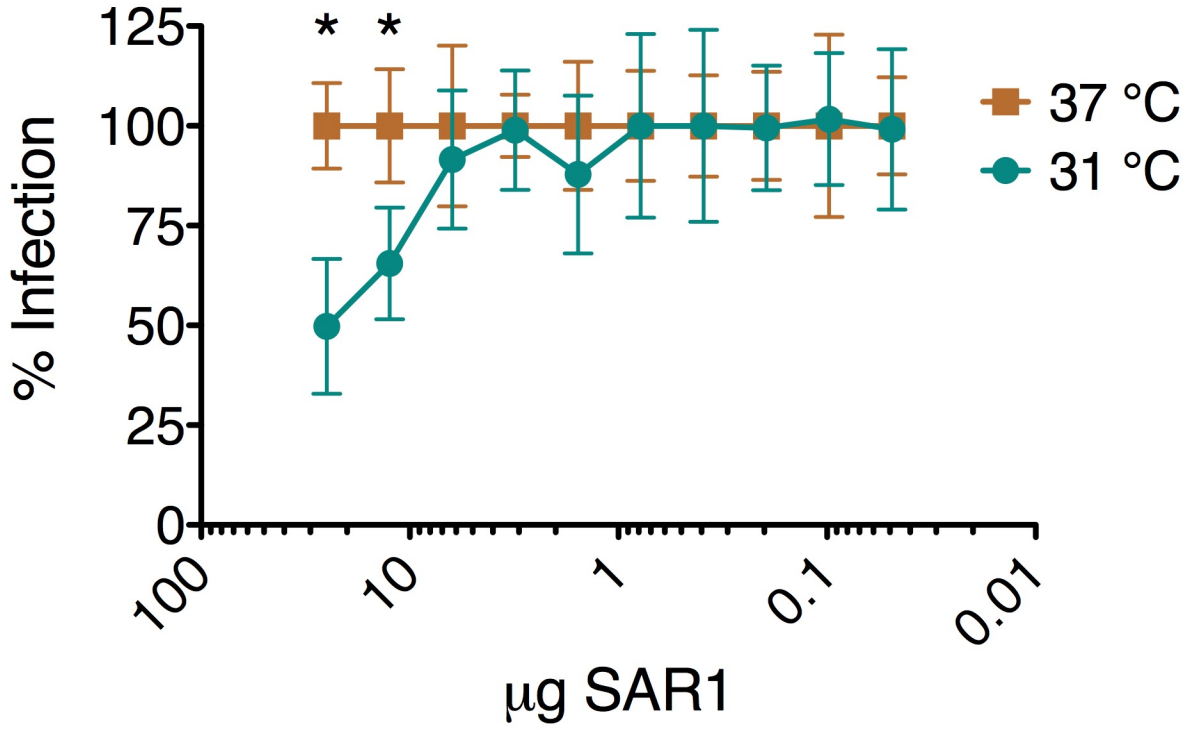

Figure 35. Post-attachment neutralization of HIV-1 89.6 by anti-CTT antibody

Anti-CTT (Kennedy epitope-specific) monoclonal antibody SAR1 was used to determine post-attachment neutralization (PAN) activity at 37 °C and 31 °C. SAR1 did not exhibit PAN at 37 °C, but there was a statistically significant reduction in viral infection when SAR1 was tested for PAN at 31 °C. * indicates statistical significance at $p < 0.05$.

a SAR1 concentration of 25 µg/ml. These results indicate that CTT sequences that are not natively exposed on the virus become exposed when the virus-cell fusion process is delayed by incubation at lower temperatures.

5.3.5 Prediction of HIV gp41 sequences to act as an in-plane membrane anchor

Based on previous results from our lab (Steckbeck et al., 2010), it was expected that the VSV-G tag in CTT3 would be accessible to antibody staining. The reactivity of a majority of the CTT sequence was, however, unexpected. In essence, the reactivity in Figure 31B demonstrates that the CTT is nearly entirely exposed on the cell surface, making Env in that case a monotopic protein. For that to occur, the region traditionally thought of as the MSD must not span the membrane, instead functioning to anchor Env to the cell surface by acting as either an in-plane anchoring sequence, or as a reentrant loop. This topological arrangement is not without precedent in viral proteins, as the Erns proteins from the Pestivirus genus of Flaviviridae are membrane anchored through in-plane helices (Fetzer et al., 2005; Sapay et al., 2006b; Tews and Meyers, 2007). Predictive programs addressing the ability of a protein sequence to act as either an in-plane anchor or reentrant loop were used to examine this idea further. TOPCONS analysis of the 89.6 gp41 sequence revealed no evidence for a propensity to form reentrant loops (data not shown). On the other hand, analysis using AmphipaSeek (Combet et al., 2000; Sapay et al., 2006a) indicated sequence with strong potential to exist as an in-plane membrane anchor. The sequence ^{672}FDITNWLWYIRLFIMIVG^{689} scored as a strong positive as an in-plane membrane anchoring sequence when tested

under conditions of high specificity and low sensitivity (Figure 36). This sequence contains the C-terminal end of the MPER and the N-terminal seven residues of the traditional MSD. These results demonstrate that Env has sequence characteristics similar to those found in in-plane membrane anchors and may demonstrate a propensity towards alternative topologies.

5.4 DISCUSSION

The current data extend previously published work from our lab on the cell surface exposure of the Kennedy epitope to a study of the topology of the entirety of the CTT sequence. In the current study, we have demonstrated a differing membrane topology for the HIV gp41 CTT on the cell surface compared to that which is present on the surface of infectious virions. On the cell surface, much of the CTT appears to be exposed to antibody binding on the extracellular side of the membrane, while on the virion, no exposure of CTT sequences is detected. These results reiterate the idea that the topology of the CTT of gp41 is more complex than the prevailing model of a completely cytoplasmic tail. The data also appear to conflict with published studies that demonstrate functional interactions between the Env CTT sequences and intracellular proteins (Byland et al., 2007; Freed and Martin, 1995, 1996; Lopez-Vergès et al., 2006; Murakami and Freed, 2000a, b; Ohno et al., 1997). The question then becomes how does the current data fit into the universe of CTT trafficking and functional studies that reinforce the traditional cytoplasmic CTT?

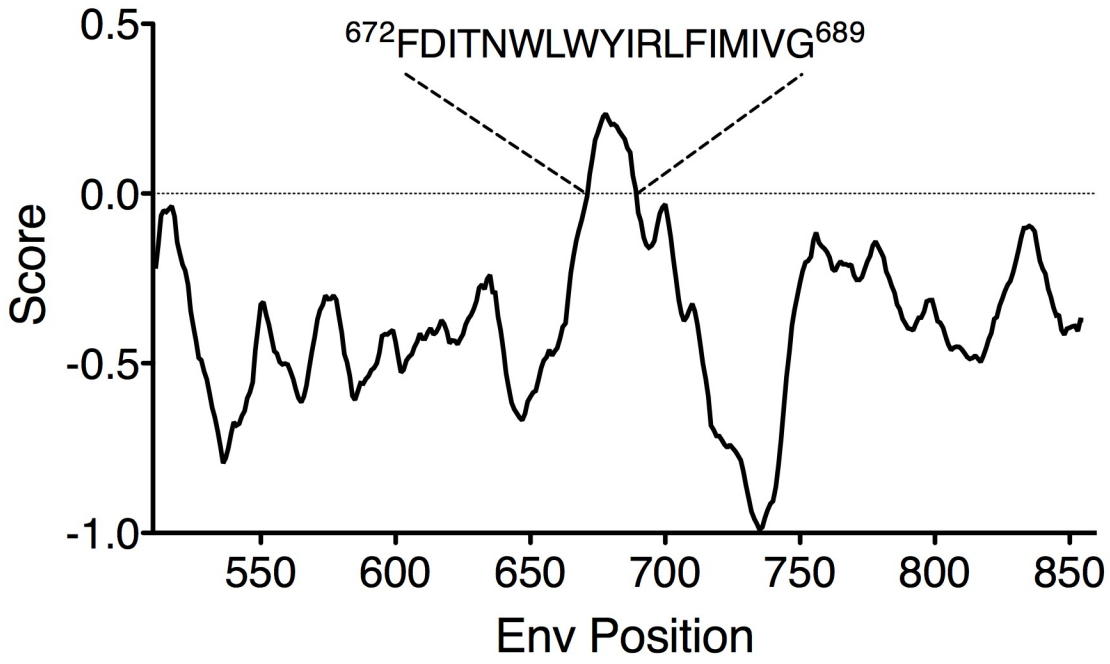

Figure 36. In-plane anchor prediction for HIV gp41

The amino acid sequence of HIV-1 89.6 gp41 was subjected to in-plane anchor prediction using the AmphipaSeek methodology. The score is presented as a moving average over seven residues. A score above zero is considered positive for the propensity of a sequence to serve as an in-plane membrane anchor. The results were not dependent on the window size (7 - 15 residues).

The simplest explanation for the discrepancy between the current data and the traditional model is that the CTT-exposed Env is either misfolded or an alternative fold. Misfolding of the Env could be due to protein overexpression in transfected cells. However, in general, overexpression of membrane proteins that overwhelm the synthetic machinery tend to lead to the cessation of new protein synthesis concomitant with the induction of the unfolded protein response (UPR) and ER-associated degradation (ERAD) (Brodsky, 2007; Griffith et al., 2003; Kaufman et al., 2002; Meusser et al., 2005; Nakatsukasa et al., 2008; Wagner et al., 2006). If the UPR and ERAD cannot successfully reduce the accumulation of misfolded protein, the cell ultimately undergoes apoptosis (Kaufman et al., 2002). As the current results were determined on intact live cells, the observed CTT exposure is not thought to be an artifact of misfolded protein.

A second explanation is that the observed CTT-exposed Env is the result of an alternative folding of the Env with respect to its membrane topology. The idea of multiple folds for membrane proteins is not without precedent. For example, the large envelope glycoprotein of hepatitis B virus is initially inserted in the membrane during biosynthesis with a cytoplasmic N-terminus (Lambert and Prange, 2001; Lambert and Prange, 2003). Following post-translational maturation, the N-terminus is translocated across the membrane in approximately 50% of the molecules by the insertion of another membrane helix (Lambert and Prange, 2001; Lambert and Prange, 2003). It is possible that during HIV Env biosynthesis a similar process occurs, resulting in a population of Env with both cytoplasmic and extracellular CTT sequences. Under this assumption, the CTT-exposed Env may be similar to the flavivirus Erns proteins that are attached to the

membrane by an in-plane membrane anchor (Brass et al., 2007; Sapay et al., 2006b). It is important to note that our data do not exclude the presence of cell-surface Env with cytoplasmically-localized CTT sequences; the data only demonstrate that CTT-exposed Env exists on the surface of Env-expressing cells. It is possible that both cytoplasmic and extracellularly-localized CTT exists at the same time in the same cell, and that CTT-exposed Env accumulates on the surface of the cell because it cannot be endocytosed due to extracellular localization of the functional endocytic signals (Byland et al., 2007), leading to enhanced detection.

In addressing the extent of surface-exposed CTT, we found that a substantial portion of the Env on the cell surface was CTT-exposed. Under our experimental conditions, prior binding by anti-VSV-G (CTT) antibody reduced anti-HA (gp120) staining by up to 50%. It will be critical to recapitulate these cell surface CTT topology results in the context of a natural viral infection. If CTT sequences are exposed on the cell surface during the course of a viral infection, the CTT may then become an important target for neutralizing antibodies. Previous results have demonstrated that anti-Kennedy epitope MAbs can mediate post-attachment neutralization of viral infection (Heap et al., 2005; Reading et al., 2003). In addition, the CTT exhibits greater sequence conservation than gp120 (Steckbeck et al., 2011). These results, in light of the cell-surface CTT exposure demonstrated here, suggest that the *in vivo* exposure of CTT sequences on the surface of infected cells might provide the immune system exposure necessary to generate an anti-HIV antibody response. It is important to remember that early studies demonstrated an abundance of anti-CTT antibodies in patient sera (Chanh et al., 1986; Ho et al., 1987).

In contrast to the exposure of the CTT on the cell surface, we were not able to detect CTT exposure on the virion. The fact that we were able to make infectious pseudoviruses suggests that the Env incorporated into particles is properly functional, a result that was not predictable due to the 11-residue sequence replacement required for epitope tagging. Production of infectious pseudoviruses in and of itself suggests that functional redundancy might be encoded in the CTT, a concept also suggested by previous results demonstrating a need to mutate two endocytic sequences in the CTT to fully abrogate trafficking of Env to late endosomes (Byland et al., 2007).

Observed topological differences in the CTT between the cell and virion surfaces leads to questions regarding the mechanism by which the observed differences occur. There are two potential explanations: (1) the CTT-exposed Env is incorporated into the budding virion, and undergoes topological rearrangements to localize to the interior of the viral membrane; or (2) the CTT-exposed Env is excluded from the budding virion. While (1) is possible, we believe that (2) is the more plausible explanation.

A number of studies have demonstrated a functional interaction of the CTT with proteins localized exclusively in the cellular cytoplasm (Byland et al., 2007; Freed and Martin, 1995, 1996; Jiang and Aiken, 2007; Kol et al., 2007; Lopez-Vergès et al., 2006; Ohno et al., 1997; Wyma et al., 2004; Wyma et al., 2000). The simplest argument for the exclusion of CTT-exposed from the budding virion is that in this arrangement the CTT cannot interact with known intracellular partners, and thus does not traffic properly into the viral budding site. However, this simple explanation is sufficient to explain the exclusion of CTT-exposed Env from the virion only to the extent that Env incorporation is an active, regulated process. Evidence from CTT-deleted Env constructs suggests

that active Env incorporation is not always necessary. It has been previously demonstrated that viruses encoding a truncated CTT of only 17 amino acids past the putative membrane spanning domain retain full infectivity *in vitro* (Chakrabarti et al., 1989; Hirsch et al., 1989; Kodama et al., 1989). Since CTT-deleted Env cannot be actively incorporated into the budding virion through interactions with intracellular partners, it must be incorporated passively (Checkley et al., 2011). This then begs the question, "Why is the CTT-exposed Env not passively incorporated into budding virions?"

The answer may have to do with the differences in cholesterol and phospholipid content between cellular and viral membranes, particularly with regard to changes in lipid membranes material properties imposed by cholesterol. Previous reports have demonstrated that the cholesterol to total phospholipid ratio of the HIV virion is increased relative to the parental cell by approximately 2-2.5 fold (Aloia et al., 1988; Aloia et al., 1993; Brügger et al., 2006). Additionally, the levels of saturated lipids in the viral membrane is increased relative to the cellular membrane by 3.6 fold, resulting from both an enrichment of saturated lipids combined with a reduction in the levels of di- and polyunsaturated lipid species, particularly for dipalmitoyl-phosphatidylcholine (Brügger et al., 2006). The observed cholesterol and saturated lipid enrichment in the viral membrane is important as recent studies have demonstrated that increased levels of cholesterol strongly affects the bending modulus, K_C, of membranes in which both phospholipid acyl chains are unsaturated by increasing the energy required to bend the membrane (Pan et al., 2008; Pan et al., 2009). This effect becomes significant in the case of protein sorting mediated by the differential effect on lipid bilayer curvature

induced by transbilayer versus in-plane protein insertion. Transmembrane helices distort both leaflets of the membrane approximately equally with respect to the lipids, and therefore would not be expected to induce any membrane curvature. In-plane insertion, on the other hand, leads to membrane curvature in laterally uncoupled monolayers depending on the depth of insertion (Campelo et al., 2008). Therefore, with respect to the cholesterol and unsaturated phospholipid-enriched membrane microdomains on the cell surface from which HIV is thought to bud, inclusion of an in-plane helix relative to a transmembrane helix would be thermodynamically unfavorable due to the proposed increase in microdomain K_C. This model (see Figure 37) could allow for the observed passive incorporation of CTT-deleted Env, while allowing for the exclusion of CTT- exposed Env from the membrane of a budding virion.

While the above model proposes a means by which CTT-exposed Env is excluded from budding sites, the transient exposure of CTT sequences as demonstrated by use of the PAN assay adds an additional level of complexity to the topic of CTT topology. Transient exposure of CTT sequences during membrane fusion may provide a mechanistic rationale for the previously demonstrated conservation of arginine relative to lysine in the CTT (Steckbeck et al., 2011). In order to become transiently exposed during the virus-cell fusion process, CTT sequences would need to

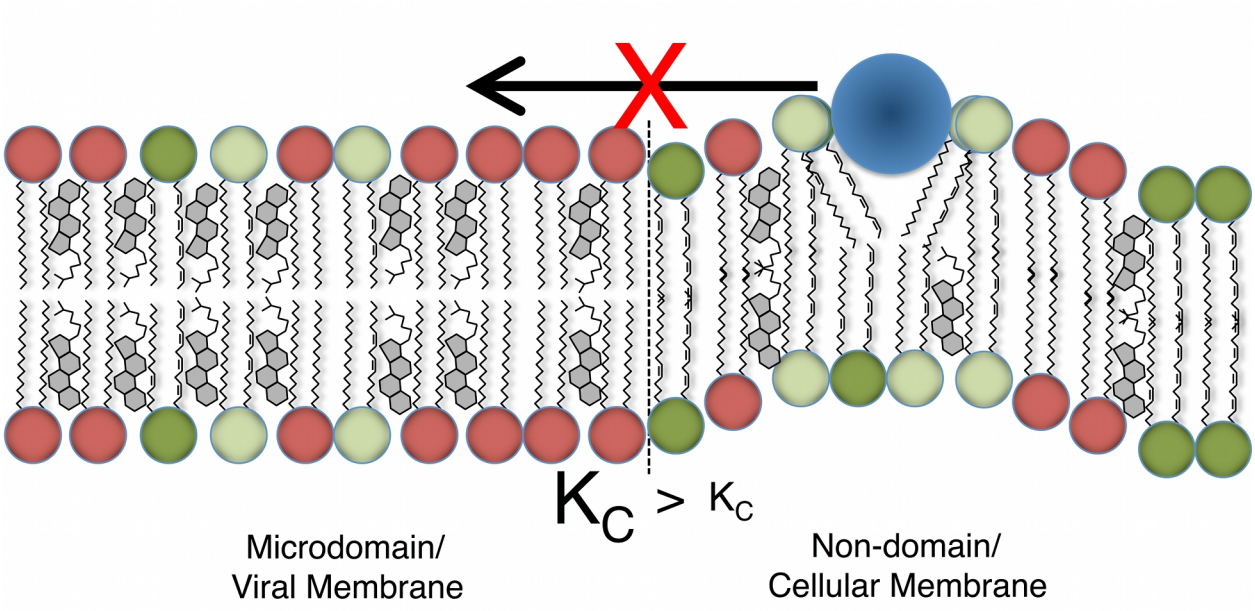

Figure 37. Schematic model for exclusion of CTT-exposed Env from viral assembly and budding sites

A physical model is proposed for the apparent inability of CTT-exposed to segregate into viral budding sites. Lipids with green headgroups have unsaturated acyl chains, while red headgroups have saturated acyl chains and are enriched, along with cholesterol (gray), in membrane microdomains and the viral membrane. The blue circle represents a helix partially inserted into the membrane interface. K_C is the bending modulus of the membrane.

traverse the lipid bilayer from their intravirion localization. Arginine has been demonstrated to facilitate translocation of soluble small peptides and large proteins through lipid bilayers (Futaki, 2005; Inomata et al., 2009; Mitchell et al., 2000; Tung and Weissleder, 2003). It is possible that the functional consequence of arginine conservation is to facilitate the movement of CTT sequences through the membrane during virus-cell fusion. The functional requirement for CTT translocation remains to be mechanistically defined.

The results presented here introduce a new concept to the topology of the CTT, in particular that the CTT has the potential to exist in two distinct states that can differ between Env-expressing cells and the virion surface. Combined with the apparently transient exposure of CTT sequences during virus-cell fusion, the current studies highlight that the CTT sequence is more complex, both functionally and structurally, than the current intracytoplasmic model suggests.

6.0 GENERAL DISCUSSION OF RESULTS AND FUTURE RESEARCH DIRECTIONS

6.1 SUMMARY OF MAJOR FINDINGS

The goal of this dissertation project was to provide an experimentally-determined topology of the HIV gp41 CTT. Inherent to this goal was finding an amenable methodology that would provide a reasonable means of determining the "true" topology of the CTT (whatever it might be), on both Env-expressing cells as well as on the virion surface. An additional goal was to define the conservation of CTT sequence and structural elements to provide a commentary on the overall relevance of the CTT as well as to provide the first broad measure of these traits in the CTT literature. In this regard, through the use of physicochemical property calculations and computational folding simulations combined with experimental secondary structural confirmation, remarkable structure and property conservation was observed for CTT domains in spite of relatively low phylogenetic relatedness and the resulting high levels of genetic diversity. In parallel, the experimental topology determination provided additional evidence for the existence of a complex topology for the CTT beyond that of the currently-accepted model of Env as strictly a type I integral membrane protein, including transient exposure of CTT sequences during the membrane fusion process. These results provide for the

first time a consistent systematic topological determination for the CTT on the surface of both intact Env-expressing cells and viral particles. Collectively, in light of the surrounding CTT literature, these results suggest that the CTT sequence has important functional consequences beyond its currently described roles in Env virion incorporation and maturation including a functionally-necessary rearrangement during membrane fusion that necessitates conservation of sequence and structural properties. These findings will be discussed in more detail below.

6.1.1 Physicochemical and structural characteristics of the CTT are widely conserved across genetically distinct viral sequences

6.1.1.1 Summary of CTT conservation

The findings from Chapter 3 demonstrated a remarkable conservation of physicochemical and structural properties of CTT sequences despite a high level of sequence divergence and genetic dissimilarity. In particular, the LLP regions of the CTT displayed consistent hydrophobicity, net charge, and hydrophobic moment across consensus sequences from eight clades/groups while demonstrating similar levels of sequence variability to the non-LLP regions, suggesting that these physicochemical characteristics are specifically conserved in otherwise variable sequences. Additionally, the CTT displayed a surprising conservation of arginine residues relative to lysine, particularly in the LLP regions (see section 3.3.2).

From a structural standpoint, the LLP regions of the CTT also displayed a high level of conservation. When consensus sequences of the LLP regions were subjected to computational folding, the structural diversity was low for each LLP region within a

clade/group, as evidenced by few structural clusters, in general (see Section 3.3.3). In addition, for each LLP region the structural similarity between clades/groups was demonstrated to be quite high (again, see Section 3.3.3, Figure 21). This predicted structural conservation was confirmed experimentally, where peptide analogs of the consensus sequences (synthetic) and the lowest average divergence sequences (actual sequences) from clades B, C, and group O were examined using CD spectroscopy in SDS micelles and TFE. The CD spectroscopy results confirmed the conservation of helical character in the peptides in both membrane-mimetic environments.

6.1.1.2 Implications of CTT property and structural conservation

The conservation of the physicochemical properties is particularly interesting for the LLP regions in light of their proposed membrane associating characteristics (Chernomordik et al., 1994; Fujii et al., 1992; Kliger and Shai, 1997; Srinivas et al., 1992), where the hydrophobic moment and charge may complement each other. The hydrophobic moment has been proposed as a measure of the tendency of a sequence to prefer the chemically-complex interfacial boundary between the hydrocarbon membrane interior and the aqueous phase (Eisenberg et al., 1984). The preference for the chemically complex membrane- water interface is suggestive of a role for the LLP regions as membrane-anchoring sequences. The (partial) insertion of the LLP sequences into the interfacial region would have a physical influence on the local membrane environment through introduction of local curvature stress, as well as by affecting the lateral pressure profile in the membrane interior (Booth et al., 2001; Cantor, 1997; Cantor, 1999a, b; Tillman and Cascio, 2003; van den Brink-van der Laan et al., 2004). These physical alterations of the membrane environment by interfacially-localized LLP peptides may

provide an explanation for the modulation of Env ectodomain conformation induced by point mutations in the LLP (Kalia et al., 2005). While the mutations introduced by Kalia et al. maintain the amphipathic potential of the LLP2 region, they alter the net charge of the region from +3 to -1 (Kalia et al., 2005). The resulting overall negative charge of the region may impact its association with the negatively charged inner membrane leaflet by introducing charge-charge incompatibilities with the negatively charged phosphatidylserine headgroups. Altered association (i.e. reduced/abrogated association) of LLP2 with the inner leaflet of the membrane would lead to a change in the local lateral pressure profile of the membrane, which in turn could lead to conformational changes in the protein by altering the arrangement of the transmembrane helices (Cantor, 1997; Cantor, 1999a; Cantor, 2002; Marsh, 2007; Tillman and Cascio, 2003; van den Brink-van der Laan et al., 2004). While it is difficult to experimentally validate this argument since the composition and structure of the gp41 MSD is unknown, recent structural work on full-length and CTT-deleted Env on the virion surface hint at differences in the geometry of the MSD with respect to the membrane (White et al., 2010). Comparing the approximate geometries of the protein near the membrane interface (the best that can be done under these circumstances) reveals that the full-length Env adopts a roughly cylindrical shape, while the CTT-deleted Env protein exits the membrane in a pyramidal arrangement (c.f. Figure 10). Assuming similar lipid content between the full-length and CTT-deleted viruses (which may be an invalid assumption as the CTT may interact with and encourage the assimilation of specific types of lipids into the budding virion), the only difference then would be the presence/absence of the CTT, and resulting subsequent differences in

membrane curvature and lateral pressure, which, as indicated above, could account for differences in transmembrane helix packing suggested by the recent structural work. Likewise, as these two Env proteins are thought to represent distinct functional states (with the full-length representing the "resting" or native state of Env, and the CTT-deleted representing the receptor-bound, or "activated" Env), the effect of lateral pressure changes could, in theory, play a significant role in modulating the conformation of membrane-bound Env (Cantor, 1997; Cantor, 1999a; Cantor, 2002). Interestingly, the CTT has recently been demonstrated to modulate the conformation of gp120 and the gp41 ectodomain in virus particles, where Env with a CTT deletion displayed a more open state similar to the CD4-bound state (Joyner et al., 2011; White et al., 2010). Thus, there exists the possibility that the CTT, and the LLP regions in particular, evolves under pressure to maintain the "native" conformation of Env by maintaining specific consistent interactions with the lipid membrane.

The potential implications of arginine conservation with respect to LLP2 have been discussed in detail in Section 3.4. Briefly, studies demonstrating a transient exposure of LLP2 sequences during the cell-cell fusion process are suggestive of a mechanism to prefer arginine conservation relative to replacement with lysine. Arginine-rich peptides have been demonstrated to cross cellular membranes while peptides with lysine instead of arginine do not (Futaki, 2005; Mitchell et al., 2000; Tung and Weissleder, 2003). In addition, arginine-rich peptides have been demonstrated to deliver soluble proteins into the cytoplasm of live cells (Inomata et al., 2009). The observed apparent traversing of the membrane by LLP2 during the fusion process would seem, then, to require the presence of arginine.

Implications for the overall observed conservation of arginine in the CTT become more interesting when the current topology results are taken into account. The data presented in Chapters 4 and 5 demonstrate that the topology of the CTT is (can be) distinct between the cellular and virion surface ("can be" as a function of the fact that our cellular data does not preclude the existence of virion-like CTT in the cell). This discovery of a cellular topology that is distinct from the virion topology is suggestive of either an alternative, stable topology that is (apparently) functionally irrelevant to the virion, or a dynamic rearrangement of the topology from the extracellular (lumenal or exposed) state to the internally-localized (intracytoplasmic) form (It is worth noting at this point that these two statements are not wholly incompatible, although some conjecture is necessary to align the two; as a membrane-bound protein, it is possible that Env interacts with a number of proteins during its biosynthesis. There exists, then, the possibility that an unknown interactant could stabilize the exposed topology in a subset of the synthesized Env population that may normally undergo dynamic topological rearrangement during the biosynthetic process).

A potential topological rearrangement will be addressed later in a proposed Env model, but for the sake of argument, if CTT-exposed Env must "flip" sides of the membrane (crossing the hydrocarbon interior in the process) to reach a different topology, the presence of arginine relative to lysine would be energetically preferred according to the biological hydrophobicity scale (Hessa et al., 2005a). A movement of arginine-rich sequences across a membrane when captive as part of a larger protein complex (as opposed to soluble peptide fragments) is perhaps best demonstrated in the voltage-sensor paddle of the potassium-gated ion channel KvAP, where the arginine-

rich S4 helix (with four arginine residues) is proposed to cross the membrane hydrocarbon interior in response to changes in the membrane electrical field (Jiang et al., 2003). Interestingly, this same helical segment has demonstrated potential to insert into the membrane as a transmembrane helix in spite of the four arginine residues (Hessa et al., 2005b).

In all, the findings from the study of CTT conservation led to the identification of sequence and structural elements that contributed to the ability to propose a topological model for the CTT that has mechanistic and functional support from other membrane protein systems, and is consistent with membrane protein topological principles.

6.1.2 Distinct CTT topologies on the cell and virion surface

6.1.2.1 Summary of CTT cellular and viral topology studies

The findings from Chapters 4 and 5 demonstrate that distinct topologies exist for the CTT between the surface of Env-expressing cells and viral particles. In Chapter 4, a particularly controversial antibody epitope in the CTT, the Kennedy epitope, was examined on cells and virions to provide an assessment of its exposure on each. This study introduced the use of an epitope-tagging technique whereby the VSV-G epitope tag was used to replace CTT sequences in order to take advantage of a high-affinity, well-characterized commercial monoclonal antibody. The results from this study demonstrated that the Kennedy epitope was exposed to antibody binding on the surface of Env-expressing cells as determined by flow cytometry, but was not apparently accessible on the virion surface in an immunoprecipitation assay. The immunoprecipitation results were confirmed by SPR spectroscopy to measure the

relative affinity of the anti-Kennedy epitope antibody to ensure that the observed lack of virion binding was not due to a high dissociation-rate of the antibody-virion complex.

The study presented in Chapter 5 extended the use of the VSV-G epitope tag to the entirety of the CTT sequence. Here, the VSV-G epitope tag was used to serially replace sequences along the entirety of the CTT, and cell surface exposure of each was determined by flow cytometry. Additionally, and importantly, the VSV-G-tagged Env constructs were used to create pseudoviral particles in an attempt to provide a consistent measurement across cellular and virion Env presentation. The VSV-G-tagged *env* were co-transfected into 293T cells with pSG3ΔEnv to make pseudoviral particles that were determined to be infectious using the TZM-bl assay, suggesting that all the VSV-G-tagged Env constructs were functional. These VSV-G tagged pseudovirions were then used to determine apparent exposure of the VSV-G epitope tag by immunoprecipitation. The results from these studies indicated that a majority of the CTT is accessible to antibody binding on the surface of Env-expressing cells. In contrast, no CTT sequences were detected on the pseudovirion surface. Finally, the use of a post- attachment neutralization assay (PAN) allowed a demonstration of a transient exposure of CTT sequence, as virus and anti-CTT MAb incubated at 31 °C resulted in viral neutralization while incubation at 37 °C did not.

6.1.2.2 Implications of different cellular and virion CTT topologies

The results presented here provided the first systematic evaluation of CTT topology on both the cell and virion surface in an experimentally-consistent manner. The first important observation comes from the infectivity results using the VSV-G-tagged pseudoviral particles. The serial replacement of CTT sequences by the VSV-G epitope

tag was expected to yield some constructs that would be non-infectious, as the replacement of CTT sequences would be expected to alter important functional domains, particularly the demonstrated MA interaction (Freed and Martin, 1995, 1996; Murakami and Freed, 2000a, b). This interaction would have been expected to be altered in VSV-G constructs CTT8, CTT9, and CTT10 as they each replaced a portion of the sequence demonstrated to be important for MA interaction and Env incorporation (Murakami and Freed, 2000a, b). Since all constructs did yield infectious virus and displayed similar Gag:Env ratios, the results suggest that Env has evolved functional redundancies into the CTT. This is not unprecedented for Env, even in the CTT. Previous studies have demonstrated the existence of two functionally and mechanistically distinct endocytic motifs in the CTT, with locations near the N- and C-terminal ends of the CTT sequence. The N-terminally-localized YxxΦ interacts with the clathrin adaptor protein 2 (AP-2) to mediate Env endocytosis from the plasma membrane, while the C-terminal dileucine motif interacts with AP-1 to effect Env endocytosis independent of the AP-2 (Byland et al., 2007; Checkley et al., 2011). Importantly, mutation of both motifs is required to abolish Env cycling to late endosomes from the cell surface (Byland et al., 2007). Additional functional redundancies exist in the CTT in the form of motifs that regulate or enhance partitioning of Env to cholesterol-rich membrane microdomains. Here, acylation of endogenous cysteine residues, the presence of a cholesterol recognition amino acid consensus (CRAC) motif, and helicity of LLP1 have all been proposed as mechanisms by which the CTT directs Env to the cholesterol-rich microdomains from which HIV is thought to bud (Greaves and Chamberlain, 2007; Rousso et al., 2000; Yang et al., 1995; Yang et al., 2010).

Functional incorporation of VSV-G-tagged Env into infectious pseudoviral particles suggests that there are redundancies in the mechanism by which Env is incorporated into viral particles, either in the MA-mediated incorporation pathway alone, or through the use of multiple incorporation pathways.

Another implication of a distinct cellular CTT-exposed topology is the potential for the CTT to act as an immunological decoy. This idea is conceptually appealing and enjoys support from numerous published studies. Early studies that characterized the Kennedy epitope as a potential neutralizing epitope also demonstrated that sera from HIV-infected individuals were reactive to the Kennedy epitope peptide (Chanh et al., 1986; Ho et al., 1987). This observed sera reactivity is suggestive of the fact that the CTT can serve as a potent immunogen during the course of viral infection. The cell surface-exposed CTT topology demonstrated here suggests a mechanism by which the previously-observed immunogenicity may occur. CTT exposure would present non-self antigens on the surface of HIV-infected cells. The induction of antibodies to the surface-exposed CTT would provide the immune system with a target that is bereft of the other immune decoy mechanisms inherent to Env, including high levels of glycosylation and conformational masking (Roux and Taylor, 2007). Antibodies to the Kennedy epitope in the CTT are of ambiguous utility, with *in vitro* neutralization demonstrated only in a minority of studies at the physiologically-relevant 37 °C (Chanh et al., 1986; Cleveland et al., 2003), with most studies requiring a delay of the fusion process by reducing temperature to ≦ 31 °C to enable neutralization (Cheung et al., 2005; Heap et al., 2005; Reading et al., 2003). If the exposure of the CTT during virus-cell fusion is a transient process, as is suggested by these and other studies (Viard et al., 2008), the induction of

antibodies to the CTT would result in antibodies with little to no anti-HIV activity. In addition to serving as a mechanism to "distract" the immune response, this may explain the relatively high levels of sequence variation observed in the CTT as demonstrated in Chapter 3. In this regard, data from our collaborators have demonstrated an apparent immunodominance of CTT sequences when added to vaccine immunogens containing other membrane interactive segments of gp41 (personal communication). In total, these results suggest that the cell-surface exposure of CTT sequences results in a directing of the immune system toward Env epitopes that are not exposed on the virion and have been demonstrated to be functionally ambiguous *in vivo.*

The demonstrated ability of anti-Kennedy epitope antibodies to neutralize virus infection in a post-attachment manner also opens up the possibility of targeting the CTT to inhibit viral infection. The apparent transient exposure of the Kennedy epitope during the fusion process presents a number of problems affecting antibody-mediated neutralization that could be overcome through the use of small(er) inhibitors. While antibody-antigen binding is generally a high affinity process (on the order of nanomolar K_D, or lower; the affinity of current anti-CTT monoclonal antibodies for gp41 are not known), the architecture of the fusion site is likely to present a major inhibitory influence on antibody binding to CTT sequences. The initial contact distance between biological membranes during the fusion process is on the order of 10 - 20 nanometers (nm) which decreases to 2 - 3 nm prior to the formation of the hemifusion stalk (Chernomordik et al., 2006). As it is thought that the major temporal delays in HIV fusion occur at the CD4 (followed by rapid formation of the pre-hairpin intermediate, leaving the large distance between membranes) and co-receptor binding steps (followed by rapid formation of the

six-helix bundle, leaving a small distance between membranes) (Gallo et al., 2003), under physiological conditions a potentially neutralizing antibody would have a small timeframe to reach into the fusion site and bind the transiently-exposed Kennedy epitope, as the coreceptor binding step occurs with a 500-fold faster apparent on-rate than the CD4 binding interaction (Chang et al., 2005). As the incubation temperature of the PAN assay (31 °C) is thought to delay the progression of the gp41 complex from the prehairpin intermediate to the formation of the six-helix bundle (Golding et al., 2002), and neutralization by anti-Kennedy epitope monoclonal antibodies only occurs at this temperature, the assumption must be that the CTT undergoes dynamic rearrangement during this time, which as indicated above is brief at 37 °C. The 2 - 3 nm distance at the point just before the formation of the hemifusion stalk would exclude antibody, which has molecular measurements on the order of 5x5x10 nm. A small molecule, on the other hand, could be designed to mimic the neutralizing effect of antibody in a size that would be much more amenable to interacting in the narrow confines of membranes undergoing the fusion process. The dynamic exposure, then, of sequences that make the CTT a potential neutralizing epitope also make the CTT a target for small molecule drug development and inhibition.

A final implication of a cellular CTT-exposed topology is that there would then be the potential for the CTT to act as a target for cytotoxic T-lymphocytes (CTLs). While this would appear to be in direct conflict with the concept that endocytic signals exist in the CTT in order to effect the active removal of Env from the cell surface so as not to “broadcast” the cell’s infected state, time may be on the side of the virus. Assuming the observed exposure is a result of accumulation of CTT-exposed Env at the cell surface,

and that the same process would take place in an infected cell, there should be sufficient time for virion assembly and budding to ensure the propagation of the infection before the accumulation of enough CTT-exposed Env to result in CTL recognition and cell lysis.

6.2 A COMPREHENSIVE WORKING MODEL FOR THE CTT

A new comprehensive model for CTT topology must take into account experimentally-determined information including both explicit topology data as well as implicit functional data. In addition, a proposed model must make theoretical sense and must conform to the known principles for membrane protein insertion and topology. Finally, a proposed model should ideally provide testable hypotheses where it makes assumptions based on predictive results. As such, following a brief theoretical revisitation of CTT topology prediction, we propose a comprehensive model for CTT topology that addresses all current and published functional and topological data.

6.2.1 Revisiting the Membrane Association of gp41 - A Theoretical Study

The experimental data presented here have demonstrated surface accessibility of CTT sequences on intact Env-expressing cells with the same sequences being undetectable on the virion surface. These results, combined with the calculated propensity for a gp41 sequence to serve as a potential in-plane anchor, prompted an *ab initio* theoretical study of the association of gp41 with the membrane. Key to this study are recent

developments in the understanding of membrane protein biogenesis, chief among which was the recent development of a biologically-derived hydrophobicity scale for protein insertion into the ER membrane that has provided the best evidence to date that the insertion of membrane proteins into the lipid bilayer is a thermodynamically-driven partitioning process (Hessa et al., 2005a; Hessa et al., 2007). The important principle underlying this finding is that as a thermodynamic process, the insertion of hydrophobic segments is an equilibrium event between insertion and translocation of the segment in question, dependent on ΔG^{app}, where $\Delta G^{app} = 0$ results in a 1:1 ratio of insertion:translocation into the ER lumen (Hessa et al., 2005a; White and von Heijne, 2008). In essence, the goal is to start at the translocon and build a theoretical model for gp41 membrane association.

The protein translation and translocation machinery of the cell has no knowledge of the history or bias inherent to the study of the topology of gp41. It "merely" detects and properly partitions protein regions on the basis of their hydrophobicity into either the lipid bilayer (for hydrophobic sequences with appropriate properties) or into the lumen of the ER (for ultimately soluble, non-cytoplasmic proteins) (von Heijne, 2006; White and von Heijne, 2008). The first step was to reexamine the Env sequence using the biological hydrophobicity scale to understand the theoretical basis of its membrane association since Env is translated as a polyprotein and cleaved into gp120 and gp41 post-translationally and after trafficking from the ER to the Golgi (Checkley et al., 2011). The initial scan identified four regions as potential transmembrane helices in the gp160 sequence: (i) positions 12-30, with $\Delta G^{app} = 0.417$ kcal/mol (the signal peptide); positions 514-542, with $\Delta G^{app} = 0.883$ kcal/mol (the fusion peptide); positions 678-705 with ΔG^{app}

= -4.28 kcal/mol (N-terminal to, and including the traditional MSD sequence); and positions 749-778 (including residues prior to and including the N-terminal 14 amino acids of LLP2), with ΔG^{app} = 0.22 kcal/mol. When converted to K^{app}, the predicted ΔG^{app} reveals the relative extent to which a helix should be inserted into the membrane as compared to translocated into the lumen of the ER (ΔG^{app} = -RT ln K^{app}; R = universal gas constant; T = absolute temperature; K^{app} = inserted/non-inserted). It was not unexpected to find membrane-insertion propensity for both the signal peptide as well as the fusion peptide as the role of each is known to involve membrane interaction. As expected, the traditional MSD sequence was predicted to insert as a membrane spanning helix. Perhaps unexpected, however, was that residues N-terminal to the MSD were included as part of the predicted membrane-inserting region. A comparison of the ΔG^{app} for both the traditional MSD (-0.88 kcal/mol) and the predicted region (-4.28 kcal/mol), reveals that inclusion of the N-terminal residues results in an approximately 282-fold increase in the K^{app} (1214 versus 4.3, where K^{app} is the ratio of inserted:non-inserted), leading to a much greater predicted insertion efficiency. It is interesting to note that no known MAbs react with the sequences that are part of the predicted transmembrane region (from the LANL immunology antibody database, www.hiv.lanl.gov/content/immunology/maps/ab/gp160.html).

Perhaps unexpected was the identification of an additional potential transmembrane helix in the CTT, near and including part of the LLP2 region. The predicted ΔG^{app} for this region, at 0.22 kcal/mol, makes this sequence very interesting from a membrane interaction standpoint. Conversion to K^{app} reveals that the inserted:non-inserted ratio is 0.69, suggesting that nearly 40% of the time this region is

predicted to insert into the membrane as a transmembrane helix. It will be important to experimentally validate the predicted membrane insertion efficiency of this region.

Predictions using the biological hydrophobicity scale provide insight into the potential for protein membrane association during translation. It does not provide explicit information on the final folded protein, and in fact, it is becoming clear that many membrane proteins may rearrange post-translationally into a structure different from that which is inserted by the translocon (Kauko et al., 2010). As such, post-translational trimerization of Env may provide an opportunity for rearrangement of the transmembrane regions.

6.2.2 A Proposed Topological Model for the HIV gp41 CTT

Env is translated as a monomeric unit that is directed to the ER membrane for translation and membrane insertion by recognition of the N-terminal signal peptide by the signal recognition complex on the translating ribosome. While the *env* mRNA is being translated into protein by the ribosome, the nascent peptide chain is being evaluated cotranslationally by the translocation complex, and the translocon in particular, for transmembrane sequences. Sequences displaying the correct properties, chief among them being sufficient hydrophobicity, are shuttled into the membrane environment by what is thought to be a thermodynamically-driven partitioning process (Hessa et al., 2005a; von Heijne, 2006; White and von Heijne, 2008). An important point to consider before going further is that the ER membrane has symmetric lipid distribution between the bilayer leaflets (van Meer et al., 2008) and displays no electrical membrane potential (Dowhan and Bogdanov, 2011). Theoretically, then, there

should be no charge/lipid influence on membrane protein topology during translocation (it is important to note that we believe those factors play a role later in the biosynthetic process, particularly upon trafficking to the Golgi). The comprehensive model is presented in Figure 38. At this first step, Env gp160 is predicted to display a potential for a complex topology. The biological hydrophobicity scale predicts strong membrane insertion of a 28 residue segment that includes the traditional MSD sequence, labeled MSD1 in Figure 38. Conversion of ΔG^{app} to K^{app} results in an inserted:non-inserted ratio of 1214, indicating that MSD1 is predicted to insert as a transmembrane region in the vast majority of the translated Env. There would be, however, a very small population (< 1 in 1000 translated gp160 polyproteins) where this region would not insert as a transmembrane helix, and assuming that the remainder of the sequence translocated, this would give rise to the topology observed on the cell surface, with nearly the entirety of the CTT accessible to antibody binding, and the predicted in-plane anchor sequence (which is very similar to the predicted transmembrane sequence) along with the inaccessible LLP2 region anchoring Env to the membrane. The observation of this very small population would be enhanced at the cell surface due to the fact that the functional endocytic signals would be located on the opposite side of the membrane from the endocytic machinery, allowing an accumulation of exposed Env on the cell surface; Env with a single membrane spanning domain would be rapidly endocytosed, as previously described (Byland et al., 2007). This accumulation at the cell surface may account for the topological results presented in Chapters 4 and 5. Transmembrane insertion of the second region, labeled MSD2, (positions 749-778) would sequester the C-terminal 73 amino acids in ~40% of this already small population

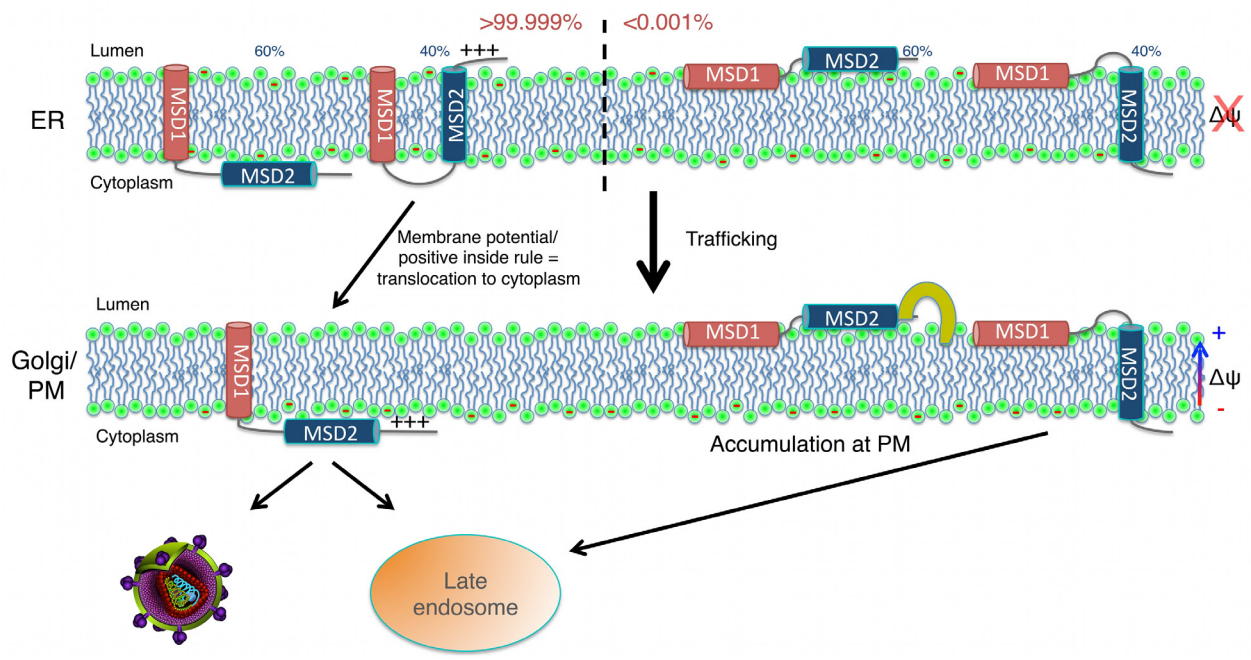

Figure 38. Proposed topological model for the HIV gp41 CTT

A proposed gp41 CTT topological model during the protein biosynthetic process, from the initial protein translation and translocation into the ER, trafficking to the Golgi and plasma membrane (PM), and finally incorporation into viral particles or recycling to the late endosome. The Env ectodomain (gp120 and gp41 ectodomain) are located on the lumenal side of the membrane. MSD regions are those predicted using ΔGPred with the indicated percentages being the relative insertion efficiency based on predicted K^{App} (red for MSD1 and blue for MSD2). This proposed model is consistent with the topological data presented here as well as previously published functional data.

(-) - negatively charged phospholipid headgroups, predominantly phosphatidylserine; yellow "hook" represents proposed Env interacting partner that holds the CTT in the exposed orientation

into the cytoplasm, but this too would be endocytosed from the cell surface due to the presence of the extreme C-terminal dileucine endocytic motif (Byland et al., 2007). As suggested in Section 6.1.1.2, an unknown Env interactant could "hold" the CTT-exposed Env in this conformation throughout the biosynthetic process, resulting in the observations presented here in Chapters 4 and 5. This Env would not be expected to incorporate into the budding virion, as discussed in Section 5.4.

Returning to the main predicted transmembrane domain, in the vast majority of cases it is predicted to insert as a transmembrane helix, which is in line with current thoughts on the Env MSD. In addition, however, the second predicted transmembrane region at positions 749-778 is predicted to insert into the membrane with ~40% efficiency. Assuming insertion of the first transmembrane sequence, transmembrane insertion of this second sequence would result in the lumenal (ER) exposure of the 73 C-terminal amino acids in ~40% of the translated Env.

At this point, for the majority of the translated Env (gp160 polyprotein) there is a predicted 60/40 distribution of one transmembrane region (with entirely cytoplasmic CTT) and two transmembrane regions (with C-terminal 73 residues of the CTT lumenally-localized). Before leaving the ER and trafficking to the Golgi, Env undergoes a shift from monomers to trimers (Checkley et al., 2011), and here I propose that trimerization results in a "shifting" of the predicted transmembrane sequences to resemble the traditionally-held MSD sequences and bringing the conserved arginine residue to the mid-span position (Figure 39). This proposed mechanism of insertion followed by a sequence shift upon trimerization would allow for easier (and

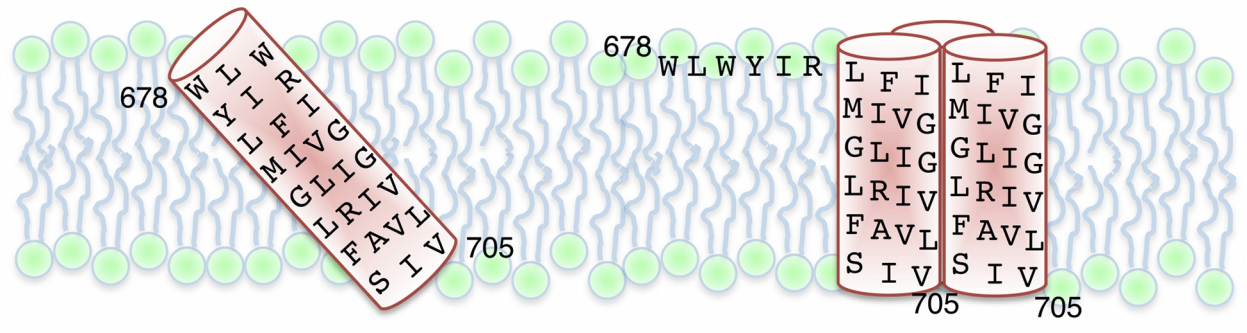

Figure 39. Proposed MSD sequence rearrangement upon trimerization

The predicted MSD for HIV gp41 (left) contains an additional six residues compared to the generally accepted (traditional) MSD sequence (right). As a monomer (as transcribed and translocated), the predicted MSD has a more favorable ΔG^{app} (-4.28 kcal/mol) than the traditional sequence (-0.88 kcal/mol). Trimerization could result in a more favorable free energy of insertion for the traditional sequence by allowing intrahelix interaction of the midspan arginine, as suggested by MD simulations (Kim et al., 2009).

energetically-favorable) snorkeling of the arginine in the translocon-inserted form that upon trimerization and movement into the membrane interior would serve to stabilize the three-helix bundle as proposed (Kim et al., 2009).

Trafficking to the Golgi introduces a new environment and new forces to the translated Env polyprotein. In the Golgi, the membrane leaflets have an asymmetric lipid distribution, with the negatively charged phosphatidylserine and the zwitterionic phosphatidylethanolamine localizing to the cytoplasmic leaflet (Dowhan, 1997; van Meer et al., 2008). Additionally, an electrochemical gradient exists across the Golgi membrane such that the electrochemical potential of the membrane is oriented positive in (cytoplasmic side is negative) (Glickman et al., 1983; Llopis et al., 1998; Rudnick, 1986). As the trimerized Env incorporates into the Golgi membrane, this new environment and force would be expected to play a role in the membrane topology, as has been described (Bogdanov et al., 2008; Dowhan and Bogdanov, 2009, 2011; von Heijne, 2006). The traditional Env with the cytoplasmically-localized CTT would not need to rearrange in response to the new environment and forces, as the overall positively charged nature of the sequence would be situated to interact favorably with the negatively charged lipid headgroups as well as be in harmony with the electrochemical potential of the membrane. The other 40% of Env with two transmembrane domains, however, would likely undergo topological rearrangement as the 73 amino acids predicted to be exposed carry a +5 charge. Here, the predicted transmembrane domain would act as a hinge region, allowing for the translocation of the final 73 amino acids into the cytoplasm from the lumen, brining the structure in line with the positive-inside rule (Bogdanov et al., 2008; von Heijne, 1986). This

rearrangement would also consolidate the majority of Env into the traditional model, with an entirely cytoplasmic CTT. (Aside: The presumption at this step for the CTT-exposed Env discussed above must be that whatever is keeping the sequence lumenally (extracellularly) localized must be doing so throughout the biosynthetic process, thus preventing translocation of the sequence to the cytoplasm in the Golgi (also, incidentally, suggesting that such a factor would then have to associate with Env in the ER or during the trafficking process to the Golgi)). This rearrangement would result in almost totally cytoplasmically-localized Env (with the minor population discussed above) and would allow for all the observed functional interactions previously described, including interactions with Gag and the endocytic machinery.

6.3 FUTURE DIRECTIONS

I would guess that any good dissertation project, while answering the proposed questions will, in the advisor's best hope, generate additional questions to be answered by those who come after (and thus spins the science "circle of life"). In this regard, this project was no different. We have identified a number of interesting possibilities for the next generation to advance and investigate the structure of the CTT and the structural implications of the CTT on Env ectodomains.

6.3.1 Did the epitope mapping technique provide a non-native topology for the CTT?

A natural hesitation regarding the use of sequence substitutions to map protein topology would be that the substituted sequence will likely not perfectly replicate the physicochemical properties or topological signals inherent to the native sequence. While we believe our approach was valid, and that the cell surface topology we reported is correct and relevant (particularly as the epitope substitutions ultimately resulted in infectious viral particles), it is always useful to provide verification in a manner that will absolutely minimize the impact of the methodology on the protein. As such, the logical step would be to revisit the use of single cysteine mutants to map the topology of the CTT. Replacement of single cysteines would be expected to have a very small impact on the topology of a protein, and has been used in many systems to map protein topology (Bogdanov et al., 2002; Bogdanov et al., 2008; Bogdanov et al., 2005; Cai et al., 2001; Cai et al., 1999; Klein-Seetharaman et al., 1999, 2001; Wang et al., 2002; Zhang et al., 2003). Most importantly, this would allow for the determination of CTT topology in HIV infected cells as opposed to cells overexpressing Env. While overexpression would not be expected to alter the thermodynamic equilibrium of the different topological forms, it would increase the relative size of each population and might make the CTT-exposed topology more easily observed than it would be in virally-infected cells. As such, the use of single cysteine mutants, which have been demonstrated to be infectious, would allow for topological determination of the CTT under relevant physiological conditions. The difficulty of this technique would occur in

preparing purified virus in sufficient quantities away from cellular microvesicles that would obscure interpretation of CTT cysteine labeling.

6.3.2 Why is arginine so highly conserved in the CTT?

In earlier sections, speculations were proposed regarding why arginine is so highly conserved in the LLP domains, especially relative to lysine. This is the first question to come out of the current research, and perhaps the easiest to address. Earlier published studies addressed the impact of mutating two arginine residues in LLPs 1 and 2 to glutamate, which is a drastic mutation, altering the amino acid from a positive to a negative charge (Kalia et al., 2003, 2005). The question arising from the current studies is "what would happen if the highly conserved arginine residues were replaced with lysine?" As has been discussed earlier in section 6.1.1.b, lysine rarely replaces arginine in the CTT, especially at the highly conserved positions in the LLP regions. The natural experiments, then, are to replace the highly conserved arginine residues in the LLP regions with lysine by site-directed mutagenesis and determine the effects on virion Env incorporation, Env-mediated membrane fusion, and ultimately viral infectivity. Based on the current understanding of the impact of the CTT on Env function and the speculations of why arginine is conserved we would expect that lysine replacements would not impact levels of Env virion incorporation, as lysine substitutions would retain the amphipathic nature of the LLP regions and those positions are not known to play a potential role in MA interactions. It is more likely that substituting lysine for arginine, at least in LLP2, would have an impact on Env-mediated fusogenicity in the context of cell-cell fusion, since even glutamate replacement had no effect on viral infectivity (Kalia et

al., 2003). It will also be of interest to determine the impact of arginine to lysine substitutions in LLP1 as mutations to glutamate in this region led to decreased Env incorporation, decreased Env-mediated cell-cell fusion, and decreased viral infectivity (Kalia et al., 2003). An additional point of interest would be to examine the impact of lysine substitutions on the regulation of overall Env conformation, as glutamate substitutions in LLP2 resulted in alterations in Env conformation and viral neutralization sensitivity to conformationally-dependent monoclonal antibodies (Kalia et al., 2005). If the regulation of Env conformation is simply a matter of membrane interaction of the CTT sequences, then lysine substitution would not be expected to have a large impact as it should maintain the amphiphilic potential and the interfacial partitioning preference of the LLP regions. These proposed substitutions and experiments should provide valuable information on the functions impacting the observed conservation of arginine in the CTT.

6.3.3 Can the CTT be a target for small molecule drugs as HIV inhibitors?

As we have demonstrated, the CTT can be a target for antibody-mediated neutralization by inhibiting virus-cell fusion under the proper conditions, as demonstrated in Section 5.3 and discussed in Section 6.1.2.b. Additionally, LLP2 is transiently exposed to antibody binding during the cell-cell fusion process (Lu et al., 2008) while being membrane-associated both pre- and post-fusion (Viard et al., 2008). These findings make the CTT a potentially attractive target for small molecule drug development, and suggest that binding the exposed CTT sequences during the fusion process and preventing reassociation with the membrane could lead to a new class of virus

inhibitors. Fortunately, the CTT provides an amenable target in that it contains numerous highly conserved sequences and structural properties in the areas demonstrated to undergo transient exposure. For example, the Kennedy epitope contains a number of charged amino acids that are highly conserved across all clades/ groups, and LLP2 is structurally conserved and contains a number of arginine residues that are also highly conserved across all clades/groups, as demonstrated in Chapter 3.

The logical first step would be to screen existing chemical libraries for compounds that bind to peptides representing the Kennedy epitope and LLP2. While recombinant gp41 would obviously be a better option, thus far no one has demonstrated a method to purify gp41 in a functional state. SPR-based screening would be a good method for this due to its ability to provide information on the rates and affinity of the interaction. Small molecules that were determined to bind to either the Kennedy epitope or LLP2 could then be used as a starting point to either derive similar compounds with different functional groups, or more likely be used to identify existing compounds with structural or chemical similarity for further screening. Compounds with the highest affinities for the Kennedy epitope/LLP2 could then be used to examine their effect on viral infectivity. Here an adaptation of the TZM-bl assay used currently for antibody-mediated neutralization would seem to make the most sense, as the readout involves the production of luciferase upon virus infection, leading to a quick and easy screen for compounds that result in decreased luciferase production.

6.3.4 Do CTT sequences act as immunological decoys?

Since CTT-exposed Env is not apparently incorporated into budding virions, an idea that is attractive from a viral evolutionary standpoint is that it functions as an immunological decoy and "distracts" the immune response. As has been alluded to above, unpublished data from collaborators suggest that this may be the case, at least in the context of attempting to derive a MPER-directed immune response using DNA vaccinations. However, it will be important to determine the extent and level of immunogenicity of the CTT more thoroughly. In an effort to address this topic, DNA vaccinations of animals with chimeric influenza hemagglutinin (HA) ectodomain fused to HIV gp41 MSD-CTT sequences are currently being performed by another member of the lab. Serum from vaccinated animals will be tested for reactivity to HIV Env, with the expectation that any reactivity will by necessity be directed to the CTT, as HA and Env are not cross-reactive. If reactivity is demonstrated to HIV Env, the serum will be further evaluated for standard or post-attachment neutralization to determine any functional properties.

6.3.5 What is the measured potential for predicted transmembrane region insertion?

The predicted transmembrane regions determined using the biological hydrophobicity scale were used, along with the predicted ΔG^{app}, to derive a topological model for the CTT that is consistent with topological mapping and functional studies. It will be of great importance to measure the actual ΔG^{app} to determine to what extent the predicted

regions insert into the membrane as transmembrane regions. This would be a critical first step towards verifying the proposed CTT topological model (c.f. Figure 38).

6.4 CLOSING REMARKS

In summary, we present here data demonstrating that the topology of the CTT can be markedly different than that of a type I membrane protein, which is the predominant current model for Env CTT topology. At various times in the process of trying to understand the data and their implications, we have often found answers in the principles of the physical chemistry of protein-lipid interactions. As such, the focus of the project slowly but steadily veered from a virological perspective to one based in chemical principles, particularly with regards to the interpretation of the results. I strongly feel that this type of perspective shift, whether through a philosophical drilling-down (e.g. virologist understanding and utilizing physical chemistry, or vice-versa) on the part of existing research groups or greater cross-discipline research and understanding (e.g. close and long-standing collaborations between virologists and chemists/physicists), will be critical to uncovering what is quickly becoming a more complicated world of protein topology. And if structure dictates function (axiom alert), then the understanding of principles surrounding protein topology and its effects on the biology of the protein would be of benefit to virologists, and not relegate Env to being an interesting model system to study protein topology.

Finally, we believe the model proposed here provides a comprehensive description of the topology of the CTT that provides a theoretical explanation for the

observed differences in cellular and virion CTT topology and is consistent with data regarding functional interactions of the CTT with intracellular partners as well as experimentally-determined topology data. It will be with great interest that I continue to watch the evolution of the understanding of Env CTT topology, regardless of my future pursuits.

BIBLIOGRAPHY

Allan, J.S., Coligan, J.E., Barin, F., McLane, M.F., Sodroski, J.G., Rosen, C.A., Haseltine, W.A., Lee, T.H., Essex, M., 1985. Major glycoprotein antigens that induce antibodies in AIDS patients are encoded by HTLV-III. Science (New York, NY) 228, 1091-1094.

Aloia, R.C., Jensen, F.C., Curtain, C.C., Mobley, P.W., Gordon, L.M., 1988. Lipid composition and fluidity of the human immunodeficiency virus. Proceedings of the National Academy of Sciences of the United States of America 85, 900-904.

Aloia, R.C., Tian, H., Jensen, F.C., 1993. Lipid composition and fluidity of the human immunodeficiency virus envelope and host cell plasma membranes. Proceedings of the National Academy of Sciences of the United States of America 90, 5181-5185.

Anand, R., 2000. Probing the topology of the glutamate receptor GluR1 subunit using epitope-Tag insertions. Biochemical and biophysical research communications 276, 157-161.

Andersen, O.S., Koeppe, R.E., 2007. Bilayer thickness and membrane protein function: an energetic perspective. Annual review of biophysics and biomolecular structure 36, 107-130.

Balasubramaniam, M., Freed, E.O., 2011. New insights into HIV assembly and trafficking. Physiology (Bethesda, Md.) 26, 236-251.

Barré-Sinoussi, F., Chermann, J.C., Rey, F., Nugeyre, M.T., Chamaret, S., Gruest, J., Dauguet, C., Axler-Blin, C., Vézinet-Brun, F., Rouzioux, C., Rozenbaum, W., Montagnier, L., 1983. Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS). Science (New York, NY) 220, 868-871.

Beck, D.L., 2004. Health, wealth and the Chinese oedipus. Society 41, 64-68.

Bernsel, A., von Heijne, G., 2005. Improved membrane protein topology prediction by domain assignments. Protein science : a publication of the Protein Society 14, 1723-1728.

Bieniasz, P.D., 2006. Late budding domains and host proteins in enveloped virus release. Virology 344, 55-63.

Blobel, G., 1980. Intracellular protein topogenesis. Proceedings of the National Academy of Sciences of the United States of America 77, 1496-1500.

Bogdanov, M., Heacock, P.N., Dowhan, W., 2002. A polytopic membrane protein displays a reversible topology dependent on membrane lipid composition. The EMBO journal 21, 2107-2116.

Bogdanov, M., Xie, J., Heacock, P., Dowhan, W., 2008. To flip or not to flip: lipid-protein charge interactions are a determinant of final membrane protein topology. The Journal of cell biology 182, 925-935.

Bogdanov, M., Zhang, W., Xie, J., Dowhan, W., 2005. Transmembrane protein topology mapping by the substituted cysteine accessibility method (SCAM(TM)): application to lipid-specific membrane protein topogenesis. Methods (San Diego, Calif) 36, 148-171.

Boge, M., Wyss, S., Bonifacino, J.S., Thali, M., 1998. A membrane-proximal tyrosine-based signal mediates internalization of the HIV-1 envelope glycoprotein via interaction with the AP-2 clathrin adaptor. J Biol Chem 273, 15773-15778.

Booth, P.J., Templer, R.H., Meijberg, W., Allen, S.J., Curran, A.R., Lorch, M., 2001. In vitro studies of membrane protein folding. Critical reviews in biochemistry and molecular biology 36, 501-603.

Bosch, M.L., Earl, P.L., Fargnoli, K., Picciafuoco, S., Giombini, F., Wong-Staal, F., Franchini, G., 1989. Identification of the fusion peptide of primate immunodeficiency viruses. Science (New York, NY) 244, 694-697.

Brass, V., Pal, Z., Sapay, N., Deléage, G., Blum, H.E., Penin, F., Moradpour, D., 2007. Conserved determinants for membrane association of nonstructural protein 5A from hepatitis C virus and related viruses. Journal of virology 81, 2745-2757.

Brodsky, J.L., 2007. The protective and destructive roles played by molecular chaperones during ERAD (endoplasmic-reticulum-associated degradation). The Biochemical journal 404, 353-363.

Brügger, B., Glass, B., Haberkant, P., Leibrecht, I., Wieland, F.T., Kräusslich, H.-G., 2006. The HIV lipidome: a raft with an unusual composition. Proceedings of the National Academy of Sciences of the United States of America 103, 2641-2646.

Byeon, I.-J.L., Meng, X., Jung, J., Zhao, G., Yang, R., Ahn, J., Shi, J., Concel, J., Aiken, C., Zhang, P., Gronenborn, A.M., 2009. Structural convergence between Cryo-EM and NMR reveals intersubunit interactions critical for HIV-1 capsid function. Cell 139, 780-790.

Byland, R., Vance, P.J., Hoxie, J.A., Marsh, M., 2007. A conserved dileucine motif mediates clathrin and AP-2-dependent endocytosis of the HIV-1 envelope protein. Molecular biology of the cell 18, 414-425.

Caffrey, M., Cai, M., Kaufman, J., Stahl, S.J., Wingfield, P.T., Covell, D.G., Gronenborn, A.M., Clore, G.M., 1998. Three-dimensional solution structure of the 44 kDa ectodomain of SIV gp41. The EMBO journal 17, 4572-4584.

Cai, K., Klein-Seetharaman, J., Altenbach, C., Hubbell, W.L., Khorana, H.G., 2001. Probing the dark state tertiary structure in the cytoplasmic domain of rhodopsin: proximities between amino acids deduced from spontaneous disulfide bond formation between cysteine pairs engineered in cytoplasmic loops 1, 3, and 4. Biochemistry 40, 12479-12485.

Cai, K., Klein-Seetharaman, J., Farrens, D., Zhang, C., Altenbach, C., Hubbell, W.L., Khorana, H.G., 1999. Single-cysteine substitution mutants at amino acid positions 306-321 in rhodopsin, the sequence between the cytoplasmic end of helix VII and the palmitoylation sites: sulfhydryl reactivity and transducin activation reveal a tertiary structure. Biochemistry 38, 7925-7930.

Campelo, F., McMahon, H.T., Kozlov, M.M., 2008. The hydrophobic insertion mechanism of membrane curvature generation by proteins. Biophysical journal 95, 2325-2339.

Cantor, R.S., 1997. Lateral Pressures in Cell Membranes: A Mechanism for Modulation of Protein Function. The journal of physical chemistry B 101, 1723-1725.

Cantor, R.S., 1999a. The influence of membrane lateral pressures on simple geometric models of protein conformational equilibria. Chemistry and physics of lipids 101, 45-56.

Cantor, R.S., 1999b. Lipid composition and the lateral pressure profile in bilayers. Biophysical journal 76, 2625-2639.

Cantor, R.S., 2002. Size distribution of barrel-stave aggregates of membrane peptides: influence of the bilayer lateral pressure profile. Biophysical journal 82, 2520-2525.

Cao, G., Cheng, S., Whitley, P., von Heijne, G., Kuhn, A., Dalbey, R.E., 1994. Synergistic insertion of two hydrophobic regions drives Sec-independent membrane protein assembly. The Journal of biological chemistry 269, 26898-26903.

Chakrabarti, L., Emerman, M., Tiollais, P., Sonigo, P., 1989. The cytoplasmic domain of simian immunodeficiency virus transmembrane protein modulates infectivity. Journal of virology 63, 4395-4403.

Chan, D.C., Fass, D., Berger, J.M., Kim, P.S., 1997. Core structure of gp41 from the HIV envelope glycoprotein. Cell 89, 263-273.

Chan, W.-E., Lin, H.-H., Chen, S.S.-L., 2005. Wild-type-like viral replication potential of human immunodeficiency virus type 1 envelope mutants lacking palmitoylation signals. Journal of virology 79, 8374-8387.

Chang, M.I., Panorchan, P., Dobrowsky, T.M., Tseng, Y., Wirtz, D., 2005. Single-molecule analysis of human immunodeficiency virus type 1 gp120-receptor interactions in living cells. Journal of virology 79, 14748-14755.

Chanh, T.C., Dreesman, G.R., Kanda, P., Linette, G.P., Sparrow, J.T., Ho, D.D., Kennedy, R.C., 1986. Induction of anti-HIV neutralizing antibodies by synthetic peptides. The EMBO journal 5, 3065-3071.

Checkley, M.A., Luttge, B.G., Freed, E.O., 2011. HIV-1 Envelope Glycoprotein Biosynthesis, Trafficking, and Incorporation. Journal of molecular biology 410, 582-608.

Chernomordik, L., Chanturiya, A.N., Suss-Toby, E., Nora, E., Zimmerberg, J., 1994. An amphipathic peptide from the C-terminal region of the human immunodeficiency virus envelope glycoprotein causes pore formation in membranes. Journal of virology 68, 7115-7123.

Chernomordik, L.V., Zimmerberg, J., Kozlov, M.M., 2006. Membranes of the world unite! The Journal of cell biology 175, 201-207.

Cheung, L., McLain, L., Hollier, M.J., Reading, S.A., Dimmock, N.J., 2005. Part of the C-terminal tail of the envelope gp41 transmembrane glycoprotein of human immunodeficiency virus type 1 is exposed on the surface of infected cells and is involved in virus-mediated cell fusion. The Journal of general virology 86, 131-138.

Chohan, B., Lang, D., Sagar, M., Korber, B., Lavreys, L., Richardson, B., Overbaugh, J., 2005. Selection for human immunodeficiency virus type 1 envelope glycosylation variants with shorter V1-V2 loop sequences occurs during transmission of certain genetic subtypes and may impact viral RNA levels. Journal of virology 79, 6528-6531.

Cleveland, S.M., McLain, L., Cheung, L., Jones, T.D., Hollier, M., Dimmock, N.J., 2003. A region of the C-terminal tail of the gp41 envelope glycoprotein of human immunodeficiency virus type 1 contains a neutralizing epitope: evidence for its exposure on the surface of the virion. The Journal of general virology 84, 591-602.

Cocquerel, L., Op de Beeck, A., Lambot, M., Roussel, J., Delgrange, D., Pillez, A., Wychowski, C., Penin, F., Dubuisson, J., 2002. Topological changes in the

transmembrane domains of hepatitis C virus envelope glycoproteins. The EMBO journal 21, 2893-2902.

Combet, C., Blanchet, C., Geourjon, C., Deléage, G., 2000. NPS@: network protein sequence analysis. Trends in biochemical sciences 25, 147-150.

Craigo, J.K., Zhang, B., Barnes, S., Tagmyer, T.L., Cook, S.J., Issel, C.J., Montelaro, R.C., 2007. Envelope variation as a primary determinant of lentiviral vaccine efficacy. Proceedings of the National Academy of Sciences of the United States of America 104, 15105-15110.

Das, S., Hahn, Y., Walker, D.A., Nagata, S., Willingham, M.C., Peehl, D.M., Bera, T.K., Lee, B., Pastan, I., 2008. Topology of NGEP, a prostate-specific cell:cell junction protein widely expressed in many cancers of different grade level. Cancer research 68, 6306-6312.

Demirov, D.G., Freed, E.O., 2004. Retrovirus budding. Virus research 106, 87-102.

Dorairaj, S., Allen, T.W., 2007. On the thermodynamic stability of a charged arginine side chain in a transmembrane helix. Proceedings of the National Academy of Sciences of the United States of America 104, 4943-4948.

Dougherty, D.A., 1996. Cation-pi interactions in chemistry and biology: a new view of benzene, Phe, Tyr, and Trp. Science (New York, NY) 271, 163-168.

Dowhan, W., 1997. Molecular basis for membrane phospholipid diversity: why are there so many lipids? Annual review of biochemistry 66, 199-232.

Dowhan, W., Bogdanov, M., 2009. Lipid-Dependent Membrane Protein Topogenesis. Annual review of biochemistry 78, 515-540.

Dowhan, W., Bogdanov, M., 2011. Lipid-protein interactions as determinants of membrane protein structure and function. Biochemical Society transactions 39, 767-774.

Dragic, T., Trkola, A., Thompson, D.A., Cormier, E.G., Kajumo, F.A., Maxwell, E., Lin, S.W., Ying, W., Smith, S.O., Sakmar, T.P., Moore, J.P., 2000. A binding pocket for a small molecule inhibitor of HIV-1 entry within the transmembrane helices of CCR5. Proceedings of the National Academy of Sciences of the United States of America 97, 5639-5644.

Drummond, A., Ashton, B., Buxton, S., Cheung, M., Cooper, A., Duran, C., Field, M., Heled, J., Kearse, M., Markowitz, S., Moir, R., Stones-Havas, S., Sturrock, S., Thierer, T., Wilson, A., Geneious, http://www.geneious.com/, 5.0.4 ed.

Dubay, J.W., Roberts, S.J., Brody, B., Hunter, E., 1992. Mutations in the leucine zipper of the human immunodeficiency virus type 1 transmembrane glycoprotein affect fusion and infectivity. Journal of virology 66, 4748-4756.

Earl, P.L., Doms, R.W., Moss, B., 1990. Oligomeric structure of the human immunodeficiency virus type 1 envelope glycoprotein. Proceedings of the National Academy of Sciences of the United States of America 87, 648-652.

Earl, P.L., Koenig, S., Moss, B., 1991. Biological and immunological properties of human immunodeficiency virus type 1 envelope glycoprotein: analysis of proteins with truncations and deletions expressed by recombinant vaccinia viruses. Journal of virology 65, 31-41.

Edinger, A.L., Mankowski, J.L., Doranz, B.J., Margulies, B.J., Lee, B., Rucker, J., Sharron, M., Hoffman, T.L., Berson, J.F., Zink, M.C., Hirsch, V.M., Clements, J.E., Doms, R.W., 1997. CD4-independent, CCR5-dependent infection of brain capillary endothelial cells by a neurovirulent simian immunodeficiency virus strain. Proceedings of the National Academy of Sciences of the United States of America 94, 14742-14747.

Edwards, T.G., Wyss, S., Reeves, J.D., Zolla-Pazner, S., Hoxie, J.A., Doms, R.W., Baribaud, F., 2002. Truncation of the cytoplasmic domain induces exposure of conserved regions in the ectodomain of human immunodeficiency virus type 1 envelope protein. Journal of virology 76, 2683-2691.

Eisenberg, D., Weiss, R.M., Terwilliger, T.C., 1984. The hydrophobic moment detects periodicity in protein hydrophobicity. Proceedings of the National Academy of Sciences of the United States of America 81, 140-144.

Eisenberg, D., Wesson, M., 1990. The most highly amphiphilic alpha-helices include two amino acid segments in human immunodeficiency virus glycoprotein 41. Biopolymers 29, 171-177.

Elofsson, A., von Heijne, G., 2007. Membrane protein structure: prediction versus reality. Annual review of biochemistry 76, 125-140.

Engelman, D.M., 2005. Membranes are more mosaic than fluid. Nature 438, 578-580.

Engelman, D.M., Steitz, T.A., 1981. The spontaneous insertion of proteins into and across membranes: the helical hairpin hypothesis. Cell 23, 411-422.

Epand, R.M., 2008. Proteins and cholesterol-rich domains. Biochimica et biophysica acta 1778, 1576-1582.

Evans, D.T., Desrosiers, R.C., 2001. Immune evasion strategies of the primate lentiviruses. Immunological reviews 183, 141-158.

Fetzer, C., Tews, B.A., Meyers, G., 2005. The carboxy-terminal sequence of the pestivirus glycoprotein E(rns) represents an unusual type of membrane anchor. Journal of virology 79, 11901-11913.

Finzi, A., Orthwein, A., Mercier, J., Cohen, E.A., 2007. Productive human immunodeficiency virus type 1 assembly takes place at the plasma membrane. Journal of virology 81, 7476-7490.

Freed, E.O., Delwart, E.L., Buchschacher, G.L., Panganiban, A.T., 1992. A mutation in the human immunodeficiency virus type 1 transmembrane glycoprotein gp41 dominantly interferes with fusion and infectivity. Proceedings of the National Academy of Sciences of the United States of America 89, 70-74.

Freed, E.O., Martin, M.A., 1995. Virion incorporation of envelope glycoproteins with long but not short cytoplasmic tails is blocked by specific, single amino acid substitutions in the human immunodeficiency virus type 1 matrix. Journal of virology 69, 1984-1989.

Freed, E.O., Martin, M.A., 1996. Domains of the human immunodeficiency virus type 1 matrix and gp41 cytoplasmic tail required for envelope incorporation into virions. Journal of virology 70, 341-351.

Freed, E.O., Myers, D.J., Risser, R., 1989. Mutational analysis of the cleavage sequence of the human immunodeficiency virus type 1 envelope glycoprotein precursor gp160. Journal of virology 63, 4670-4675.

Freed, E.O., Myers, D.J., Risser, R., 1990. Characterization of the fusion domain of the human immunodeficiency virus type 1 envelope glycoprotein gp41. Proceedings of the National Academy of Sciences of the United States of America 87, 4650-4654.

Fujii, G., Horvath, S., Woodward, S., Eiserling, F., Eisenberg, D., 1992. A molecular model for membrane fusion based on solution studies of an amphiphilic peptide from HIV gp41. Protein science : a publication of the Protein Society 1, 1454-1464.

Futaki, S., 2005. Membrane-permeable arginine-rich peptides and the translocation mechanisms. Advanced drug delivery reviews 57, 547-558.

Gallo, R.C., Salahuddin, S.Z., Popovic, M., Shearer, G.M., Kaplan, M., Haynes, B.F., Palker, T.J., Redfield, R., Oleske, J., Safai, B., 1984. Frequent detection and isolation of cytopathic retroviruses (HTLV-III) from patients with AIDS and at risk for AIDS. Science (New York, NY) 224, 500-503.

Gallo, S.A., Finnegan, C.M., Viard, M., Raviv, Y., Dimitrov, A., Rawat, S.S., Puri, A., Durell, S., Blumenthal, R., 2003. The HIV Env-mediated fusion reaction. Biochimica et biophysica acta 1614, 36-50.

Gangupomu, V.K., Abrams, C.F., 2010. All-Atom Models of the Membrane-Spanning Domain of HIV-1 gp41 from Metadynamics. Biophysical journal 99, 3438-3444.

Ganser-Pornillos, B.K., Yeager, M., Sundquist, W.I., 2008. The structural biology of HIV assembly. Current opinion in structural biology 18, 203-217.

Glickman, J., Croen, K., Kelly, S., Al-Awqati, Q., 1983. Golgi membranes contain an electrogenic H+ pump in parallel to a chloride conductance. The Journal of cell biology 97, 1303-1308.

Goder, V., Junne, T., Spiess, M., 2004. Sec61p contributes to signal sequence orientation according to the positive-inside rule. Molecular biology of the cell 15, 1470-1478.

Goder, V., Spiess, M., 2001. Topogenesis of membrane proteins: determinants and dynamics. FEBS letters 504, 87-93.

Goder, V., Spiess, M., 2003. Molecular mechanism of signal sequence orientation in the endoplasmic reticulum. The EMBO journal 22, 3645-3653.

Golding, H., Zaitseva, M., de Rosny, E., King, L.R., Manischewitz, J., Sidorov, I., Gorny, M.K., Zolla-Pazner, S., Dimitrov, D.S., Weiss, C.D., 2002. Dissection of human immunodeficiency virus type 1 entry with neutralizing antibodies to gp41 fusion intermediates. Journal of virology 76, 6780-6790.

Graschopf, A., Bläsi, U., 1999. Molecular function of the dual-start motif in the lambda S holin. Molecular microbiology 33, 569-582.

Greaves, J., Chamberlain, L.H., 2007. Palmitoylation-dependent protein sorting. The Journal of cell biology 176, 249-254.

Greenwood, A.I., Pan, J., Mills, T.T., Nagle, J.F., Epand, R.M., Tristram-Nagle, S., 2008. CRAC motif peptide of the HIV-1 gp41 protein thins SOPC membranes and interacts with cholesterol. Biochimica et biophysica acta 1778, 1120-1130.

Griffith, D.A., Delipala, C., Leadsham, J., Jarvis, S.M., Oesterhelt, D., 2003. A novel yeast expression system for the overproduction of quality-controlled membrane proteins. FEBS letters 553, 45-50.

Haffar, O.K., Dowbenko, D.J., Berman, P.W., 1988. Topogenic analysis of the human immunodeficiency virus type 1 envelope glycoprotein, gp160, in microsomal membranes. The Journal of cell biology 107, 1677-1687.

Harley, C.A., Holt, J.A., Turner, R., Tipper, D.J., 1998. Transmembrane protein insertion orientation in yeast depends on the charge difference across transmembrane segments, their total hydrophobicity, and its distribution. The Journal of biological chemistry 273, 24963-24971.

Haynes, B.F., Alam, S.M., 2008. HIV-1 hides an Achilles' heel in virion lipids. Immunity 28, 10-12.

Heap, C.J., Reading, S.A., Dimmock, N.J., 2005. An antibody specific for the C-terminal tail of the gp41 transmembrane protein of human immunodeficiency virus type 1 mediates post-attachment neutralization, probably through inhibition of virus-cell fusion. The Journal of general virology 86, 1499-1507.

Henderson, R., Unwin, P.N., 1975. Three-dimensional model of purple membrane obtained by electron microscopy. Nature 257, 28-32.

Hessa, T., Kim, H., Bihlmaier, K., Lundin, C., Boekel, J., Andersson, H., Nilsson, I., White, S.H., von Heijne, G., 2005a. Recognition of transmembrane helices by the endoplasmic reticulum translocon. Nature 433, 377-381.

Hessa, T., Meindl-Beinker, N.M., Bernsel, A., Kim, H., Sato, Y., Lerch-Bader, M., Nilsson, I., White, S.H., von Heijne, G., 2007. Molecular code for transmembrane-helix recognition by the Sec61 translocon. Nature 450, 1026-1030.

Hessa, T., White, S.H., von Heijne, G., 2005b. Membrane insertion of a potassium-channel voltage sensor. Science (New York, NY) 307, 1427.

Hill, C.P., Worthylake, D., Bancroft, D.P., Christensen, A.M., Sundquist, W.I., 1996. Crystal structures of the trimeric human immunodeficiency virus type 1 matrix protein: implications for membrane association and assembly, PNAS 93, 3099-3104.

Hirsch, V.M., Edmondson, P., Murphey-Corb, M., Arbeille, B., Johnson, P.R., Mullins, J.I., 1989. SIV adaptation to human cells. Nature 341, 573-574.

Ho, D.D., Sarngadharan, M.G., Hirsch, M.S., Schooley, R.T., Rota, T.R., Kennedy, R.C., Chanh, T.C., Sato, V.L., 1987. Human immunodeficiency virus neutralizing antibodies recognize several conserved domains on the envelope glycoproteins. Journal of virology 61, 2024-2028.

Hollier, M.J., Dimmock, N.J., 2005. The C-terminal tail of the gp41 transmembrane envelope glycoprotein of HIV-1 clades A, B, C, and D may exist in two conformations: an analysis of sequence, structure, and function. Virology 337, 284-296.

Huang, C.-c., Tang, M., Zhang, M.-Y., Majeed, S., Montabana, E., Stanfield, R.L., Dimitrov, D.S., Korber, B., Sodroski, J., Wilson, I.A., Wyatt, R., Kwong, P.D., 2005. Structure of a V3-containing HIV-1 gp120 core. Science (New York, NY) 310, 1025-1028.

Humphrey, W., 1996. VMD: Visual molecular dynamics. Journal of Molecular Graphics 14, 33-38.

Inomata, K., Ohno, A., Tochio, H., Isogai, S., Tenno, T., Nakase, I., Takeuchi, T., Futaki, S., Ito, Y., Hiroaki, H., Shirakawa, M., 2009. High-resolution multi-dimensional NMR spectroscopy of proteins in human cells. Nature 458, 106-109.

Jiang, Aiken, 2007. Maturation-Dependent HIV-1 Particle Fusion Requires a Carboxyl-Terminal Region of the gp41 Cytoplasmic Tail. Journal of virology.

Jiang, Y., Ruta, V., Chen, J., Lee, A., MacKinnon, R., 2003. The principle of gating charge movement in a voltage-dependent K+ channel. Nature 423, 42-48.

Jing, W., Demcoe, A.R., Vogel, H.J., 2003. Conformation of a bactericidal domain of puroindoline a: structure and mechanism of action of a 13-residue antimicrobial peptide. Journal of bacteriology 185, 4938-4947.

Jones, D.T., 1999. Protein secondary structure prediction based on position-specific scoring matrices. Journal of molecular biology 292, 195-202.

Jouvenet, N., Bieniasz, P.D., Simon, S.M., 2008. Imaging the biogenesis of individual HIV-1 virions in live cells. Nature 454, 236-240.

Joyner, A.S., Willis, J.R., Crowe, J.E., Aiken, C., 2011. Maturation-Induced Cloaking of Neutralization Epitopes on HIV-1 Particles. PLoS pathogens 7, e1002234.

Kabsch, W., 1976. A solution for the best rotation to relate two sets of vectors. Acta Crystallographica Section A 32, 922-923.

Kabsch, W., 1978. A discussion of the solution for the best rotation to relate two sets of vectors. Acta Crystallographica Section A 34, 827-828.

Kalia, V., Sarkar, S., Gupta, P., Montelaro, R.C., 2003. Rational site-directed mutations of the LLP-1 and LLP-2 lentivirus lytic peptide domains in the intracytoplasmic tail of human immunodeficiency virus type 1 gp41 indicate common functions in cell-cell fusion but distinct roles in virion envelope incorporation. Journal of virology 77, 3634-3646.

Kalia, V., Sarkar, S., Gupta, P., Montelaro, R.C., 2005. Antibody neutralization escape mediated by point mutations in the intracytoplasmic tail of human immunodeficiency virus type 1 gp41. Journal of virology 79, 2097-2107.

Karlsson Hedestam, G.B., Fouchier, R.A.M., Phogat, S., Burton, D.R., Sodroski, J., Wyatt, R.T., 2008. The challenges of eliciting neutralizing antibodies to HIV-1 and to influenza virus. Nature reviews Microbiology 6, 143-155.

Kaufman, R.J., Scheuner, D., Schröder, M., Shen, X., Lee, K., Liu, C.Y., Arnold, S.M., 2002. The unfolded protein response in nutrient sensing and differentiation. Nature reviews Molecular cell biology 3, 411-421.

Kauko, A., Hedin, L.E., Thebaud, E., Cristobal, S., Elofsson, A., von Heijne, G., 2010. Repositioning of Transmembrane α-Helices during Membrane Protein Folding. Journal of Molecular Biology 397, 190-201.

Kennedy, R.C., Henkel, R.D., Pauletti, D., Allan, J.S., Lee, T.H., Essex, M., Dreesman, G.R., 1986. Antiserum to a synthetic peptide recognizes the HTLV-III envelope glycoprotein. Science (New York, NY) 231, 1556-1559.

Khorana, H.G., Gerber, G.E., Herlihy, W.C., Gray, C.P., Anderegg, R.J., Nihei, K., Biemann, K., 1979. Amino acid sequence of bacteriorhodopsin. Proceedings of the National Academy of Sciences of the United States of America 76, 5046-5050.

Killian, J.A., von Heijne, G., 2000. How proteins adapt to a membrane-water interface. Trends in biochemical sciences 25, 429-434.

Kim, J.H., Hartley, T.L., Curran, A.R., Engelman, D.M., 2009. Molecular dynamics studies of the transmembrane domain of gp41 from HIV-1. Biochimica et biophysica acta 1788, 1804-1812.

Kitrinos, K.M., Hoffman, N.G., Nelson, J.A.E., Swanstrom, R., 2003. Turnover of env variable region 1 and 2 genotypes in subjects with late-stage human immunodeficiency virus type 1 infection. Journal of virology 77, 6811-6822.

Klein-Seetharaman, J., Hwa, J., Cai, K., Altenbach, C., Hubbell, W.L., Khorana, H.G., 1999. Single-cysteine substitution mutants at amino acid positions 55-75, the sequence connecting the cytoplasmic ends of helices I and II in rhodopsin: reactivity of the sulfhydryl groups and their derivatives identifies a tertiary structure that changes upon light-activation. Biochemistry 38, 7938-7944.

Klein-Seetharaman, J., Hwa, J., Cai, K., Altenbach, C., Hubbell, W.L., Khorana, H.G., 2001. Probing the dark state tertiary structure in the cytoplasmic domain of rhodopsin: proximities between amino acids deduced from spontaneous disulfide bond formation between Cys316 and engineered cysteines in cytoplasmic loop 1. Biochemistry 40, 12472-12478.

Kliger, Y., Shai, Y., 1997. A leucine zipper-like sequence from the cytoplasmic tail of the HIV-1 envelope glycoprotein binds and perturbs lipid bilayers. Biochemistry 36, 5157-5169.

Kodama, T., Wooley, D.P., Naidu, Y.M., Kestler, H.W., Daniel, M.D., Li, Y., Desrosiers, R.C., 1989. Significance of premature stop codons in env of simian immunodeficiency virus. Journal of virology 63, 4709-4714.

Kol, N., Shi, Y., Tsvitov, M., Barlam, D., Shneck, R.Z., Kay, M.S., Rousso, I., 2007. A stiffness switch in human immunodeficiency virus. Biophys J 92, 1777-1783.

Kondo, N., Miyauchi, K., Meng, F., Iwamoto, A., Matsuda, Z., 2010. Conformational changes of the HIV-1 envelope protein during membrane fusion are inhibited by the replacement of its membrane-spanning domain. The Journal of biological chemistry 285, 14681-14688.

Korber, B., Gaschen, B., Yusim, K., Thakallapally, R., Kesmir, C., Detours, V., 2001. Evolutionary and immunological implications of contemporary HIV-1 variation. British medical bulletin 58, 19-42.

Kotov, A., Zhou, J., Flicker, P., Aiken, C., 1999. Association of Nef with the human immunodeficiency virus type 1 core. Journal of virology 73, 8824-8830.

Krepkiy, D., Mihailescu, M., Freites, J.A., Schow, E.V., Worcester, D.L., Gawrisch, K., Tobias, D.J., White, S.H., Swartz, K.J., 2009. Structure and hydration of membranes embedded with voltage-sensing domains. Nature 462, 473-479.

Krishnakumar, S.S., London, E., 2007. The control of transmembrane helix transverse position in membranes by hydrophilic residues. Journal of molecular biology 374, 1251-1269.

Kwong, P.D., Wyatt, R., Robinson, J., Sweet, R.W., Sodroski, J., Hendrickson, W.A., 1998. Structure of an HIV gp120 envelope glycoprotein in complex with the CD4 receptor and a neutralizing human antibody. Nature 393, 648-659.

Kyte, J., Doolittle, R.F., 1982. A simple method for displaying the hydropathic character of a protein. Journal of molecular biology 157, 105-132.

Lambert, C., Mann, S., Prange, R., 2004. Assessment of determinants affecting the dual topology of hepadnaviral large envelope proteins. The Journal of general virology 85, 1221-1225.

Lambert, C., Prange, R., 2001. Dual topology of the hepatitis B virus large envelope protein: determinants influencing post-translational pre-S translocation. The Journal of biological chemistry 276, 22265-22272.

Lambert, C., Prange, R., 2003. Chaperone action in the posttranslational topological reorientation of the hepatitis B virus large envelope protein: Implications for translocational regulation. Proceedings of the National Academy of Sciences of the United States of America 100, 5199-5204.

Lee, A.G., 2004. How lipids affect the activities of integral membrane proteins. Biochimica et biophysica acta 1666, 62-87.

Leonard, C.K., Spellman, M.W., Riddle, L., Harris, R.J., Thomas, J.N., Gregory, T.J., 1990. Assignment of intrachain disulfide bonds and characterization of potential glycosylation sites of the type 1 recombinant human immunodeficiency virus envelope glycoprotein (gp120) expressed in Chinese hamster ovary cells. The Journal of biological chemistry 265, 10373-10382.

Levy, J.A., 2009. HIV pathogenesis: 25 years of progress and persistent challenges. AIDS (London, England) 23, 147-160.

Liu, J., Bartesaghi, A., Borgnia, M., Sapiro, G., Subramaniam, S., 2008. Molecular architecture of native HIV-1 gp120 trimers. Nature.

Liu, S., Kondo, N., Long, Y., Xiao, D., Iwamoto, A., Matsuda, Z., 2010. Membrane topology analysis of HIV-1 envelope glycoprotein gp41. Retrovirology 7, 100.

Llopis, J., McCaffery, J.M., Miyawaki, A., Farquhar, M.G., Tsien, R.Y., 1998. Measurement of cytosolic, mitochondrial, and Golgi pH in single living cells with green fluorescent proteins. Proceedings of the National Academy of Sciences of the United States of America 95, 6803-6808.

Lopez-Vergès, S., Camus, G., Blot, G., Beauvoir, R., Benarous, R., Berlioz-Torrent, C., 2006. Tail-interacting protein TIP47 is a connector between Gag and Env and is required for Env incorporation into HIV-1 virions. Proceedings of the National Academy of Sciences of the United States of America 103, 14947-14952.

Lu, L., Zhu, Y., Huang, J., Chen, X., Yang, H., Jiang, S., Chen, Y.-H., 2008. Surface exposure of the HIV-1 env cytoplasmic tail LLP2 domain during the membrane fusion process: interaction with gp41 fusion core. The Journal of biological chemistry 283, 16723-16731.

Lu, M., Blacklow, S.C., Kim, P.S., 1995. A trimeric structural domain of the HIV-1 transmembrane glycoprotein. Nature structural biology 2, 1075-1082.

Luciw, P.A., Fields, B.N., Knipe, D.M., Howley, P.M., 2002. Fields' Virology, Human Immunodeficiency Viruses and Their Replication, 3 ed. Lippincott-Raven, Philadelphia, pp. 1881-1952.

Ma, J., Dougherty, D., 1997. The Cation-π Interaction. Chemical reviews 97, 1303-1324.

MacCallum, J.L., Bennett, W.F.D., Tieleman, D.P., 2007. Partitioning of amino acid side chains into lipid bilayers: results from computer simulations and comparison to experiment. The Journal of general physiology 129, 371-377.

MacCallum, J.L., Bennett, W.F.D., Tieleman, D.P., 2008. Distribution of amino acids in a lipid bilayer from computer simulations. Biophysical journal 94, 3393-3404.

Marsh, D., 2007. Lateral pressure profile, spontaneous curvature frustration, and the incorporation and conformation of proteins in membranes. Biophysical journal 93, 3884-3899.

Masciotra, S., Owen, S.M., Rudolph, D., Yang, C., Wang, B., Saksena, N., Spira, T., Dhawan, S., Lal, R.B., 2002. Temporal relationship between V1V2 variation,

macrophage replication, and coreceptor adaptation during HIV-1 disease progression. AIDS (London, England) 16, 1887-1898.

Massiah, M., Starich, M., Paschall, C., 1994. Three-dimensional Structure of the Human Immunodeficiency Virus Type 1 Matrix Protein. Journal of molecular ....

Maupetit, J., Derreumaux, P., Tuffery, P., 2009. PEP-FOLD: an online resource for de novo peptide structure prediction. Nucleic acids research 37, W498-503.

Maupetit, J., Derreumaux, P., Tufféry, P., 2010. A fast method for large-scale De Novo peptide and miniprotein structure prediction. Journal of computational chemistry 31, 726-738.

Maupetit, J., Tuffery, P., Derreumaux, P., 2007. A coarse-grained protein force field for folding and structure prediction. Proteins 69, 394-408.

McCune, J.M., Rabin, L.B., Feinberg, M.B., Lieberman, M., Kosek, J.C., Reyes, G.R., Weissman, I.L., 1988. Endoproteolytic cleavage of gp160 is required for the activation of human immunodeficiency virus. Cell 53, 55-67.

McElrath, M.J., Haynes, B.F., 2010. Induction of immunity to human immunodeficiency virus type-1 by vaccination. Immunity 33, 542-554.

Meusser, B., Hirsch, C., Jarosch, E., Sommer, T., 2005. ERAD: the long road to destruction. Nature cell biology 7, 766-772.

Miller, M.A., Cloyd, M.W., Liebmann, J., Rinaldo, C.R., Islam, K.R., Wang, S.Z., Mietzner, T.A., Montelaro, R.C., 1993. Alterations in cell membrane permeability by the lentivirus lytic peptide (LLP-1) of HIV-1 transmembrane protein. Virology 196, 89-100.

Miller, M.A., Garry, R.F., Jaynes, J.M., Montelaro, R.C., 1991. A Structural Correlation Between Lentivirus Transmembrane Proteins and Natural Cytolytic Peptides. AIDS Research and Human Retroviruses 7, 511-519.

Mitchell, D.J., Kim, D.T., Steinman, L., Fathman, C.G., Rothbard, J.B., 2000. Polyarginine enters cells more efficiently than other polycationic homopolymers. The journal of peptide research : official journal of the American Peptide Society 56, 318-325.

Miyauchi, K., Komano, J., Yokomaku, Y., Sugiura, W., Yamamoto, N., Matsuda, Z., 2005. Role of the specific amino acid sequence of the membrane-spanning domain of human immunodeficiency virus type 1 in membrane fusion. Journal of virology 79, 4720-4729.

Montefiori, D.C., Robinson, W.E., Mitchell, W.M., 1988. Role of protein N-glycosylation in pathogenesis of human immunodeficiency virus type 1. Proceedings of the National Academy of Sciences of the United States of America 85, 9248-9252.

Morariu, V.K., Daphne, Srinivasan, B.S., Dale, Raykar, V.B., Yoshua, Duraiswami, R.B., Leon, Davis, L., 2008. Automatic online tuning for fast Gaussian summation. Advances in Neural ..., Boston, MA, pp. 1113-1120.

Morita, E., Sundquist, W.I., 2004. Retrovirus budding. Annual review of cell and developmental biology 20, 395-425.

Muñoz-Barroso, I., Salzwedel, K., Hunter, E., Blumenthal, R., 1999. Role of the membrane-proximal domain in the initial stages of human immunodeficiency virus type 1 envelope glycoprotein-mediated membrane fusion. Journal of virology 73, 6089-6092.

Murakami, T., 2008. Roles of the Interactions Between Env and Gag Proteins in the HIV-1 Replication Cycle. Microbiol Immunol 52, 287-295.

Murakami, T., Freed, E.O., 2000a. Genetic evidence for an interaction between human immunodeficiency virus type 1 matrix and alpha-helix 2 of the gp41 cytoplasmic tail. Journal of virology 74, 3548-3554.

Murakami, T., Freed, E.O., 2000b. The long cytoplasmic tail of gp41 is required in a cell type-dependent manner for HIV-1 envelope glycoprotein incorporation into virions. Proceedings of the National Academy of Sciences of the United States of America 97, 343-348.

Myszka, D.G., 1999. Improving biosensor analysis. Journal of molecular recognition : JMR 12, 279-284.

Myszka, D.G., Sweet, R.W., Hensley, P., Brigham-Burke, M., Kwong, P.D., Hendrickson, W.A., Wyatt, R., Sodroski, J., Doyle, M.L., 2000. Energetics of the HIV gp120- CD4 binding reaction. Proceedings of the National Academy of Sciences of the United States of America 97, 9026-9031.

Nakatsukasa, K., Huyer, G., Michaelis, S., Brodsky, J.L., 2008. Dissecting the ER-associated degradation of a misfolded polytopic membrane protein. Cell 132, 101-112.

Newman, J.T., Sturgeon, T.J., Gupta, P., Montelaro, R.C., 2007. Differential functional phenotypes of two primary HIV-1 strains resulting from homologous point mutations in the LLP domains of the envelope gp41 intracytoplasmic domain. Virology 367, 102-116.

Norholm, M.H., Shulga, Y.V., Aoki, S., Epand, R.M., von Heijne, G., 2011. Flanking residues help determine whether a hydrophobic segment adopts a monotopic or bitopic topology in the endoplasmic reticulum membrane. J Biol Chem 286, 25284-25290.

Obermeyer, T., Fraisl, P., DiRusso, C.C., Black, P.N., 2007. Topology of the yeast fatty acid transport protein Fat1p: mechanistic implications for functional domains on

the cytosolic surface of the plasma membrane. Journal of lipid research 48, 2354-2364.

Ohno, H., Aguilar, R.C., Fournier, M.C., Hennecke, S., Cosson, P., Bonifacino, J.S., 1997. Interaction of endocytic signals from the HIV-1 envelope glycoprotein complex with members of the adaptor medium chain family. Virology 238, 305-315.

Ovchinnikov, Y.A., Abdulaev, N.G., Feigina, M.Y., Kiselev, A.V., Lobanov, N.A., 1977. Recent findings in the structure-functional characteristics of bacteriorhodopsin. FEBS letters 84, 1-4.

Ovchinnikov, Y.A., Abdulaev, N.G., Feigina, M.Y., Kiselev, A.V., Lobanov, N.A., 1979. The structural basis of the functioning of bacteriorhodopsin: an overview. FEBS letters 100, 219-224.

Palmer, C., Balfe, P., Fox, D., May, J.C., Frederiksson, R., Fenyö, E.M., McKeating, J.A., 1996. Functional characterization of the V1V2 region of human immunodeficiency virus type 1. Virology 220, 436-449.

Pan, J., Mills, T.T., Tristram-Nagle, S., Nagle, J.F., 2008. Cholesterol perturbs lipid bilayers nonuniversally. Physical review letters 100, 198103.

Pan, J., Tristram-Nagle, S., Nagle, J.F., 2009. Effect of cholesterol on structural and mechanical properties of membranes depends on lipid chain saturation. Physical review E, Statistical, nonlinear, and soft matter physics 80, 021931.

Parent, L., Wills, J., Resh, M., 1994. Identification of a membrane-binding domain within the amino-terminal region of human immunodeficiency virus type 1 Gag protein which interacts with acidic phospholipids. Journal of virology.

Phillips, R., Ursell, T., Wiggins, P., Sens, P., 2009. Emerging roles for lipids in shaping membrane-protein function. Nature 459, 379-385.

Pieniazek, D., Yang, C., Lal, R., 1998. Phylogenetic analysis of gp41 envelope of HIV-1 groups M, N, and O strains provides an alternate region for subtype determination, in: Korber, B., Foley, B., McCutchan, F., Mellors, J., Hahn, B., Sodroski, J., Kuiken, C. (Eds.), Human Retroviruses and AIDS. Los Alamos National Laboratory.

Pinter, A., Honnen, W.J., Tilley, S.A., Bona, C., Zaghouani, H., Gorny, M.K., Zolla-Pazner, S., 1989. Oligomeric structure of gp41, the transmembrane protein of human immunodeficiency virus type 1. Journal of virology 63, 2674-2679.

Platt, E.J., Bilska, M., Kozak, S.L., Kabat, D., Montefiori, D.C., 2009. Evidence that ecotropic murine leukemia virus contamination in TZM-bl cells does not affect the outcome of neutralizing antibody assays with human immunodeficiency virus type 1. Journal of virology 83, 8289-8292.

Platt, E.J., Wehrly, K., Kuhmann, S.E., Chesebro, B., Kabat, D., 1998. Effects of CCR5 and CD4 cell surface concentrations on infections by macrophagetropic isolates of human immunodeficiency virus type 1. Journal of virology 72, 2855-2864.

Pornillos, O., Ganser-Pornillos, B., Kelly, B., Hua, Y., Whitby, F., Sundquist, W., Hill, C., Stout, C., Yeager, M., 2009. ScienceDirect - Cell : X-Ray Structures of the Hexameric Building Block of the HIV Capsid. Cell.

Reading, S.A., Heap, C.J., Dimmock, N.J., 2003. A novel monoclonal antibody specific to the C-terminal tail of the gp41 envelope transmembrane protein of human immunodeficiency virus type 1 that preferentially neutralizes virus after it has attached to the target cell and inhibits the production of infectious progeny. Virology 315, 362-372.

Rossio, J.L., Esser, M.T., Suryanarayana, K., Schneider, D.K., Bess, J.W., Vasquez, G.M., Wiltrout, T.A., Chertova, E., Grimes, M.K., Sattentau, Q., Arthur, L.O., Henderson, L.E., Lifson, J.D., 1998. Inactivation of human immunodeficiency virus type 1 infectivity with preservation of conformational and functional integrity of virion surface proteins. Journal of virology 72, 7992-8001.

Rousso, I., Mixon, M.B., Chen, B.K., Kim, P.S., 2000. Palmitoylation of the HIV-1 envelope glycoprotein is critical for viral infectivity. Proc Natl Acad Sci U S A 97, 13523-13525.

Roux, K.H., Taylor, K.A., 2007. AIDS virus envelope spike structure. Current opinion in structural biology 17, 244-252.

Rowell, J.F., Stanhope, P.E., Siliciano, R.F., 1995. Endocytosis of endogenously synthesized HIV-1 envelope protein. Mechanism and role in processing for association with class II MHC. Journal of immunology (Baltimore, Md : 1950) 155, 473-488.

Rudnick, G., 1986. ATP-driven H+ pumping into intracellular organelles. Annual review of physiology 48, 403-413.

Rushlow, K., Olsen, K., Stiegler, G., Payne, S.L., Montelaro, R.C., Issel, C.J., 1986. Lentivirus genomic organization: the complete nucleotide sequence of the env gene region of equine infectious anemia virus. Virology 155, 309-321.

Russell, R., 1992. Multiple protein sequence alignment from tertiary structure comparison: Assignment of global and residue confidence levels - Russell - 2004 - Proteins: Structure, Function, and Bioinformatics - Wiley Online Library. Proteins: Structure 14, 309-323.

Sagar, M., Wu, X., Lee, S., Overbaugh, J., 2006. Human immunodeficiency virus type 1 V1-V2 envelope loop sequences expand and add glycosylation sites over the course of infection, and these modifications affect antibody neutralization sensitivity. Journal of virology 80, 9586-9598.

Salzwedel, K., West, J.T., Hunter, E., 1999. A conserved tryptophan-rich motif in the membrane-proximal region of the human immunodeficiency virus type 1 gp41 ectodomain is important for Env-mediated fusion and virus infectivity. Journal of virology 73, 2469-2480.

Sapay, N., Guermeur, Y., Deléage, G., 2006a. Prediction of amphipathic in-plane membrane anchors in monotopic proteins using a SVM classifier. BMC bioinformatics 7, 255.

Sapay, N., Montserret, R., Chipot, C., Brass, V., Moradpour, D., Deléage, G., Penin, F., 2006b. NMR structure and molecular dynamics of the in-plane membrane anchor of nonstructural protein 5A from bovine viral diarrhea virus. Biochemistry 45, 2221-2233.

Scarlatti, G., Tresoldi, E., Björndal, A., Fredriksson, R., Colognesi, C., Deng, H.K., Malnati, M.S., Plebani, A., Siccardi, A.G., Littman, D.R., Fenyö, E.M., Lusso, P., 1997. In vivo evolution of HIV-1 co-receptor usage and sensitivity to chemokine-mediated suppression. Nature medicine 3, 1259-1265.

Schibli, D.J., Montelaro, R.C., Vogel, H.J., 2001. The membrane-proximal tryptophan-rich region of the HIV glycoprotein, gp41, forms a well-defined helix in dodecylphosphocholine micelles. Biochemistry 40, 9570-9578.

Schulz, G.E., 2000. beta-Barrel membrane proteins. Current opinion in structural biology 10, 443-447.

Shacklett, B.L., Weber, C.J., Shaw, K.E., Keddie, E.M., Gardner, M.B., Sonigo, P., Luciw, P.A., 2000. The intracytoplasmic domain of the Env transmembrane protein is a locus for attenuation of simian immunodeficiency virus SIVmac in rhesus macaques. Journal of virology 74, 5836-5844.

Shang, L., Hunter, E., 2010. Residues in the membrane-spanning domain core modulate conformation and fusogenicity of the HIV-1 envelope glycoprotein. Virology 404, 158-167.

Shang, L., Yue, L., Hunter, E., 2008. Role of the membrane-spanning domain of human immunodeficiency virus type 1 envelope glycoprotein in cell-cell fusion and virus infection. Journal of virology 82, 5417-5428.

Shi, W., Bohon, J., Han, D.P., Habte, H., Qin, Y., Cho, M.W., Chance, M.R., 2010. Structural characterization of HIV gp41 with the membrane-proximal external region. The Journal of biological chemistry 285, 24290-24298.

Shioda, T., Oka, S., Xin, X., Liu, H., Harukuni, R., Kurotani, A., Fukushima, M., Hasan, M.K., Shiino, T., Takebe, Y., Iwamoto, A., Nagai, Y., 1997. In vivo sequence variability of human immunodeficiency virus type 1 envelope gp120: association of V2 extension with slow disease progression. Journal of virology 71, 4871-4881.

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

Sreerama, N., Woody, R.W., 2004. On the analysis of membrane protein circular dichroism spectra. Protein science : a publication of the Protein Society 13, 100-112.

Srinivas, S.K., Srinivas, R.V., Anantharamaiah, G.M., Segrest, J.P., Compans, R.W., 1992. Membrane interactions of synthetic peptides corresponding to amphipathic helical segments of the human immunodeficiency virus type-1 envelope glycoprotein. The Journal of biological chemistry 267, 7121-7127.

Starcich, B.R., Hahn, B.H., Shaw, G.M., McNeely, P.D., Modrow, S., Wolf, H., Parks, E.S., Parks, W.P., Josephs, S.F., Gallo, R.C., 1986. Identification and characterization of conserved and variable regions in the envelope gene of HTLV-III/LAV, the retrovirus of AIDS. Cell 45, 637-648.

Steckbeck, J.D., Craigo, J.K., Barnes, C.O., Montelaro, R.C., 2011. Highly conserved structural properties of the C-terminal tail of HIV-1 gp41 protein despite substantial sequence variation among diverse clades: implications for functions in viral replication. The Journal of biological chemistry 286, 27156-27166.

Steckbeck, J.D., Grieser, H.J., Sturgeon, T., Taber, R., Chow, A., Bruno, J., Murphy-Corb, M., Montelaro, R.C., Cole, K.S., 2006. Dynamic evolution of antibody populations in a rhesus macaque infected with attenuated simian immunodeficiency virus identified by surface plasmon resonance. Journal of medical primatology 35, 248-260.

Steckbeck, J.D., Orlov, I., Chow, A., Grieser, H., Miller, K., Bruno, J., Robinson, J.E., Montelaro, R.C., Cole, K.S., 2005. Kinetic rates of antibody binding correlate with neutralization sensitivity of variant simian immunodeficiency virus strains. Journal of virology 79, 12311-12320.

Steckbeck, J.D., Sun, C., Sturgeon, T.J., Montelaro, R.C., 2010. Topology of the C-terminal tail of HIV-1 gp41: differential exposure of the Kennedy epitope on cell and viral membranes. PloS one 5, e15261.

Sun, Z.-Y.J., Oh, K.J., Kim, M., Yu, J., Brusic, V., Song, L., Qiao, Z., Wang, J.-H., Wagner, G., Reinherz, E.L., 2008. HIV-1 broadly neutralizing antibody extracts its epitope from a kinked gp41 ectodomain region on the viral membrane. Immunity 28, 52-63.

Swofford, D.L., PAUP: Phylogenetic Analysis Using Parsimony and Other Methods. Sinauer Associates, Sunderlan, MA.

Tews, B.A., Meyers, G., 2007. The pestivirus glycoprotein Erns is anchored in plane in the membrane via an amphipathic helix. The Journal of biological chemistry 282, 32730-32741.

Tillman, T.S., Cascio, M., 2003. Effects of membrane lipids on ion channel structure and function. Cell biochemistry and biophysics 38, 161-190.

Tomita, M., Marchesi, V.T., 1975. Amino-acid sequence and oligosaccharide attachment sites of human erythrocyte glycophorin. Proceedings of the National Academy of Sciences of the United States of America 72, 2964-2968.

Tristram-Nagle, S., Chan, R., Kooijman, E., Uppamoochikkal, P., Qiang, W., Weliky, D.P., Nagle, J.F., 2010. HIV fusion peptide penetrates, disorders, and softens T-cell membrane mimics. Journal of molecular biology 402, 139-153.

Tristram-Nagle, S., Nagle, J.F., 2007. HIV-1 fusion peptide decreases bending energy and promotes curved fusion intermediates. Biophysical journal 93, 2048-2055.

Tung, C.-H., Weissleder, R., 2003. Arginine containing peptides as delivery vectors. Advanced drug delivery reviews 55, 281-294.

Ulmschneider, M.B., Sansom, M.S.P., Di Nola, A., 2005. Properties of integral membrane protein structures: derivation of an implicit membrane potential. Proteins 59, 252-265.

van den Brink-van der Laan, E., Killian, J.A., de Kruijff, B., 2004. Nonbilayer lipids affect peripheral and integral membrane proteins via changes in the lateral pressure profile. Biochimica et biophysica acta 1666, 275-288.

van Meer, G., Voelker, D.R., Feigenson, G.W., 2008. Membrane lipids: where they are and how they behave. Nature reviews Molecular cell biology 9, 112-124.

Viard, M., Ablan, S.D., Zhou, M., Veenstra, T.D., Freed, E.O., Raviv, Y., Blumenthal, R., 2008. Photoinduced reactivity of the HIV-1 envelope glycoprotein with a membrane-embedded probe reveals insertion of portions of the HIV-1 Gp41 cytoplasmic tail into the viral membrane. Biochemistry 47, 1977-1983.

von Heijne, G., 1986. The distribution of positively charged residues in bacterial inner membrane proteins correlates with the trans-membrane topology. The EMBO journal 5, 3021-3027.

von Heijne, G., 1992. Membrane protein structure prediction. Hydrophobicity analysis and the positive-inside rule. Journal of molecular biology 225, 487-494.

von Heijne, G., 1999. Recent advances in the understanding of membrane protein assembly and structure. Quarterly reviews of biophysics 32, 285-307.

von Heijne, G., 2006. Membrane-protein topology. Nature reviews Molecular cell biology 7, 909-918.

von Heijne, G., Blomberg, C., 1979. Trans-membrane translocation of proteins. The direct transfer model. European journal of biochemistry / FEBS 97, 175-181.

Wagner, S., Bader, M.L., Drew, D., de Gier, J.-W., 2006. Rationalizing membrane protein overexpression. Trends in biotechnology 24, 364-371.

Wang, X., Bogdanov, M., Dowhan, W., 2002. Topology of polytopic membrane protein subdomains is dictated by membrane phospholipid composition. The EMBO journal 21, 5673-5681.

Weissenhorn, W., Dessen, A., Harrison, S.C., Skehel, J.J., Wiley, D.C., 1997. Atomic structure of the ectodomain from HIV-1 gp41. Nature 387, 426-430.

Welsch, S., Keppler, O.T., Habermann, A., Allespach, I., Krijnse-Locker, J., Kräusslich, H.-G., 2007. HIV-1 buds predominantly at the plasma membrane of primary human macrophages. PLoS pathogens 3, e36.

West, J.T., Johnston, P.B., Dubay, S.R., Hunter, E., 2001. Mutations within the putative membrane-spanning domain of the simian immunodeficiency virus transmembrane glycoprotein define the minimal requirements for fusion, incorporation, and infectivity. Journal of virology 75, 9601-9612.

White, S.H., von Heijne, G., 2004. The machinery of membrane protein assembly. Current opinion in structural biology 14, 397-404.

White, S.H., von Heijne, G., 2008. How translocons select transmembrane helices. Annual review of biophysics 37, 23-42.

White, S.H., Wimley, W.C., 1998. Hydrophobic interactions of peptides with membrane interfaces. Biochimica et biophysica acta 1376, 339-352.

White, S.H., Wimley, W.C., 1999. Membrane protein folding and stability: physical principles. Annual review of biophysics and biomolecular structure 28, 319-365.

White, T.A., Bartesaghi, A., Borgnia, M.J., Meyerson, J.R., de la Cruz, M.J.V., Bess, J.W., Nandwani, R., Hoxie, J.A., Lifson, J.D., Milne, J.L.S., Subramaniam, S., 2010. Molecular Architectures of Trimeric SIV and HIV-1 Envelope Glycoproteins on Intact Viruses: Strain-Dependent Variation in Quaternary Structure. PLoS pathogens 6, e1001249.

Willey, R.L., Rutledge, R.A., Dias, S., Folks, T., Theodore, T., Buckler, C.E., Martin, M.A., 1986. Identification of conserved and divergent domains within the envelope gene of the acquired immunodeficiency syndrome retrovirus. Proceedings of the National Academy of Sciences of the United States of America 83, 5038-5042.

Wimley, W.C., Creamer, T.P., White, S.H., 1996. Solvation energies of amino acid side chains and backbone in a family of host-guest pentapeptides. Biochemistry 35, 5109-5124.

Wimley, W.C., White, S.H., 1996. Experimentally determined hydrophobicity scale for proteins at membrane interfaces. Nature structural biology 3, 842-848.

Wu, X., Yang, Z.-Y., Li, Y., Hogerkorp, C.-M., Schief, W.R., Seaman, M.S., Zhou, T., Schmidt, S.D., Wu, L., Xu, L., Longo, N.S., McKee, K., O'Dell, S., Louder, M.K., Wycuff, D.L., Feng, Y., Nason, M., Doria-Rose, N., Connors, M., Kwong, P.D., Roederer, M., Wyatt, R.T., Nabel, G.J., Mascola, J.R., 2010. Rational design of envelope identifies broadly neutralizing human monoclonal antibodies to HIV-1. Science (New York, NY) 329, 856-861.

Wyma, D.J., Jiang, J., Shi, J., Zhou, J., Lineberger, J.E., Miller, M.D., Aiken, C., 2004. Coupling of human immunodeficiency virus type 1 fusion to virion maturation: a novel role of the gp41 cytoplasmic tail. Journal of virology 78, 3429-3435.

Wyma, D.J., Kotov, A., Aiken, C., 2000. Evidence for a stable interaction of gp41 with Pr55(Gag) in immature human immunodeficiency virus type 1 particles. Journal of virology 74, 9381-9387.

Xie, J., Bogdanov, M., Heacock, P., Dowhan, W., 2006. Phosphatidylethanolamine and monoglucosyldiacylglycerol are interchangeable in supporting topogenesis and function of the polytopic membrane protein lactose permease. The Journal of biological chemistry 281, 19172-19178.

Yang, C., Spies, C.P., Compans, R.W., 1995. The human and simian immunodeficiency virus envelope glycoprotein transmembrane subunits are palmitoylated. Proceedings of the National Academy of Sciences of the United States of America 92, 9871-9875.

Yang, P., Ai, L.-S., Huang, S.-C., Li, H.-F., Chan, W.-E., Chang, C.-W., Ko, C.-Y., Chen, S.S.-L., 2010. The Cytoplasmic Domain of Human Immunodeficiency Virus Type 1 Transmembrane Protein gp41 Harbors Lipid Raft Association Determinants. Journal of virology 84, 59-75.

Yuan, T., Mietzner, T.A., Montelaro, R.C., Vogel, H.J., 1995. Characterization of the calmodulin binding domain of SIV transmembrane glycoprotein by NMR and CD spectroscopy. Biochemistry 34, 10690-10696.

Yue, L., Shang, L., Hunter, E., 2009. Truncation of the membrane-spanning domain of human immunodeficiency virus type 1 envelope glycoprotein defines elements required for fusion, incorporation, and infectivity. Journal of virology 83, 11588-11598.

Zanetti, G., Briggs, J.A.G., Grünewald, K., Sattentau, Q.J., Fuller, S.D., 2006. Cryo-electron tomographic structure of an immunodeficiency virus envelope complex in situ. PLoS pathogens 2, e83.

Zhang, W., Bogdanov, M., Pi, J., Pittard, A.J., Dowhan, W., 2003. Reversible topological organization within a polytopic membrane protein is governed by a change in membrane phospholipid composition. The Journal of biological chemistry 278, 50128-50135.

Zhang, W., Campbell, H.A., King, S.C., Dowhan, W., 2005. Phospholipids as determinants of membrane protein topology. Phosphatidylethanolamine is required for the proper topological organization of the gamma-aminobutyric acid permease (GabP) of Escherichia coli. The Journal of biological chemistry 280, 26032-26038.

Zhou, T., Xu, L., Dey, B., Hessell, A.J., Van Ryk, D., Xiang, S.-H., Yang, X., Zhang, M.-Y., Zwick, M.B., Arthos, J., Burton, D.R., Dimitrov, D.S., Sodroski, J., Wyatt, R., Nabel, G.J., Kwong, P.D., 2007. Structural definition of a conserved neutralization epitope on HIV-1 gp120. Nature 445, 732-737.

Zhu, P., Liu, J., Bess, J., Chertova, E., Lifson, J.D., Grisé, H., Ofek, G.A., Taylor, K.A., Roux, K.H., 2006. Distribution and three-dimensional structure of AIDS virus envelope spikes. Nature 441, 847-852.

Zhu, P., Winkler, H., Chertova, E., Taylor, K.A., Roux, K.H., 2008. Cryoelectron tomography of HIV-1 envelope spikes: further evidence for tripod-like legs. PLoS pathogens 4, e1000203.

Zwick, M.B., 2005. The membrane-proximal external region of HIV-1 gp41: a vaccine target worth exploring. AIDS (London, England) 19, 1725-1737.