

**DESIGN, PRODUCTION, AND FUNCTIONAL AND STRUCTURAL
CHARACTERIZATION OF HIV-1 ENVELOPE PROTEIN GP41 ECTODOMAIN**

By

Koyeli Banerjee

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

Chemistry – Doctor of Philosophy
Quantitative Biology – Dual Major

2014

ABSTRACT

DESIGN, PRODUCTION, AND FUNCTIONAL AND STRUCTURAL CHARACTERIZATION OF HIV-1 ENVELOPE PROTEIN GP41 ECTODOMAIN

By

Koyeli Banerjee

HIV is an enveloped virus which causes acquired immune deficiency syndrome in humans. HIV infects the host T-cells and macrophages and deteriorates the whole immune system. The virus initiates the infection process by membrane fusion of the viral and host cell by the membrane protein gp160. Gp160 forms a trimer structure with each gp160 molecule comprising two non-covalently bonded subunit - gp120 and gp41. Gp120 binds to the host cell receptors and coreceptors, undergoes conformational change and moves away to expose gp41. The fusion between the viral and host cell membranes is catalyzed by the ectodomain of the gp41 protein. A large number of antibody binding sites are located in the gp120 and gp41 sequence. Also, gp120 and gp41 are capable of eliciting neutralizing antibodies *in vivo*. However, there are a large number of drawbacks for the neutralization of the HIV – (a) fast mutation of HIV strains to evade immune response, (b) non-exposure of most of the antibody epitope regions due to heavy glycosylation, and (c) non-neutralization of a broad range of HIV strains due to mutation. Due to these difficulties, there is no vaccine for HIV developed yet.

My research work focused on production of large gp41 ectodomain constructs and biophysical and structural characterization. The *N*-terminal fusion peptide (FP) and *C*-terminal membrane proximal external region (MPER) are critical for fusion and are postulated to bind to the host cell and HIV membranes, respectively. One significant finding was observation of synergy

between the N-terminal FP, gp41 ectodomain core region, and C-terminal MPER in vesicle fusion. The oligomeric state of gp41 ectodomain has been studied. Conditions are found with predominant monomer or hexamer but not trimer and these may be oligomeric states during fusion. Monomer gp41 ectodomain is hyperthermostable and has helical hairpin conformation. A new HIV fusion model has been proposed through this work, where: (1) hemifusion of membranes is catalyzed by folding of gp41 ectodomain monomers into hairpins; and (2) subsequent fusion steps are catalyzed by assembly into a hexamer with FPs in an antiparallel β -sheet arrangement.

Another reason of interest was that MPER is a highly conserved region which can bind to few broadly neutralizing antibodies effective against various strains of HIV. Two of these antibodies bind to our designed and produced gp41 ectodomain constructs and therefore, adds a step forward towards investigation of the gp41 ectodomain as an immunogen in HIV vaccine development.

Copyright by
KOYELI BANERJEE
2014

To my Father, Late Nityananda Pramanik

ACKNOWLEDGEMENTS

- I would like to thank my advisor, Dr. David Weliky for his enormous support and guidance in research and life. I would be highly obliged to Dr. Weliky for guiding and encouraging me through this great journey of learning and making it a splendid experience. I was fortunate for I was given freedom to explore various techniques in my research project and learn and share innovative ideas. Dr. Weliky has always been like a pathfinder whenever I felt directionless in the broad field of research. I would also thank him for teaching me to think critically and suggestions to interact with the scientific community efficiently. I am grateful to him for his patience and molding me into a responsible scientist.
- I would like to acknowledge members of my guidance committee including Dr. James Geiger, Dr. Gary Blanchard, and Dr. Daniel Jones. I would like to express my sincerest gratitude to all my committee members for the helpful discussions and support throughout the whole journey in the Graduate school.
- I am grateful to present and past members of Weliky-group. I would like to thank my previous group members Dr. Kelly Sackett, Dr. Matthew J. Nethercott, Dr. Scott Schimck, and Dr. Erica P. Vogel for helping me to adapt in the new lab during my initial days and discussing with me whenever required. I would like to thank Dr. Erica P. Vogel, Dr. Matthew J. Nethercott and Dr. Kelly Sackett for teaching me various techniques performed in our lab. I am grateful to Erica for her patience to train me in different molecular biology and protein biochemistry techniques. I am also grateful to my present lab members Dr. Charles Gabrys, Li Xie, Punsisi Upeka Ratnayake, Ujjayini Ghosh, Lihui Jia, Shuang Liang, and Robert Wolfe for being so friendly and helpful. . It was a great learning experience interacting with Li regarding NMR techniques. I

would also like to extend my appreciation to Punsisi for the helpful discussions regarding protein purification methods.

- My journey in the graduate school would not have been possible without the support and encouragement of my family; my Mom, in-laws, my sister, uncles and aunts, my husband and my son. My Mom, Bela Pramanik, has always been very encouraging to shape my career path. She has always taught me to be organized and make every single day productive. I would like to thank my sister and her family for enormous support and good wishes. Special thanks go out to my in laws and my uncles and aunts for their best wishes.
- I am grateful to Dr. Anindya Chanda, Satyaki Sengupta, and Dr. Iwan Setiawan as I would have never survived in graduate school without their kind friendship. Thank you all for the wonderful memories.
- My husband Rahul has been a tremendous source of encouragement and motivation. He always motivated me to go forward no matter what the difficulty is. He always supported me in adverse situations and encouraged me to be optimistic. I find short of word to say thank you to him for supporting my passion for higher education. It has been such a great experience discussing scientific issues with him and learn to think critically. Together we came to MSU with our toddler son, Raunak who is now eight years old. My son has been very understanding always. During the most difficult days of graduate life, he has been such a great source of rejuvenation. You both will remain the greatest source of inspiration for me.

TABLE OF CONTENTS

LIST OF TABLES	xi
LIST OF FIGURES	xii
KEY TO ABBREVIATIONS	xix
CHAPTER 1: Human Immunodeficiency Virus Membrane Fusion Protein Gp41	1
1.1. Introduction	2
1.1.1. Fusion Peptide	4
1.1.2. N-terminal Helix Region-C-terminal Helix Region	5
1.1.3. Membrane Proximal External Region	7
1.1.4. Transmembrane Domain	12
1.1.5. Cytoplasmic Endodomain	13
1.2. Possible Mechanism of Membrane Fusion	13
1.3. Mechanism of Fusion of HIV-1 and Host Cell Membranes Caused by Gp41	22
1.3.1. Model I	23
1.3.2. Model II	24
1.3.3. Caveats of Both the Models	25
1.4. Stoichiometry of Fusion proteins at the Site of Membrane Fusion	26
REFERENCES	27
CHAPTER 2: Materials and Methods	35
2.1. Site Directed Mutagenesis	36
2.2. Transformation	37
2.3. Expression of Gp41 HP Construct	38
2.4. Detection of Expression of HM using Solid State NMR of Inclusion Bodies	40
2.5. Solubilization of HM	43
2.6. Purification of Gp41 Ectodomain Protein Constructs	46
2.6.1. HP	46
2.6.2. HM	46
2.7. Western Blot Analysis	49
2.8. Peptide Synthesis of N-terminal Fusion Peptide Region of Gp41 Ectodomain	49
2.9. Production of Full-length Gp41 Ectodomain Constructs by Native Chemical Ligation	51
2.10. Stabilization of Gp41 Ectodomain Protein Constructs	54
2.11. Determination of Secondary Structure of Gp41 Ectodomain Constructs	55
2.11.1. Circular Dichroism	55
2.11.2. Solid State Nuclear Magnetic Resonance	55
2.12. Determination of Oligomeric State(s) by Size Exclusion Chromatography	57
2.13. Antibody Binding Studies by Direct Immunoprecipitation	59

2.14.	Vesicle Fusion Assays by Fluorescence Spectroscopy	60
	REFERENCES	62
CHAPTER 3: Structural and Functional Characterization of Gp41 Ectodomain		65
3.1.	Introduction	66
3.2.	Experimental Procedures	70
3.2.1.	HP and HM Inserts	70
3.2.2.	HP and HM Expression, Solubilization and Purification	72
3.2.3.	Synthesis and Purification of FP-HM	73
3.2.4.	Size Exclusion Chromatography	75
3.2.5.	CD Spectroscopy	75
3.2.6.	Western Blots	75
3.2.7.	Immunoprecipitation	75
3.2.8.	Protein Induced Vesicle Fusion	76
3.3.	Results	77
3.3.1.	High Yield Protein Production	77
3.3.2.	Solubility Only at pH 3 or With 6M GuHCl	80
3.3.3.	Monomer or Hexamer and Not Trimer	81
3.3.4.	Hyperthermostable Hairpin Structure of the Monomer	84
3.3.5.	Hairpin Monomer With MPER is Highly Fusogenic	88
3.3.6.	Synergy of FP, Hairpin, and MPER with the GP41 Copy Number of a Virion	91
3.3.7.	Folded Hairpin Ectodomain With MPER Binds to bNAbs	94
3.4.	Discussion	97
	REFERENCES	113
CHAPTER 4: Construction of HIV-1 Gp41 Pre-Hairpin Intermediate		120
4.1.	Specific Aims	121
4.2.	Introduction	121
4.3.	Approach	123
4.3.1.	Mutagenesis of CHR	123
4.3.2.	Verification of Formation of Pre-Hairpin Intermediate	125
4.4.	Materials and Method	127
4.4.1.	Construction of Mutant HP and HM	127
4.4.2.	Experimental Procedure	128
4.5.	Results	129
4.6.	Discussion	135
4.7.	Future Work	136
	REFERENCES	137
CHAPTER 5: Future Work		140
	REFERENCES	143

APPENDICES	145
APPENDIX I: Designing of FP-HP by Native Chemical Ligation	146
APPENDIX II: Determination of Secondary Structure by Solid State NMR of Membrane Associated FP of Gp41 Ectodomain Construct FP-HP	150
APPENDIX III: Study of Binding of Designed HIV Pre-Hairpin Intermediate and Membrane Fusion Inhibitor C34 Peptide	154
APPENDIX IV: Sample Preparation of Uniformly Labeled Bacterially Expressed HIV-1 Gp41 Hairpin Ectodomain	161
APPENDIX V: Location of Experimental Data Files	165
REFERENCES	168

LIST OF TABLES

Table 2.1. Solubilization of inclusion bodies containing HM in various solubilizing agents	45
--	----

LIST OF FIGURES

Figure 1.1. (Left) Infection model of HIV, (Right) Electron microscopy pictures of HIV infection process, (a) Binding of HIV and host cell, (b) Hemifusion of viral and host cell membrane, (c) Pore formation, and (d) complete fusion and entry of viral genetic material into the host cell.

3

Figure 1.2. (A) Schematic diagram of the HIV-1 gp41 and (B) gp41 ectodomain amino acid sequence color coded according to the different domains.

4

Figure 1.3. (A) Six-helix bundle of gp41 comprising NHR and CHR and (B) NHR and CHR showing heptad repeat amino acid positions as orange and yellow respectively. N36 is the 36 residues in the NHR and C34 is the 34 residues in the CHR.

7

Figure 1.4. (A) NMR structure of MPER in DPC micelle showing the kink at residue W672-N674. (B) Crystal structure of monomer NHR-CHR-MPER showing MPER as a continuous helix. The MPER is from L662-I675. The NHR and CHR are denoted as HR1 and HR2 respectively.

9

Figure 1.5. (A) Model of MPER only (bottom) and MPER bound to 4E10 (top). The residues W672-W676 shown in blue and N671 and W680 shown in magenta undergoes conformational change upon binding to the antibody. (B) Model of MPER bound to antibody Z13e1 showing conformational change of MPER helices with change in ψ values

11

Figure 1.6. Stages of membrane fusion

21

Figure 1.7. Proposed energy profile of membrane fusion caused by HIV-1 gp41 envelope protein. The maximum energy barrier for stalk formation is calculated as ~ 25 kcal/mol (Kuzmin, P. I. et al. PNAS (2001), 98). The energy barrier to reach hemifusion and pore formation is lower than energy barrier for stalk formation. The energy barrier between stalk and hemifusion stage is due to activation energy required to allow the lipid molecule acyl chains to move away from their hydrophobic environment and avoidance of water molecules. After hemifusion, there is complete fusion between the two membranes which lead to pore formation. In the whole process $H_{\text{Two membrane}} = H_{\text{Fusion pore}}$

22

Figure 1.8. Model I: Schematic diagram of membrane fusion caused by gp41. (1) Trimer gp120+gp41, (2) dissociation of gp120 from gp41 and exposure of gp41 FP, (3) SHB folding of gp41 and hemifusion of viral and host cell membrane, and (4) pore formation with close apposition of FP and TM of gp41. 'A' denotes viral membrane and 'F' denotes FP region

24

Figure 1.9. Model II. Stages of fusion: (a) gp120+gp41 trimer, (b) gp41 PHI formation with FP insertion into host membrane, (c) hemifusion caused by PHI, and (d) SHB formation

25

Figure 2.1. Amino acid and DNA sequence of HP construct

38

Figure 2.2. SDS-PAGE gel showing RP-HPLC purified HP in glacial acetic acid at pH 3.0	40
Figure 2.3. SDS-PAGE gel showing IMAC purified HP in the presence of urea at pH 7.4	40
Figure 2.4. Amino acid and DNA sequence of HM	41
Figure 2.5. Chemical shift corresponding to ^{13}CO - Leu	42
Figure 2.6. Solubilization of inclusion bodies in various solubilizing agents	44
Figure 2.7. SDS-PAGE gel showing low solubilization and negligible binding of HM with 2Gly-4His tag in 8M urea with Co^{2+} resin	47
Figure 2.8. SDS-PAGE gel showing inefficient binding of HM with 4Gly-4His tag in 3xCMC Sarkosyl with Co^{2+} resin	48
Figure 2.9. SDS-PAGE gel showing binding of HM with 2Gly-4His tag in 6M GuHCl with Co^{2+} resin and elution with high purity. The bands near 25 kD are dimer bands, confirmed by western blot with anti-His antibody	48
Figure 2.10. Schematic representation of <i>t</i> -Boc solid phase peptide synthesis	51
Figure 2.11. Mechanism of native chemical ligation. The FP23 that is synthesized by <i>t</i> -Boc SPPS is shown having a thiol on the C-terminus. To make the reaction more rapid, a catalyst MPAA is used. MPAA undergoes thioester exchange, and the Hairpin with the N-terminal Cys can easily form thioester-linked intermediate. Finally the more stable product, i.e., the peptide bond is formed	53
Figure 2.12. The REDOR pulse sequence	57
Figure 2.13. SEC A_{280} of MW standards loaded on (A) Superdex 75 and (B) Superdex 200 columns. The protein standards used were (A) Conalbumin (75 kDa), Ovalbumin (44 kDa), Carbonic Anhydrase (29 kDa), Ribonuclease A (13.7 kDa), and Aprotinin (6.5 kDa) and (B) Thyroglobulin (669 kDa), Ferritin (440 kDa), Aldolase (158 kDa), Conalbumin (75 kDa), Ovalbumin (44 kDa), Carbonic Anhydrase (29 kDa), Ribonuclease A (13.7 kDa), and Aprotinin (6.5 kDa). The V_o is the void volume and corresponds to 2 MDa Blue Dextran. K_{av} vs $\log_{10}(\text{MW})$ for the (C) Superdex 75 and (D) Superdex 200 columns. The $K_{av} = (V_e - V_o)/(V_c - V_o)$ where V_e and V_c are the elution and column volumes, respectively	58
Figure 3.1. (A) Schematic diagram of HIV gp41 where FP $\equiv$ fusion peptide, NHR $\equiv$ N-heptad repeat; CHR $\equiv$ C-heptad repeat, MPER $\equiv$ membrane-proximal external region, TM $\equiv$ transmembrane domain, and Endo $\equiv$ endodomain. (B) Amino acid sequences of HP, HM, and FP-HM. The epitopes of the 2F5 and 4E10 bNAbs are marked	71
Figure 3.2. DNA sequence of the HM insert	71

Figure 3.3. MALDI mass spectra of (A) the HM and (B) FP-HM peaks from the first HPLC purification, Fig. 4A. Inset is SDS-PAGE where the middle lane is the FP-HM peak of Fig. 4A and the right lane is purified HM

74

Figure 3.4. (A) First-round RP-HPLC profile of FP+HM ligation. (B) Second-round RP-HPLC of FP-HM peak from panel A

74

Figure 3.5. (A) ^{13}C SSNMR spectra of a cell pellet harvested from centrifugation of a cell lysis in PBS. The *E. coli* cells contained a plasmid with the HM insert and expression was induced for two hours in minimal medium containing ^{13}CO -Leu. Any expressed HM will therefore be ^{13}CO -Leu labeled. The SSNMR acquisition parameters included a 9.4 T magnetic field, 4 mm diameter rotor, 8 kHz magic angle spinning frequency, and ~ 1 day of signal averaging. The ratio of integrated isotropic ^{13}CO intensity (peaked at ~ 175 ppm) to integrated aliphatic intensity (0 – 90 ppm region) translates to expression of ~ 300 mg HM/L culture. (B) SDS-PAGE of purified HM (MW = 13.7 kDa) and HP (MW = 11.6 kDa). (C) MALDI MS of FP-HM ligation product twice-purified by RP-HPLC. The experimental ratio $[m/z(\text{FP-HM}):m/z(\text{HM})]$ is 1.1522 and matches the ratio of expected MWs, $[\text{MW}(\text{FP-HM}):\text{MW}(\text{HM})] = 1.1522$

79

Figure 3.6. SEC A_{280} profiles for HP with (A) 1 mg/mL loading, Superdex 200 column and (B) 10 mg/mL loading, Superdex 75 column. The loading and running buffer was 50 mM sodium formate pH 3.2, 150 mM NaCl, and 0.2 mM TCEP. The blue numbers are the MW standards in kDa and their respective elution volumes are the downward blue arrows. There are accompanying SEC profiles of the MW standards as well as plots of K_{av} vs $\log_{10}(\text{MW})$ (Chapter 2, Figure 11)

82

Figure 3.7. SEC A_{280} profiles from a semi-preparative Superdex 200 column for HP and HM in buffers containing 6 M GuHCl and 150 mM NaCl. The pH 3.2 buffer was 50 mM sodium formate with 0.2 mM TCEP and the pH 7.4 buffer was 10 mM phosphate with 2 mM DTT. For each run, the loading and running buffers were the same and the loading stock solution had a 1 mg/mL protein concentration. The blue numbers are the MW standards in kDa and their respective elution volumes are the downward blue arrows. The most prominent peaks in the profiles correspond to hexamer masses and vertical dashed lines show the elution volume offset between the HP and HM hexamers. Smaller peaks most consistent with dodecamer and 24-mer masses are identified by shaded regions. For HM, there is also a small peak associated with the V_o void volume, i.e. aggregates with $\text{MW} \geq 2$ MDa

83

Figure 3.8. Evidence for thermodynamic equilibrium between hexamers and larger oligomers. SEC with A_{280} detection was run for HP and HM in buffer containing 50 mM sodium formate at pH 3.2 with 150 mM NaCl, 6 M GuHCl, and 0.2 mM TCEP (Figure 3.7). The hexamer peak was collected and loaded onto the Superdex 75 column for SEC with this same buffer. The loading solution contained (A) 0.3 or (B) 0.2 mg protein/mL. Each blue number is the MW of a standard in kDa and the blue downward arrows is its elution volume from Figure 2.13, Chapter 2. The resultant profiles for (A) HP and (B) HM show a mixture of hexamers and larger oligomers similar to the profiles of the initial run (Figure 3.7). This similarity supports this mixture as the thermodynamic equilibrium state for the protein

84

Figure 3.9. (A) CD spectra of HP and HM in pH 3.2 buffer. All spectra show θ_{222} indicative of a significant fraction of α helical structure. (B) θ_{222} vs temperature for HM at pH 3.2. No unfolding transition is observed. (C) CD spectra of WT and mutant HP at pH 3.2. For a SHB trimer, the D632A mutation could disrupt the inter-monomer salt bridge with K574 and the W628A mutation could disrupt the intra-monomer hydrophobic interaction with W571. More positive CD signal for the double but not the single mutant is consistent with monomer rather than SHB trimer HP at pH 3.2. (D) CD spectra of HP and HM in buffer with 6 M GuHCl. The spectra show θ_{222} indicative of a significant fraction of α helical structure. The [protein] $\approx$ 15 μ M ($\sim$ 0.2 mg/mL), the pH 3.2 buffers contained 50 mM formate with 0.2 mM TCEP, and the pH 7.4 buffers contained 10 mM phosphate and 2 mM DTT

87

Figure 3.10. Vesicle fusion at 37 °C. Protein was added just prior to the increase in fusion with $\sim$ 5 s assay dead time. For panels A and B, the vesicles were negatively-charged and for panels C and D, the vesicles were positively-charged. The calculated protein charge is +16, -2, and -4 at respective pH values 3.2, 7.4, and 9.0. The protein:lipid mole ratios were: (A) 1:700, (B) 1:75, and (C, D) 1:300, with Chol not included in the lipid quantity. The data overall support the importance of attractive protein/vesicle attractive energy for fusion at low protein:lipid ratios. The protein stock contained 20 μ M protein ($\sim$ 0.25 mg/L) in 10 mM formate at pH 3.2 with 0.2 mM TCEP which are condition for predominant monomeric protein. For panels A and B, [POPC] = 120 μ M, [POPG] = 30 μ M, and [Chol] = 75 μ M, and for panels C and D, the vesicles contained [POPC] = 120 μ M, [DOTAP] = 30 μ M, and [Chol] = 75 μ M. Vesicle buffers were 10 mM formate at pH 3.2, 5 mM HEPES/10 mM MES at pH 7.4, and 10 mM Tris-HCl at pH 9.0

90

Figure 3.11. Vesicle fusion at 37 °C. A 7.5 μ L aliquot of protein was added just prior to the increase in fusion and the final total volume was 1200 μ L. The POPC:POPG:Chol vesicle composition and buffers were the same as Fig. 6. The protein stock buffer was 10 mM formate at pH 3.2 with 6 M GuHCl and 0.2 mM TCEP and following addition of protein stock, the assay buffer contained 40 mM GuHCl. For the pH 3.2 assays, the stock [protein] = 20 μ M ($\sim$ 0.25 mg/mL), and for the pH 7.4 assays, the stock [protein] = 160 μ M ($\sim$ 2 mg/mL). The [protein]:[POPC+POPG] = 1:1200 for the pH 3.2 assays and 1:150 for the pH 7.4 assays

92

Figure 3.12. Dose response of vesicle fusion induced by HM and FP-HM at pH 3.2. The assay conditions were the same as Fig. 7A except that the protein stock concentrations were 20, 10, and 5 μ M for [protein]:[total lipid] = 1:1200, 1:2400, and 1:4800, respectively

93

Figure 3.13. Vesicle fusion induced by GuHCl at (A) pH 3.2 or (B) pH 7.4. A 7.5 μ L aliquot containing 6 M GuHCl was added to a vesicle suspension with final total [GuHCl] = 40 mM in a volume of 1200 μ L. The assay conditions were the same as Fig. 7 except that there was no protein in the aliquot that contained GuHCl

93

Figure 3.14. Western blots showing binding of HM to the 4E10 and 2F5 bNAbs and weaker binding of HP to 2F5. The left-side arrows are MWs in kDa

95

Figure 3.15. SDS-PAGE of the immunoprecipitation of HP and HM with the 4E10 and 2F5 bNAbs. There is stronger binding of HM to 4E10 and 2F5 and weaker binding of HP to 2F5. The protein stock buffer contained 50 mM sodium formate at pH 3.2 with 150 mM NaCl. For the right three lanes of the panel A gel, the buffer also contained 6 M GuHCl whereas for the left three lanes of panel A, the GuHCl was absent. Binding was done in PBS at pH 7.4 and for the right lanes, the diluted [GuHCl] = 40 mM. The center lane is the MW ladder in kDa 96

Figure 3.16. SDS-PAGE of the immunoprecipitation of HP and HM with IgG. Protein binding to IgG is not observed. The protein stock buffer contained 50 mM sodium formate at pH 3.2 with 150 mM NaCl. For the right two lanes of the gel, the buffer also contained 6 M GuHCl. The antibody binding was done in PBS at pH 7.4 and for the right lanes of the gel, the diluted [GuHCl] = 40 mM. The inner lanes are MW ladders in kDa 97

Figure 3.17. Working structural models of the gp41 ectodomain with Fig. 1 color coding. Monomeric HP is modeled as the hairpin structure of a monomer in the SHB. HM and FP-HM hexamers are modeled as two SHB trimers that contact at their *N*-/*C*-terminal interfaces. This model correlates with the observed intermolecular antiparallel β sheet structure of FPs 100

Figure 3.18. The monomeric elution peak of protein GB1 in PBS in the presence of 6M GuHCl at (A) pH 3.2 and (B) pH 7.4. The elution volume corresponds to the molecular weight of ~6.3 kDa and ~9 kDa at pH 3.2 and 7.4 respectively. The size exclusion chromatography was done at Superdex 75 column. The blue arrows show the elution volumes of MW standards with the MWs given in kDa 103

Figure 3.19. Model for membrane fusion that includes the gp41 ectodomain monomer and hexamer. The different regions of gp41 are color coded according to Fig. 1 and the TM and endodomains are not shown. One of the monomers is not displayed in steps 3-5. The initial gp41 structure of step 1 is based on the ~5 Å gp140 structures and the final SHB structure of step 7 is based on high-resolution structures 109

Figure 4.1. HIV fusion model. (a) Native state of envelope protein before fusion, (b) Pre-Hairpin intermediate of gp41, and (c) Final Hairpin state of gp41 122

Figure 4.2. Helical wheel representation of NHR and CHR. Three NHR helices and one CHR helix are represented as helical wheel projections. It is an end-on-view of the NHR-CHR complex looking down the axis of the trimer. The NHR helices interact with other NHR helices at the “a” and “d” positions (colored orange). Positions of the NHR and the CHR helices that occupy the interhelical space between two NHR helices and one CHR helix are shaded yellow. The residues which interact among themselves are colored accordingly 125

Figure 4.3. (Left) Far-UV CD analysis of Hairpin showing 100% helicity with the mean residue molar ellipticity at 222nm (θ_{222}) is $-33,000 \text{ deg cm}^2 \text{ dmol}^{-1}$. For inset, abscissa is the temperature scale in °C, and the ordinate is θ_{222} , showing the thermal stability of Hairpin with temperature. (Right) Thermal stability of Hairpin (●) and N70 (+). The melting temperature of N70 is 45 °C 126

Figure 4.4. Amino acid sequence of HP protein construct, DNA sequence of HP WT and HP M3
128

Figure 4.5. The SHB trimer of gp41 comprising of inner core NHR and outer CHR. Each monomer is colored differently. *Top*, residues shown forms intermonomer salt bridge, and *bottom*, shows residues forming hydrophobic interaction between intramonomer strands
130

Figure 4.6. CD spectra of WT and mutant HP at pH 3.2. For a SHB trimer, the D632A mutation disrupts the inter-monomer salt bridge with K574 and the W628A mutation disrupts the intra-monomer hydrophobic interaction with W571. The [protein] $\approx 15 \mu\text{M}$ ($\sim 0.2 \text{ mg/mL}$), the pH 3.2 buffers contained 50 mM formate with 0.2 mM TCEP
132

Figure 4.7. CD spectra of HP WT and HP M3 at pH 3.2. W631R mutation disrupt the intra-monomer hydrophobic interaction with W571. The [protein] $\approx 10 \mu\text{M}$ ($\sim 0.2 \text{ mg/mL}$), the pH 3.2 buffers contained 50 mM formate with 0.2 mM TCEP
133

Figure 4.8. CD spectra of HM WT and HM M3 at pH 3.2. For a SHB trimer, W631R mutation along with W628A disrupts the intra-monomer hydrophobic interaction with W571. The [protein] $\approx 10 \mu\text{M}$ ($\sim 0.2 \text{ mg/mL}$), the pH 3.2 buffers contained 50 mM formate with 0.2 mM TCEP
134

Figure 4.9. Estimation of thermal stability of mutant HM. *Left*, measurement of thermal unfolding of HM WT and HM M3 with mean residue molar ellipticity as a function of temperature at 222nm and pH 3.2, and *right*, CD spectra of thermally denatured protein after refolding
135

Figure AI-1. SDS-PAGE gel of Hairpin (non His-tag) purified by RP-HPLC. MW of Hairpin (non His-tag) is 10.7 kDa. Lanes: 1- Hairpin; 2- Standard
147

Figure AI-2. CD spectra of the refolded FP-HP
149

Figure AI-3. SDS-PAGE gel comparing FP-HP and Hairpin. MW of FP-HP is 12.8kDa, and Hairpin (non His-tag) is 10.7kDa. Lanes: 1- Standard, 2- FP-HP, 3, 4 –Hairpin
149

Figure AII-1. (Left) The structure of part of the FP showing the positions of the ^{13}C and the ^{15}N labels on Leu (L7) and Phe (F8) respectively. (Right) REDOR subtracted signal for L7F8 in DTPC:DTPG:Chol vesicles. The CO peak of L7 was referenced to the methylene peak of adamantane at 40.5 ppm. The chemical shifts of 178.5 and 173.7 ppm correspond to α -helix and β -sheet respectively
152

Figure AIII-1. (A) Schematic diagram of gp41 showing different domains – fusion peptide (fp), NHR (Heptad repeat 1), CHR (Heptad repeat 2), transmembrane (tm) and endodomain (cyto). The C34 peptide is shown to compare the sequence length with CHR. The three residues that bind into the groove are shown in red. (B) Down the axis view of SHB structure of gp41 ectodomain, (C) sidewise view of SHB showing the position of CHR on the NHR core grooves. (D)

The packing of the C34 on the N36 (NHR) grooves. The N36 is represented as a molecular surface and the three residues of C34 which fit into the binding pocket are shown in red 155

Figure AIII-2. Circular dichroism spectra of (A) 5 μ M of HM M3 and (B) 5 μ M C34 at pH 3.0 158

Figure AIII-3. Secondary structure of HM M3 + C34, 5 μ M each. The incubation was performed overnight at 4°C. The CD was performed at pH 3.0 159

Figure AIII-4. The secondary structure HM M3 + C34, 5 μ M each. The protein was heated for 2 min and the C34 was added and incubated overnight at 4°C 159

Figure AIV-1. MAS cross polarization signal for ^{13}C labeled inclusion body containing HP protein. The CO peak of HP was referenced to the methylene peak of adamantane at 40.5 ppm 164

KEY TO ABBREVIATIONS

AUC	analytical ultracentrifugation
BCA	bicinchoninic assay
bNAbs	broadly neutralizing antibody
CD	circular dichroism
Chol	cholesterol
CHR	C-terminal helix region
DOTAP	1,2-dioleoyl-3-trimethylammonium-propane (chloride salt)
FP	fusion peptide
gp160	glycoprotein 160
gp140	glycoprotein 140
gp120	glycoprotein 120
gp41	glycoprotein 41
GuHCl	guanidine hydrochloride
IB	inclusion body
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HIV	Human immune deficiency virus
IPTG	isopropyl β -D-1-thiogalactopyranoside
LB	Luria-Bertani
MES	2-(<i>N</i> -morpholino)ethanesulfonic acid
MPAA	S-Trityl- β -mercaptopropionic acid
MPER	membrane-proximal external region
SHB	six-helix bundle

NHR	<i>N</i> -terminal helix region
<i>N</i> -NBD-DPPE	<i>N</i> -(7-nitro-2,1,3-benzoxadiazol-4-yl) (ammonium salt) dipalmitoylphosphatidylethanolamine
<i>N</i> -Rh-DPPE	<i>N</i> -(lissamine rhodamine B sulfonyl) (ammonium salt) dipalmitoylphosphatidylethanolamine
PHI	pre-hairpin intermediate
POPC	1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine
POPG	1-palmitoyl-2-oleoyl-sn-glycero-3-[phospho-rac-(1-glycerol)] (sodium salt)
RP	recombinant protein
RP-HPLC	reversed-phase HPLC
SEC	size-exclusion chromatography
SSNMR	solid-state nuclear magnetic resonance
TCEP	Tris(2-carboxyethyl) phosphine hydrochloride

CHAPTER 1

Human Immunodeficiency Virus Membrane Fusion Protein Gp41

1.1. Introduction:

Acquired Immune Deficiency Syndrome (AIDS) is a disease of the human immune system caused by the Human Immunodeficiency Virus (HIV). According to World Health Organization, there were 1.1 million people living with HIV till the year 2012 in the United States and 1.6 million annual deaths worldwide.

The viral entry to the host cell is postulated to occur in two ways. (1) Fusion of the viral and the host cell membrane and then release of the viral genetic material into the host cell which eventually integrates with the host genome (1,2). (2) The second is the endocytic pathway. The virus enters the cell cytoplasm through the endosomes and then complete fusion occurs followed by delivery of viral genetic material into the cell cytoplasm (3,4). The mechanism of viral entry through fusion of cell membrane is believed to be more prevalent based on the evidence that there is shorter lag between viral membrane protein-host cell receptor coreceptor complex and cell-cell fusion (5) compared to delayed endosomal fusion of HIV (3).

The process of membrane fusion is mediated by HIV membrane glycoprotein gp160. Gp160 is comprised of non-covalently associated outer receptor binding subunit, gp120 and an inner fusion protein, gp41(1). The membrane spanning gp41 participates in the fusion of the viral membrane with the host cell membrane and this membrane fusion creates a pore which can transfer the viral genetic material into the host cell (Figure 1.1).

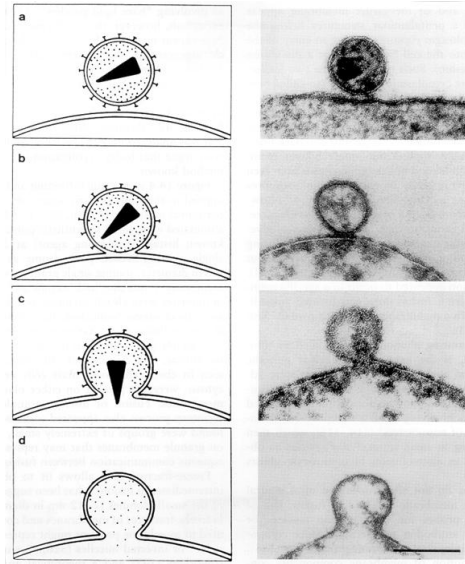


Figure 1.1. (Left) Infection model of HIV, (Right) Electron microscopy pictures of HIV infection process, (a) Binding of HIV and host cell, (b) Hemifusion of viral and host cell membrane, (c) Pore formation, and (d) complete fusion and entry of viral genetic material into the host cell(6).

The gp160 in HIV membrane is activated for fusion by the interactions with CD4 (receptor) and CXCR4 or CCR5 (coreceptor) of the T-cell or macrophage cell membrane. Gp120 binds to the CD4 receptor and coreceptors. The binding results in a conformational change in gp120 and as a result gp120 dissociates with gp41 exposed(1). The gp41 plays a significant role in fusion between viral and host cell membranes. Gp41 consists of different regions such as ectodomain with amino acid residues 512-683, comprising of N-terminal fusion peptide (FP), an N-terminal helix region (NHR), an immunodominant loop region, a C-terminal helix region (CHR), membrane proximal ectodomain region (MPER), the transmembrane region (TM) amino acid residues 684-705, and the endodomain (cytoplasmic tail) amino acid residues 705-856 (Figure 1.2). Each region of gp41 plays a unique role.

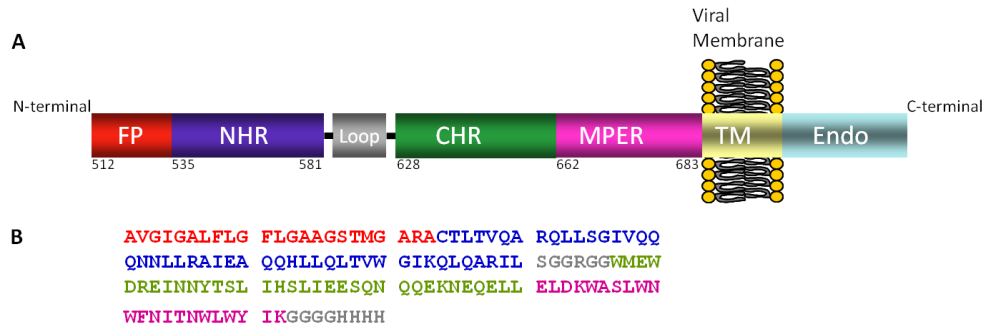


Figure 1.2. (A) Schematic diagram of the HIV-1 gp41 and (B) gp41 ectodomain amino acid sequence color coded according to the different domains.

1.1.1. Fusion Peptide

The fusion peptide (FP) region is a Gly rich hydrophobic N-terminal domain of gp41. There is no atomic resolution crystal structure of the FP region; however, some NMR structural studies have been done with FP associated with detergent micelles. The FP is unstructured in solution whereas it forms α -helical structure in the presence of detergents SDS or DPC (7-9). There have been contradictory structural studies showing FP in physiologically relevant membrane environment as β -sheet structure (10,11), whereas, another study shows that FP region acquires α -helical monomeric structure and β -sheet structure at low and high peptide:lipid concentration respectively (12). Difference in conformation of FP has been observed with the presence or absence of cholesterol in the lipid membrane. The FP adopts α -helical structure in the absence of cholesterol and β -sheet structure in the presence of cholesterol (11). Studies on FP domain indicate that oligomerization of FP may be important for fusion activity. Lipid mixing assay and analytical ultracentrifugation studies of FP_{monomer}, FP_{dimer}, and FP_{trimer} shows maximum fusion activity for the trimer FP construct (13). Mutation of the Val-2 to Glu-2 (V2E) in

the N-terminal end shows inhibition of fusion activity and syncytia formation. The effect of V2E mutation is such that even in the presence of excess wild type gp160, the membrane fusion and HIV infection is inhibited revealing the functional oligomeric property of FP domain (14). These studies suggest that oligomerization of FP is essential for fusion of viral and host cell membrane. The study of interaction of synthetic FP and PG LUVs has been done by lipid mixing assays, vesicle binding and leakage experiments. The studies suggest that wild type FP can penetrate through vesicle monolayer and cause permeabilization (12). Similar study performed with mutant V2E FP peptides showed no destabilization of PG vesicles (15). Also solid state NMR studies of FP in the lipid membranes show maximum membrane insertion depth and fusogenicity of trimer FP than compared to monomers or mutated (V2E) FP (16). Based on these studies, it could be proposed that probably FP inserts into the host membrane during the fusion process.

1.1.2. N-terminal Helix Region-C-terminal Helix Region

The high resolution crystal structure of the NHR and CHR without the loop region shows a highly helical trimer structure forming a hairpin conformation (17,18). This trimer NHR-CHR structure also known as six-helix bundle (SHB) is proposed to be the structure of gp41 core region after the fusion process is complete (Figure 1.3A). At the center of the SHB is parallel trimer NHRs held together by hydrophobic interactions primarily between the Ile or Leu residues. The amino acid sequence of N-helix region (NHR) and the C-helix region (CHR) comprise of 4-3 heptad repeats of hydrophobic residues. Heptad repeats are formed by repeats of hydrophobic residues in the 1st and 4th residue in each seven residue helical turn. The

hydrophobic residues in the 'a' and 'd' positions form the hydrophobic interactions (Figure 1.3B). The CHR helices pack on the outside grooves of the NHR trimer core in antiparallel orientation. The interaction between the NHR and CHR is predominantly hydrophobic between the 'e' and 'g' residues of NHR helices and 'a' and 'd' residues of CHR helices respectively (figure 3B). Some of the hydrophobic interactions include: (a) W628, W631 of CHR interacts with W571 of NHR, (b) I635 of CHR interacts with L565 and L568 of NHR, and (c) Q653 of CHR has intramolecular hydrogen bond with Q551 of NHR and Q653 forms intermolecular hydrogen bond with the backbone CO oxygen of V549 of NHR. There is a salt bridge between the NHR and CHR between the K574 and D632. The NHR-CHR SHB is very thermally stable with melting temperatures of $\sim 70^{\circ}$ C for shorter ectodomain constructs with hairpin structure whereas temperatures up to 110° C are observed for longer constructs (19,20). This thermostability has supported the SHB as the final gp41 structure during fusion. Very recent solution NMR structural study of NHR-CHR in DPC micelles suggests monomeric structure with very less interaction between the NHR and CHR. The $^{13}\text{C}^{\alpha}$ chemical shifts of all the residues of NHR-CHR construct was compared with the chemical shifts of the residues of the individual peptides NHR and CHR. It was observed that in the $^{13}\text{C}^{\alpha}$ chemical shift pattern of the NHR-CHR construct exactly overlapped with the individual NHR and CHR peptides. The same group conducted the study to understand the interaction of the NHR-CHR protein construct with detergent molecules by paramagnetic relaxation enhancement. It was observed that both the NHR and CHR helices are amphipathic with the hydrophobic residues lying on one side of the helix and the hydrophilic residues on the other side of each helix. Based on this observation, it is proposed that both NHR and CHR can be embedded at the lipid-water interface (21).

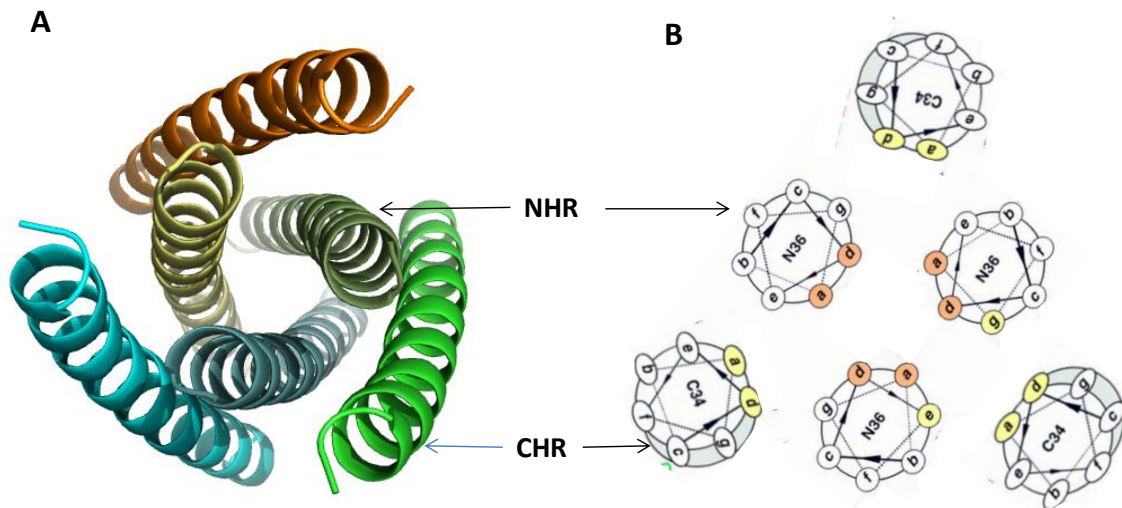


Figure 1.3. (A) Six-helix bundle of gp41 comprising NHR and CHR and (B) NHR and CHR showing heptad repeat amino acid positions as orange and yellow respectively (22). N36 is the 36 residues in the NHR and C34 is the 34 residues in the CHR.

1.1.3. Membrane Proximal External Region

Membrane proximal external region (MPER) is the last ectodomain region following CHR towards the C-terminal end comprising of the amino acids from 662 – 683. The two primary characteristics of the MPER due to which this is a region of great interest are: the amino acid residues are highly conserved (23-25), and the region comprise of epitopes for three broadly neutralizing antibodies (23,25). Broadly neutralizing antibodies are the antibodies which can stop infection caused by a broad range of HIV strains. HIV-1 can easily mutate to evade the immune response of the host. The immune response of the host is further non-functional due to non-exposure of neutralizing epitopes. The gp120 and gp41 can elicit neutralizing antibodies,

however, these antibodies are isolate specific (26). Very few broadly neutralizing antibodies are elicited by the epitopes contained in the highly conserved MPER domain. The broadly neutralizing antibodies 4E10, 2F5, and Z13 recognize epitopes in the MPER (23,25). Even though the broadly neutralizing antibodies are efficient enough to inhibit the fusion process by binding to the MPER epitopes, still there is no success by injecting the antigen into the host body to elicit antibody (26). Efforts to produce broadly neutralizing antibodies to the conserved epitopes of MPER proved to be of not much use. The conserved neutralizing epitopes have been observed to be poor immunogens as they mimic host antigens leading to depletion of immune tolerance (27).

Structural characterization of MPER peptide in lipid and detergent environment by EPR and NMR respectively shows kinked helical structure (Figure 1.4A). The kinked structure is estimated by measurement of immersion depth between specific residues in the MPER and spin labeled lipid molecules using EPR technique (28).

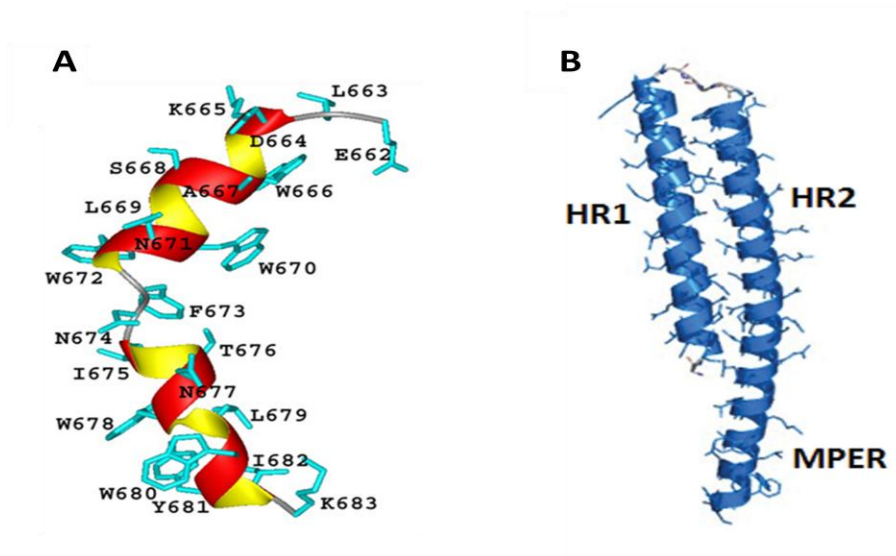


Figure 1.4. (A) NMR structure of MPER in DPC micelle showing the kink at residue W672-N674 (28). (B) Crystal structure of monomer NHR-CHR-MPER showing MPER as a continuous helix (29). The MPER is from L662-I675. The NHR and CHR are denoted as HR1 and HR2 respectively.

Recent high resolution crystal structure of MPER along with NHR-CHR shows helical structure (Figure 1.4B) (29). The kinked MPER peptide structure where the kink starts at W672, and the NHR-CHR-MPER crystal structure containing residue ⁶⁶²LELDKWASLWNWFDI⁶⁷⁵ shows complete helical structure with absence of a kink (29,30). X-ray crystal structure study of MPER when bound to broadly neutralizing antibodies 2F5 show that the N-terminal end ⁶⁶¹LELDKWAS⁶⁶⁸ within the MPER domain is in helical conformation (31). EPR and NMR structural studies show change in conformation of the MPER after binding to the antibodies. By EPR studies, the membrane insertion depth is determined for spin labeled MPER bound to 4E10 antibody with respect to POPC/POPG membranes. The difference in chemical shifts of amides of the residues of MPER in bound and unbound state suggests conformational change after binding to antibodies. Also NMR cross-saturation transfer experiments in which magnetization is

transferred from ^1H of methyl groups of the bound antibodies to the amides of residues of perdeuterated MPER further confirms the conformational change of the residues W672, F673, N674, I675, T676, N671, N677, L679, Y681, I682, and K683 of MPER in contact with the antibodies (28,30). The kink between W672-K683 changes to a continuous helical structure upon binding to antibody 4E10 (Figure 1.5A) as observed in the crystal structure of NHR-CHR-MPER (29). Upon Z13e1 binding, the dihedral angles of the MPER comprising L669-W672, and I675-Y681 changes significantly (Figure 1.5B) compared to dihedral angles of MPER in micelle bound structure in the absence of Z13e1. The backbone dihedral angles ϕ and ψ of the residues for the helix region comprising of L669-W672 and I675-Y681 changes from typical α -helical values of -57° and -47° upon binding to Z13e1. The change in ψ values of the residues L669-W672 is such that these residues adopt β -strand structure when bound to the antibody (30).

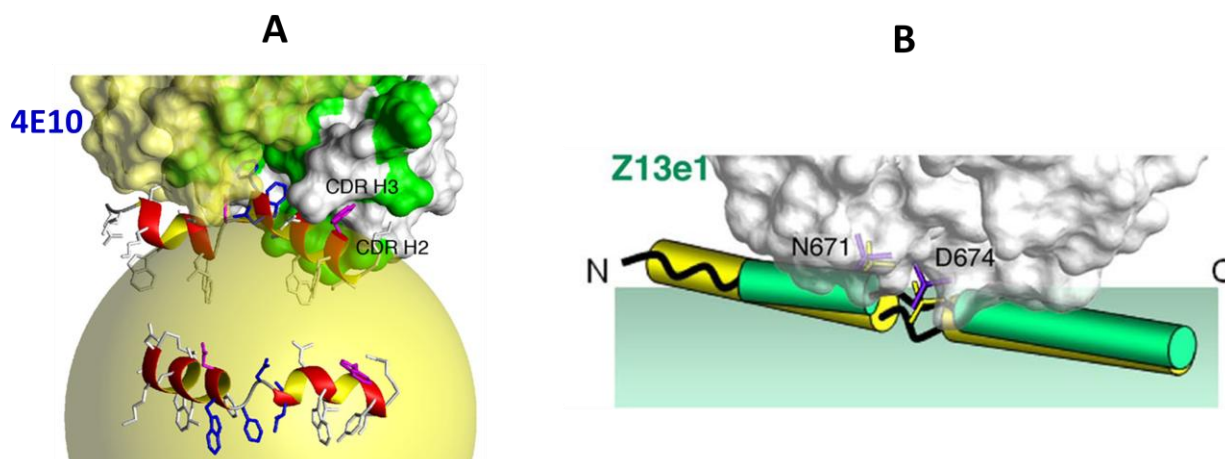


Figure 1.5. (A) Model of MPER only (bottom) and MPER bound to 4E10 (top). The residues W672-W676 shown in blue and N671 and W680 shown in magenta undergoes conformational change upon binding to the antibody (28). (B) Model of MPER bound to antibody Z13e1 showing conformational change of MPER helices with change in ψ values (30).

MPER is a Trp rich region. The mutation of the Trp residues reduces the fusogenic activity by affecting the cell-cell fusion (24). The synthetic peptide comprising part of CHR and MPER (amino acids 638-673) also known as T-20 or enfuvirtide inhibits syntitia formation and inhibits fusogenic activity of gp41 in the early stages of viral life cycle (32). Viral life cycle is comprised of two stages – (1) early stage comprise of viral attachment, transfer of viral genetic material (RNA) to the host cell, reverse transcription of viral RNA to viral DNA in the host cell cytoplasm, insertion of viral DNA to host nucleus, and integration of viral DNA to host cell DNA in the host nucleus and (2) late stage is comprised of transcription, translation, assembly of viral proteins and genetic material, budding, maturation, and new virus formation or syntitia formation with neighboring host cells.

As enfuvirtide is a peptide mimic of the CHR+MPER domain, it is proposed that enfuvirtide binds to NHR in the early steps of gp41 conformational change during membrane fusion (33). The binding of the broadly neutralizing antibodies 2F5, 4E10, and Z13 to the MPER epitopes reduces the Env mediated fusion activity (25,34,35). Therefore, binding of the antibodies or peptide enfuvirtide inhibits the fusion of viral and host cell membrane, thus blocking the very first step of viral life cycle. Studies of MPER and membrane interaction revealed membrane permeability and fusion (36-38). FP and MPER peptides interact and act synergistically to enhance membrane fusion (35,36,39,40).

1.1.4. Transmembrane Domain

The transmembrane (TM) domain anchors the gp41 in the viral membrane and comprises 22 amino acids. There is no structural detail of the TM region only, however, according to computer simulations data, it acquires α -helical structure (41) and has tendency to form a trimer (42). The effect of highly conserved TM in the viral membrane fusion has been studied which gives contradictory result. The mutations in the amino acid residues between 696-707 showed loss in fusion activity while the anchoring of the Env in the lipid bilayers remained same (43). Also substitution of the TM domain with C-terminal glycosyl- phosphatidylinositol lipid anchor reduced the fusion activity considerably (44,45). On the contrary, another study of fusion activity and infectivity by complete substitution of gp41 TM by cytoplasmic 180 residues at the C—terminal from cellular protein CD22 showed no change in function (46). Future studies on the role in fusion activity by TM domain would be interesting.

1.1.5. Cytoplasmic Endodomain

The gp41 cytoplasmic endodomain (Endo) is the C-terminal region after the TM domain and comprise of ~150 amino acids. To date there is no high resolution structural information of the cytoplasmic Endo domain. According to the computational modeled structure, gp41 Endo is a single-pass membrane protein (41). However, another monoclonal antibody binding study shows that Endo has at least one membrane spanning domain possibly exposing the epitope on the surface of the virion (47). The study of HIV-1 particle entry during virus-cell fusion by fluorescence assay showed immature virus particles were less active than mature virus particles. The inactivity of the immature virus particles is attributed to interaction of cytoplasmic Endo with matrix protein, and only dissociation of Endo from matrix protein and nucleocapsid protein leads to gp41-mediated fusion (48). Functional studies of the cytoplasmic endodomain by mutation reveal that it plays a significant role in incorporation of Env protein into the viral membrane (49,50). Interaction of cytoplasmic endodomain with viral matrix protein has been suggested by mutagenesis of the C-terminal Endo (51,52).

1.2. Possible Mechanism of Membrane Fusion

Stages of Membrane Fusion – Fusion of two membranes is a crucial step for viral entry into host cells. Membrane fusion proceeds through the following steps: (1) contact between outer leaflet membrane surface, (2) stalk formation, (3) hemifusion – fusion of outer leaflet of the two membranes, and (4) pore formation (Figure 1.4). The whole process of membrane fusion is an isothermal process (Figure 1.7). X-ray measurements have estimated the distance between biologically relevant membranes as 2-3 nm (53). Membranes with no net charge on the lipid

molecules are at equilibrium due to two interactions – Van der Waals attraction (long range) and repulsive interactions (short range). When the distance between the membranes is <3 nm, energy is required to overcome the repulsive force between the membrane outer leaflets of the two membranes (54). This repulsive force is also known as hydration force. In this membrane environment, the water molecules tend to bind to the polar head groups of the lipid molecules by hydrophilic interaction. For a membrane fusion between viral and host cell to occur, energy is required to displace the bound water molecules, also known as desolvation energy. The desolvation energy is reduced by contact between the two membranes comprising of a small area compared to flat membrane surfaces. This contact at a small area is known as stalk (Figure 1.6B). After the stalk formation, there is the stage which tends to rearrange the lipid molecules of the inner leaflet (Figure 1.6C). Energy is required to allow the lipid molecule acyl chains to move away from their hydrophobic environment and avoidance of water molecules, during the initial stages of stalk formation (Figure 1.6C and 1.7). This stage is followed by hemifusion stage (Figure 1.6D). In the hemifusion stage, the hydrophobic lipid acyl chains rearrange themselves by not getting exposed to water surface (54,55). The energy of the stalk formation is higher than hemifusion stage because more energy is required for the desolvation penalty to be paid during stalk formation (Figure 1.7). The calculated $(H_{\text{Stalk}} - H_{\text{Two Membrane}}) = \sim 25 \text{ kcal/mol}$ (54,56). Hemifusion stage is followed by pore formation (Figure 1.6E). Calorimetric experimental studies of gp41 comprising NHR-CHR shows $\Delta H_{\text{unfolding}} = \sim 65 \text{ kcal/mol}$ (57). It could be that the energy required for all the stages of membrane fusion is being supplied by the energy released during the conformational changes in the viral membrane protein. To date there is no experimental evidence of this energy exchange between membranes and gp41 protein folding, however, it

can be proposed that the folding of gp41 NHR-CHR ectodomain may be an exothermic process releasing ~65 kcal/mol of energy and that energy is utilized to overcome the repulsive force during membrane fusion. However, this hypothesis of membrane fusion caused by protein folding is not consistent with our observation of fusogenicity of already folded hairpin conformation of gp41 ectodomain. Gp41 ectodomain core sequence comprising of NHR-short loop-CHR (HP) forms stable folded hairpin conformation at pH 3.2 and also causes vesicle fusion. Therefore, it is hypothesized that the binding of the protein molecules to the lipid membrane perturbs the lipid molecules increasing the kinetic energy thereby causing local heating in the membrane. The amount of local heating is calculated as follows:

Using the equation $H = mc\Delta T$ (1) where H is Heat required

m is mass

c is specific heat

ΔT is change in temperature

There is electrostatic interaction between the protein molecules and lipid molecules in membrane. Considering each protein molecule as a cylinder, then the number of lipid molecules in contact is calculated as

For Hairpin protein, $r = 6 \text{ \AA}$, and $h = 50 \text{ \AA}$,

Surface area of each protein molecule = $2\pi r(r+h) = 2112 \text{ \AA}^2$

Possibly half of the protein molecule surface will be in contact with the lipid molecules. Area of each lipid headgroup is $\sim 70 \text{ \AA}^2$.

Therefore, number of lipid molecules in contact with protein molecule = $1056 \text{ \AA}^2 / 70 \text{ \AA}^2 = 15$

Energy associated with the perturbation of 15 lipid molecules is

$$E = (k_e z_1 z_2) / \epsilon r \dots\dots\dots(2) \quad \text{where } z = \text{charge on molecules}$$

r = distant between lipid and protein molecules

k_e = Coulomb's constant

ϵ = relative permittivity of lipid membrane

However, for ionic interaction to occur between the protein molecules and lipid molecules, we need to estimate the electrostatic energy between the charged molecules only. Only 20% of the lipid molecules in vesicles (according to our sample preparation) are negatively-charged at physiologic pH, therefore, from 15 lipid molecules only 3 will efficiently cause electrostatic interaction with charged protein molecules.

The charge of HP or HM is +2 at physiologic pH.

We are considering the protein and lipid interaction on the outer surface of the membrane leaflet, we therefore, consider the dielectric constant near the lipid head groups, which is ~ 30 (58).

$$E = [(8.968 \times 10^9 \text{ J m/C}^2) \times (1.602 \times 10^{-19})^2 \text{ C}^2 \times 3 \times 2] / 30 \times 5 \times 10^{-10} \text{ m}$$

$$E = 1 \times 10^{-19} \text{ J} \dots\dots\dots(3)$$

This energy eq. (3) is being distributed among 15 lipid molecules as the protein molecule is in contact with the same number of lipid molecules.

The volume of each lipid molecule is calculated by considering each lipid molecule as a cylinder, and radius as 2.4 Å. Therefore, Volume_{lipid} = 4.5 cm³(4)

Substituting values (3) and (4) in eqn. (1) we get:

$$\Delta T = (1 \times 10^{-19} \text{ J}) / \{[(4.5 \times 10^{22} \text{ g}) \times 15] \times [(4.2 \text{ J/K.g})]\} = 3.6 \text{ K} \sim 4 \text{ K} \dots\dots\dots(5)$$

Thus, there could be increase in temperature by ~ 4 K in the lipid membrane due to interaction with one protein molecule.

Considering the exothermic energy due to folding of HP as ~ 65 kcal/mol, we can calculate the local heating in the lipid membrane each protein molecule could cause.

$$E = 65 \text{ kcal/mol} = 4.5 \times 10^{-19} \text{ J} \dots\dots\dots (6)$$

Substituting the value from eq (6) in eq (1), we get

$$\Delta T = (4.5 \times 10^{-19} \text{ J}) / \{[(4.5 \times 10^{22} \text{ g}) \times 15] \times [(4.2 \text{ J/K.g})]\} = 15.9 \text{ K} \sim 16 \text{ K} \dots\dots\dots(7)$$

Therefore, from equations (5) and (7) we observe that the difference in temperature due to local heating caused by membrane perturbation by each protein molecule is 4 times less compared to the difference in temperature that could be caused by the energy released by folding of each protein molecule.

Due to this increase in the local temperature, the reaction rate increases (equation 6.) favoring the forward reaction towards membrane fusion.

According to Arrhenius equation:

$$k = Ae^{-E_a/RT} \dots\dots\dots (8) \quad \text{where } k \text{ is rate constant}$$

$E(A)$ is Activation energy

R is gas constant

T is Temperature in Kelvin

A is frequency factor constant

$$\ln k = \ln A - (E_a/RT) \dots\dots\dots (9)$$

For temperatures T_1 (initial) and T_2 (final), the reaction rates would be k_1 and k_2 .

Equation (9) becomes:

$$\ln (k_2/k_1) = (E_a/R)[(1/T_1) - (1/T_2)] \dots\dots\dots (10)$$

Considering the activation energy for stalk formation during membrane fusion as ~ 25 kcal/mol (56) we have:

$$E_a = 25 \text{ kcal/mol} = 1.7 \times 10^{-19} \text{ J}$$

$$T_1 = 310 \text{ K}$$

(I) Considering the $\Delta T = 4 \text{ K}$, (from eq. 4)

$$T_2 = 314 \text{ K}$$

$$R = 8.314 \text{ J/K mol}$$

Substituting all the values in eq. (10) we get

$$\ln (k_2/k_1) = (1.7 \times 10^{-19} \text{ J} / 8.314 \text{ J/K mol})[(1/310 \text{ K}) - (1/314 \text{ K})]$$

$$\ln (k_2/k_1) = [(1.7 \times 10^{-19} \text{ J} / (8.314 \text{ J/K mol}) \times (\text{mol}/6.022 \times 10^{23}))][(1/310 \text{ K}) - (1/314 \text{ K})]$$

$$\ln (k_2/k_1) = 0.51$$

$$k_2/k_1 = 1.7 \approx 2$$

$$\text{or, } k_2 = 2k_1 \dots\dots\dots(11)$$

Therefore, the reaction rate for membrane stalk formation doubles due to local heating caused by perturbation of lipid molecules by folded hairpin gp41 ectodomain.

(II) Now considering $\Delta T = 16 \text{ K}$, (from eq. 7)

$$T_2 = 326 \text{ K}$$

$$R = 8.314 \text{ J/K mol}$$

Substituting all the values in eq. (10) we get

$$\ln (k_2/k_1) = (1.7 \times 10^{-19} \text{ J} / 8.314 \text{ J/K mol})[(1/310 \text{ K}) - (1/326 \text{ K})]$$

$$\ln (k_2/k_1) = [(1.7 \times 10^{-19} \text{ J} / (8.314 \text{ J/K mol}) \times (\text{mol}/6.022 \times 10^{23}))][(1/310 \text{ K}) - (1/326 \text{ K})]$$

$$\ln (k_2/k_1) = 1.95$$

$$k_2/k_1 = 7$$

$$\text{or, } k_2 = 7k_1 \dots\dots\dots(12)$$

Therefore, the rate after local heating increases by 7 times the initial rate of membrane fusion.

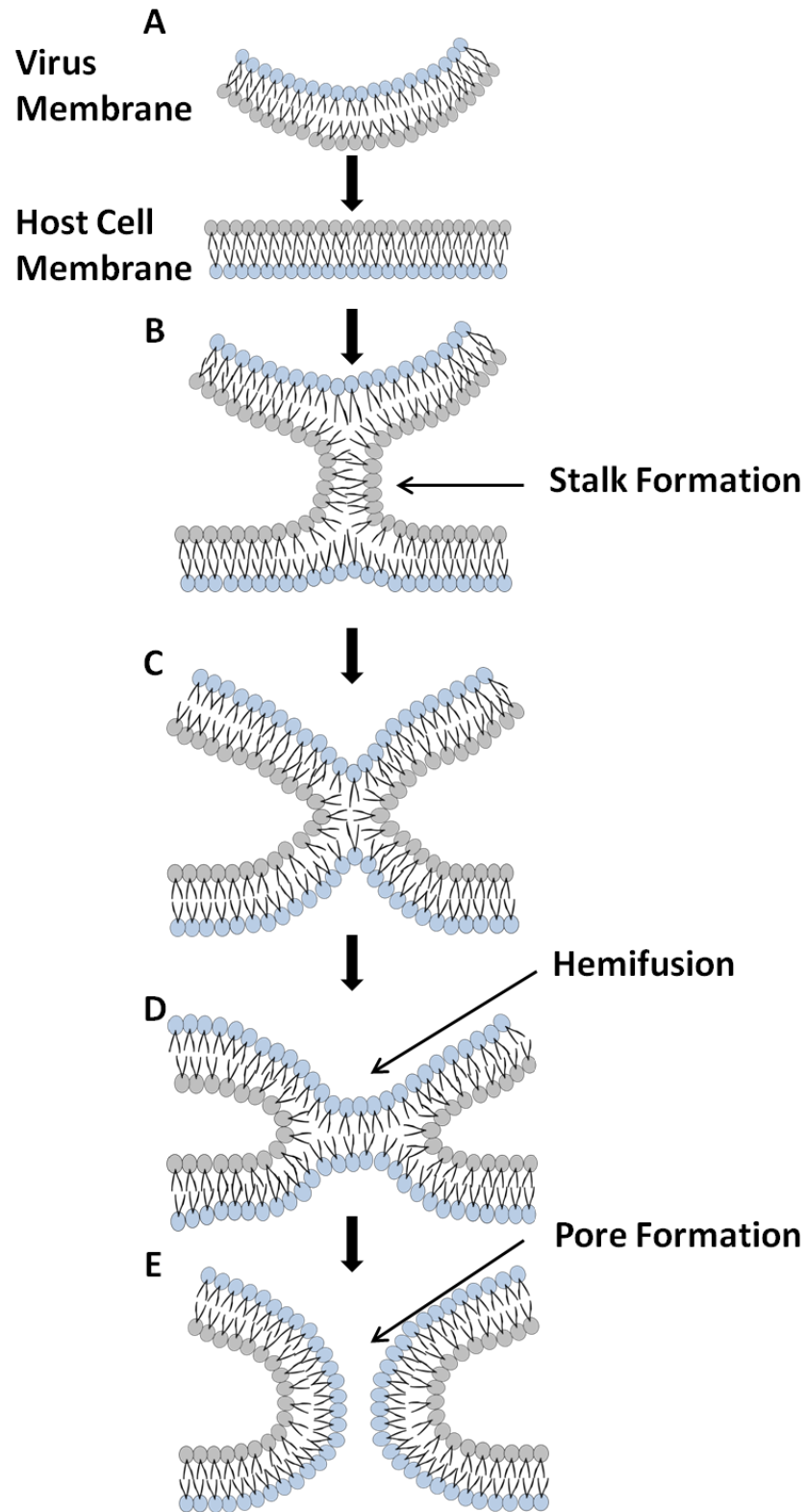


Figure 1.6. Stages of membrane fusion.

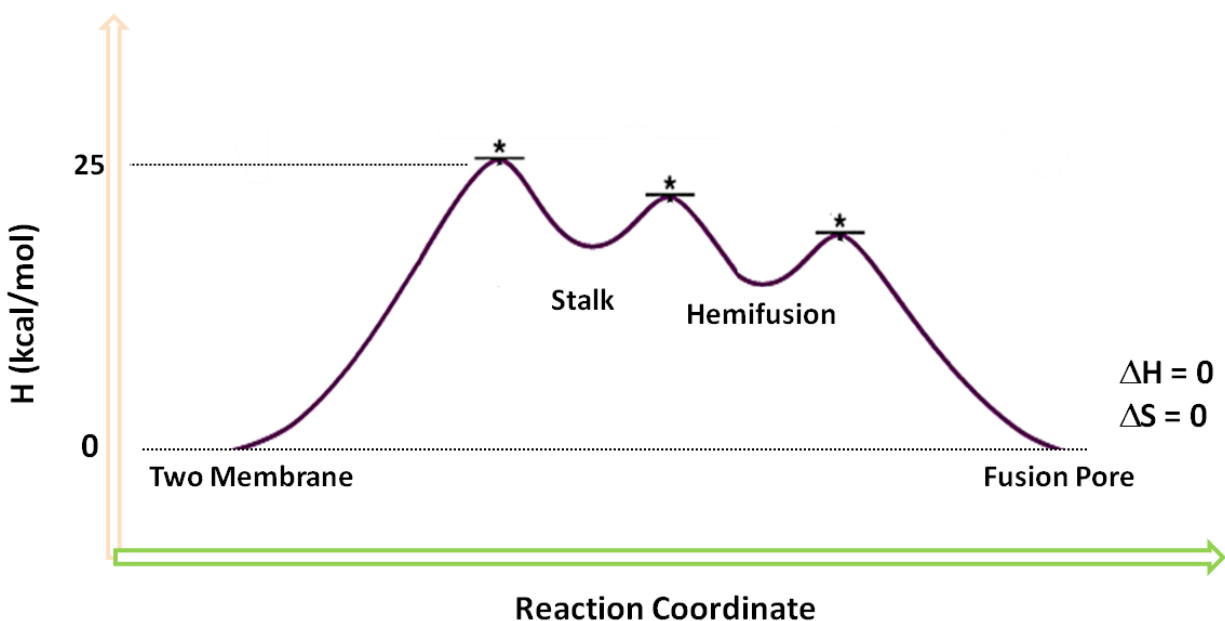


Figure 1.7. Proposed energy profile of membrane fusion caused by HIV-1 gp41 envelope protein. The maximum energy barrier for stalk formation is calculated as ~ 25 kcal/mol (Kuzmin, P. I. et al. PNAS (2001), 98). The energy barrier to reach hemifusion and pore formation is lower than energy barrier for stalk formation. The energy barrier between stalk and hemifusion stage is due to activation energy required to allow the lipid molecule acyl chains to move away from their hydrophobic environment and avoidance of water molecules. After hemifusion, there is complete fusion between the two membranes which lead to pore formation. In the whole

$$\text{process } H_{\text{Two membrane}} = H_{\text{Fusion pore}}$$

1.3. Mechanism of Fusion of HIV-1 and Host Cell Membranes Caused by Gp41

HIV-1 and host cell membrane fuse to enable the transfer of viral genetic material into the host cell and further use the host cell machinery to complete the life cycle and further propagation (59). The Env glycoprotein i.e. gp160 plays a significant role to mediate the fusion between the

two membranes. Gp160 comprise of non-covalently associated gp120 and gp41. Gp120 has strong affinity for CD4 receptors on T-cells or macrophages (60) and co-receptors CXCR4/CCR5 (1). After binding, gp120 undergoes conformational change and dissociates to fully expose fusion efficient gp41 (61). Recent cryoelectron tomography study of virion particles with Env bound to HIV neutralizing proteins showed conformational change of gp120 still attached to gp41 (62). Another cryoelectron tomography study of conformational change of gp120 upon binding with CD4 receptor and coreceptor ligand 17b showed gp140 (gp120+gp41 ectodomain truncated) intact structure (63). However, this gp140 structure is stabilized by disulfide bonds between gp120 and gp41 residues. Therefore, to date there is no clear structural evidence whether gp120 completely dissociates from gp41. Based on biochemical assays (61) it is proposed that after dissociation, gp41 plays the main role in the occurrence of fusion between the viral and host cell membrane. There are two proposed models for the fusion of membrane caused by gp41.

1.3.1. Model I

After the dissociation of gp120, gp41 undergoes a conformational change with exposure of FP towards the host membrane (Figure 1.8 in brackets). In this elongated intermediate stage, the gp41s are as a trimer bundle. This is followed by formation of SHB of the NHR and CHR with close apposition of N-terminal FP and C-terminal TM regions which consequently brings the two membranes close to each other. Viral and host cell membrane hemifusion followed by complete fusion, and the gp41 FP and TM now lie in the same membrane (Figure 1.6) (18). This model of membrane fusion caused by gp41 satisfies the energy requirements for the fusion

process to take place by the energy released due to folding of gp41 from intermediate state to final SHB folded state.

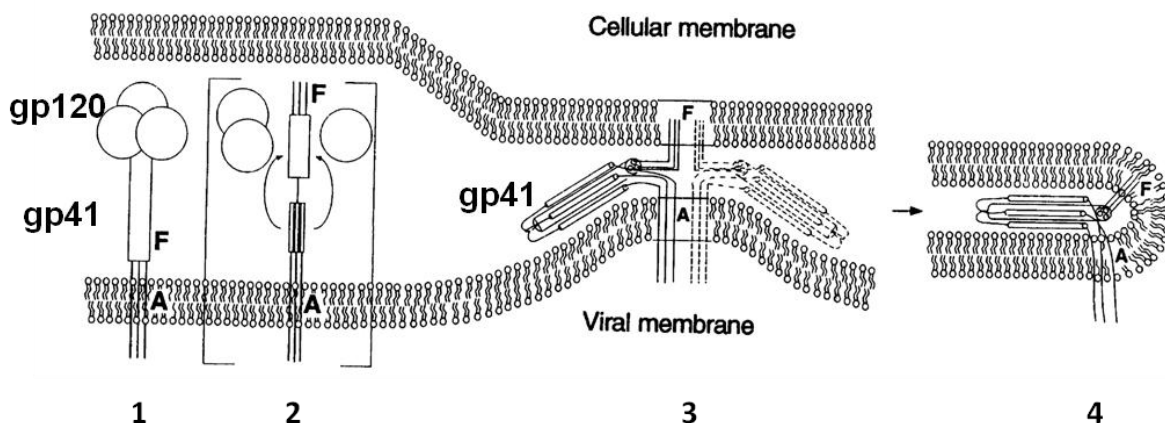


Figure 1.8. Model I: Schematic diagram of membrane fusion caused by gp41 (64). (1) Trimer gp120+gp41, (2) dissociation of gp120 from gp41 and exposure of gp41 FP, (3) SHB folding of gp41 and hemifusion of viral and host cell membrane, and (4) pore formation with close apposition of FP and TM of gp41. 'A' denotes viral membrane and 'F' denotes FP region.

1.3.2. Model II

According to this model, after the gp120 moves away from gp41, the hemifusion of membrane outer leaflet starts caused by membrane insertion of FP in PHI state. The hemifusion state is followed by fusion pore formation by folding of PHI to SHB (Figure 1.9) (65). The experimental evidence supporting the pore formation before SHB formation is shown by closing of pores in cell/cell fusion assays after the addition of NHR or CHR peptides (66). Experimental study of vesicle fusion assay of PHI like model at physiological pH and SHB show rapid and negligible fusion respectively (67).

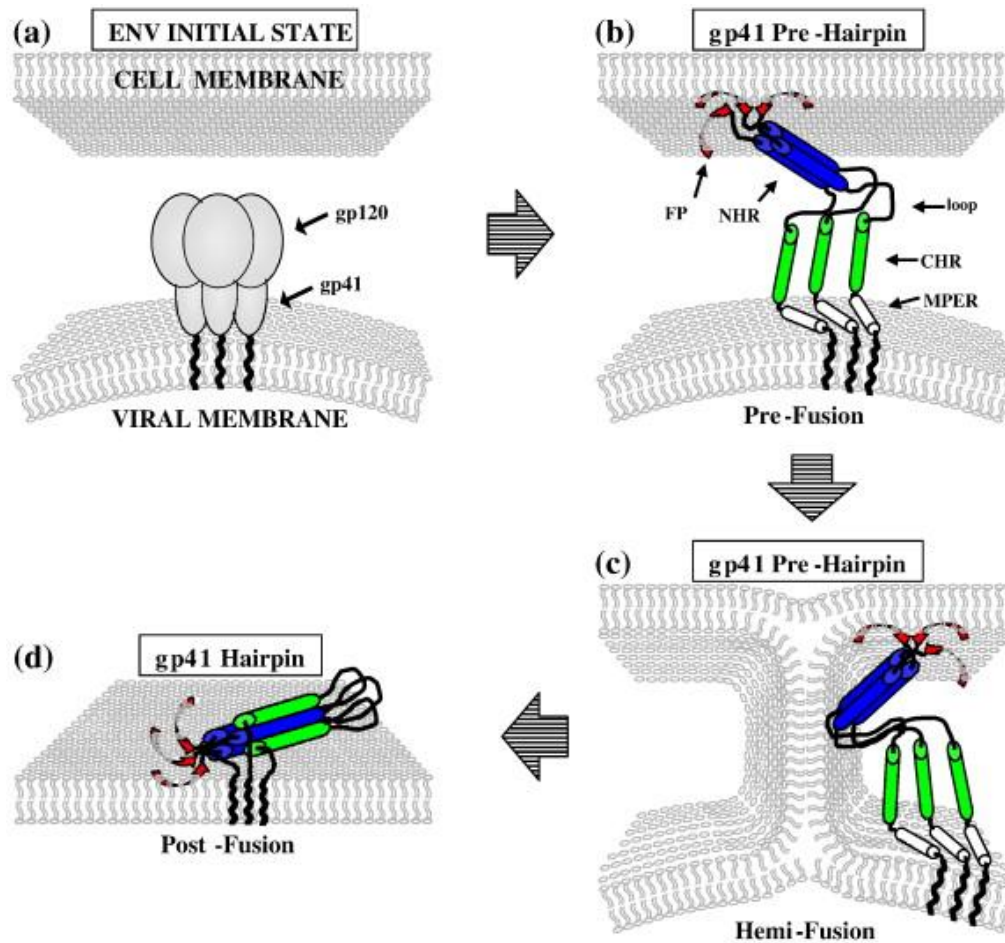


Figure 1.9. Model II. Stages of fusion: (a) gp120+gp41 trimer, (b) gp41 PHI formation with FP insertion into host membrane, (c) hemifusion caused by PHI, and (d) SHB formation (65).

1.3.3. Caveats of Both the Models

(1) In model I, the sequence of the stages of the protein folding is not consistent with the experimental observations of inhibition of fusion and syncytia formation by binding of NHR or CHR peptides (66) or CHR+MPER peptide (32). (2) In both the models I and II, it has been presumed that the PHI is a trimer structure with NHR forming the core. There is no structural information of the PHI state of gp41. Recent x-ray crystal and cryo-EM structure of gp140

(gp120+gp41 without the FP and MPER) presumably is the state prior to membrane fusion show loosely packed NHR domains. The gp140 trimer structure could be stabilized by disulfide bonds between gp120 and gp41 domains and I559P mutation in the NHR to keep the helices intact (68-71).

1.4. Stoichiometry of Fusion Proteins at the Site of Membrane Fusion

Structural and functional studies of gp120 and gp41 have led to understand the most probable fusion mechanism of viral and host cell membrane fusion. Based on fusion mechanism studies, a number of antiretroviral drugs are being used for the prevention of HIV infection. However, HIV strains mutate and eventually become resistant against these drugs. Mathematical estimation of number of trimers per virion concludes that there are 5-8 functional trimers per virion that effectively participate for viral entry (72,73). Cryoelectron microscopy tomography study on wild-type HIV-1 virions show ~14 Env protein spikes clustered on the viral membrane surface (74). Experimental studies on number of Env trimers which may be responsible for anchoring or binding of the viral Env to the host receptors have been done by cryo-electron tomography. According to cryo-electron tomography study, the Env trimers form a claw like structure when the HIV-1 particles interact with CD-4 positive cells. There are ~5-7 Env trimers which come in contact with target host cells (75). Recent solid state NMR studies of membrane associated FP in gp41 ectodomain FP-HP shows antiparallel β -sheet arrangement. It is proposed that this antiparallel arrangement is possible by interleaved FP strands of two gp41 trimers (65).

REFERENCES

REFERENCES

1. White, J. M., Delos, S. E., Brecher, M., and Schornberg, K. (2008) Structures and Mechanisms of Viral Membrane Fusion Proteins: Multiple Variations on a Common Theme. *Critical Reviews in Biochemistry and Molecular Biology* **43**, 189-219
2. Klasse, P. J. (2012) The molecular basis of HIV entry. *Cell Microbiol* **14**, 1183-1192
3. Miyauchi, K., Kim, Y., Latinovic, O., Morozov, V., and Melikyan, G. B. (2009) HIV Enters Cells via Endocytosis and Dynamin-Dependent Fusion with Endosomes. *Cell* **137**, 433-444
4. Melikyan, G. B. (2014) HIV entry: a game of hide-and-fuse? *Curr Opin Virol* **4**, 1-7
5. Mkrtchyan, S. R., Markosyan, R. M., Eadon, M. T., Moore, J. P., Melikyan, G. B., and Cohen, F. S. (2005) Ternary complex formation of human immunodeficiency virus type 1 Env, CD4, and chemokine receptor captured as an intermediate of membrane fusion. *Journal of Virology* **79**, 11161-11169
6. Grewe, C., Beck, A., and Gelderblom, H. R. (1990) Hiv - Early Virus-Cell Interactions. *J Acq Immun Def Synd* **3**, 965-974
7. Jaronec, C. P., Kaufman, J. D., Stahl, S. J., Viard, M., Blumenthal, R., Wingfield, P. T., and Bax, A. (2005) Structure and dynamics of micelle-associated human immunodeficiency virus gp41 fusion domain. *Biochemistry* **44**, 16167-16180
8. Gabrys, C. M., and Weliky, D. P. (2007) Chemical shift assignment and structural plasticity of a HIV fusion peptide derivative in dodecylphosphocholine micelles. *Bba-Biomembranes* **1768**, 3225-3234
9. Li, Y. L., and Tamm, L. K. (2007) Structure and plasticity of the human immunodeficiency virus gp41 fusion domain in lipid micelles and bilayers. *Biophys J* **93**, 876-885
10. Nieva, J. L., Nir, S., Muga, A., Goni, F. M., and Wilschut, J. (1994) Interaction of the Hiv-1 Fusion Peptide with Phospholipid-Vesicles - Different Structural Requirements for Fusion and Leakage. *Biochemistry* **33**, 3201-3209
11. Zheng, Z., Yang, R., Bodner, M. L., and Weliky, D. P. (2006) Conformational flexibility and strand arrangements of the membrane-associated HIV fusion peptide trimer probed by solid-state NMR spectroscopy. *Biochemistry* **45**, 12960-12975
12. Rafalski, M., Lear, J. D., and Degrado, W. F. (1990) Phospholipid Interactions of Synthetic Peptides Representing the N-Terminus of Hiv Gp41. *Biochemistry* **29**, 7917-7922

13. Yang, R., Prorok, M., Castellino, F. J., and Weliky, D. P. (2004) A Trimeric HIV-1 Fusion Peptide Construct Which Does Not Self-Associate in Aqueous Solution and Which Has 15-Fold Higher Membrane Fusion Rate. *JACS Communications* **126**, 14722-14723
14. FREED, E., DELWART, E. L., GARY L. BUCHSCHACHER, J., and PANGANIBAN, A. T. (1992) A mutation in the human immunodeficiency virus type 1 transmembrane glycoprotein gp41 dominantly interferes with fusion and infectivity. *Proc Natl Acad Sci USA* **89**, 70-74
15. Pereira, F. B., Goni, F. M., and Nieva, J. L. (1995) Liposome Destabilization Induced by the Hiv-1 Fusion Peptide Effect of a Single Amino-Acid Substitution. *Febs Lett* **362**, 243-246
16. Qiang, W., Sun, Y., and Weliky, D. P. (2009) A strong correlation between fusogenicity and membrane insertion depth of the HIV fusion peptide. *P Natl Acad Sci USA* **106**, 15314-15319
17. Chan, D. C., and Kim, P. S. (1998) HIV Entry and Its Inhibition. *Cell* **93**, 681-684
18. W, W., A, D., SC, H., JJ, S., and DC, W. (1997) Atomic structure of the ectodomain from HIV-1 gp41. *Nature* **387**, 426-430
19. Lu, M., Ji, H., and Shen, S. (1999) Subdomain folding and biological activity of the core structure from human immunodeficiency virus type 1 gp41: implications for viral membrane fusion. *J. Virol.* **73**, 4433-4438
20. Sackett, K., Nethercott, M. J., Epand, R. F., Epand, R. M., Kindra, D. R., Shai, Y., and Weliky, D. P. (2010) Comparative analysis of membrane-associated fusion peptide secondary structure and lipid mixing function of HIV gp41 constructs that model the early pre-hairpin intermediate and final hairpin conformations. *J. Mol. Biol.* **397**, 301-315
21. Roche, J., Louis, J. M., Grishaev, A., Ying, J. F., and Bax, A. (2014) Dissociation of the trimeric gp41 ectodomain at the lipid-water interface suggests an active role in HIV-1 Env-mediated membrane fusion. *Proc Natl Acad Sci USA* **111**, 3425-3430
22. Chan, D. C., Fass, D., Berger, J. M., and Kim, P. S. (1997) Core structure of gp41 from the HIV envelope glycoprotein. *Cell* **89**, 263-273
23. Muster, T., Steindl, F., Purtscher, M., Trkola, A., Klima, A., Himmler, G., Ruker, F., and Katinger, H. (1993) A Conserved Neutralizing Epitope on Gp41 of Human-Immunodeficiency-Virus Type-1. *Journal of Virology* **67**, 6642-6647
24. Salzwedel, K., West, J. T., and 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

25. Zwick, M. B., Labrijn, A. F., Wang, M., Spenlehauer, C., Saphire, E. O., Binley, J. M., Moore, J. P., Stiegler, G., Katinger, H., Burton, D. R., and Parren, P. W. H. I. (2001) Broadly neutralizing antibodies targeted to the membrane-proximal external region of human immunodeficiency virus type 1 glycoprotein gp41. *Journal of Virology* **75**, 10892-10905
26. Montero, M., van Houten, N. E., Wang, X., and Scott, J. K. (2008) The membrane-proximal external region of the human immunodeficiency virus type 1 envelope: Dominant site of antibody neutralization and target for vaccine design (vol 72, pg 54, 2008). *Microbiol Mol Biol R* **72**, 378-378
27. Kelsoe, G., Verkoczy, L., and Haynes, B. F. (2014) Immune System Regulation in the Induction of Broadly Neutralizing HIV-1 Antibodies. *Vaccines* **2**, 1-14
28. Sun, Z. Y. J., Oh, K. J., Kim, M. Y., Yu, J., Brusic, V., Song, L. K., Qiao, Z. S., Wang, J. H., Wagner, G., and 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
29. Shi, W., Bohon, J., Han, D. P., Habte, H., Qin, Y., Cho, M. W., and Chance, M. R. (2010) Structural Characterization of HIV gp41 with the Membrane-proximal External Region. *Journal of Biological Chemistry* **285**, 24290–24298
30. Song, L., Sun, Z.-Y. J., Coleman, K. E., Zwick, M. B., Gach, J. S., Wang, J.-h., Reinherz, E. L., Wagner, G., and Kim, M. (2009) Broadly neutralizing anti-HIV-1 antibodies disrupt a hinge-related function of gp41 at the membrane interface. *PNAS* **106**, 9057–9062
31. Julien, J. P., Bryson, S., Nieva, J. L., and Pai, E. F. (2008) Structural Details of HIV-1 Recognition by the Broadly Neutralizing Monoclonal Antibody 2F5: Epitope Conformation, Antigen-Recognition Loop Mobility, and Anion-Binding Site. *Journal of Molecular Biology* **384**, 377-392
32. Wild, C. T., Shugars, D. C., Greenwell, T. K., Mcdanal, C. B., and Matthews, T. J. (1994) Peptides Corresponding to a Predictive Alpha-Helical Domain of Human-Immunodeficiency-Virus Type-1 Gp41 Are Potent Inhibitors of Virus-Infection. *P Natl Acad Sci USA* **91**, 9770-9774
33. Matthews, T., Salgo, M., Greenberg, M., Chung, J., DeMasi, R., and Bolognesi, D. (2004) Enfuvirtide: The first therapy to inhibit the entry of HIV-1 into host CD4 lymphocytes. *Nat Rev Drug Discov* **3**, 215-225
34. Zwick, M. B., Jensen, R., Church, S., Wang, M., Stiegler, G., Kunert, R., Katinger, H., and Burton, D. R. (2005) Anti-Human Immunodeficiency Virus Type 1 (HIV-1) Antibodies 2F5 and 4E10 Require Surprisingly Few Crucial Residues in the Membrane-Proximal External Region of Glycoprotein gp41 To Neutralize HIV-1. *Journal of Virology* **79**, 1252–1261

35. Lorizate, M., Gomara, M. J., de la Torre, B. G., Andreu, D., and Nieva, J. L. (2006) Membrane-transferring sequences of the HIV-1 gp41 ectodomain assemble into an immunogenic complex. *Journal of Molecular Biology* **360**, 45-55
36. Suarez, T., Gallaher, W. R., Agirre, A., Goni, F. M., and Nieva, J. L. (2000) Membrane interface-interacting sequences within the ectodomain of the human immunodeficiency virus type 1 envelope glycoprotein: Putative role during viral fusion. *Journal of Virology* **74**, 8038-8047
37. Apellaniz, B., Nir, S., and Nieva, J. L. (2009) Distinct Mechanisms of Lipid Bilayer Perturbation Induced by Peptides Derived from the Membrane-Proximal External Region of HIV-1 gp41. *Biochemistry* **48**, 5320-5331
38. Apellaniz, B., Garcia-Saez, A. J., Nir, S., and Nieva, J. L. (2011) Destabilization exerted by peptides derived from the membrane-proximal external region of HIV-1 gp41 in lipid vesicles supporting fluid phase coexistence. *Bba-Biomembranes* **1808**, 1797-1805
39. Bellamy-McIntyre, A. K., Lay, C. S., Bar, S., Maerz, A. L., Talbo, G. H., Drummer, H. E., and Pountourios, P. (2007) Functional links between the fusion peptide-proximal polar segment and membrane-proximal region of human immunodeficiency virus gp41 in distinct phases of membrane fusion. *Journal of Biological Chemistry* **282**, 23104-23116
40. Lorizate, M., de la Arada, I., Huarte, N., Sanchez-Martinez, S., de la Torre, B. G., Andreu, D., Arrondo, J. L. R., and Nieva, J. L. (2006) Structural analysis and assembly of the HIV-1 gp41 amino-terminal fusion peptide and the pretransmembrane amphipathic-at-interface sequence. *Biochemistry* **45**, 14337-14346
41. Gallaher, W. R., Ball, J. M., Garry, R. F., Griffin, M. C., and Montelaro, R. C. (1989) A General-Model for the Transmembrane Proteins of Hiv and Other Retroviruses. *Aids Res Hum Retrov* **5**, 431-440
42. Kim, J. H., Hartley, T. L., Curran, A. R., and Engelman, D. M. (2009) Molecular dynamics studies of the transmembrane domain of gp41 from HIV-1. *Bba-Biomembranes* **1788**, 1804-1812
43. Owens, R. J., Burke, C., and Rose, J. K. (1994) Mutations in the Membrane-Spanning Domain of the Human-Immunodeficiency-Virus Envelope Glycoprotein That Affect Fusion Activity. *Journal of Virology* **68**, 570-574
44. Salzwedel, K., Johnston, P. B., Roberts, S. J., Dubay, J. W., and Hunter, E. (1993) Expression and Characterization of Glycophospholipid-Anchored Human-Immunodeficiency-Virus Type-1 Envelope Glycoproteins. *Journal of Virology* **67**, 5279-5288
45. Weiss, C. D., and White, J. M. (1993) Characterization of Stable Chinese-Hamster Ovary Cells Expressing Wild-Type, Secreted, and Glycosylphosphatidylinositol-Anchored

- Human-Immunodeficiency-Virus Type-1 Envelope Glycoprotein. *Journal of Virology* **67**, 7060-7066
46. Wilk, T., Pfeiffer, T., Bukovsky, A., Moldenhauer, G., and Bosch, V. (1996) Glycoprotein incorporation and HIV-1 infectivity despite exchange of the gp160 membrane-spanning domain. *Virology* **218**, 269-274
 47. Cleveland, S. M., McLain, L., Cheung, L., Jones, T. D., Hollier, M., and 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. *J Gen Virol* **84**, 591-602
 48. Wyma, D. J., Jiang, J. Y., Shi, J., Zhou, J., Lineberger, J. E., Miller, M. D., and 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
 49. Piller, S. C., Dubay, J. W., Derdeyn, C. A., and Hunter, E. (2000) Mutational analysis of conserved domains within the cytoplasmic tail of gp41 from human immunodeficiency virus type 1: Effects on glycoprotein incorporation and infectivity. *Journal of Virology* **74**, 11717-11723
 50. Yu, X. F., Yuan, X., Mclane, M. F., Lee, T. H., and Essex, M. (1993) Mutations in the Cytoplasmic Domain of Human-Immunodeficiency-Virus Type-1 Transmembrane Protein Impair the Incorporation of Env Proteins into Mature Virions. *Journal of Virology* **67**, 213-221
 51. Freed, E. O., and 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
 52. Dorfman, T., Mammano, F., Haseltine, W. A., and Gottlinger, H. G. (1994) Role of the Matrix Protein in the Virion Association of the Human-Immunodeficiency-Virus Type-1 Envelope Glycoprotein. *Journal of Virology* **68**, 1689-1696
 53. Rand, R. P., and Parsegian, V. A. (1989) Hydration Forces between Phospholipid-Bilayers. *Biochim Biophys Acta* **988**, 351-376
 54. Cohen, F. S., and Melikyan, G. B. (2004) The energetics of membrane fusion from binding, through hemifusion, pore formation, and pore enlargement. *J Membrane Biol* **199**, 1-14
 55. Chernomordik, L. V., Zimmerberg, J., and Kozlov, M. M. (2006) Membranes of the world unite! *J Cell Biol* **175**, 201-207

56. Kuzmin, P. I., Zimmerberg, J., Chizmadzhev, Y. A., and Cohen, F. S. (2001) A quantitative model for membrane fusion based on low-energy intermediates. *Proc Natl Acad Sci USA* **98**, 7235-7240
57. Sackett, K., Nethercott, M. J., Epand, R. F., Epand, R. M., Kindra, D. R., Shai, Y., and Weliky, D. P. (2010) Comparative Analysis of Membrane-Associated Fusion Peptide Secondary Structure and Lipid Mixing Function of HIV gp41 Constructs that Model the Early Pre-Hairpin Intermediate and Final Hairpin Conformations. *Journal of Molecular Biology* **397**, 301–315
58. Zhou, F., and Schulten, K. (1994) Molecular-Dynamics Studies of a Dipe Membrane Bilayer and Phospholipase-A2 Dipe Membrane Complex. *Biophys J* **66**, A399-A399
59. Grove, J., and Marsh, M. (2011) The cell biology of receptor-mediated virus entry. *J Cell Biol* **195**, 1071-1082
60. Sattentau, Q. J., and Weiss, R. A. (1988) The Cd4 Antigen - Physiological Ligand and Hiv Receptor. *Cell* **52**, 631-633
61. Moore, J. P., Mckeating, J. A., Weiss, R. A., and Sattentau, Q. J. (1990) Dissociation of Gp120 from Hiv-1 Virions Induced by Soluble Cd4. *Science* **250**, 1139-1142
62. Meyerson, J. R., Tran, E. E., Kuybeda, O., Chen, W., Dimitrov, D. S., Gorlani, A., Verrips, T., Lifson, J. D., and Subramaniam, S. (2013) Molecular structures of trimeric HIV-1 Env in complex with small antibody derivatives. *Proc Natl Acad Sci U S A* **110**, 513-518
63. Harris, A., Borgnia, M. J., Shi, D., Bartesaghi, A., He, H., Pejchal, R., Kang, Y. K., Depetris, R., Marozsan, A. J., Sanders, R. W., Klasse, P. J., Milne, J. L. S., Wilson, I. A., Olson, W. C., Moore, J. P., and Subramaniam, S. (2011) Trimeric HIV-1 glycoprotein gp140 immunogens and native HIV-1 envelope glycoproteins display the same closed and open quaternary molecular architectures. *Proc Natl Acad Sci USA* **108**, 11440–11445
64. Weissenhorn, W., Dessen, A., Harrison, S. C., Skehel, J. J., and Wiley, D. C. (1997) Atomic structure of the ectodomain from HIV-1 gp41. *Nature* **387**, 426-430
65. Sackett, K., Nethercott, M. J., Zheng, Z., and Weliky, D. P. (2014) Solid-State NMR Spectroscopy of the HIV gp41 Membrane Fusion Protein Supports Intermolecular Antiparallel β Sheet Fusion Peptide Structure in the Final Six-Helix Bundle State. *Journal of Molecular Biology* **426**, 1077-1094
66. Markosyan, R. M., Cohen, F. S., and Melikyan, G. B. (2003) HIV-1 Envelope Proteins Complete Their Folding into Six-helix Bundles Immediately after Fusion Pore Formation. *Molecular Biology of the Cell* **14**, 926–938

67. Sackett, K., Nethercott, M. J., Shai, Y., and Weliky, D. P. (2009) Hairpin Folding of HIV gp41 Abrogates Lipid Mixing Function at Physiologic pH and Inhibits Lipid Mixing by Exposed gp41 Constructs. *Biochemistry* **48**, 2714–2722
68. Julien, J. P., Cupo, A., Sok, D., Stanfield, R. L., Lyumkis, D., Deller, M. C., Klasse, P. J., Burton, D. R., Sanders, R. W., Moore, J. P., Ward, A. B., and Wilson, I. A. (2013) Crystal Structure of a Soluble Cleaved HIV-1 Envelope Trimer. *Science* **342**, 1477-1483
69. Harris, A. K., Bartesaghi, A., Milne, J. L. S., and Subramaniam, S. (2013) HIV-1 Envelope Glycoprotein Trimers Display Open Quaternary Conformation When Bound to the gp41 Membrane-Proximal External-Region-Directed Broadly Neutralizing Antibody Z13e1. *Journal of Virology* **87**, 7191–7196
70. Lyumkis, D., Julien, J. P., de Val, N., Cupo, A., Potter, C. S., Klasse, P. J., Burton, D. R., Sanders, R. W., Moore, J. P., Carragher, B., Wilson, I. A., and Ward, A. B. (2013) Cryo-EM Structure of a Fully Glycosylated Soluble Cleaved HIV-1 Envelope Trimer. *Science* **342**, 1484-1490
71. Sanders, R. W., Vesanen, M., Schuelke, N., Master, A., Schiffner, L., Kalyanaraman, R., Paluch, M., Berkhout, B., Maddon, P. J., Olson, W. C., Lu, M., and Moore, J. P. (2002) Stabilization of the soluble, cleaved, trimeric form of the envelope glycoprotein complex of human immunodeficiency virus type 1. *Journal of Virology* **76**, 8875-8889
72. Klasse, P. J. (2007) Modeling how many envelope glycoprotein trimers per virion participate in human immunodeficiency virus infectivity and its neutralization by antibody. *Virology* **369**, 245-262
73. Magnus, C., and Regoes, R. R. (2012) Analysis of the Subunit Stoichiometries in Viral Entry. *Plos One* **7**, e33441
74. Zhu, P., Liu, J., Jr, J. B., Chertova, E., Lifson, J. D., Grise', H., Ofek, G. A., Taylor, K. A., and Roux, K. H. (2006) Distribution and three-dimensional structure of AIDS virus envelope spikes. *Nature* **441**, 847-852
75. Sougrat, R., Bartesaghi, A., Lifson, J. D., Bennett, A. E., Bess, J. W., Zabransky, D. J., and Subramaniam, S. (2007) Electron Tomography of the Contact between T Cells and SIV/HIV-1: Implications for Viral Entry. *PLoS Pathogens* **3**, 0570-0581

CHAPTER 2

Materials and Methods

2.1. Site Directed Mutagenesis

Our aim was to design the gp41 Hairpin protein with the MPER included. For the addition of the MPER residues, insertion site directed mutagenesis was performed. Insertion site directed mutagenesis is initiated with a short primer which contains the required mutation. Primers were designed complementary to the template DNA sequence near the insertion site to enable hybridization of the primer to the gene of interest. Two primers i.e. forward and reverse were designed complementary to the gene of interest DNA double strand. The primers with the added nucleotides for insertion mutation elongate using *DNA polymerase*. The amplification of the required DNA was performed by polymerase chain reaction (PCR).

The various steps of PCR comprised of;

1. Initialization – The initialization is done by a hot start at ~95 °C for 5 min. This is done to activate the thermally stable *Pfu* DNA polymerase.
2. Denaturation – This short step of heating (~95 °C) is done for DNA melting of the template DNA of the template DNA by breaking the H-bonds between the nucleotides of the double strand DNA. In this step single strand template DNA is obtained.
3. Annealing – In this step the primers anneal with the single strand DNA. To encourage binding of the primers to the template, the temperature is lowered, typically ~5 °C of the melting temperature of the primers. It was taken care that the primers were designed to be GC rich to enable better binding to the template DNA. After primer-template DNA hybrid formation, synthesis of the whole DNA starts.

4. Elongation – After binding of the primer to the template DNA, the DNA polymerase elongates the DNA complementary to the template DNA by adding dNTPs. The optimum temperature is typically 72 °C.
5. Final Elongation – This step is also done at 72 °C to make sure that all the single stranded DNA are completely synthesized.

While performing insertion mutation, various annealing temperatures were applied during PCR. Additions of 3, 6, or 9 nucleotides could be done at one time. To verify whether the desired DNA has been produced or not, the final PCR product is run on agarose gel (electrophoresis) and compared with the DNA ladder.

2.2. Transformation

The PCR produced plasmid with the insertion of nucleotides was allowed to be taken up by artificially produced competent BL21(DE3) cells. The plasmid includes the Ampicillin resistant gene, the *lac* operon, and the gene which encodes gp41 HP with MPER. After the transformation, the cells were allowed to grow on agar plate. After overnight growth at 37 °C, typically one colony was grown in LB. Thus, several 100 mL flasks of LB were inoculated with single colony and grown overnight at 37 °C. Then the cells were centrifuged and plasmid extraction was done. The extracted plasmids were then sent for DNA sequencing to verify the mutation occurrence.

2.3. Expression of Gp41 Ectodomain HP Construct

The gp41 HP plasmid comprised of NHR, short engineered loop, CHR, and a 6-His tag. The insert was in the pGEM-t vector and Figure 2.1 displays the amino acid sequence with the DNA sequence. The amino acid sequences are from the HXB2 laboratory strain of the virus and are described using Env residue numbering. HP includes the NHR residues 535(M535C)-581, a non-native loop (SGGRGG), followed by CHR residues 628-666 and a non-native H₆ tag.

CTLTVQARQLLSGIVQQQNNLLRAIEAQQHLLQLTVWGIKQLQARILSG
GRGGWMEWDREINNYTSLIHSLIEESQNQQEKNEQELLELDKWHHHHH
H

atg tgc acg ctg acg gta cag gcc aga caa tta ttg tct ggt ata gtg cag cag cag aac aat ttg ctg agg
gct att gag gcg caa cag cat ctg ttg caa ctc aca gtc tgg ggc atc aag cag ctc cag gca aga atc ctg
tct ggt ggc cgt ggc ggt tgg atg gag tgg gac aga gaa att aac aat tac aca agc tta ata cac tcc tta
att gaa gaa tcg caa aac cag caa gaa aag aat gaa caa gaa tta ttg gaa tta gat aaa tgg cat cac cat
cac cat cac tga

Figure 2.1. Amino acid and DNA sequence of HP construct.

The expression system was BL21(DE3) *E. coli* cells. The cells were harvested in LB and induced with 2 mM IPTG for over-expression of the protein of interest. The HP protein has a tendency to get expressed in the form of inclusion bodies in the cell cytoplasm. Therefore, the purification was further preceded by inclusion body isolation and solubilization. The cells were collected by centrifugation of the culture at 9000g. To isolate the inclusion body, the cells were lysed using

sonicator. To prevent the protein of interest from degradation, protease inhibitor cocktail tablet was added in the stock lysis solution. At first the cells were lysed in PBS with 50% amplitude of sonication, and 0.5 s on and 0.5 s off for 1 min. This cycle was repeated 5 times with an alternating rest period of 1 min. This was done to avoid any heating of the protein sample. The inclusion bodies' comprising the HP protein did not dissolve in PBS, therefore, a milky white solution was obtained. The inclusion bodies were then collected by centrifugation at 48000g. This step of PBS solubilization is beneficial to dissolve the unwanted bacterial cell proteins. Thus, three rounds of PBS solubilization were done by lysis to eliminate most of the unwanted bacterial protein impurities. After three rounds of sonication in PBS, the cells were lysed in PBS with 6M GuHCl. The presence of 6M GuHCl solubilized the inclusion bodies completely. The sonication conditions used were 70 % amplitude with 0.8s on and 0.2s off for 1 min, with 1 min rest period. This cycle was repeated for 5 times. After the sonication, the solution appeared clear indicating the solubilization of the inclusion bodies in PBS with 6M GuHCl. The protein solution was then collected by centrifugation at 48000g which eliminated the cell membrane and debris as a small pellet. All the above mentioned procedure was done at pH 7.4. The HP protein inclusion bodies could be solubilized in various other conditions and pH as well; glacial acetic acid at pH 3.0, PBS with 8M Urea at pH 7.4, and PBS with 1% SDS at pH 7.4. The solubility of HP was concluded after observation of clear lysate solution after lysis in the respective buffers. The solubilized and purified SDS-PAGE gels of HP in glacial acetic acid (Figure 2.2) and 8M urea are shown below (Figure 2.3). For low pH in glacial acetic acid purification, RP-HPLC was used and for physiologic pH urea purification, IMAC was used.

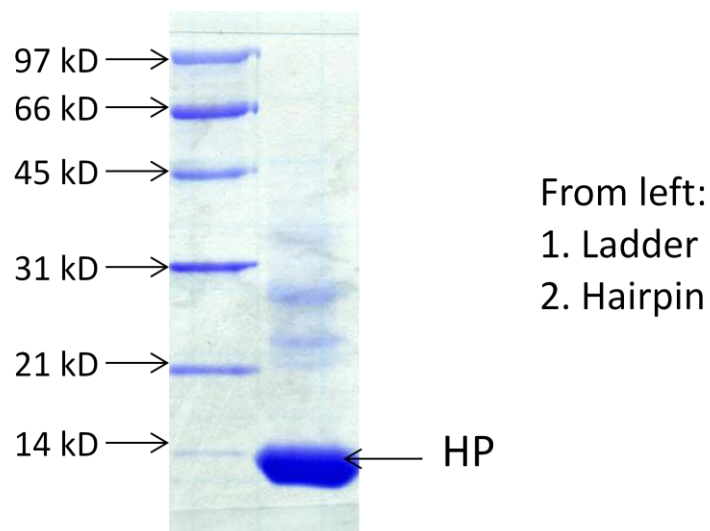


Figure 2.2. SDS-PAGE gel showing RP-HPLC purified HP in glacial acetic acid at pH 3.0

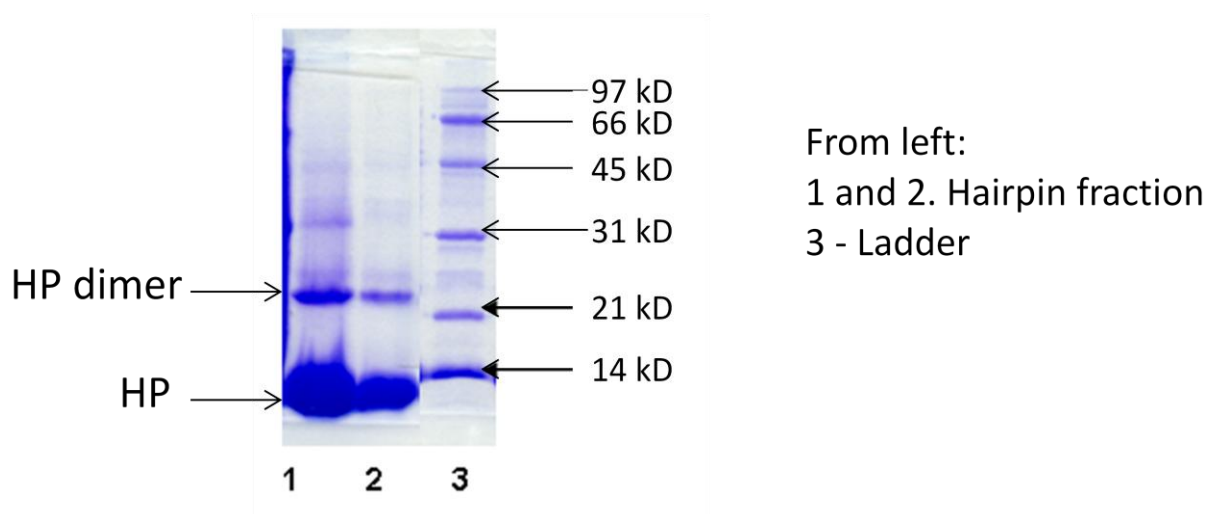


Figure 2.3. SDS-PAGE gel showing IMAC purified HP in the presence of urea at pH 7.4.

2.4. Detection of Expression of HM using Solid State NMR of Inclusion Bodies

The gp41 HM plasmid includes the NHR residues 535(M535C)-581, a non-native loop (SGGRGG), followed by CHR and MPER residues 628-683, and a non-native GGGGHHHH tag. The amino acid residue numbering is based on the HXB2 Env sequence numbering.

CTLTVQARQLLSGIVQQQNNLLRAIEAQQHLLQLTVWGIKQLQARILSG
GRGGWMEWDREINNYTSLIHSLIEESQNQKEKNEQELLELDKWASLWN
WFNITNWLWYIKGGGGHHHH

atg tgc acg ctg acg gta cag gcc aga caa tta ttg tct ggt ata gtg cag cag cag aac aat ttg ctg agg
gct att gag gcg caa cag cat ctg ttg caa ctc aca gtc tgg ggc atc aag cag ctc cag gca aga atc ctg
tct ggt ggc cgt ggc ggt tgg atg gag tgg gac aga gaa att aac aat tac aca agc tta ata cac tcc tta
att gaa gaa tcg caa aac cag caa gaa aag aat gaa caa gaa tta ttg gaa tta gat aaa tgg gcg tcg ctg
tgg aac tgg ttt aac atc acg aac tgg ctc tgg tac atc aag ggt ggt ggc ggc cat cac cat cac tga

Figure 2.4. Amino acid and DNA sequence of HM.

Initially it was not understood whether the HM protein was expressed or not. To understand the expression of HM, we implemented solid state NMR on the HM inclusion bodies. The cells expressing HM were grown in minimal media with addition of ^{13}C labeled Leu. The Addition of ^{13}C -Leu during the expression period resulted in ^{13}C -labeling of HM. Cells were lysed in PBS followed by centrifugation of the lysate and the harvested pellet was enriched in any HM IBs. The ^{13}C NMR spectrum of this pellet showed a prominent ^{13}C feature consistent with ~300 mg HM/L in IBs (Figure 2.5). The main bottleneck(s) to purified HM were therefore low solubilization of HM IBs and/or purification losses.

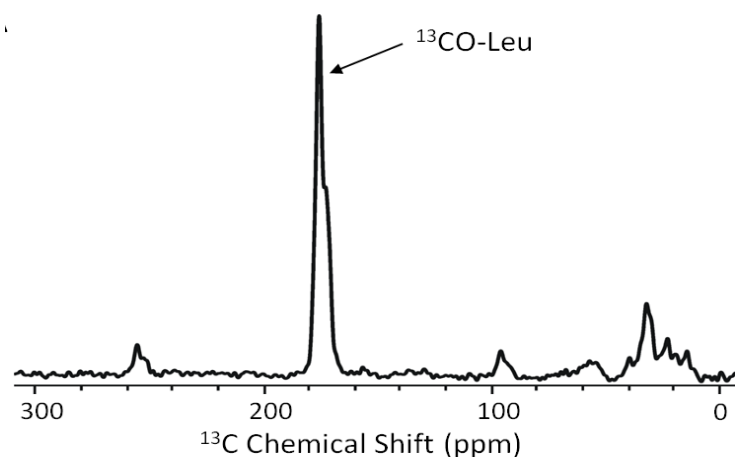


Figure 2.5. Chemical shift corresponding to ^{13}C -Leu.

Solid state NMR sample preparation and experimental conditions: BL21(DE3) cells expressing HM protein were grown overnight in 50 mL flask at 37 °C. After overnight growth, the cells were collected by centrifugation at 9000 g at 4 °C for 15 min. The cells were then resuspended into M9 minimal media also containing 250 μM of 50% v/v glycerol as carbon source, and 100 μL 1.0 M MgSO_4 . After one hour of growth, the cells were induced with 2 mM IPTG for over-expression of HM. Also, 10 mg each of all the essential amino acids were added and 10 mg of ^{13}C Leu was added to label the HM at Leu residues. Another addition of the amino acids was done after one hour. Then the cells were allowed to grow for three hours more, totaling growth time to be six hours. The cells were then collected by centrifugation. The cells were lysed with sonication at 60% amplitude with 0.6s on and 0.4s off. Lysing the cells in PBS at pH 7.4 formed a milky white solution with inclusion bodies suspension. The inclusion bodies were collected by centrifugation at 48000 g. The labeled HM inclusion body sample was then frozen in liquid N_2 and packed in 4 mm rotor using packing tool.

The NMR data were obtained with a 9.4T, triple resonance magic angle spinning (MAS) probe with a 4 mm rotor. The temperature was maintained at -20 °C with the flow of liquid N₂. Experimental parameters were as follows: (1) 8 kHz MAS frequency, and (2) 5 μ s 1H $\pi/2$ pulse, 2ms cross-polarization time with 50 kHz 1H field, and 70-80 kHz ramped 13C field. The spectrum was externally referenced to the methylene carbon of adamantane at 40.5 ppm.

2.5. Solubilization of HM

The solubilization steps of HM inclusion bodies were same as described in section 2.3 for HP solubilization. HM is most efficiently soluble in PBS with 6M GuHCl at pH 7.4. Other solubilizing agents at different pHs did not solubilize HM from inclusion bodies very well; PBS with 8M Urea at pH 7.4, 10 mM Tris-HCl at pH 9.0, glacial acetic acid at pH 3.0, PBS containing 1% SDS, 2X CMC of non-ionic detergents like triton, N-decyl- β -D-maltopyranoside, N-lauroyl sarcosine, and N-octyl- β -D-1-glucopyranoside (Table 2.1). Presence of SDS made HM fully soluble, however, the protein aggregated with the attempt of removal of SDS in the later stages of purification and refolding. The solubility of HM was checked by SDS-PAGE of the lysate in the respective buffers. The SDS-PAGE gel did not show clear difference between the solubility of various solubilizing agents (Figure 2.6). However, after sonification in the respective buffers the following were observed: (1) a clear solution in 1% SDS, (2) partially clear solution in 0.5% Sarkosyl, (3) cloudy solution in the presence of 0.4% triton and (4) cloudy solution in the presence of 8M urea. The extent of solubility of the protein construct was not determined by measuring the absorbance at 280 nm. Based on the visual clarity of the protein solution in respective buffers, the determination of best buffer for maximum solubilization was done.

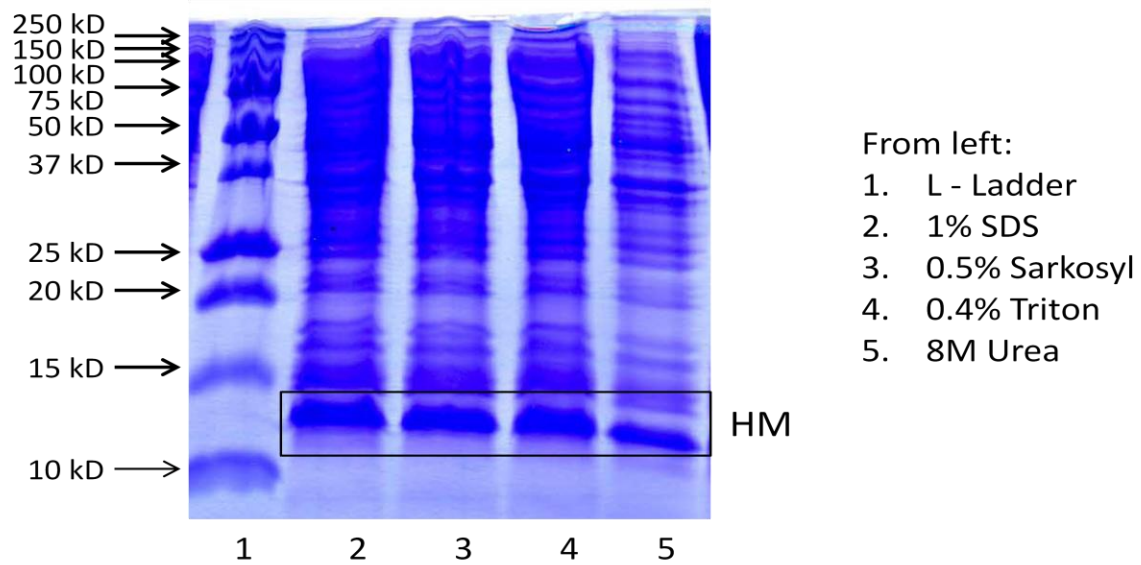


Figure 2.6. Solubilization of inclusion bodies in various solubilizing agents.

Solubilizing agent	Solubility	Purification
PBS	Not Soluble	
Tris	Not Soluble	
Urea	Soluble	Did not bind to His-select resin
SDS	Very Soluble	Protein precipitated out with the removal of SDS
N-Decyl- β -D-maltopyranoside	Sparingly	Did not bind to His-select resin
N-lauroyl sarcosine	Negligible	
N-Octyl β -D-1-glucopyranoside	Negligible	
Triton	Negligible	
Glacial acetic acid	Soluble	Low yield in RP-HPLC purification due to binding of the protein molecules to the column. The protein eluted out with varied retention times at high acetonitrile concentration.

Table 2.1. Solubilization of inclusion bodies containing HM in various solubilizing agents.

2.6. Purification of Gp41 Ectodomain Protein Constructs

2.6.1. HP

After solubilization as explained in section 2.5, the lysate containing HP in PBS + 6M GuHCl in the presence of protease inhibitor was allowed to mix in His-select cobalt (Co^{2+}) resin for 4 hours at 4 °C. HP has a C-terminal six residue His-tag for binding to the charged metal ion in the resin. To avoid non-specific binding to the resin, 1 mM imidazole was added to the buffer. After the binding of the protein with the Co^{2+} resin, the protein was isolated from the resin by gravity purification in a column at room temperature. To remove the impurities i. e. the other bacterial proteins, a gradient of imidazole concentration was used sequentially. Then finally at high imidazole concentration, i. e. 250 mM, the HP protein was eluted out. The concentration of eluted protein was measuring the absorbance of the Trp residues at 280 nm. The yield of HP was ~50 mg/L of culture.

2.6.2. HM

The purification of HM was done in the same procedure as HP explained in section 2.6.1. However, it was observed that HM did not bind to the Co^{2+} resin as efficiently as HP. This led us to believe that probably the six residue His-tag was not exposed for proper binding with the charged metal ion in the resin. At first a short two residue Gly tag was added before the N-terminal end of His-tag. However, it was observed that the protein construct still did not bind to the resin (Figure 2.7). Therefore, we added a four residue Gly-tag before the His-tag, i. e. at the N-terminal of the His-tag by site directed insertion mutation. With the addition of the Gly-tag, we observed that the protein HM bound very tightly to the resin and could not be eluted out easily by 500 mM imidazole containing elution buffer. To avoid this tight binding and the low

yield, we deleted two His residues from the six residue His-tag by site directed deletion mutation. Thus the HM protein sequence has the four residues Gly-tag and six residue His-tag. At this point, we observed that it was still challenging to solubilize HM in solubilizing agents like 3xCMC Sarkosyl (Figure 2.8). We observed that only 6M GuHCl efficiently solubilized HM and resulted in efficient purification (Figure 2.9). This Gly-His-tag helped to bind the protein to the resin and could be eluted out with high yield of protein. The yield HM was ~22 mg/L of culture.

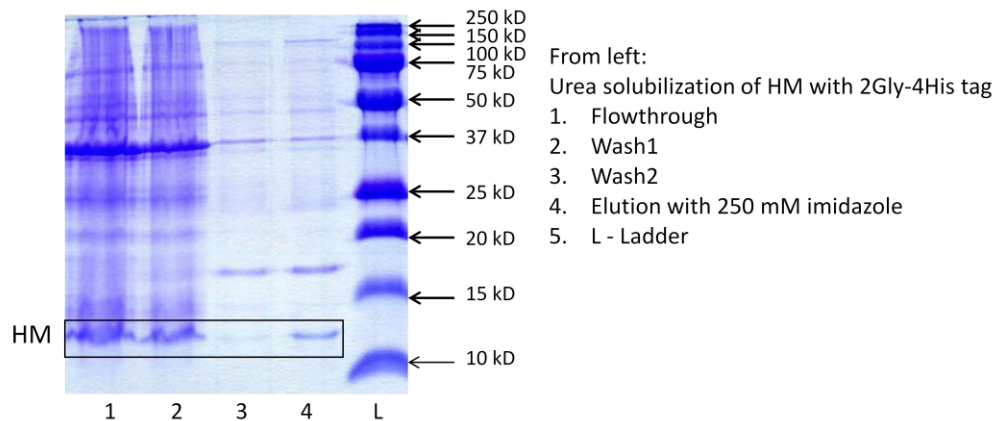


Figure 2.7. SDS-PAGE gel showing low solubilization and negligible binding of HM with 2Gly-4His tag in 8M urea with Co^{2+} resin.

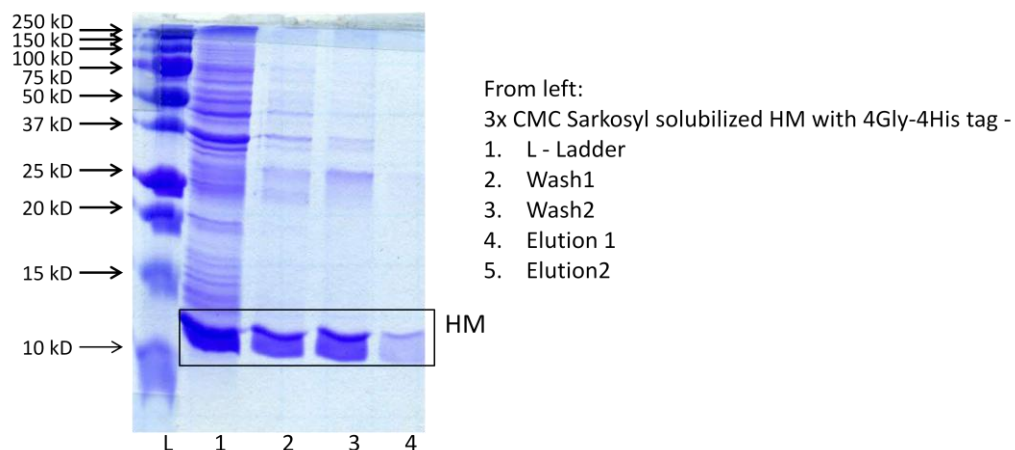


Figure 2.8. SDS-PAGE gel showing inefficient binding of HM with 4Gly-4His tag in 3xCMC

Sarkosyl with Co^{2+} resin.

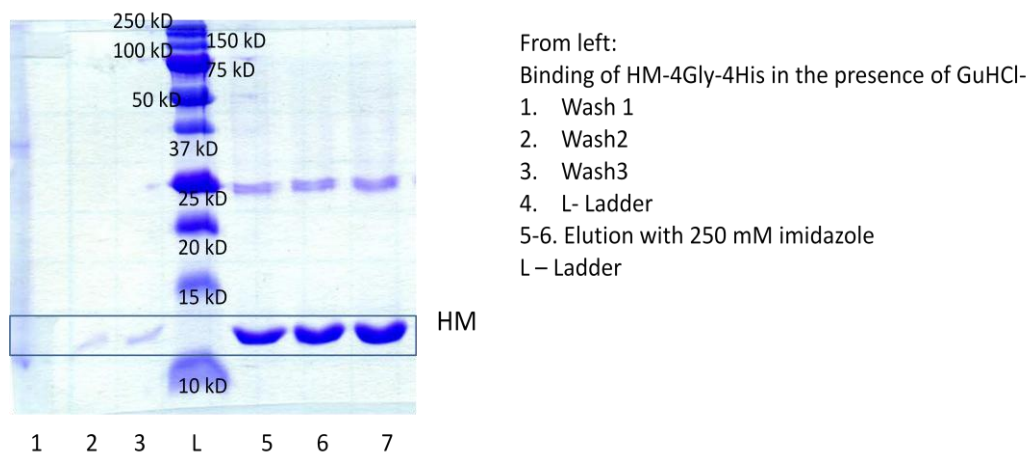


Figure 2.9. SDS-PAGE gel showing binding of HM with 2Gly-4His tag in 6M GuHCl with Co^{2+} resin and elution with high purity. The bands near 25 kD are dimer bands, confirmed by Western blot with anti-His antibody.

2.7. Western Blot Analysis

The protein purified was verified by Western blot technique. It is a very sensitive technique as it can detect presence of specific protein in a heterogeneous mixture of proteins. In this method, the protein sample is first separated by SDS-PAGE. Then the protein bands are transferred to nitrocellulose membrane. The membrane with protein bands was allowed to stain with antibodies which can specifically bind to the protein of interest. In our work, the HP and HM protein samples has a His-tag at the C-terminus. Therefore, we used anti-His-HRP conjugated antibody. The HRP conjugate facilitates the observation by chemiluminescence by activating the chemiluminescent substrate. To observe binding of the HIV-1 broadly neutralizing antibodies 4E10 and 2F5 to the MPER epitopes in HM and partial epitope in HP, the antibodies were allowed to bind to the protein of interest by Western blot. The antibodies 4E10 and 2F5 were not conjugated to HRP as mention earlier in this section. Therefore, a secondary antibody Goat Anti-Hamster IgG (H+L) HRP conjugate was used to bind to the 4E10 and 2F5.

2.8. Peptide Synthesis of N-terminal Fusion Peptide Region of Gp41 Ectodomain

Previous studies have shown that there is less protein expressed with the FP region included in FP-HP construct (1) . The FP region is a hydrophobic domain and thus could easily get aggregated in the solution. Also the FP has affinity to bind to the membrane and so binding of FP to *E.coli* membranes during purification process leads to more difficulties in the isolation process. Therefore, the ectodomain construct of gp41 can be designed by ligating (details of ligation in section. 2.9) the chemically synthesized FP with the expressed HM. The wild type sequence of HIV-1 HXB2 strain envelope protein from 512-534 (S534A)-thioester linker (FP23 linker) sequence of FP was synthesized by *t*-Boc synthesis. Solid phase peptide synthesis known

as – *tert*-Butoxycarbonyl (*t*-Boc) method can be done based on the principles of the work of Merrifield (2,3). According to this method there is sequential building of peptide chains with the peptides being anchored to a solid insoluble resin support. To the resin (solid phase support) the first amino acid is covalently bonded with a C-terminal linker group and the N-terminus is protected. This is followed by series of washes to remove the contaminants and deprotect the N-terminus of the attached amino acid. Then a new amino acid is added with an unprotected C-terminal to get attached to the N-terminal of the previous amino acid and build up the sequence. This whole cycle is repeated until the required sequence is created (2). The C-terminal was modified consisting of a thioester. The whole process of *t*-Boc synthesis is shown in Figure 2.10. The prepared peptide can be cleaved from the resin by breaking the covalent bond between the resin and the first amino acid using hydrofluoric acid. The synthesized peptide was sent for cleavage from the resin to Midwest Bio-tech Inc. After the cleavage, it was purified by RP-HPLC and MW was verified by mass spectrometry.

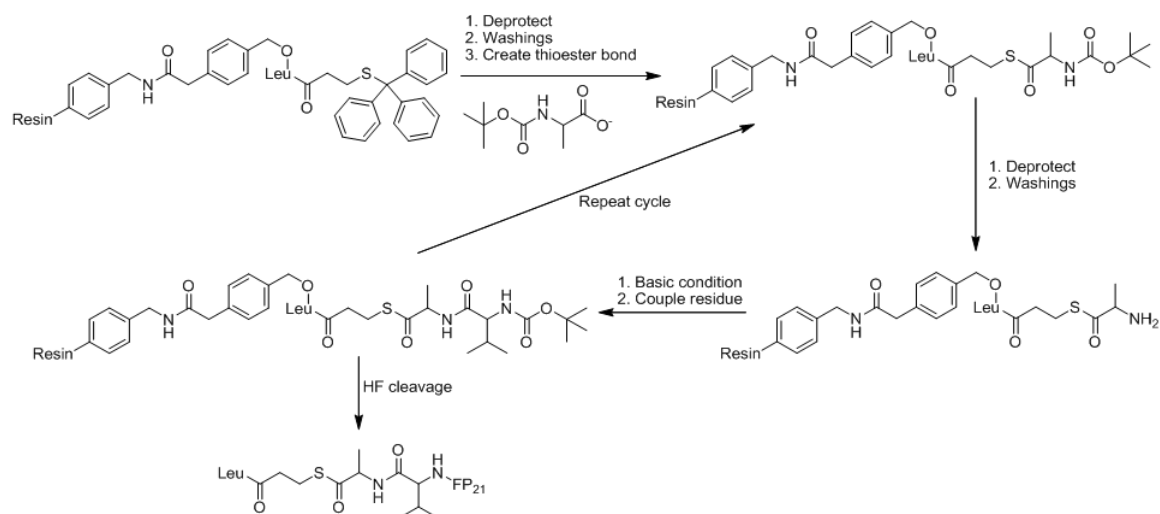


Figure 2.10. Schematic representation of *t*-Boc solid phase peptide synthesis.

From 1 g of Leu-Pam resin, 750 mg of crude peptide was obtained, and after an initial purification we expect to obtain 240 mg of pure peptide from this synthesis. There are some disadvantages of solid phase peptide synthesis, such as, as the chain grows in length, the residues tend to form secondary structure and fold. This might hinder the free N-terminal of the attached amino acid where the next residue would be added.

2.9. Production of Full-Length Gp41 Ectodomain Constructs by Native Chemical Ligation

Native chemical ligation (NCL) is an efficient technique to prepare larger protein constructs (1). This method allows the synthesis of larger constructs by allowing ligation of synthetic peptides to an expressed protein (4,5). The peptide could be synthesized with *t*-Boc solid phase peptide synthesis (details in the section 2.8). In this synthesis the resin used consists of thioester which would get covalently bonded to the C-terminal amino acid. The principle (4) on which native

chemical ligation works is the synthetic peptide with the thioester at the α -carboxyl group undergoes nucleophilic attack by the side chain of the N-terminal Cys residue of the amino terminal of the expressed Hairpin protein. The thioester ligated product initially undergoes rapid intramolecular reaction due to the favorable five-membered ring geometric arrangement of the α -amino group of the second peptide (expressed HM). This finally yields a product with a native peptide bond at the site of ligation. Both the ligated peptides (synthetic peptide + expressed protein) are in unprotected form and needs no further manipulation. There are several factors on which the rate of ligation depends: (1) the nature of the thiol leaving group, and (2) side chain steric conditions (3,5). The alkyl thioesters which are used in solid phase peptide synthesis are comparatively less reactive. This encourages the use of catalyst in this reaction and the rate of the reaction depends on the choice of the catalyst also. It has been shown that 4-mercaptophenyl acetic acid (MPAA) catalyses the ligation reaction to promote the *in situ* formation of the more reactive thioester moiety (3,5) (Figure 2.11).

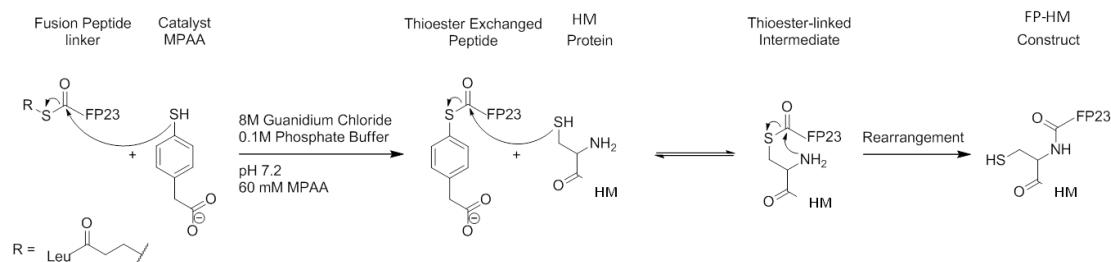


Figure 2.11. Mechanism of native chemical ligation. The FP23 that is synthesized by t-Boc SPPS is shown having a thiol on the C-terminus. To make the reaction more rapid, a catalyst MPAA is used. MPAA undergoes thioester exchange, and the Hairpin with the N-terminal Cys can easily form thioester-linked intermediate. Finally the more stable product, i.e., the peptide bond is formed.

In order to obtain the FP-HM construct, the expressed HM protein was ligated with FP. The FP was dissolved in the ligation buffer (8M Guanidine chloride, 0.1 M phosphate buffer, and pH-7). The peptide was reacted with 4-mercaptophenyl acetic acid (MPAA) catalyst to modify the thioester linked to the C-terminus of the fusion peptide (5,6) and the mixture was incubated for 30 min. Similarly, HM was dissolved in the ligation buffer including Tris(2-carboxyethyl)-phosphine (TCEP) (2mM) to maintain cysteine residue in reduced state, and the two solutions were mixed together. The reaction was allowed to proceed for 2 days in an inert (Ar) atmosphere and ambient temperature. This was followed by purification by RP-HPLC with a C4 column (6). Three peaks comprising of FP23, HM, and FP-HM were collected and MW was verified by mass spectrometry. Acetonitrile was removed from the eluted FP-HM fraction and allowed to concentrate.

2.10. Stabilization of Gp41 Ectodomain Protein Constructs

The refolding and stabilization of the purified protein were done by dialyzing against 10 mM or 50 mM sodium formate buffer at pH 3.2 containing 150 mM NaCl + 0.2 mM TCEP reducing agent at 4°C. The HP and HM protein were diluted to ~0.1 mg/mL concentration and then subjected to dialysis for removal of the 6M GuHCl solubilizing agent. After at least three buffer exchanges, the protein sample was concentrated to ~1 mg/mL. The solubility of HP and HM at low pH 3.2 without any solubilizing agent was verified by centrifugation and then measuring the absorbance of the supernatant at 280 nm. The secondary structure of the refolded protein constructs were verified by measuring the mean residue molar ellipticity by circular dichroism (CD). HP and HM protein constructs were also dialyzed against phosphate buffers ranging from pH 5-7.4 and Tris-HCl buffer at pH 9.0. However, the proteins aggregated indicating no solubility. Aggregation of the protein constructs was concluded by centrifuging the sample in respective buffer at 16000g for 5 min and measuring the protein concentration in the supernatant. One of our aims was to attain the gp41 ectodomain constructs HP and HM in physiological pH. The solubility of HP and HM was also tested in the presence of detergent like n-dodecylphosphocholine (DPC). It was observed that even in the presence of DPC, HP and HM were soluble at low pH ~4.0 and not at pH 7.4. We therefore studied the effect of solubilizing agent to solubilize at physiological pH. For this purpose 6M GuHCl was used. In the presence of 6M GuHCl, both the HP and HM were soluble at physiological pH 7.4 as well as low pH 3.2. The global secondary structure of the HP and HM showed folded structure i. e. α -helical. The α -helical curve has expected minima at 222 nm and the minimum at 208 nm is distorted due to the GuHCl effect.

2.11. Determination of Secondary Structure of Gp41 Ectodomain Constructs

2.11.1. Circular Dichroism

CD is a very useful spectroscopic tool to verify global secondary structure of large biological molecules. In this technique the difference in absorption of left and right circularly polarized light by chiral molecules is recorded and plotted against the wavelength. Percent helicity can be calculated from the mean residue molar ellipticity (θ_{222}) at wavelength 222 nm (minima). θ_{222} of -33,000 deg cm² dmol⁻¹ at wavelength 222 nm is considered to be 100% (7). The thermal stability of the protein constructs can also be measured by observing the CD signal at 222 nm as a function of temperature with 2 °C steps between 10 °C – 100 °C.

2.11.2. Solid State Nuclear Magnetic Resonance

X-ray crystallography and solution state NMR are powerful techniques which generally give high resolution structures. There are many x-ray crystallography (8-12) and liquid state NMR (13) structures of different parts of the gp41. However, there are no structures of the full length gp41 in physiologically relevant membrane environment. X-ray crystallography and solution state NMR generally do not allow the study of membrane proteins in membrane environment. Gp41 is a membrane protein and it would be ideal to study the structure in natural membrane environment. Solid state NMR is an ideal technique that can be used for studying membrane proteins structure in membrane mimetic environment. Our group has been successful in

determining structures of the fusion peptide of the gp41 in membranes by solid state NMR (14-17).

Analysis of the chemical shifts of the α -carbon, β -carbon, or the carbonyl carbon (CO) for a specific amino acid residue can reveal the structural information for that particular residue in the context of the protein. The chemical shift of each of these carbons is correlated with the likely residue conformation to get the information whether the residue is in the helical, and β -strand conformation (18).

Rotational echo double resonance (19) (REDOR) is a high-resolution solid state NMR experiment which can be used for observing the chemical shift of a specific carbonyl carbon in the entire protein molecule. In order to observe a specific site in the protein, a unique sequential pair should be chosen. In our case, two schemes of labeling will be performed. (i) Labeling of synthesized FP: For example, to have ^{13}C labeled leucine and ^{15}N labeled phenylalanine in the FP, accordingly labeled amino acids can be used during peptide synthesis. (ii) Labeling of expressed HP-MPER: For instance, ^{13}C labeling all of the carbonyl carbons of aspartic acid and ^{15}N labeling all of the amide nitrogens of lysine is possible. Whenever there is an Asp-Lys in the protein sequence, there will be two NMR active nuclei connected to each other. A particular REDOR pulse sequence (Figure 2.12) would allow us to have acquisition of signal (S_0) from all the ^{13}C in the protein sample. At half rotor period a ^{15}N π -pulse is given and a dipolar dephased signal (S_1) is produced. The difference signal ($S_0 - S_1$) is due to $^{13}\text{C} - ^{15}\text{N}$ dipolar coupling through space. Thus selective labeling process coupled with REDOR pulse program provides a powerful tool for predicting the secondary structure of a particular amino acid in the entire protein. This

is attained by comparing the chemical shift obtained from the protein sample and the databases of the ^{13}C chemical shifts (18).

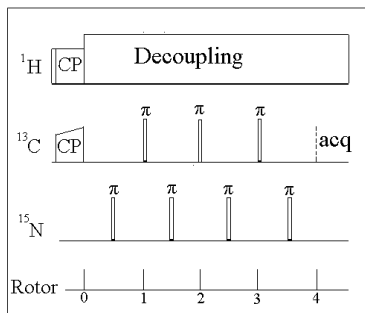


Figure 2.12. The REDOR pulse sequence.

2.12. Determination of Oligomeric State(s) by Size Exclusion Chromatography

The oligomeric property of protein molecules can be determined by Size Exclusion Chromatography (SEC). The chromatographic method SEC is based on separation of particles based on the size or the molecular weight. The sample is allowed to pass through a matrix typically made up of polysaccharide cross-linking forming pores of various sizes. The matrix is the stationary phase and the sample to be characterized is dissolved in the mobile phase. The criterion for the matrix material is that it should not interact with the sample in the mobile phase. The rate of flow of the sample depends on the size of the particles. The protein molecules of larger size elute earlier and the smaller molecules elutes later. The elution curve in the chromatogram resembles Gaussian distribution. The calibration of the matrix columns were done by using protein samples of known molecular weight as mobile phase (Figure 2.13).

The calculations for calibration are as follows:

$$\text{Partition coefficient } K_{av} = (V_e - V_o)/(V_c - V_o) \quad \text{where } V_e = \text{elution volume}$$

V_c = geometric column volume

V_o = column void volume

The linear calibration curve is plotted as K_{av} as a function of $\text{Log}_{10}(\text{MW})$.

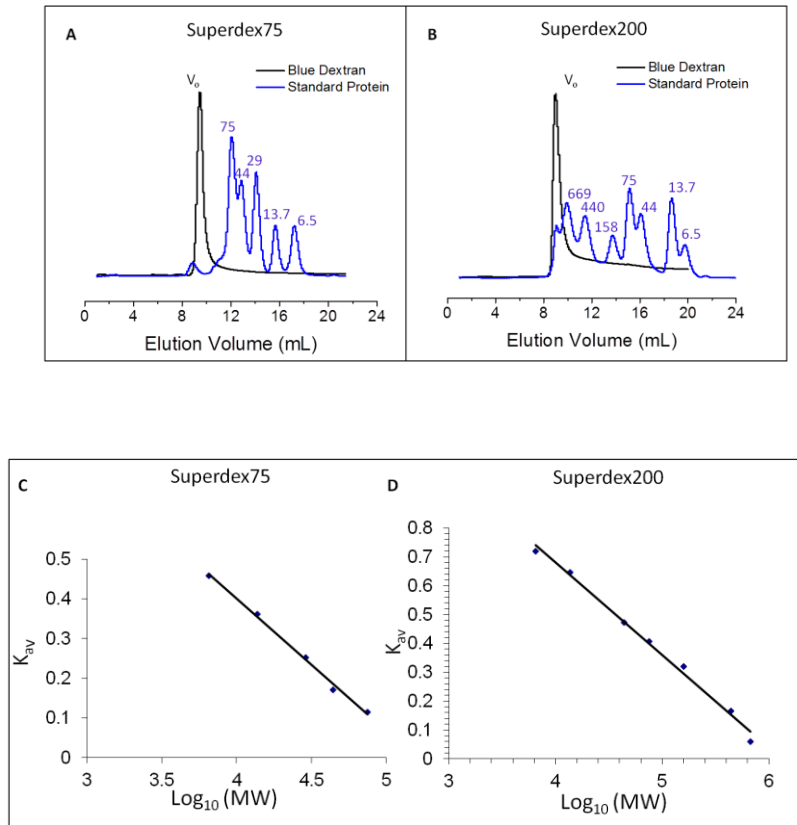


Figure 2.13. SEC A_{280} of MW standards loaded on (A) Superdex 75 and (B) Superdex 200 columns. The protein standards used were (A) Conalbumin (75 kDa), Ovalbumin (44 kDa), Carbonic Anhydrase (29 kDa), Ribonuclease A (13.7 kDa), and Aprotinin (6.5 kDa) and (B) Thyroglobulin (669 kDa), Ferritin (440 kDa), Aldolase (158 kDa), Conalbumin (75 kDa), Ovalbumin (44 kDa), Carbonic Anhydrase (29 kDa), Ribonuclease A (13.7 kDa), and Aprotinin (6.5 kDa). The V_o is the void volume and corresponds to 2 MDa Blue Dextran. K_{av} vs $\text{log}_{10}(\text{MW})$ for the (C) Superdex 75 and (D) Superdex 200 columns. The $K_{av} = (V_e - V_o)/(V_c - V_o)$ where V_e and V_c are the elution and column volumes, respectively.

2.13. Antibody binding studies by Direct Immunoprecipitation

We had used the technique immunoprecipitation to detect the binding of specific broadly neutralizing antibodies to epitopes in the proteins HP and HM. Immunoprecipitation is a technique which can separate a protein antigen-antibody complex from a solution. The antigen-antibody complex is immobilized on a solid phase and later collected. We had implemented the technique of Direct Immunoprecipitation to ensure that the protein antigen binds specifically to the immobilized antibodies. In direct immunoprecipitation the antibodies are first allowed to bind to the protein A/G beads. The beads used were magnetic beads with covalently linked protein G on them. Protein G covalently binds to the antibody heavy chain. The protein of interest is then allowed to bind to the immobilized antibody. This also verifies whether the epitope for the antibodies is present in the protein of interest or not. We have used 4E10 and 2F5, the broadly neutralizing antibodies among various strains of HIV to bind to HP and HM constructs of HIV-1. The experimental steps and conditions are as follows:

Binding of Antibodies – The protein G magnetic beads were washed with PBS + 1% Tween20 at pH 7.4 and then antibody was added to bind to the beads. It was observed that even though excess of antibody was added, there was still the presence of some unbound beads which could bind to the protein molecules leading to misinterpretation of results. We therefore, allowed IgG to bind to the protein G beads after the binding of 4E10 or 2F5. With this step, the unbound beads were covered with IgG. IgG binds to protein G very efficiently. The control experiments were done with only IgG.

Wash – The beads were then washed to remove the unbound antibodies.

Binding of Protein – The protein samples HP or HM were added to allow binding to the antibodies by gently mixing for 2 hours. The whole sample volume of 700 μ L was maintained at pH 7.4.

Extraction of Protein – The protein was extracted from bound antibody-protein complex immobilized on beads by collecting the beads and boiling in SDS buffer for $\sim$ 5 min. The extracted protein sample was then run on 15% SDS-PAGE gel to verify the presence of our protein of interest.

2.14. Vesicle fusion Assays using Fluorescence Spectroscopy

POPC: POPG:Chol 8:2:5 μ mole was dissolved in chloroform followed by chloroform removal with nitrogen gas and vacuum pumping. The film was homogenized by freeze-thaw cycles in $\sim$ 1 mL buffer and extruded through 100 nm diameter pores to form unilamellar vesicles. Fluorescently labeled vesicles were similarly prepared except that the mixture also contained 2 mole% of the fluorescent lipid *N*-NBD-PE and 2 mole% of the quenching lipid *N*-Rh-PE. Labeled and unlabeled vesicles were mixed in 1:9 ratio with total [lipid + Chol] $\approx$ 230 μ M. Experiments were also done using positively-charged vesicles with POPC:DOTAP:Chol in same mole ratio. Fluorescence of the stirring vesicle solution was measured at 37 $^{\circ}$ C with 467 nm excitation, 530 nm detection, and 1 s time increment. After measurement of the baseline fluorescence F_0 , a protein aliquot was added and marked time $t = 0$. Vesicle fusion was reflected in the increased fluorescence $\Delta F(t) = F(t) - F_0$ due to longer distances between fluorescent and quenching lipids in a fused (labeled+unlabeled) vesicle relative to the initial labeled vesicle. The dead-time in the assay was $\sim$ 5 s and asymptotic fluorescence (ΔF_f) was reached after $\sim$ 600 s. The maximum

fluorescence change (ΔF_{max}) was detected after addition of 12 μL 10% Triton X-100 which solubilized the vesicles. Percent fusion was $M(t) = [\Delta F(t)/\Delta F(\max)] \times 100$. Comparison among assay replicates showed $\delta(M_f)/M_f \approx 0.02$.

We wanted to compare vesicle fusion induced by HP, HM, and FP-HM where all proteins were in the same stock buffer conditions. FP-HM was not soluble without 6 M GuHCl, so the chosen stock conditions were 10 mM sodium formate at pH 3.2, 6 M GuHCl, and 0.2 mM TCEP. The effects of the stock pH and GuHCl were minimized by always adding 7.5 μL of stock into a final total volume of 1200 μL . Therefore the final conditions were maintained as physiological, e.g. pH 7.4 with very low $[\text{GuHCl}] = 0.04 \text{ M}$. Vesicle fusion was performed for final pHs of 3.2 and 7.4 with respective stock $[\text{protein}] = 20 \text{ }\mu\text{M}$ and $160 \text{ }\mu\text{M}$, chosen so that (1) for all proteins, $M_f < 100\%$ with no light scattering; and (2) for at least one protein, M_f was appreciably greater than 0%.

REFERENCES

REFERENCES

1. Sackett, K., Wexler-Cohen, Y., and Shai, Y. (2006) Characterization of the HIVN-terminal fusion peptide-containing region in context of key gp41 fusion conformations. *Journal of Biological Chemistry* **281**, 21755-21762
2. Merrifield, R. B. (1963) Solid Phase Peptide Synthesis .1. Synthesis of a Tetrapeptide. *J Am Chem Soc* **85**, 2149-&
3. Hackeng, T. M., Griffin, J. H., and Dawson, P. E. (1999) Protein synthesis by native chemical ligation: Expanded scope by using straightforward methodology. *Proc Natl Acad Sci USA* **96**, 10068-10073
4. Dawson, P. E., Muir, T. W., Clarklewis, I., and Kent, S. B. H. (1994) Synthesis of proteins by native chemical ligation. *Science* **266**, 776-779
5. Johnson, E. C. B., and Kent, S. B. H. (2006) Insights into the mechanism and catalysis of the native chemical ligation reaction. *J Am Chem Soc* **128**, 6640-6646
6. Sackett, K., Nethercott, M. J., Shai, Y., and Weliky, D. P. (2009) Hairpin Folding of HIV gp41 Abrogates Lipid Mixing Function at Physiologic pH and Inhibits Lipid Mixing by Exposed gp41 Constructs. *Biochemistry* **48**, 2714–2722
7. Greenfield, N. and Fasman, G. D. (1969) Computed Circular Dichroism Spectra for Evaluation of Protein Conformation. *Biochemistry* **8**, 4108-&
8. Liu, J., Deng, Y. Q., Dey, A. K., Moore, J. P., and Lu, M. (2009) Structure of the HIV-1 gp41 Membrane-Proximal Ectodomain Region in a Putative Prefusion Conformation. *Biochemistry* **48**, 2915-2923
9. Chan, D. C., Fass, D., Berger, J. M., and Kim, P. S. (1997) Core structure of gp41 from the HIV envelope glycoprotein. *Cell* **89**, 263-273
10. Weissenhorn, W., Dessen, A., Harrison, S. C., Skehel, J. J., and Wiley, D. C. (1997) Atomic structure of the ectodomain from HIV-1 gp41. *Nature* **387**, 426-430
11. Shi, W., Bohon, J., Han, D. P., Habte, H., Qin, Y., Cho, M. W., and Chance, M. R. (2010) Structural Characterization of HIV gp41 with the Membrane-proximal External Region. *Journal of Biological Chemistry* **285**, 24290–24298
12. Buzon, V., Natrajan, G., Schibli, D., Campelo, F., Kozlov, M. M., and Weissenhorn, W. (2010) Crystal Structure of HIV-1 gp41 Including Both Fusion Peptide and Membrane Proximal External Regions. *Plos Pathogens* **6**

13. Song, L., Sun, Z.-Y. J., Coleman, K. E., Zwick, M. B., Gach, J. S., Wang, J.-h., Reinherz, E. L., Wagner, G., and Kim, M. (2009) Broadly neutralizing anti-HIV-1 antibodies disrupt a hinge-related function of gp41 at the membrane interface. *PNAS* **106**, 9057–9062
14. Sackett, K., Nethercott, M. J., Epand, R. F., Epand, R. M., Kindra, D. R., Shai, Y., and Weliky, D. P. (2010) Comparative Analysis of Membrane-Associated Fusion Peptide Secondary Structure and Lipid Mixing Function of HIV gp41 Constructs that Model the Early Pre-Hairpin Intermediate and Final Hairpin Conformations. *Journal of Molecular Biology* **397**, 301–315
15. Bodner, M. L., Gabrys, C. M., Parkanzky, P. D., Yang, J., Duskin, C. A., and Weliky, D. P. (2004) Temperature dependence and resonance assignment of C-13 NMR spectra of selectively and uniformly labeled fusion peptides associated with membranes. *Magn Reson Chem* **42**, 187-194
16. Yang, J., and Weliky, D. P. (2003) Solid-state nuclear magnetic resonance evidence for parallel and antiparallel strand arrangements in the membrane-associated HIV-1 fusion peptide. *Biochemistry* **42**, 11879-11890
17. Yang, J., Parkanzky, P. D., Bodner, M. L., Duskin, C. A., and Weliky, D. P. (2002) Application of REDOR subtraction for filtered MAS observation of labeled backbone carbons of membrane-bound fusion peptides. *J Magn Reson* **159**, 101-110
18. Zhang, H. Y., Neal, S., and Wishart, D. S. (2003) RefDB: A database of uniformly referenced protein chemical shifts. *J Biomol Nmr* **25**, 173-195
19. Gullion, T. (1998) Introduction to rotational-echo, double-resonance NMR. *Concept Magnetic Res* **10**, 277-289

CHAPTER 3

Structural and Functional Characterization of Gp41 Ectodomain

3.1 Introduction:

HIV is enveloped by a membrane obtained during viral budding from an infected host cell. An early step in infection of another cell is fusion of the viral and cell membranes with accompanying release of the viral nucleocapsid into the cytoplasm (1). The HIV membrane includes a gp160 glycoprotein complex comprised of two noncovalently associated subunits, gp120 and gp41 (2). Gp41 is a ~350-residue monotopic integral membrane protein with a ~180-residue ectodomain (Figure 3.1). Gp120 is bound to the gp41 ectodomain. HIV targets lymphocytes via binding of gp120 to cell receptors and gp120 moves away from or dissociates from gp41. Gp41 then undergoes large conformational changes with accompanying catalysis of membrane fusion. To our knowledge, the gp120/cell receptor complex is only for target cell identification and gp41 is the only fusion protein. Much of our biophysical understanding of fusion has therefore come from studies of gp41 with an emphasis on its ectodomain which can contact the outer leaflets of both the viral and cell membranes. Mutagenesis has demonstrated that there are two ~20-residue regions of the gp41 ectodomain that play key roles in fusion likely through membrane interaction (3,4). The *N*-terminal fusion peptide (FP) and *C*-terminal membrane-proximal external region (MPER) are postulated to bind to the host cell and viral membranes, respectively. The fusion significances of the FP and MPER have been supported by observation of vesicle fusion induced by FP or MPER peptides (5,6). There is an intervening *N*-heptad repeat (NHR), loop, and *C*-heptad repeat (CHR) between the FP and MPER domains (Figure 3.1).

Electron micrographs of virions show clusters of three gp160, i.e. three gp120 and three gp41 molecules. These clusters are likely the initial protein state before any changes in membrane topology due to membrane fusion (7). The extraviral region of gp160 is termed gp140 and is comprised of gp120 and the gp41 ectodomain without the gp41 TM and gp41 endodomain. WT Gp140 is typically monomeric but gp140 trimers can be stabilized via mutations and a gp120/gp41 ectodomain cross-link (8,9). There are ~ 5 Å structures of such gp140 trimers that likely represent the protein state prior to membrane fusion (10-12). The structure includes a loose bundle of three parallel NHR helices and three CHR helices forming a tripod. The monomer structure is NHR-helix/ 70° -turn/CHR-helix. The FP and MPER are not in the structure. There are also atomic-resolution structures of segments of the gp41 ectodomain typically without the FP and MPER and without gp120 (13-16). These show NHR-helix/ 180° -turn/CHR-helix hairpin structure as well as assembly of three molecules into a six-helix bundle (SHB) with the three NHRs forming parallel coiled-coil structure on the bundle interior and the three CHRs packing antiparallel to the NHRs on the bundle exterior. Melting temperatures of $\sim 70^\circ$ C are observed for shorter ectodomain constructs with hairpin structure whereas temperatures up to 110° C are observed for longer constructs (17,18). This thermostability has supported the SHB as the final gp41 structure during fusion.

In addition to the initial NHR/ 70° -turn/CHR bent structure and the final NHR/ 180° -turn/CHR SHB structure, a “pre-hairpin intermediate” (PHI) structure has been proposed to form after removal of the gp120s. Each PHI gp41 has a fully extended (no turn) structure, i.e. NHR-helix/ 0° -turn/CHR-helix and there are separate NHR and CHR trimer helical bundles. To our knowledge, the existence of the gp41 PHI is only supported by functional studies, in particular

inhibition of membrane fusion and HIV infection with NHR or CHR+MPER peptides (19,20). These peptides are proposed to bind to the CHR and NHR bundles of the PHI, respectively, and to inhibit the PHI→SHB structural transition. CHR+MPER peptides are a clinically-prescribed HIV treatment (21).

There are also distinct membrane structures during fusion (22). The separate viral and host cell membranes first merge into a hemifusion intermediate characterized by intermembrane lipid mixing and no contents mixing. This is followed by breaking the hemifusion barrier and formation of a small pore through which small species (e.g. atomic ions) can pass. The fusion pore then expands to create a single membrane enclosing the cell and the viral capsid.

There are little data about the relative timing of gp41 and membrane structural changes. One common model I has been: (1) gp120 receptor binding followed by gp120 removal; (2) formation of extended PHI gp41 trimer followed by FP insertion into the host cell membrane; (3) PHI→SHB trimer folding that brings the two membranes close together; (4) hemifusion; (5) initial pore formation; and (6) fusion pore expansion (23). The appealing intuitive aspect of this model is that some of the free energy released during PHI→SHB folding is used to form membrane intermediates. However, the relative timings of this model are not supported by the observation that CHR+MPER peptides inhibit fusion up to the final fusion pore expansion step (20). Because the peptides are presumed to bind to the PHI trimer but not the SHB trimer, these data suggest an alternative model II: (1) gp120 receptor binding followed by gp120 removal; (2) formation of extended PHI gp41 trimer followed by FP insertion into the host cell

membrane; (3) hemifusion; (4) initial pore formation; (5) PHI→SHB trimer folding; and (6) fusion pore expansion (24).

In the present work, we show that the gp41 ectodomain can form stable hairpin monomers as well as stable hexamers that are likely composed of two SHB trimers. CHR+MPER inhibitor peptides likely bind to the monomer but not the trimer or hexamer. These findings are the basis of a new model III (Figure 3.19): (1) gp120 receptor binding followed by gp120 removal; (2) dissociation of gp41 ectodomain into monomers and formation of extended PHI gp41 ectodomain monomer followed by FP insertion into the host cell membrane; (3) PHI→hairpin monomer folding that brings the two membranes close together; (4) hemifusion; (5) initial pore formation; (6) hairpin monomer→SHB trimer→hexamer ectodomain assembly; and (7) fusion pore expansion. Like model I and unlike model II, the new model III retains the appealing coupling of the PHI→SHB transition to initial steps of membrane fusion. There are more reasonable coordinated changes of the ectodomain and membrane topologies for PHI→SHB monomer folding of model III than for PHI→SHB trimer folding of models I and II. Finally, the discovery of stable hexamers correlates with other data supporting a requirement of multiple gp160 trimers for membrane fusion and HIV infection (25,26).

Much of our understanding of the FP and MPER regions of the gp41 ectodomain has been based on studies of vesicle fusion induced by peptides (27-29). Rapid (~5 s) fusion typically requires 500-1000 peptides per ~100 nm diameter vesicle which is much higher than the ~30 gp160 per virion and suggests that there are aspects of viral fusion unaccounted for in the peptide studies (30). A reasonable hypothesis with some supporting data is that the rest of the

ectodomain plays an important role in fusion. In the present work, we demonstrate efficient vesicle fusion with only ~15 gp41 per vesicle provided that the FP, hairpin, and MPER are included in the large gp41 ectodomain construct. To our knowledge, this is the first demonstration of fusion synergy between the FP, hairpin, MPER in a large gp41 ectodomain construct. Such synergy correlates with postulated binding of FP and MPER regions to host cell and viral membranes, respectively, and also with postulated FP/MPER interaction in a folded hairpin structure.

The MPER is the epitope of several broadly-neutralizing antibodies (bNAbs) that prevent infection by diverse isolates of HIV (4,31). There has consequently been continued effort to develop a HIV vaccine with a MPER immunogen. The ectodomain with MPER and hairpin structure is a candidate immunogen in part because of the stability of this structure. However, there is disagreement in the literature about the antigenicity of the hairpin, i.e. whether bNAbs bind well to the MPER in this structure (32-34). The present study shows such binding for the hairpin ectodomain which is initially in a monomer or hexamer state.

3.2 Experimental Procedures

3.2.1. HP and HM inserts

Amino acid (Figure 3.1B) sequences are for the HXB2 laboratory strain of HIV and are described using gp160 numbering. HP and HM include the NHR residues 535(M535C)-581 and a non-native loop (SGGRGG). HP includes CHR residues 628-666 and a non-native H₆ tag. HM includes CHR+MPER residues 628-683, a non-native G₄ spacer, and a non-native H₄ tag. The HP insert in the pGEM-t vector without the H₆ tag has been previously described and the H₆ tag was then

added via PCR (35). The HM insert was generated from the HP insert via multiple rounds of PCR.

The HM DNA sequence is shown in figure 3.2.

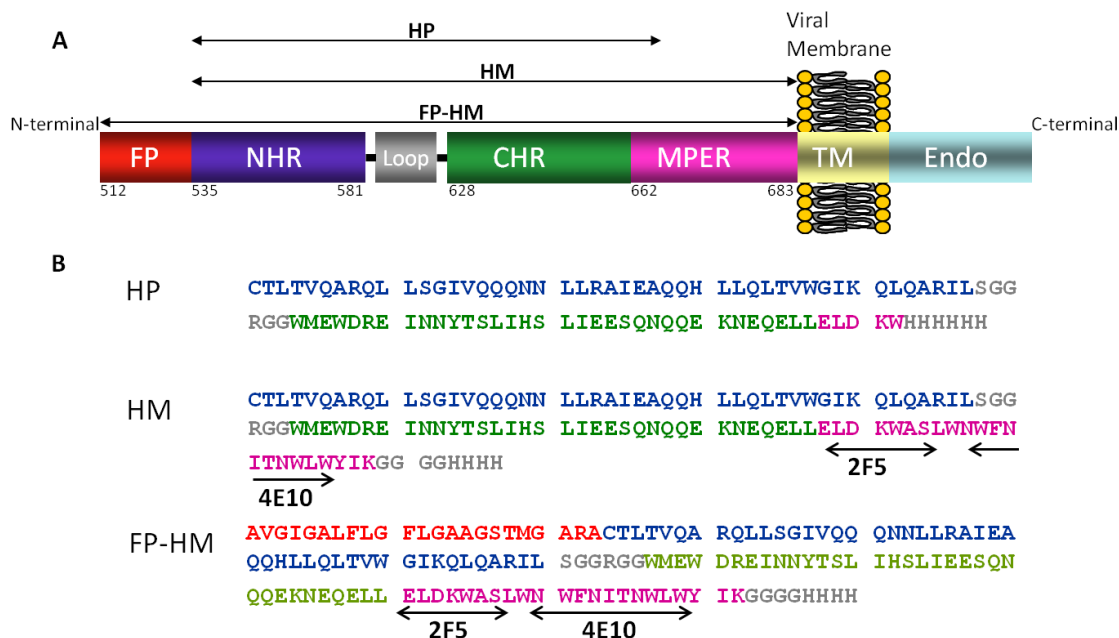


Figure 3.1. (A) Schematic diagram of HIV gp41 where FP ≡ fusion peptide, NHR ≡ N-heptad

repeat; CHR ≡ C-heptad repeat, MPER ≡ membrane-proximal external region, TM ≡

transmembrane domain, and Endo ≡ endodomain. (B) Amino acid sequences of HP, HM, and

FP-HM. The epitopes of the 2F5 and 4E10 bNAbs are marked.

```

Atg tgc acg ctg acg gta cag gcc aga caa tta ttg tct ggt ata gtg cag cag cag aac aat ttg ctg
agg gct att gag gcg caa cag cat ctg ttg caa ctc aca gtc tgg ggc atc aag cag ctc cag gca
aga atc ctg tct ggt ggc cgt ggc ggt tgg atg gag tgg gac aga gaa att aac aat tac aca agc tta
ata cac tcc tta att gaa gaa tcg caa aac cag caa gaa aag aat gaa caa gaa tta ttg gaa tta gat
aaa tgg gcg tcg ctg tgg aac tgg ttt aac atc acg aac tgg ctc tgg tac atc aag ggt ggt ggc
ggg cat cac cat cac cat cac tga

```

Figure 3.2. DNA sequence of the HM insert

3.2.2. HP and HM Expression, Solubilization, and Purification

The expression host was *E. coli*, BL21(DE3) strain. The typical protocol began with addition of 1 mL of bacterial glycerol stock to 50 mL of LB medium. After overnight growth at 37 °C, the 50 mL culture was added to 1 L of fresh LB medium. Growth was continued for two hours with a final OD₆₀₀ of ~0.8. Protein expression was induced with addition of 2 mM IPTG and continued for 6 hours at 37 °C. “Pellet I” (~9 g) was harvested by centrifugation at 9000g. The next steps were done at 4 °C using 30 mL PBS at pH 7.4 with protease inhibitor cocktail. 3 g of pellet I was suspended in PBS and lysed by tip sonication in an ice bath. The lysate was centrifuged at 48000g and the resultant “pellet II” was sonicated in PBS with subsequent centrifugation. SDS-PAGE showed that the resultant “pellet III” had a high mass fraction of recombinant protein (RP). Pellet III was effectively solubilized by sonication in PBS + 6M GuHCl and the RP in this solution was purified at ambient temperature by affinity chromatography with Co²⁺ resin. The resin suspension solutions were PBS (pH 7.4) + 6M GuHCl + imidazole and resin was isolated using gravity filtration. After initial protein binding with 1 mM imidazole, weakly-bound proteins were removed using sequential washes with 5 mM imidazole (2×), 10 mM imidazole (2×), and 20 mM imidazole (2×). The RP was eluted using 250 mM imidazole (4×) and the purified yields of HP and HM were ~50 and ~15 mg/L culture, respectively, as determined by A₂₈₀. Elutions were: (1) diluted to ~0.1 mg RP/mL in PBS + 6 M GuHCl + 2 mM DTT; (2) dialyzed at 4 °C against 50 mM sodium formate buffer (pH 3.2) + 150 mM NaCl + 0.2 mM TCEP reducing agent; and (3) concentrated to ~1 mg/mL. HP and HM were aggregated in the pH 5 – 9 range in the absence of GuHCl. The newly constructed HM MW was also confirmed by MALDI (Figure

3.3A). The theoretical MW of HM is 13738 Da and according to MALDI it was obtained as 13728 Da. The error of the observed mass from the theoretical mass is 0.073%.

3.2.3. Synthesis and Purification of FP-HM

FP23 (residues 512-534 (S534A)-(thioester linker)) was synthesized manually by t-boc chemistry, purified by RP-HPLC, and quantitated with the BCA assay Purity was >95% by MALDI MS. Native chemical ligation between FP23 and HM was done with 1.2 mM FP23, 1.2 mM HM, 0.1 M sodium phosphate buffer at pH 7.2, 8 M GuHCl, 60 mM MPAA, 2 mM TCEP, 1 day time, and ambient temperature. The ligation products were purified by RP-HPLC. The MW of the ligated product FP-HM was verified by SDS-PAGE and MALDI experiments (Figure 3.3B and inset and 3.4A). The theoretical value of MW of FP-HM is 15829 Da and observed value was 15815 Da. The error of the observed mass from the theoretical mass is 0.088%. The difference in mass of HM and FP-HM from the theoretical value is probably due to the calibration error. The second round of purification by RP-HPLC shows only FP-HM peak (Figure 3.4B). However, the yield was very low, therefore, worked with the 1st round purified product containing ~40% of HM mixed in FP-HM.

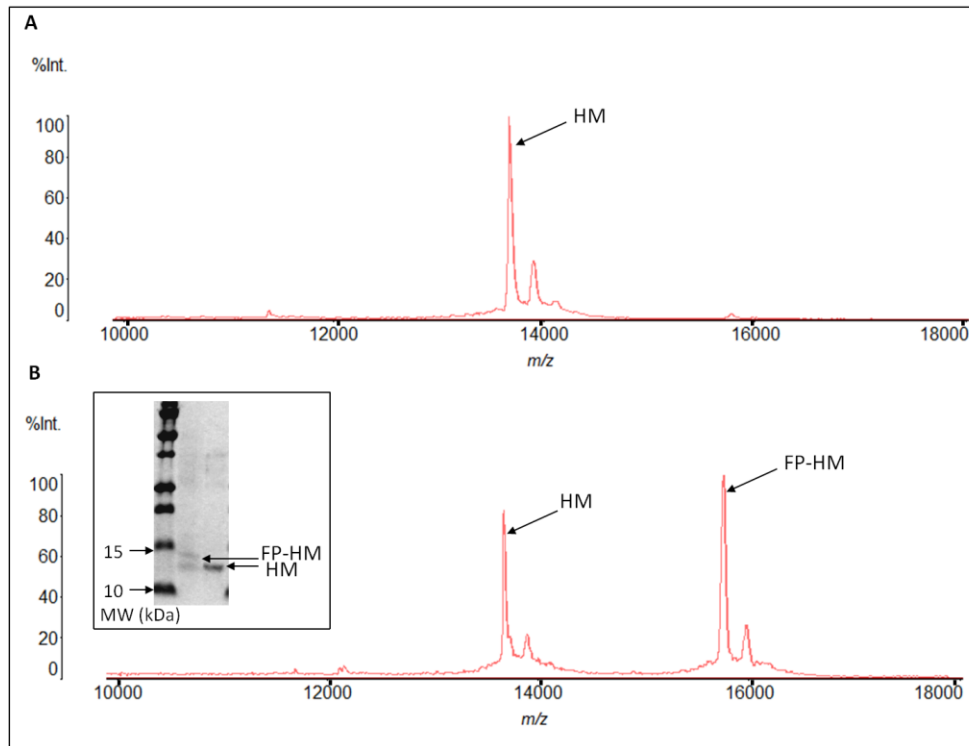


Figure 3.3. MALDI mass spectra of (A) the HM and (B) FP-HM peaks from the first HPLC purification. *Inset*: SDS-PAGE where the middle lane is the FP-HM peak of figure 3.3B and the right lane is purified HM.

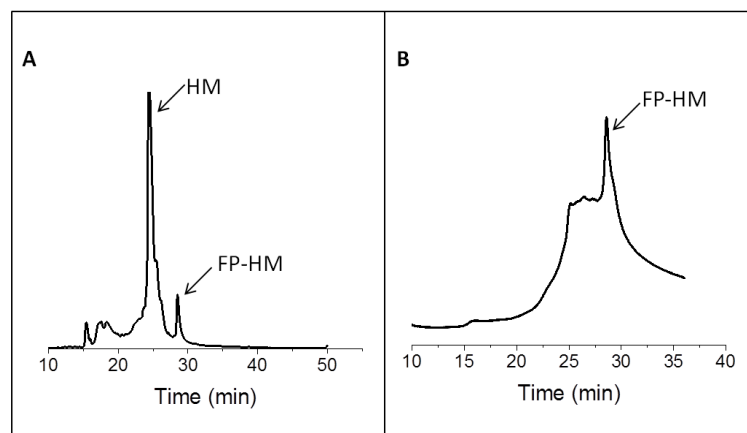


Figure 3.4. (A) First-round RP-HPLC profile of FP+HM ligation. (B) Second-round RP-HPLC of FP-HM peak from panel A.

3.2.4. Size-Exclusion Chromatography (SEC)

A DuoFlow Pathfinder 20 instrument (Bio-Rad) was used with a flow rate of 0.3 mL/min, A_{280} detection, and Tricorn semi-preparative columns (GE Technologies).

3.2.5. CD Spectroscopy

A Chirascan spectrometer (Applied Photophysics) was used with a quartz cuvette with 1 mm pathlength. There were 0.5 nm steps with 1.5 s per step. For each sample, three scans were averaged. The final spectrum was the (RP + buffer) – buffer difference spectrum.

3.2.6. Western Blots.

Purified RP in SDS buffer (~0.5 mg/mL) was boiled followed by SDS-PAGE (~5 µg RP per lane) and transfer to a nitrocellulose membrane and incubation in a 10 mL solution containing antibody (5 µg), TBST at pH 7.4, and 5% w/v nonfat dry milk. Subsequent incubation with Goat Anti-Hamster IgG (H+L) HRP conjugate secondary antibody (1 µg) was followed by development with SuperSignal West Pico chemiluminescent substrate.

3.2.7. Immunoprecipitation

5 µg quantities of RP and antibodies were used from 1 mg/mL stocks. Solutions other than RP stock contained PBS at pH 7.4. bNAb was incubated with Protein G magnetic beads followed by 3× wash removal of free bNAb. Because RP binds to unpassivated beads but not IgG, the beads were incubated with IgG and then washed. The beads were then incubated with RP in 700 µL

solution for 1 hour and washed. Bound RP and antibodies were removed by boiling in SDS sample buffer and analyzed by SDS-PAGE.

3.2.8. Protein-Induced Vesicle Fusion

Lipid:Chol (1.0:0.5 μ mole) was dissolved in chloroform followed by chloroform removal with nitrogen gas and vacuum pumping. The film was homogenized by freeze-thaw cycles in $\sim$ 1 mL buffer and extruded through 100 nm diameter pores to form unilamellar vesicles. Fluorescently labeled vesicles were similarly prepared except that the mixture also contained 2 mole% of the fluorescent lipid *N*-NBD-PE and 2 mole% of the quenching lipid *N*-Rh-PE. Labeled and unlabeled vesicles were mixed in 1:9 ratio with total [lipid+Chol] $\approx$ 230 μ M. Fluorescence of the stirring vesicle solution was measured at 37 $^{\circ}$ C with 467 nm excitation, 530 nm detection, and 1 s time increment. After measurement of the baseline fluorescence F_0 , a protein aliquot was added and marked time $t = 0$. Vesicle fusion was reflected in the increased fluorescence $\Delta F(t) = F(t) - F_0$ due to longer distances between fluorescent and quenching lipids in a fused (labeled+unlabeled) vesicle relative to the initial labeled vesicle. The dead-time in the assay was $\sim$ 5 s and asymptotic fluorescence (ΔF_f) was usually reached by $\sim$ 600 s. The maximum fluorescence change (ΔF_{max}) was detected after addition of 12 μ L 10% Triton X-100 which solubilized the vesicles. Percent fusion was $M(t) = \{\Delta F(t) / \Delta F_{max}\} \times 100$. Comparison among assay replicates showed $\delta(M_f) / M_f \approx 0.02$.

We wanted to compare vesicle fusion induced by HP, HM, and FP-HM where all proteins were in the same stock buffer conditions. FP-HM was not soluble without 6 M GuHCl, so the chosen stock conditions were 10 mM sodium formate at pH 3.2, 6 M GuHCl, and 0.2 mM TCEP. The

effects of the stock pH and GuHCl were minimized by always adding 7.5 μL of stock into a final total volume of 1200 μL with final $[\text{GuHCl}] = 40 \text{ mM}$. Vesicle fusion was done for final pHs of 3.2 and 7.4 with respective stock $[\text{protein}] = 20 \text{ }\mu\text{M}$ and $160 \text{ }\mu\text{M}$, chosen so that (1) for all proteins, $M_f < 100\%$ with no light scattering; and (2) for at least one protein, M_f was appreciably greater than 0%.

3.3. Results

3.3.1. High-yield Protein Production

Cells that had expressed HP were lysed in PBS but SDS-PAGE of the soluble lysate did not show a clear HP band. It was therefore concluded that most of the HP was in inclusion bodies (IBs). After an additional lysis in PBS, different solubilization conditions were tested for “pellet III” as defined in the Experimental section. Similar intensity HP bands were observed in SDS-PAGE of the lysates from glacial acetic acid, 1% w/v SDS, 8 M urea, or 6 M GuHCl. Purification of the 6 M GuHCl lysate yielded HP with high purity (Figure 3.2B). The most intense band was the HP monomer and there were also reproducible weaker dimer bands confirmed to be HP by Western blotting with anti-His tag antibody. The purified yield was $\sim 50 \text{ mg HP/L culture}$.

The first HM construct had a H_6 C-terminal tag without glycines and the initial efforts to solubilize and purify the protein were unsuccessful as judged by no clear band in SDS-PAGE. It was unclear whether the main problem was low expression, low solubilization, or purification losses. HM expression prior to solubilization was then quantitated with a recently-developed solid-state NMR (SSNMR) method (36). Addition of ^{13}CO -Leu during the expression period resulted in ^{13}CO -labeling of HM. Cells were lysed in PBS followed by centrifugation of the lysate

and the harvested pellet was enriched in any HM IBs. The ^{13}C NMR spectrum of this pellet showed a prominent ^{13}CO feature consistent with ~300 mg HM/L in IBs (Figure 3.5A). The main bottleneck(s) to purified HM were therefore low solubilization of HM IBs and/or purification losses.

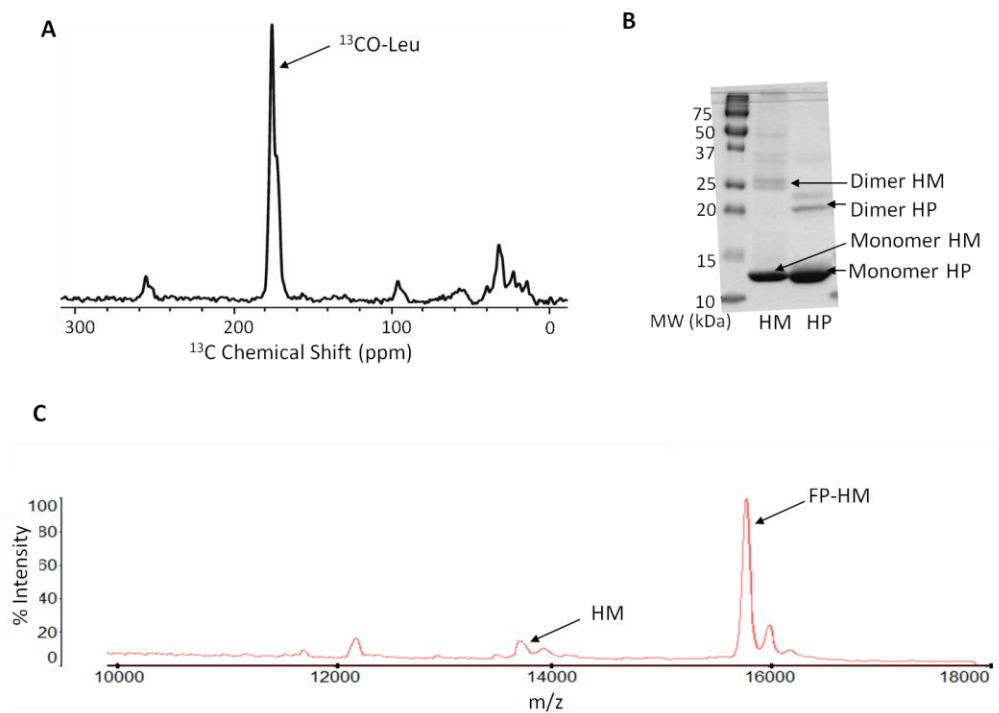


Figure 3.5. (A) ^{13}C SSNMR spectra of a cell pellet harvested from centrifugation of a cell lysis in PBS. The *E. coli* cells contained a plasmid with the HM insert and expression was induced for two hours in minimal medium containing $^{13}\text{CO-Leu}$. Any expressed HM will therefore be $^{13}\text{CO-Leu}$ labeled. The SSNMR acquisition parameters included a 9.4 T magnetic field, 4 mm diameter rotor, 8 kHz magic angle spinning frequency, and ~1 day of signal averaging. The ratio of integrated isotropic ^{13}CO intensity (peaked at ~175 ppm) to integrated aliphatic intensity (0 – 90 ppm region) translates to expression of ~300 mg HM/L culture. (B) SDS-PAGE of purified HM (MW = 13.7 kDa) and HP (MW = 11.6 kDa). (C) MALDI MS of FP-HM ligation product twice-purified by RP-HPLC. The experimental ratio $[m/z(\text{FP-HM}):m/z(\text{HM})]$ is 1.1522 and matches the ratio of expected MWs, $[\text{MW}(\text{FP-HM}):\text{MW}(\text{HM})] = 1.1522$.

A systematic search was carried out to find conditions for solubilization of HM IBs. Two sequential lyses were done in PBS to solubilize other proteins. Assessment of solubilization of

HM IBs in pellet III was done by: (1) visual reduction in pellet size; and (2) SDS-PAGE of the solution. Many conditions that solubilized HP IBs did not solubilize HM IBs. Appreciable HM solubilization was only achieved with 1% SDS or 6 M GuHCl and the latter additive was chosen for solubilization and purification. HM with H₆ tag did not bind the Co²⁺ resin whereas HM with a G₄H₆ tag bound so tightly that there was no elution with 250 mM imidazole. HM with a G₄H₄ tag both bound tightly to the resin with 20 mM imidazole and also eluted from the resin with 250 mM imidazole. The purified yield was ~15 mg HM/L culture and SDS-PAGE showed dominant monomer and weaker dimer bands (Figure 3.5B).

FP-HM was produced by native chemical ligation between FP and HM and purified by RP-HPLC. For one round of purification, MS intensities showed FP-HM:HM≈1.2 which correlated with SDS-PAGE. For two rounds of RP-HPLC, the FP-HM:HM≈10 (Figure 3.5C). However, there was too little FP-HM, so the FP-HM+HM mixture after only one round was used for subsequent experiments.

3.3.2. Solubility Only at pH 3 or With 6 M GuHCl.

Solubility in a particular buffer was examined by: (1) dialyzing the protein (~0.2 mg/mL) against the buffer for one day; (2) centrifugation at 16000g for 5 minutes; and (3) measuring protein concentration in the supernatant. A protein was typically either soluble by the criterion [final concentration]/[initial concentration] > 0.8 or poorly soluble with obvious precipitation.

None of the proteins were soluble in: (1) 10 mM sodium formate at pH 5.0; (2) 5 mM HEPES/10 mM MES at pH 7.4; (3) PBS at pH 7.4 with or without 0.1% (w/v) non-ionic detergent (Triton X-

100, *N*-lauroylsarcosine, *n*-decyl- β -D-maltopyranoside, and *n*-dodecylphosphocholine were tested); or (4) 10 mM Tris at pH 9.0. Both HP and HM were soluble in 10 or 50 mM sodium formate buffer at pH 3.2. However, FP-HM was not soluble under these conditions. All proteins were soluble with 6 M GuHCl at either pH 3.2 (50 mM sodium formate) or pH 7.4 (10 mM sodium phosphate). There was solubility either with or without 150 mM NaCl. At pH 7.4, there was no appreciable solubility with [GuHCl] < 6 M.

3.3.3. Monomer or Hexamer and Not Trimer.

Oligomerization was probed by SEC for conditions that satisfied the above solubility criteria. The typical protein loading was 1 mg/mL and there was ~10-fold dilution in the column. For a particular buffer, the dominant oligomeric state was either monomer or hexamer and never trimer. For example, for HP at pH 3.2 without GuHCl, the SEC profile is consistent with a monomer (Figure 3.6A). This is significant as CD, calorimetric, and vesicle fusion studies of gp41 ectodomain constructs have often been done with protein at pH 3 (18,35,37). It has typically been assumed that there are SHB trimers; however, the SEC only shows monomers. For 10 mg HP/mL loading concentration, the SEC profile also shows a major fraction monomer with much smaller fractions of trimers, hexamers, and larger aggregates (Figure 3.6B). SEC was attempted for HM at pH 3.2 without GuHCl but the protein bound to the Superdex column material. For buffer with 6 M GuHCl at either pH 3.2 or pH 7.4, both HP and HM show major fraction hexamer with much smaller fractions assigned to dodecamers, 24-mers, and aggregates with MW > 2 MDa (Figure 3.7). Although cell physiology does not include 6 M GuHCl, it is the only condition we found that resulted in a defined small oligomer state. The hexamer is also the

dominant species when non-native protein tags are added to the ectodomain construct (38). We therefore used 6 M GuHCl as one initial solubilization condition in subsequent experiments. The major hexamer population and minor populations of larger oligomers likely represent a thermodynamic equilibrium state. This was evidenced by collecting the hexamer fraction, concentrating to 1 mg/mL, and then reloading this fraction for a second round of SEC. The resultant SEC profiles for HP and HM are similar to those initially observed (Figure 3.9).

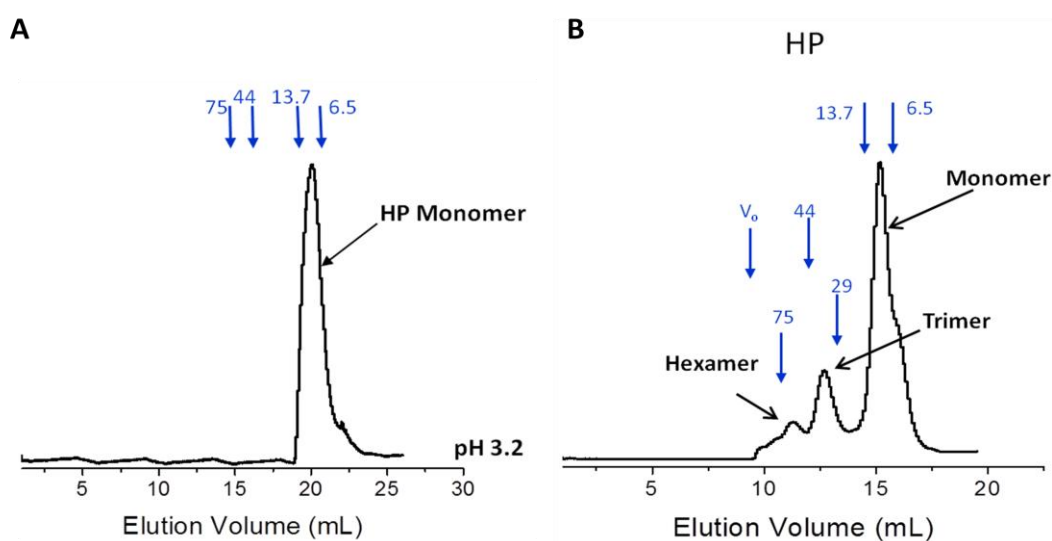


Figure 3.6. SEC A₂₈₀ profiles for HP with (A) 1 mg/mL loading, Superdex 200 column and (B) 10 mg/mL loading, Superdex 75 column. The loading and running buffer was 50 mM sodium formate pH 3.2, 150 mM NaCl, and 0.2 mM TCEP. The blue numbers are the MW standards in kDa and their respective elution volumes are the downward blue arrows. There are accompanying SEC profiles of the MW standards as well as plots of K_{av} vs $\log_{10}(\text{MW})$ (Chapter 2, Figure 2.13).

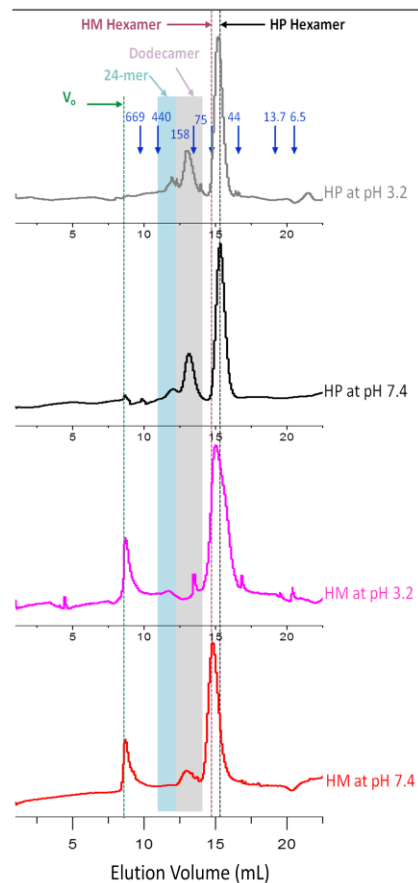


Figure 3.7. SEC A_{280} profiles from a semi-preparative Superdex 200 column for HP and HM in buffers containing 6 M GuHCl and 150 mM NaCl. The pH 3.2 buffer was 50 mM sodium formate with 0.2 mM TCEP and the pH 7.4 buffer was 10 mM phosphate with 2 mM DTT. For each run, the loading and running buffers were the same and the loading stock solution had a 1 mg/mL protein concentration. The blue numbers are the MW standards in kDa and their respective elution volumes are the downward blue arrows. The most prominent peaks in the profiles correspond to hexamer masses and vertical dashed lines show the elution volume offset between the HP and HM hexamers. Smaller peaks most consistent with dodecamer and 24-mer masses are identified by shaded regions. For HM, there is also a small peak associated with the V_0 void volume, i.e. aggregates with $MW \geq 2$ MDa.

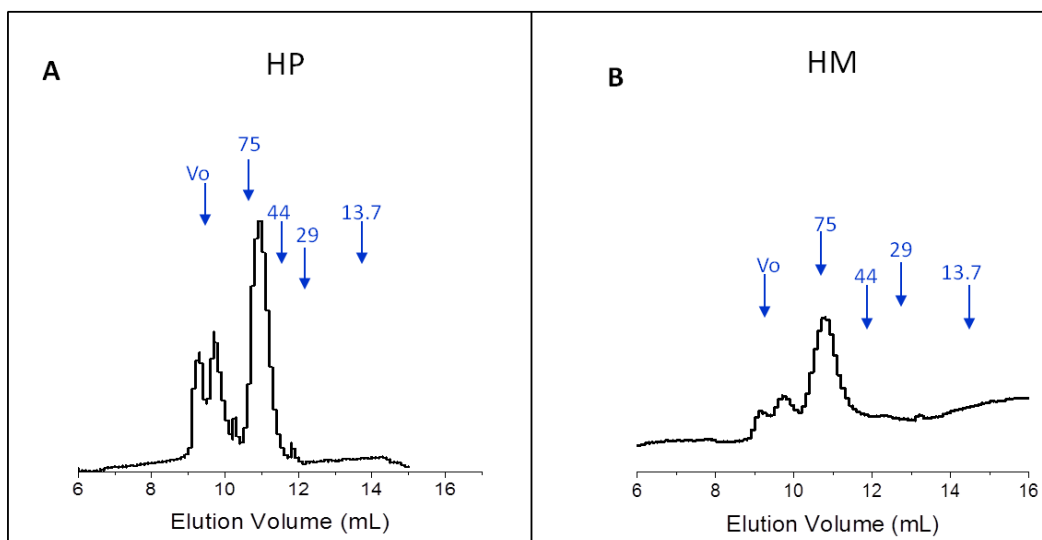


Figure 3.8. Evidence for thermodynamic equilibrium between hexamers and larger oligomers. SEC with A_{280} detection was run for HP and HM in buffer containing 50 mM sodium formate at pH 3.2 with 150 mM NaCl, 6 M GuHCl, and 0.2 mM TCEP (Figure 3.7). The hexamer peak was collected and loaded onto the Superdex 75 column for SEC with this same buffer. The loading solution contained (A) 0.3 or (B) 0.2 mg protein/mL. Each blue number is the MW of a standard in kDa and the blue downward arrows is its elution volume from Figure 2.13, Chapter 2. The resultant profiles for (A) HP and (B) HM show a mixture of hexamers and larger oligomers similar to the profiles of the initial run (Figure 3.7). This similarity supports this mixture as the thermodynamic equilibrium state for the protein.

3.3.4. Hyperthermostable Hairpin Structure of the Monomer

CD spectra of HP and HM at pH 3.2 without GuHCl have the 208/222 nm minima characteristic of α helical structure and the θ_{222} values are consistent with dominant helical structure (Figure 3.10A). HP is predominantly a monomer in this buffer so the data support a highly helical monomer. There is a small fraction of trimer for high SEC loading and the CD spectrum of this

fraction is also consistent with dominant helical structure. For HM, there was a small linear decrease in $|\theta_{222}|$ between 20 and 95 °C (Figure 3.10B) that is similar to the previously observed temperature-dependence for HP and other large ectodomain constructs at pH 3. This correlates with calorimetrically-determined T_m 's of ~110 °C (18).

CD spectra for the D632A and D632A/W628A mutants of HP provide additional support for the helical monomer model (Figure 3.9C). In the atomic-resolution SHB trimer structure, D632 and W574 form an inter-monomer CHR/NHR salt bridge and W628 and W571 form an intra-monomer hydrophobic interaction (39). There is negligible difference between the WT and D632A CD spectra whereas there is about two-times less magnitude molar ellipticity in the D632A/W628A spectrum that is consistent with a significant loss in helicity. These data support hairpin monomer structure of HP which is similar to a monomer unit of the SHB trimer structure (Figure 3.17A). The T_m of ~110 °C corresponds to unfolding of hairpin monomer rather SHB trimer.

For HP and HM in 6 M GuHCl at either pH 3.2 or pH 7.4, there is a 222 nm minimum consistent with helical structure with significant absorption/interference below 215 nm from the GuHCl (Figure 3.9D). The hexamer is the dominant oligomeric state under these conditions and may be a dimer of two SHB trimers. For HP in 6 M GuHCl at pH 7 between 20 and 90 °C, there is a linear decrease in $|\theta_{222}|$ comparable to the figure 3.9B change without GuHCl. This supports thermostability of the hairpin monomer structure within the hexamer.

The ligations are done in 8 M GuHCl at pH 7 and ambient temperature so it is interesting to consider the structural properties in this condition. For HP in 8 M GuHCl at 20 °C, the $|\theta_{222}|$ is

about 60% of the value in 6 M GuHCl. In addition, the $|\theta_{222}|$ in 8 M GuHCl exhibits a small linear decrease between 20 and 70 °C (Figure 2-7C in Matthew J. Nethercott's Dissertation). These data are consistent with a dimer of SHB trimers in 8 M GuHCl.

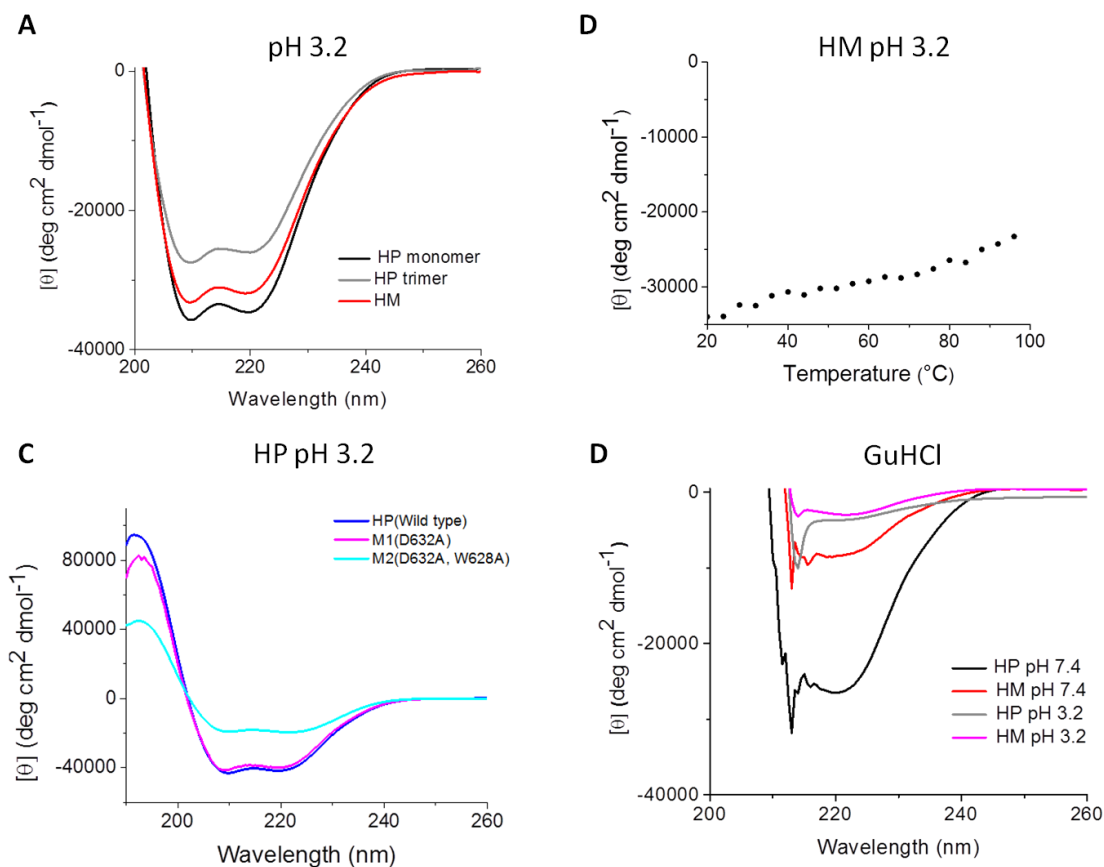


Figure 3.9. (A) CD spectra of HP and HM in pH 3.2 buffer. All spectra show θ_{222} indicative of a significant fraction of α helical structure. (B) θ_{222} vs temperature for HM at pH 3.2. No unfolding transition is observed. (C) CD spectra of WT and mutant HP at pH 3.2. For a SHB trimer, the D632A mutation could disrupt the inter-monomer salt bridge with K574 and the W628A mutation could disrupt the intra-monomer hydrophobic interaction with W571. More positive CD signal for the double but not the single mutant is consistent with monomer rather than SHB trimer HP at pH 3.2. (D) CD spectra of HP and HM in buffer with 6 M GuHCl. The spectra show θ_{222} indicative of a significant fraction of α helical structure. The [protein] $\approx$ 15 μ M ($\sim$ 0.2 mg/mL), the pH 3.2 buffers contained 50 mM formate with 0.2 mM TCEP, and the pH 7.4 buffers contained 10 mM phosphate and 2 mM DTT.

3.3.5. Hairpin Monomer With MPER is Highly Fusogenic

Earlier work has shown that peptides corresponding to the FP and the MPER regions induce vesicle fusion (27,29). For the present study, we investigated the effects of these regions as part of the gp41 ectodomain including the hairpin region. The typical vesicle composition was POPC:POPG:Chol (8:2:5) and reflected the major fraction PC, minor fraction negatively charged lipid, and mole fraction Chol in the membranes of HIV and its host cells (40). Fusion was studied at final pHs of 3.2 and 7.4 which was motivated by previous studies with HP showing much greater fusion of negatively charged vesicles at pH 3.2 relative to pH 7.4 (41). A second vesicle composition was POPC:DOTAP:Chol (8:2:5) and contained positively-charged DOTAP rather than negatively charged POPG. Comparison of results for the two compositions and two pH's provides insight into the role of protein/vesicle electrostatics because the POPG and DOTAP are respectively negatively and positively charged at both pH's whereas the protein is positively charged at pH 3.2 and negatively charged at pH 7.4.

HP and HM were solubilized as monomers at pH 3.2 and induced rapid fusion of negatively-charged POPG vesicles at this pH with low 1:700 protein:total lipid mole ratio (Figure 3.10A). At pH 7.4, comparable fusion required an order-of-magnitude higher 1:75 ratio (Figure 3.10B). For 1:150 ratio, there was negligible fusion which is consistent with earlier work. For positively-charged DOTAP vesicles, there was little fusion at pH 3.2, measurable fusion at pH 7.4, and higher fusion at pH 9.0 (Figure 3.10C,D). The different pH dependences for POPG and DOTAP vesicles likely reflect differences in protein/vesicle electrostatic energy, i.e. (1) attractive energy between the positively-charged protein and negatively-charged POPG vesicles at low pH and

between the negatively-charged protein and positively-charged DOTAP vesicles at neutral and high pH; and (2) repulsive energy between the negatively-charged protein and negatively-charged POPG vesicles at neutral pH and between the positively-charged protein and positively-charged DOTAP vesicles at low pH. Attractive energy results in a higher fraction bound protein and repulsive energy results in a lower fraction. In the spatially restricted environment of HIV/host cell fusion, there is likely high protein/membrane binding that is similar to near-quantitative binding during the pH 3.2 vesicle fusion of POPG vesicles. Higher fusion extent for POPG vesicles at low pH relative to DOTAP vesicles at neutral and high pH may be a consequence of the different magnitudes of attractive electrostatic energy. This hypothesis is evidenced by calculated hairpin charges of about +10 at pH 3.2, -2 at pH 7.4, and -4 at pH 9.0. For many data, fusion was rapid and occurred during the ~5 s dead-time of the assay which is consistent with earlier studies. Other studies have also shown that as the fraction of charged lipid in the vesicle is decreased, the fusion rate decreases but the fusion extent increases where the latter effect is probably due to reduced intervesicle electrostatic repulsion. Fusion is enhanced with inclusion of the MPER in the construct as evidenced by the following differences for HM relative to HP with POPG vesicles: (1) at pH 3.2, higher initial fusion rate; and (2) at pH 7.4, higher final fusion extent (M_f). For pH 3.2, the final [protein] = 0.2 μ M (~0.003 mg/mL) and the hairpin monomer likely makes initial contact with the vesicle. HM-induced fusion occurs within a few seconds, so the monomer is likely a fusion-active species. The enhanced fusion with inclusion of the MPER is consistent with earlier observations of vesicle fusion induced by MPER peptides.

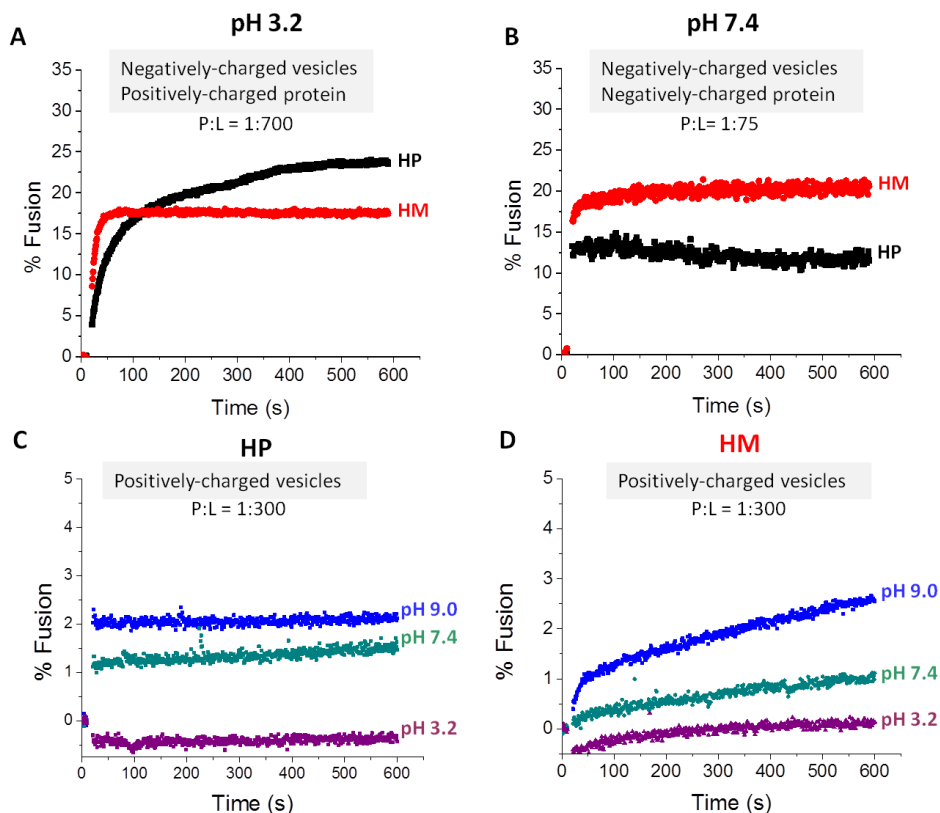


Figure 3.10. Vesicle fusion at 37 °C. Protein was added just prior to the increase in fusion with ~5 s assay dead time. For panels A and B, the vesicles were negatively-charged and for panels C and D, the vesicles were positively-charged. The calculated protein charge is +16, -2, and -4 at respective pH values 3.2, 7.4, and 9.0. The protein:lipid mole ratios were: (A) 1:700, (B) 1:75, and (C, D) 1:300, with Chol not included in the lipid quantity. The data overall support the importance of attractive protein/vesicle attractive energy for fusion at low protein:lipid ratios. The protein stock contained 20 μ M protein (~0.25 mg/L) in 10 mM formate at pH 3.2 with 0.2 mM TCEP which are condition for predominant monomeric protein. For panels A and B, [POPC] = 120 μ M, [POPG] = 30 μ M, and [Chol] = 75 μ M, and for panels C and D, the vesicles contained [POPC] = 120 μ M, [DOTAP] = 30 μ M, and [Chol] = 75 μ M. Vesicle buffers were 10 mM formate at pH 3.2, 5 mM HEPES/10 mM MES at pH 7.4, and 10 mM Tris-HCl at pH 9.0.

3.3.6. Synergy of FP, Hairpin, and MPER With the Gp41 Copy Number of a Virion

The individual and collective contributions of the FP, hairpin, and MPER to ectodomain-induced fusion were studied with HP, HM, and FP-HM solubilized as hexamers in 6 M GuHCl. The final [GuHCl] = 40 mM and in the absence of protein, only modest fusion was observed at pH 3.2 ($M_f \approx 4\%$) and no fusion was observed at pH 7.4 (Figure 3.13). The probable reason for GuHCl to induce fusion to some extent is due to reduction in electrostatic repulsion between the negatively-charged vesicles due to the presence of bound GuH^+ ions. Much greater fusion was observed with protein (Figure 3.11). Relative to HP which lacked most of the MPER, inclusion of the full MPER in HM resulted in higher M_f . Even greater M_f was induced for FP-HM which contains both the FP and MPER. Similar quantitative binding was observed in the absence and presence of the FP (41,42). Dose response was observed for HM and FP-HM (Figure 3.12). A substantial M_f of $\sim 15\%$ was observed for FP-HM even at ultra-low protein:lipid = 1:4800 that corresponds to ~ 15 proteins per ~ 100 nm diameter vesicle. This number is significant because it is comparable to the ~ 30 gp41 molecules per virion. To our knowledge, this is the first observation of rapid (~ 5 s) ectodomain-mediated vesicle fusion at the protein copy number of a virion. In some contrast, rapid FP or MPER peptide-induced vesicle fusion has typically been observed at 500-1000 peptides per vesicle. Our result highlights the importance of the whole ectodomain in fusion, including contributions from FP, hairpin, and MPER regions (43,44). The fusogenicity of FP-HM is even higher than shown in Fig. 8 because our FP-HM stock contained $\sim 40\%$ HM impurity.

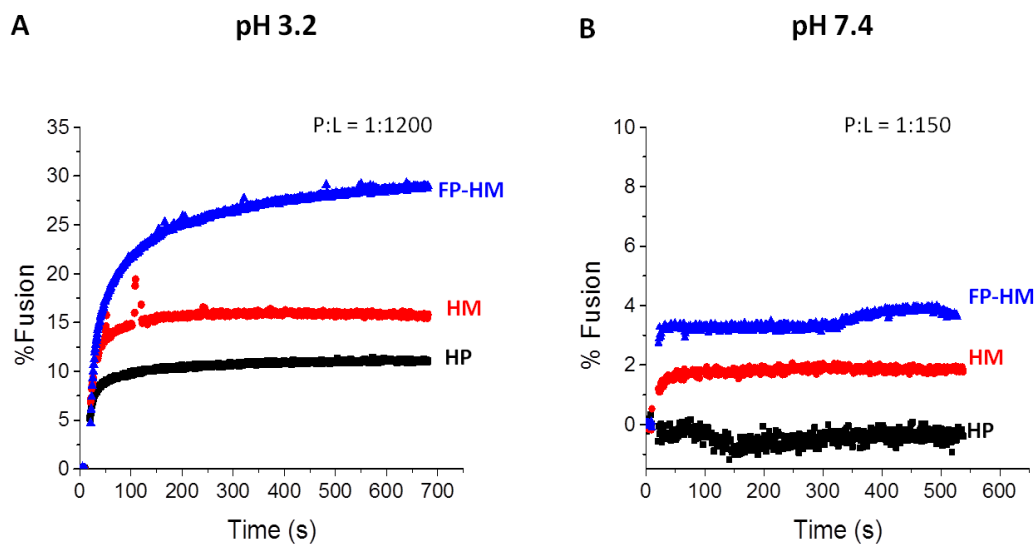


Figure 3.11. Vesicle fusion at 37 °C. A 7.5 μ L aliquot of protein was added just prior to the increase in fusion and the final total volume was 1200 μ L. The POPC:POPG:Chol vesicle composition and buffers were the same as Fig. 6. The protein stock buffer was 10 mM formate at pH 3.2 with 6 M GuHCl and 0.2 mM TCEP and following addition of protein stock, the assay buffer contained 40 mM GuHCl. For the pH 3.2 assays, the stock [protein] = 20 μ M ($\sim$ 0.25 mg/mL), and for the pH 7.4 assays, the stock [protein] = 160 μ M ($\sim$ 2 mg/mL). The [protein]:[POPC+POPG] = 1:1200 for the pH 3.2 assays and 1:150 for the pH 7.4 assays.

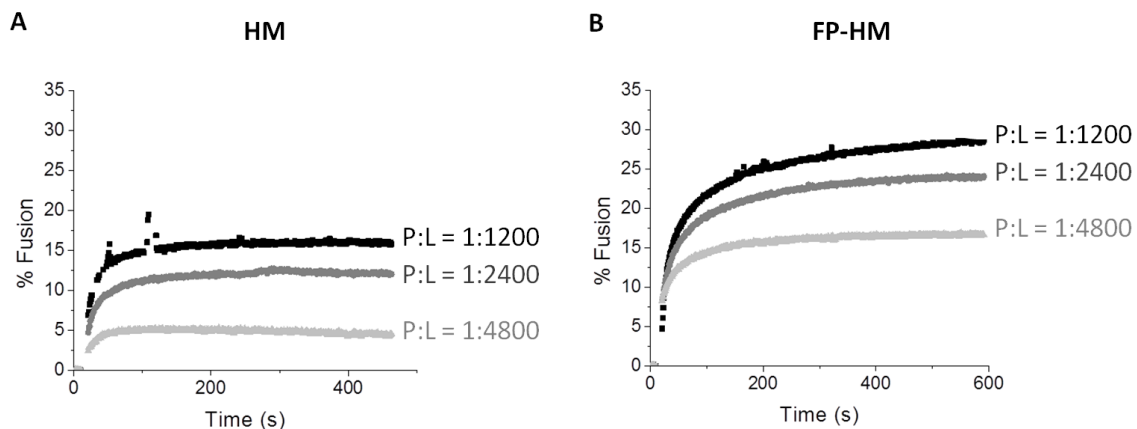


Figure 3.12. Dose response of vesicle fusion induced by HM and FP-HM at pH 3.2. The assay conditions were the same as Fig. 7A except that the protein stock concentrations were 20, 10, and 5 μM for [protein]:[total lipid] = 1:1200, 1:2400, and 1:4800, respectively.

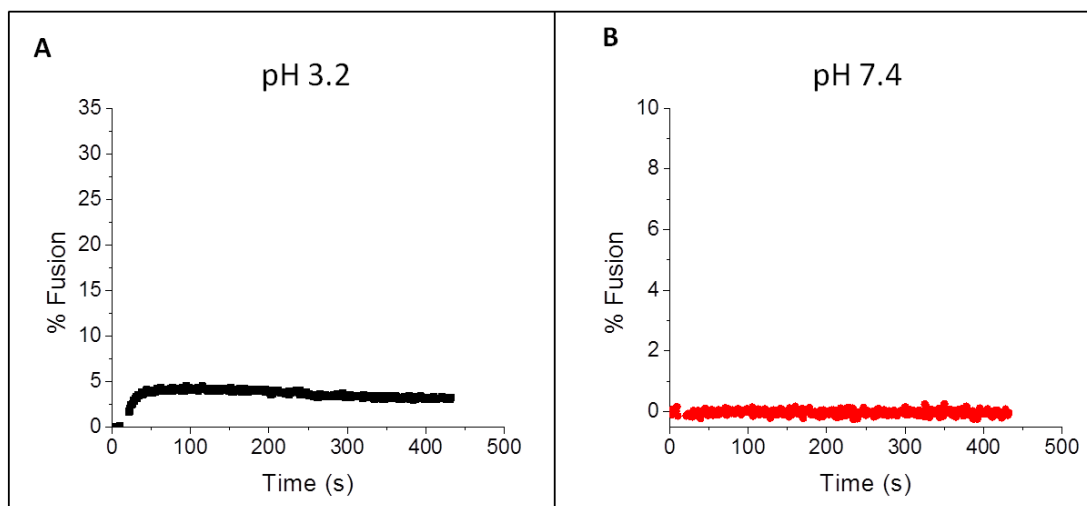


Figure 3.13. Vesicle fusion induced by GuHCl at (A) pH 3.2 or (B) pH 7.4. A 7.5 μL aliquot containing 6 M GuHCl was added to a vesicle suspension with final total [GuHCl] = 40 mM in a volume of 1200 μL . The assay conditions were the same as Figure 3.7 except that there was no protein in the aliquot that contained GuHCl.

3.3.7. Folded Hairpin Ectodomain With MPER Binds to bNAbs.

The MPER region of gp41 includes the linear epitopes of several bNAbs and there is consequently continued effort to develop a MPER-based immunogen as a HIV vaccine. HM or similar folded hairpin molecules are appealing because of their high stability but there are conflicting data about the extent to which they bind bNAbs. However, our Western blots show reproducible binding of HM to the 4E10 bNAb (Figure 3.14A). HP lacks the 4E10 epitope and serves as a negative control. Western blots also show that HM binds to the 2F5 bNAb more strongly than HP which includes only part of the 2F5 epitope (Figures 3.14B and 3.1B). In the Western blot approach, the protein is in 15% SDS prior to antibody binding. Although we do not know HM structure in this condition, HM hairpin structure is hyperthermostable and a shorter ectodomain construct without the MPER forms crystals in SDS with helical SHB structure (45). A gel shift is not observed between HP and HM (Figures 3.5B and 3.14B) and could be the result of folded rather than random coil structure. HM is predominantly monomeric in 15% SDS and may therefore bind bNAb as a hairpin monomer. The blots also show minor dimer populations binding to 2F5.

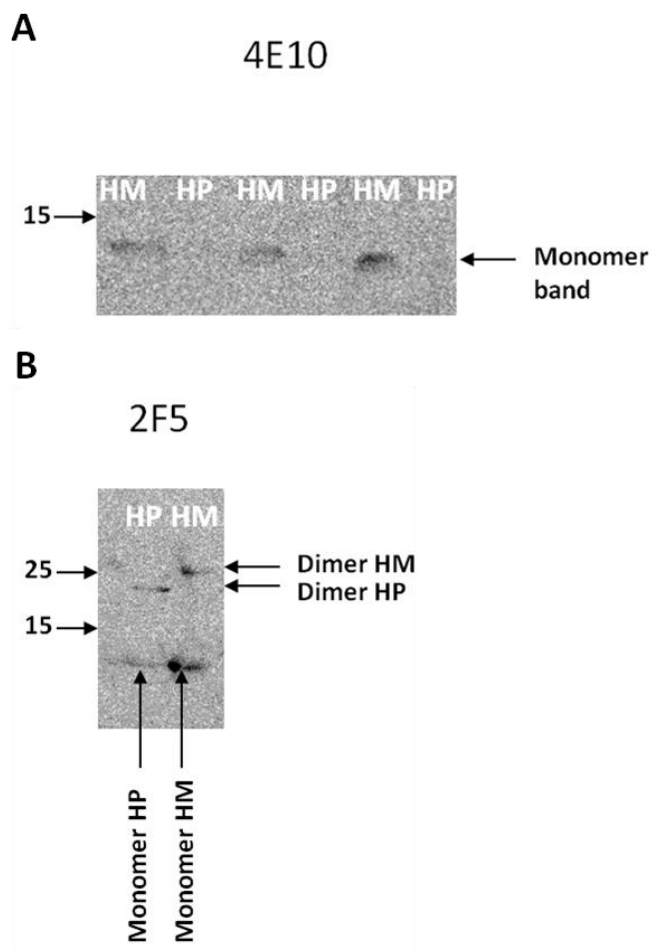


Figure 3.14. Western blots showing binding of HM to the 4E10 and 2F5 bNAbs and weaker binding of HP to 2F5. The left-side arrows are MWs in kDa.

Binding to bNAbs was probed under physiological conditions using direct immunoprecipitation. The protein stock solution contained either monomeric or hexameric protein and the binding was done at pH 7.4 with $[\text{GuHCl}] < 10 \text{ mM}$. For either initial oligomeric state, HM bound to 4E10 and 2F5 (Figure 3.15) whereas there was no binding to IgG (Figure 3.16). This is the antigenicity needed for a vaccine immunogen. HP binds 2F5 weaker than HM which is consistent with an incomplete epitope in HP.

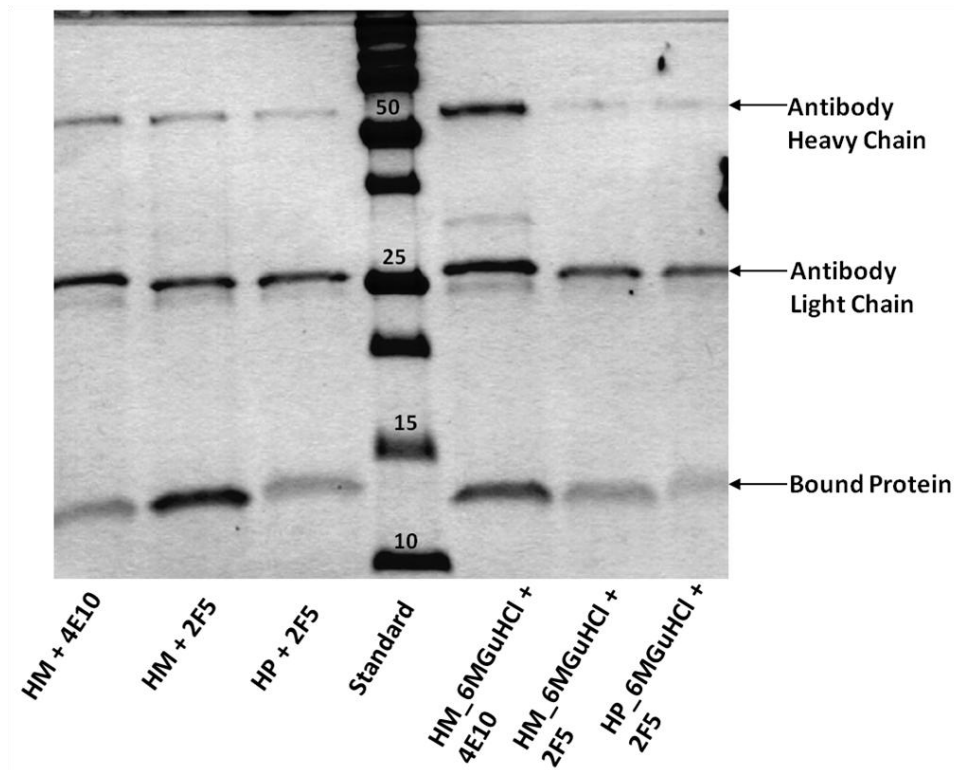


Figure 3.15. SDS-PAGE of the immunoprecipitation of HP and HM with the 4E10 and 2F5 bNAbs.

There is stronger binding of HM to 4E10 and 2F5 and weaker binding of HP to 2F5. The protein stock buffer contained 50 mM sodium formate at pH 3.2 with 150 mM NaCl. For the right three lanes of the panel A gel, the buffer also contained 6 M GuHCl whereas for the left three lanes of panel A, the GuHCl was absent. Binding was done in PBS at pH 7.4 and for the right lanes, the diluted [GuHCl] = 40 mM. The center lane is the MW ladder in kDa.

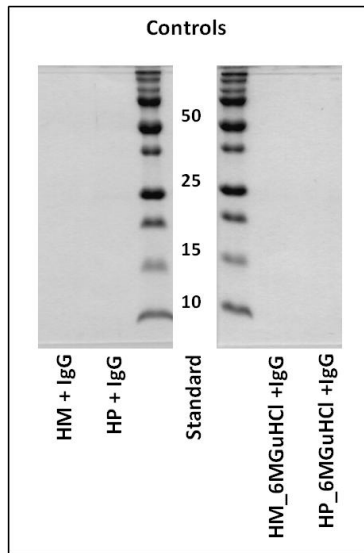


Figure 3.16. SDS-PAGE of the immunoprecipitation of HP and HM with IgG. Protein binding to IgG is not observed. The protein stock buffer contained 50 mM sodium formate at pH 3.2 with 150 mM NaCl. For the right two lanes of the gel, the buffer also contained 6 M GuHCl. The antibody binding was done in PBS at pH 7.4 and for the right lanes of the gel, the diluted [GuHCl] = 40 mM. The inner lanes are MW ladders in kDa.

3.4. Discussion

Significant findings: The ectodomain of the HIV gp41 protein plays a major role in catalyzing fusion between HIV and host cell membranes. The likely respective binding of the FP and MPER to the host cell and viral membranes are probably part of the underlying physical basis for catalysis. Although there has been significant functional and structural study of FP and MPER peptides, there have been just a few investigations of these regions as part of the larger ectodomain. The present work shows the significance of the FP, hairpin, and MPER regions for ectodomain-induced vesicle fusion including fusion with ~15 proteins per vesicle which is

comparable to the gp41 copy number of the virion. For FP or MPER peptides, there has typically only been appreciable rapid vesicle fusion at 500-1000 peptides per vesicle and to our knowledge, ours is this first observation of rapid vesicle fusion at the virion gp41 copy number. Detection of either predominant monomer or hexamer gp41 ectodomain rather than trimer gp41 ectodomain was incorporated into the new HIV/host cell membrane fusion **model III** (Figure 3.19). One significant feature is inclusion of monomer ectodomain rather than trimer ectodomain PHI→hairpin folding at step 3. The free energy of gp41 ectodomain folding is coupled to free energy of formation of membrane intermediates. Our work may also aid HIV vaccine development via an immunogen with a folded gp41 ectodomain including MPER. This immunogen has the advantage of very high stability but there is literature disagreement about ectodomain binding to bNAbs. We show bNAb binding to gp41 ectodomain protein initially prepared as a monomer or hexamer.

SSNMR is a novel and important tool in recombinant protein (RP) production: One key result that moved the work forward was SSNMR detection of HM in IBs at ~300 mg/L culture. The bottleneck to purified protein was therefore IB solubilization rather than expression and subsequent effort was focused on increasing solubilization. This result exemplifies how SSNMR is a general method to quantitate RP yield without solubilization or purification. High expression was obtained with common and inexpensive plasmids, *E. coli* strain, and growth and expression conditions. IB solubilization rather than expression may therefore be a general determinant of purified protein yield. Our yields were 15-50 mg/L and obtained with single-step affinity purification which should be advantageous in immunogen production. Our approach contrasts with earlier efforts that used protein fusion tags that had to be cleaved (46).

Monomer and hexamer rather than trimer gp41 ectodomain: For gp41 ectodomain constructs lacking solubility tags, only constructs with short NHR and CHR segments and no FP or MPER appear to form soluble trimers near pH 7 (47). For longer constructs that contain some or all of the FP and/or the MPER, e.g. HP and HM, the present and earlier studies show most protein forming large aggregates (MW>2 MDa) even in the presence of non-ionic detergents (48). There is solubility near pH 3 and many biophysical measurements have been done at this pH. High fractional helicity and $T_m > 100$ °C were considered strong evidence of SHB trimers.

One important contribution of the present study is SEC showing predominant monomer rather than trimer gp41 ectodomain at pH 3 (Figure 3.5). Our SEC monomer result is consistent with the SEC of large ectodomain constructs done by other groups (Figure 1A from Ref. 48 and Figure S1 from Ref. 51) although this monomer interpretation was typically not made by the authors of these papers. For these latter studies, the construct was either NHR+native loop+CHR, FP+NHR+native loop+CHR+MPER+TM, or short NHR+short loop+short CHR.

The gp41 ectodomain monomer is highly helical and hyperthermostable. The most plausible monomer structure is *N*-helix/180° turn/*C*-helix model (Figure 3.17A) similar to the monomer unit of the high-resolution SHB trimer structures. The trimer is stabilized by several inter-monomer NHR/NHR hydrophobic interactions and small reorientations of helices could reposition residues for favorable intra-monomer NHR/CHR interactions. A looser helical monomer hairpin structure has been observed for a short NHR+loop+CHR construct in non-ionic detergent but we favor the tight structure of figure 17A (49). Large ectodomain constructs like HP and HM have $T_m > 100$ °C either in the absence or presence of detergent and this high

thermostability likely reflects substantial interhelical contact in the folded monomer. The looser structure of much shorter ectodomain constructs may be reflected in their typical $T_m \leq 70$ °C.

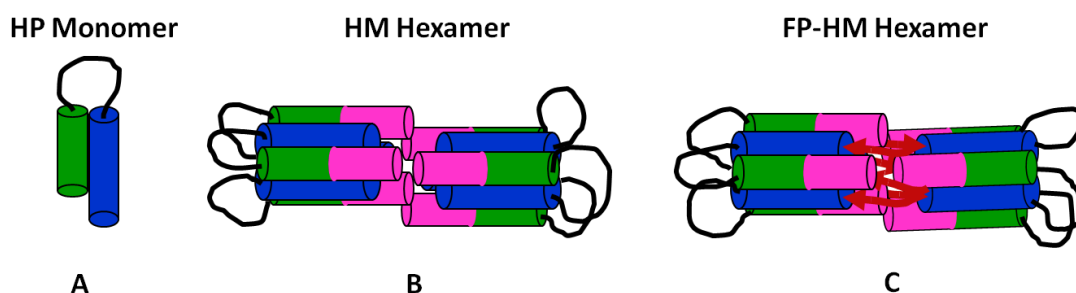


Figure 3.17. Working structural models of the gp41 ectodomain with figure 3.1 color coding.

Monomeric HP is modeled as the hairpin structure of a monomer in the SHB. HM and FP-HM hexamers are modeled as two SHB trimers that contact at their *N-/C-terminal* interfaces. This model correlates with the observed intermolecular antiparallel β sheet structure of FPs.

Measurements of large ectodomain constructs at low pH should be interpreted in the context of monomer protein. For example, the hyperthermostability detected by calorimetry and CD is likely a consequence of intra-molecular NHR/CHR interaction rather than inter-molecular NHR/NHR interaction. The stability of the ectodomain monomer would be consistent with an important role for the monomer in fusion, discussed in a subsequent section and presented in Figure 3.19. We also note that rapid vesicle fusion using stock protein at low pH is likely a consequence of membrane perturbation by hairpin monomer rather than SHB trimer. This includes fusion of negatively-charged vesicles at pH 3.2 (Figure 3.10A) and positively-charged vesicles at pH 7.4 and 9.0 (Figure 3.10C,D).

For 10 mg HP/mL SEC loading, there is a major monomer peak as well as minor trimer and oligomer/aggregate peaks (Figure 3.9B). The normalized values of monomer:trimer:larger oligomer integrations are: 0.72:0.18:0.10. These correspond to [monomer] $\approx$ 60 μ M and [trimer] $\approx$ 5 μ M using the experimentally-determined ten-fold dilution in SEC. For the association equilibrium, 3 monomer $\leftrightarrow$ trimer, the derived $K_a \approx 2 \times 10^7 M^{-2}$, and corresponds to equal mass concentrations of monomer and trimer for total HP of $\sim$ 9 mg/mL ($\sim$ 90 mg/mL loading in SEC). Earlier analytical ultracentrifugation (AUC) data have been analyzed to yield $K_a \approx 5 \times 10^{11} M^{-2}$ (14,50,51). We do not understand the discrepancy between the SEC and AUC analyses.

Protein hexamers were observed in 6 M GuHCl at either pH 3.2 or 7.4 (Figure 3.9). The data show the hexamer as the fundamental unit of oligomerization rather than the trimer, e.g. dodecamers were observed but not nonamers. A reasonable model for a hexamer is a dimer of SHB trimers with inter-trimer contact between the MPERs (Figure 3.17B). This model is consistent with the observed helical structure in the hexamer and with $T_m > 100^\circ\text{C}$. This model is also consistent with some of the crystal packing of SHB trimers (39). In addition, SSNMR studies of FP structure in the membrane-associated gp41 ectodomain show an oligomeric FP β sheet with antiparallel but not parallel arrangement of the FP strands (24). The antiparallel arrangement would be consistent with end-on arrangement of two SHB trimers (Figure 3.17C).

It is important to consider the impact that GuHCl may have on the ectodomain structure as GuHCl often denatures proteins. We think denaturation is unlikely for the ectodomain because there is substantial helical structure and $T_m > 100^\circ\text{C}$. In addition, to our knowledge, it is very

unusual that denatured protein would have a well-defined oligomeric state other than monomer. A control experiment was performed by running a size exclusion chromatography experiment of protein GB1 in the presence of 6M GuHCl at low pH 3.2 and physiologic pH 7.4. The protein GB1 construct which has been studied has a MW ~6.2 KDa. Previous spectroscopic studies have showed the complete unfolding of protein GB1 in the presence of 6M GuHCl (52). Consistent with this result, we have obtained a clear monomer peak of the protein GB1 in the presence of 6M GuHCl with the MW equal to ~6.3 kDa and ~9 kDa at pH 3.2 and 7.4 respectively (Figure 3.18). There was no other peak showing the presence of higher order oligomer. The hexamer is also the dominant oligomeric state of a modified ectodomain construct in physiologic solution, i.e. pH 7.4 without any GuHCl (38). This Soc-gp41ectoM-Fd construct includes a central gp41ectoM region which is very similar to FP-HM as well as a ~9 kDa non-native Soc protein *N*-terminal of the FP and a ~2 kDa Fd non-native protein *C*-terminal of the MPER. The Soc and Fd proteins were chosen because they form homotrimers in physiological solution. However, there was only minor trimer population of Soc-gp41ectoM-Fd in physiologic solution. Instead, the hexamer was dominant with additional smaller populations of two larger oligomers. The overall oligomer distribution of Soc-gp41ectoM-Fd in physiologic solution is very similar to that of HP and HM in 6 M GuHCl with dominant hexamer and two minor higher-order oligomers that are reasonably assigned to dodecamer and 24-mer. It is likely that there are similar hexamer structures of HP, HM, and Soc-gp41ectoM-Fd.

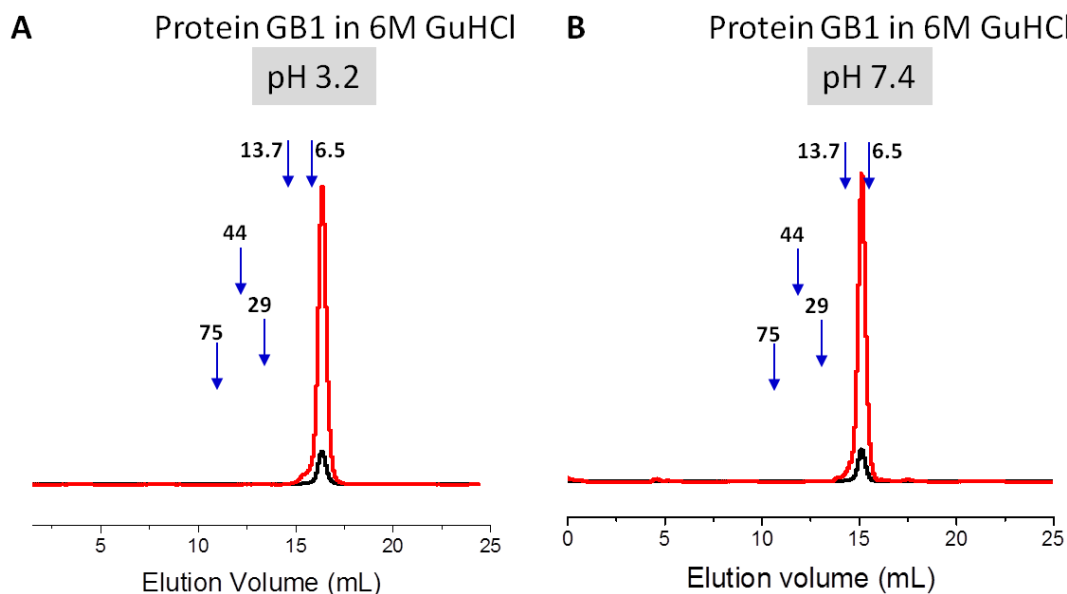


Figure 3.18. The monomeric elution peak of protein GB1 in PBS in the presence of 6M GuHCl at (A) pH 3.2 and (B) pH 7.4. The elution volume corresponds to the molecular weight of ~6.3 kDa and ~9 kDa at pH 3.2 and 7.4 respectively. The size exclusion chromatography was done at Superdex 75 column. The blue arrows show the elution volumes of MW standards with the MWs given in kDa.

At neutral pH, large ectodomain constructs form >2 MDa aggregates whereas at low pH, monomers predominate for <10 mg protein/mL and there is significant trimer population with SHB structure for >50 mg protein/mL. The pH-dependence of oligomeric state is observed with or without non-ionic detergent so our explanatory model is based on relative magnitudes of protein electrostatic vs hydrophobic effects. Trimer stabilization is primarily due to hydrophobic interaction between the three NHR helices on the SHB interior. At low pH, there is a calculated charge of ~+10 for the hairpin region of the ectodomain monomer. We posit that monomer rather than oligomers are favored at low pH because inter-molecular electrostatic repulsion

overwhelms the hydrophobic interaction. At neutral pH, the calculated charge is -2 and hydrophobic interaction is dominant. The resulting aggregates may be composed of SHB trimers.

For solutions containing GuHCl at both low and neutral pH, hexamers predominate, i.e. addition of GuHCl induces monomer→hexamer oligomerization at low pH and induces aggregate→hexamer breakup at neutral pH. To our knowledge, the specific interaction of positively-charged GuH^+ with protein molecules is not understood (53). For the ectodomain protein, we posit that the hexamer is formed from two SHB trimers (Figure 3.17B,C) and that GuH^+ interacts with COO^- groups on the SHB surfaces. The COO^- rather than the COOH state is favored at both pH's because of attractive interaction with GuH^+ . At low pH, this reduces the positive charge of the protein so that the SHB trimer does not dissociate due to electrostatic repulsion between monomers. Hexamers form because of favorable hydrophobic interaction between the FP/MPER end of one SHB trimer and the FP/MPER end of a second trimer. Aggregation beyond hexamers is disfavored because this aggregation would reduce exposed protein surface area and therefore the number of $\text{GuH}^+/\text{COO}^-$ contacts. As noted above, SOC-gp41ectoM-Fd forms a hexamer in the absence of GuHCl and we posit this is also two SHB trimers. Further aggregation is disfavored because of the solubility of the SOC and FD proteins and perhaps because of their steric bulk.

Correlation of vesicle fusion with HIV/cell fusion: One significant result is observation of vesicle fusion for ~ 15 FP-HM per vesicle (Figure 3.12B) which is comparable to the gp41 copy number of ~ 30 in the virion. To our knowledge, FP-HM thus provides the first example for which

protein-induced vesicle fusion might reasonably be directly compared to viral fusion. FP-HM is more fusogenic than shorter ectodomain constructs like HP which lacks the FP and MPER, and also more fusogenic than isolated FP or MPER peptides. This potent fusogenicity highlights the importance of the full ectodomain in fusion, with inclusion of the *N*-terminal FP, hairpin region, and *C*-terminal MPER. During HIV/cell fusion, the FP and MPER likely bind to the cell and HIV membranes, respectively, and higher fusion for FP-HM relative to HM supports synergy between these two regions, likely via membrane perturbations. Vesicle fusion is probably a better model of the earlier hemifusion (lipid mixing) step of viral fusion than the final fusion pore expansion step. Pore formation and expansion in vesicle fusion is less regulated than in HIV/cell fusion and leakage of contents out of the vesicles occurs much faster than mixing of contents between vesicles.

For conditions with attractive protein/vesicle electrostatic energy, protein/vesicle binding is quantitative and FP-independent. The observed FP-dependent fusogenicity is therefore not reasonably ascribed to FP-dependent binding. Protein binding is described by the equation $f_m = (K_m \times [L]) / \{1 + (K_m \times [L])\}$ where $f_m \equiv$ fraction bound-protein, $K_m \equiv$ equilibrium constant derived from $G_m \equiv$ binding free-energy, and $L \equiv$ lipid. For our assays, $[L] = 1.5 \times 10^{-4} M$ and protein/vesicle binding is quantitative ($f_m = 1.0$) when $K_m > 10^5 M$. The $G_m = G_{elec} + G_{hydro}$ and corresponding $K_m = K_{elec} \times K_{hydro}$ describe contributions from electrostatic and hydrophobic interactions. The FP is hydrophobic but lacks ionizable residues so it affects G_{hydro} but not G_{elec} . For negatively-charged vesicles at pH 3 and positively-charged vesicles at pH 9, the calculated $K_{elec} > 10^5 M$ and corresponds to quantitative binding of all proteins, which has also been experimentally-observed. For conditions with much smaller K_{elec} , such as negatively-charged

vesicles at neutral pH, binding may be FP-dependent and changes in fusogenicity could be related to binding differences.

High fusion by FP-HM requires a pH for which there is attractive protein/vesicle electrostatic energy with resultant quantitative binding to the vesicles (Figure 3.10). Attractive electrostatic energy is not a requirement for fusion by the ectodomain of full-length gp41 in the virion, likely because the ectodomain is restricted to the space between the viral and host cell membranes. This spatial confinement favors protein binding to the membranes even without electrostatic attraction.

For the present study, a 0.2 fraction charged lipid was used which is larger than the typical fraction in the cell membrane. However, a recent study showed efficient vesicle fusion with much smaller fractions charged lipid (42). As the fraction charged lipid is decreased, there is a corresponding increase in fusion extent that is likely due to decreased intervesicle electrostatic repulsion. Some charged lipid is needed for quantitative protein binding to the vesicles and consequently fusion. As one example, HP at protein:lipid = 1:700 (~100 HP molecules/vesicle) induced ~40% fusion of vesicles containing only 0.02 fraction negatively-charged phosphatidylserine lipid. These results, along with the higher fusogenicity of FP-HM relative to HP (Figure 3.11A), support correlation between FP-HM induced vesicle fusion and hemifusion between the HIV and cell membranes mediated by the viral ectodomain.

New fusion model with gp41 monomer and hexamer ectodomain: As noted in the Introduction, models for gp160-induced membrane fusion have typically considered changes in structure of a gp160 trimer that are temporally and energetically coupled to changes in the topologies of the

HIV and host cell membranes. Models I and II of the Introduction share the feature of trimeric ectodomain gp41 throughout fusion. For model I, concerted PHI→SHB trimer folding is coupled to subsequent hemifusion. However, this is inconsistent with fusion inhibition by CHR+MPER peptides up to the final pore expansion step and the corollary assumption of peptide binding to the PHI trimer but not the SHB trimer. Model II delays the PHI→SHB trimer folding until the pore expansion step which suggests that the SHB is only associated with fused membrane stabilization and fusion arrest. The new model III incorporates asynchronous ectodomain monomer PHI→hairpin folding followed by changes in membrane topology leading to hemifusion (Figure 3.19). The steps of membrane fusion led by gp41 according to the proposed new model are as follows - (1) the gp120 and gp41 are non-covalently associated as a trimer structure. At this stage the FP is expected to be inside the hydrophobic core of gp120 and gp41 trimer structure, as FP is a very hydrophobic domain. (2a&b) After the gp120 binds to the host cell receptors and coreceptors, gp120 moves away, and gp41 is exposed. The FP comes very close to the host membrane, which further helps FP to insert into the host membrane and stabilize. However, this close proximity of the host and viral membrane is only possible if the gp120 is still bound to the gp41 while being attached to the host cell receptors and coreceptors. On the other hand this attachment of gp120 to the gp41 is very unlikely as in the presence of gp120, the folding of gp41 into hairpin conformation and subsequent SHB formation would not be possible due to steric hindrance. (3-5) There is subsequent ectodomain monomer hairpin assembly into SHB trimers and (6-7) final assembly of two trimers into a hexamer with an antiparallel FP β sheet. During viral fusion, the *N*-terminal FP is likely attached to the cell membrane and the *C*-terminal TM is in the viral membrane. In our view, these membrane

topologies are more reasonably maintained with asynchronous folding of individual monomer ectodomains than with concerted folding of the three ectodomains of the trimer. Asynchronous ectodomain monomer folding to a hairpin is also consistent with the hyperthermostability ($T_m \approx 110^\circ\text{C}$) of the monomer hairpin.

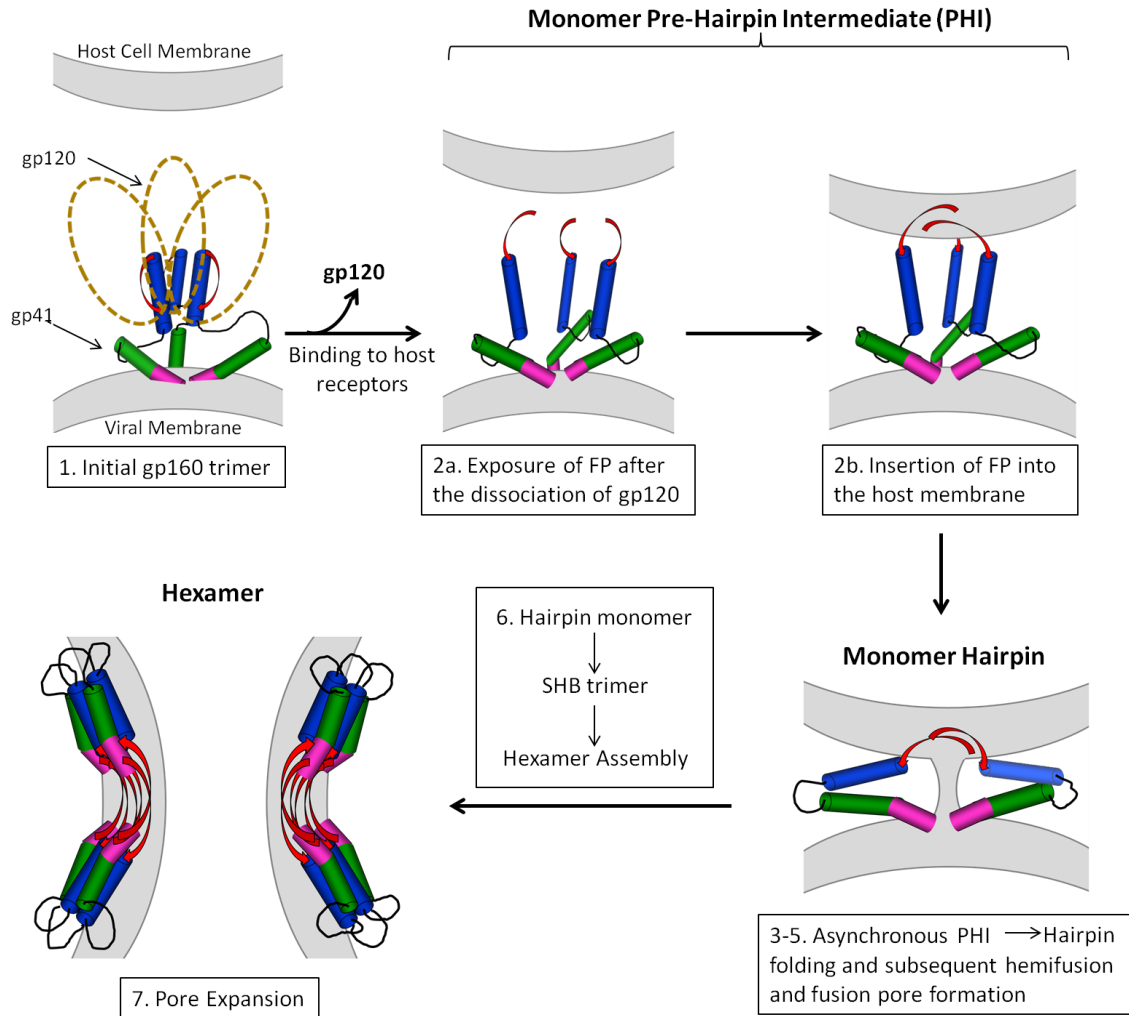


Figure 3.19. Model for membrane fusion that includes the gp41 ectodomain monomer and hexamer. The different regions of gp41 are color coded according to figure 3.1 and the TM and endodomains are not shown. One of the monomers is not displayed in steps 3-5. The initial gp41 structure of step 1 is based on the ~ 5 Å gp140 structures and the final SHB structure of step 7 is based on high-resolution structures.

There is evidence supporting the involvement of monomer hairpin gp41 ectodomain at some stage of HIV/host cell fusion. Such fusion has typically been assayed using the surrogate system of fusion between cells expressing gp160 and cells expressing CD4 and co-receptor proteins. Fusion is quantitated using the number of syncytia (fused cell aggregates). There is typically dramatic reduction in fusion with mutations impacting intra-monomer CHR/NHR interactions, e.g. no syncytia are formed with the W571R mutation (54). Membrane fusion and HIV infection are inhibited by the clinically prescribed CHR+MPER enfuvirtide therapeutic which could act by binding to exposed NHR surfaces in the PHI or hairpin ectodomain monomer but would not bind to the SHB trimer for which the three NHR helices are on the bundle interior (55). The physiological relevance of the gp41 ectodomain monomer is also supported by observation of monomer rather than trimer WT gp140 protein (8). Gp140 is a non-covalent complex between the HIV gp120 protein and the ectodomain of gp41 but lacks the TM and intraviral regions of gp41.

A role for hexameric gp41 in HIV/cell fusion is supported by the trans-dominant effect of the V513E mutation in the FP region of gp41 on gp160-mediated membrane fusion and HIV infection (56). These data have been modeled using a requirement of multiple gp160 trimers for fusion and infection (26). A membrane-associated gp41 hexamer is consistent with the antiparallel β sheet FP structure in the membrane-associated gp41 ectodomain observed by SSNMR (24). Such structure is consistent with interleaved FP strands from two trimers (Figure 3.17C).

After submission of the present paper, another paper was published showing that a shorter NHR+loop+CHR gp41 ectodomain construct is monomeric at low pH in non-ionic detergent (49). This result is consistent with our observations for the larger HP and HM constructs. We also observed that pH rather than detergent is the critical parameter for solubility. In the absence of GuHCl, both the HP and HM constructs are soluble at low pH and insoluble at neutral pH and these results are independent of the presence or absence of detergent. The short construct in detergent at low pH has a loose hairpin structure with non-interacting NHR and CHR helices which lie on the micelle surface. The low pH micelle location is consistent with our observations of protein-to-lipid headgroup contact for a large, membrane-associated ectodomain construct. At low pH, NHR+CHR backbone ^{13}C -to-lipid ^{31}P distances of 8-9 Å were observed via SSNMR spectroscopy (42). At neutral pH, there was no detectable contact which corresponds to ^{13}C - ^{31}P distances >12 Å even though the protein remained membrane-bound via the FP region. The pH-dependence of the protein-to-lipid distances is consistent with the attractive protein/membrane electrostatic energy at low pH and repulsive energy at neutral pH.

Although not included in the gp41 constructs of the present study, the TM domain is important for anchoring gp41 in the viral membrane and is likely also important in fusion pore formation (57). The TM sequence is conserved across clades of HIV and this conservation may be needed for efficient trafficking of gp160 to the infected cell membrane surface prior to viral budding (58). There are moderate effects on fusion and infection with some mutations in the TM region (59,60). It would be interesting in the future to study the oligomerization and fusion activity of gp41 constructs that include the TM domain. SEC of a full ectodomain+TM construct showed predominant monomers at pH 4 in detergent which is consistent with the SEC for ectodomain

without TM (51). There was little contact between the FP and TM regions for the large construct in detergent but some contact has been observed between synthesized FP peptide and TM peptide in membranes (61).

REFERENCES

REFERENCES

1. Grewe, C., Beck, A., and Gelderblom, H. R. (1990) HIV: early virus-cell interactions. *J. AIDS* **3**, 965-974
2. White, J. M., Delos, S. E., Brecher, M., and Schornberg, K. (2008) Structures and mechanisms of viral membrane fusion proteins: Multiple variations on a common theme. *Crit. Rev. Biochem. Mol. Biol.* **43**, 189-219
3. Durell, S. R., Martin, I., Ruyschaert, J. M., Shai, Y., and Blumenthal, R. (1997) What studies of fusion peptides tell us about viral envelope glycoprotein-mediated membrane fusion. *Mol. Membr. Biol.* **14**, 97-112
4. Montero, M., van Houten, N. E., Wang, X., and Seott, J. K. (2008) The membrane-proximal external region of the human immunodeficiency virus type 1 envelope: Dominant site of antibody neutralization and target for vaccine design. *Microbiol. Mol. Biol. Rev.* **72**, 54-84
5. Nieva, J. L., and Agirre, A. (2003) Are fusion peptides a good model to study viral cell fusion? *Biochim. Biophys. Acta* **1614**, 104-115
6. Epand, R. M. (2003) Fusion peptides and the mechanism of viral fusion. *Biochim. Biophys. Acta* **1614**, 116-121
7. Liu, J., Bartesaghi, A., Borgnia, M. J., Sapiro, G., and Subramaniam, S. (2008) Molecular architecture of native HIV-1 gp120 trimers. *Nature* **455**, 109-U176
8. Schulke, N., Vesanen, M. S., Sanders, R. W., Zhu, P., Lu, M., Anselma, D. J., Villa, A. R., Parren, P., Binley, J. M., Roux, K. H., Maddon, P. J., Moore, J. P., and Olson, W. C. (2002) Oligomeric and conformational properties of a proteolytically mature, disulfide-stabilized human immunodeficiency virus type 1 gp140 envelope glycoprotein. *J. Virol.* **76**, 7760-7776
9. Sanders, R. W., Vesanen, M., Schuelke, N., Master, A., Schiffner, L., Kalyanaraman, R., Paluch, M., Berkhout, B., Maddon, P. J., Olson, W. C., Lu, M., and Moore, J. P. (2002) Stabilization of the soluble, cleaved, trimeric form of the envelope glycoprotein complex of human immunodeficiency virus type 1. *J. Virol.* **76**, 8875-8889
10. Bartesaghi, A., Merk, A., Borgnia, M. J., Milne, J. L. S., and Subramaniam, S. (2013) Prefusion structure of trimeric HIV-1 envelope glycoprotein determined by cryo-electron microscopy. *Nat. Struct. Mol. Biol.* **20**, 1352-1357

11. Julien, J.-P., Cupo, A., Sok, D., Stanfield, R. L., Lyumukis, D., Deller, M. C., Klasse, P.-J., Burton, D. R., Sanders, R. W., Moore, J. P., Ward, A. B., and Wilson, I. A. (2013) Crystal structure of a soluble cleaved HIV-1 envelope trimer. *Science* **342**, 1477-1483
12. Lyumukis, D., Julien, J.-P., de Val, N., Cupo, A., Potter, C. S., Klasse, P.-J., Burton, D. R., Sanders, R. W., Moore, J. P., Carragher, B., Wilson, I. A., and Ward, A. B. (2013) Cryo-EM structure of a fully glycosylated soluble cleaved HIV-1 envelope trimer. *Science* **342**, 1484-1490
13. Tan, K., Liu, J., Wang, J., Shen, S., and Lu, M. (1997) Atomic structure of a thermostable subdomain of HIV-1 gp41. *Proc. Natl. Acad. Sci. U.S.A.* **94**, 12303-12308
14. Caffrey, M., Cai, M., Kaufman, J., Stahl, S. J., Wingfield, P. T., Covell, D. G., Gronenborn, A. M., and Clore, G. M. (1998) Three-dimensional solution structure of the 44 kDa ectodomain of SIV gp41. *EMBO J.* **17**, 4572-4584
15. Yang, Z. N., Mueser, T. C., Kaufman, J., Stahl, S. J., Wingfield, P. T., and Hyde, C. C. (1999) The crystal structure of the SIV gp41 ectodomain at 1.47 Å resolution. *J. Struct. Biol.* **126**, 131-144
16. Buzon, V., Natrajan, G., Schibli, D., Campelo, F., Kozlov, M. M., and Weissenhorn, W. (2010) Crystal structure of HIV-1 gp41 including both fusion peptide and membrane proximal external regions. *PLoS Pathog.* **6**, e1000880
17. Lu, M., Ji, H., and Shen, S. (1999) Subdomain folding and biological activity of the core structure from human immunodeficiency virus type 1 gp41: implications for viral membrane fusion. *J. Virol.* **73**, 4433-4438
18. Sackett, K., Nethercott, M. J., Epand, R. F., Epand, R. M., Kindra, D. R., Shai, Y., and Weliky, D. P. (2010) Comparative analysis of membrane-associated fusion peptide secondary structure and lipid mixing function of HIV gp41 constructs that model the early pre-hairpin intermediate and final hairpin conformations. *J. Mol. Biol.* **397**, 301-315
19. Wild, C. T., Shugars, D. C., Greenwell, T. K., McDanal, C. B., and Matthews, T. J. (1994) Peptides corresponding to a predictive alpha-helical domain of human immunodeficiency virus type 1 gp41 are potent inhibitors of virus infection. *Proc. Natl. Acad. Sci. U.S.A.* **91**, 9770-9774
20. Markosyan, R. M., Cohen, F. S., and Melikyan, G. B. (2003) HIV-1 envelope proteins complete their folding into six-helix bundles immediately after fusion pore formation. *Mol. Biol. Cell* **14**, 926-938
21. Matthews, T., Salgo, M., Greenberg, M., Chung, J., DeMasi, R., and Bolognesi, D. (2004) Enfuvirtide: The first therapy to inhibit the entry of HIV-1 into host CD4 lymphocytes. *Nature Rev. Drug Discovery* **3**, 215-225

22. Chernomordik, L. V., and Kozlov, M. M. (2008) Mechanics of membrane fusion. *Nat. Struct. Mol. Biol.* **15**, 675-683
23. Weissenhorn, W., Dessen, A., Harrison, S. C., Skehel, J. J., and Wiley, D. C. (1997) Atomic structure of the ectodomain from HIV-1 gp41. *Nature* **387**, 426-430
24. Sackett, K., Nethercott, M. J., Zheng, Z. X., and Weliky, D. P. (2014) Solid-state NMR spectroscopy of the HIV gp41 membrane fusion protein supports intermolecular antiparallel beta sheet fusion peptide structure in the final six-helix bundle state. *J. Mol. Biol.* **426**, 1077-1094
25. Sougrat, R., Bartesaghi, A., Lifson, J. D., Bennett, A. E., Bess, J. W., Zabransky, D. J., and Subramaniam, S. (2007) Electron tomography of the contact between T cells and SIV/HIV-1: Implications for viral entry. *PLoS Pathogens* **3**, 570-581
26. Magnus, C., Rusert, P., Bonhoeffer, S., Trkola, A., and Regoes, R. R. (2009) Estimating the stoichiometry of Human Immunodeficiency Virus entry. *J. Virol.* **83**, 1523-1531
27. Yang, J., Prorok, M., Castellino, F. J., and Weliky, D. P. (2004) Oligomeric b-structure of the membrane-bound HIV-1 fusion peptide formed from soluble monomers. *Biophys. J.* **87**, 1951-1963
28. Shnaper, S., Sackett, K., Gallo, S. A., Blumenthal, R., and Shai, Y. (2004) The C- and the N-terminal regions of glycoprotein 41 ectodomain fuse membranes enriched and not enriched with cholesterol, respectively. *J. Biol. Chem.* **279**, 18526-18534
29. Apellaniz, B., Garcia-Saez, A. J., Nir, S., and Nieva, J. L. (2011) Destabilization exerted by peptides derived from the membrane-proximal external region of HIV-1 gp41 in lipid vesicles supporting fluid phase coexistence. *Biochim. Biophys. Acta* **1808**, 1797-1805
30. Zhu, P., Chertova, E., Bess, J., Lifson, J. D., Arthur, L. O., Liu, J., Taylor, K. A., and Roux, K. H. (2003) Electron tomography analysis of envelope glycoprotein trimers on HIV and simian immunodeficiency virus virions. *Proc. Natl. Acad. Sci. U.S.A.* **100**, 15812-15817
31. Huang, J. H., Ofek, G., Laub, L., Louder, M. K., Doria-Rose, N. A., Longo, N. S., Imamichi, H., Bailer, R. T., Chakrabarti, B., Sharma, S. K., Alam, S. M., Wang, T., Yang, Y. P., Zhang, B. S., Migueles, S. A., Wyatt, R., Haynes, B. F., Kwong, P. D., Mascola, J. R., and Connors, M. (2012) Broad and potent neutralization of HIV-1 by a gp41-specific human antibody. *Nature* **491**, 406-412
32. Frey, G., Peng, H., Rits-Volloch, S., Morelli, M., Cheng, Y., and Chen, B. (2008) A fusion-intermediate state of HIV-1 gp41 targeted by broadly neutralizing antibodies. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 3739-3744

33. Wang, J., Tong, P., Lu, L., Zhou, L. L., Xu, L. L., Jiang, S. B., and Chen, Y. H. (2011) HIV-1 gp41 core with exposed membrane-proximal external region inducing broad HIV-1 neutralizing antibodies. *PLoS One* **6**, e18233
34. Dawood, R., Benjelloun, F., Pin, J. J., Kone, A., Chanut, B., Jospin, F., Lucht, F., Verrier, B., Moog, C., Genin, C., and Paul, S. (2013) Generation of HIV-1 potent and broad neutralizing antibodies by immunization with postfusion HR1/HR2 complex. *AIDS* **27**, 717-730
35. Sackett, K., Nethercott, M. J., Shai, Y., and Weliky, D. P. (2009) Hairpin Folding of HIV gp41 Abrogates Lipid Mixing Function at Physiologic pH and Inhibits Lipid Mixing by Exposed gp41 Constructs. *Biochemistry* **48**, 2714-2722
36. Vogel, E. P., and Weliky, D. P. (2013) Quantitation of recombinant protein in whole cells and cell extracts via solid-state NMR spectroscopy. *Biochemistry* **52**, 4285-4287
37. Lev, N., Fridmann-Sirkis, Y., Blank, L., Bitler, A., Epand, R. F., Epand, R. M., and Shai, Y. (2009) Conformational stability and membrane interaction of the full-length ectodomain of HIV-1 gp41: Implication for mode of action. *Biochemistry* **48**, 3166-3175
38. Gao, G. F., Wiczorek, L., Peachman, K. K., Polonis, V. R., Alving, C. R., Rao, M., and Rao, V. B. (2013) Designing a soluble near full-length HIV-1 gp41 trimer. *J. Biol. Chem.* **288**, 234-246
39. Shi, W. X., Bohon, J., Han, D. P., Habte, H., Qin, Y. L., Cho, M. W., and Chance, M. R. (2010) Structural characterization of HIV gp41 with the membrane-proximal external region. *J. Biol. Chem.* **285**, 24290-24298
40. Brugger, B., Glass, B., Haberkant, P., Leibrecht, I., Wieland, F. T., and Krasslich, H. G. (2006) The HIV lipidome: A raft with an unusual composition. *Proc. Natl. Acad. Sci. U.S.A.* **103**, 2641-2646
41. Sackett, K., TerBush, A., and Weliky, D. P. (2011) HIV gp41 six-helix bundle constructs induce rapid vesicle fusion at pH 3.5 and little fusion at pH 7.0: understanding pH dependence of protein aggregation, membrane binding, and electrostatics, and implications for HIV-host cell fusion. *Eur. Biophys. J.* **40**, 489-502
42. Ratnayake, P., Sackett, K., Nethercott, M. J., and Weliky, D. P. pH-dependent vesicle fusion induced by the ectodomain of the Human Immunodeficiency Virus membrane fusion protein gp41: Two kinetically distinct processes and fully-membrane-associated gp41 with predominant beta sheet fusion peptide conformation. *Biochim. Biophys. Acta*, , submitted
43. Suarez, T., Gallaher, W. R., Agirre, A., Goni, F. M., and Nieva, J. L. (2000) Membrane interface-interacting sequences within the ectodomain of the human immunodeficiency

- virus type 1 envelope glycoprotein: putative role during viral fusion. *J. Virol.* **74**, 8038-8047.
44. Kim, C. S., Epand, R. F., Leikina, E., Epand, R. M., and Chernomordik, L. V. (2011) The final conformation of the complete ectodomain of the HA2 subunit of Influenza Hemagglutinin can by itself drive low pH-dependent fusion. *J. Biol. Chem.* **286**, 13226-13234
 45. Shu, W., Ji, H., and Lu, M. (2000) Interactions between HIV-1 gp41 core and detergents and their implications for membrane fusion. *J. Biol. Chem.* **275**, 1839-1845
 46. Lin, C. H., Lin, C. H., Chang, C. C., Wei, T. S., Cheng, S. F., Chen, S. S. L., and Chang, D. K. (2011) An efficient production and characterization of HIV-1 gp41 ectodomain with fusion peptide in Escherichia coli system. *J. Biotechnol.* **153**, 48-55
 47. Shu, W., Liu, J., Ji, H., Radigen, L., Jiang, S. B., and Lu, M. (2000) Helical interactions in the HIV-1 gp41 core reveal structural basis for the inhibitory activity of gp41 peptides. *Biochemistry* **39**, 1634-1642
 48. Caffrey, M., Braddock, D. T., Louis, J. M., Abu-Asab, M. A., Kingma, D., Liotta, L., Tsokos, M., Tresser, N., Pannell, L. K., Watts, N., Steven, A. C., Simon, M. N., Stahl, S. J., Wingfield, P. T., and Clore, G. M. (2000) Biophysical characterization of gp41 aggregates suggests a model for the molecular mechanism of HIV-associated neurological damage and dementia. *J. Biol. Chem.* **275**, 19877-19882
 49. Roche, J., Louis, J. M., Grishaev, A., Ying, J. F., and Bax, A. (2014) Dissociation of the trimeric gp41 ectodomain at the lipid-water interface suggests an active role in HIV-1 Env-mediated membrane fusion. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 3425-3430
 50. Caffrey, M., Kaufman, J., Stahl, S., Wingfield, P., Gronenborn, A. M., and Clore, G. M. (1999) Monomer-trimer equilibrium of the ectodomain of SIV gp41: Insight into the mechanism of peptide inhibition of HIV infection. *Protein Sci.* **8**, 1904-1907
 51. Lakomek, N. A., Kaufman, J. D., Stahl, S. J., Louis, J. M., Grishaev, A., Wingfield, P. T., and Bax, A. (2013) Internal dynamics of the homotrimeric HIV-1 viral coat protein gp41 on multiple time scales. *Angew. Chem. Int. Ed.* **52**, 3911-3915
 52. Tcherkasskaya, O., Knutson, J. R., Bowley, S. A., Frank, M. K., and Gronenborn, A. M. (2000) Nanosecond dynamics of the single tryptophan reveals multi-state equilibrium unfolding of protein GB1. *Biochemistry* **39**, 11216-11226
 53. Lim, W. K., Rosgen, J., and Englander, S. W. (2009) Urea, but not guanidinium, destabilizes proteins by forming hydrogen bonds to the peptide group. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 2595-2600

54. Cao, J., Bergeron, L., Helseth, E., Thali, M., Repke, H., and Sodroski, J. (1993) Effects of amino-acid changes in the extracellular domain of the Human Immunodeficiency Virus Type-1 gp41 envelope glycoprotein. *J. Virol.* **67**, 2747-2755
55. Peisajovich, S. G., and Shai, Y. (2001) SIV gp41 binds to membranes both in the monomeric and trimeric states: Consequences for the neuropathology and inhibition of HIV infection. *J. Mol. Biol.* **311**, 249-254
56. Freed, E. O., Delwart, E. L., Buchschacher, G. L., Jr., and Panganiban, A. T. (1992) A mutation in the human immunodeficiency virus type 1 transmembrane glycoprotein gp41 dominantly interferes with fusion and infectivity. *Proc. Natl. Acad. Sci. U.S.A.* **89**, 70-74
57. Kemble, G. W., Danieli, T., and White, J. M. (1994) Lipid-anchored influenza hemagglutinin promotes hemifusion, not complete fusion. *Cell* **76**, 383-391
58. Miyauchi, K., Curran, A. R., Long, Y. F., Kondo, N., Iwamoto, A., Engelman, D. M., and Matsuda, Z. (2010) The membrane-spanning domain of gp41 plays a critical role in intracellular trafficking of the HIV envelope protein. *Retrovirology* **7**, 95
59. Wilk, T., Pfeiffer, T., Bukovsky, A., Moldenhauer, G., and Bosch, V. (1996) Glycoprotein incorporation and HIV-1 infectivity despite exchange of the gp160 membrane-spanning domain. *Virology* **218**, 269-274
60. Miyauchi, K., Curran, R., Matthews, E., Komano, J., Hoshino, T., Engelman, D. M., and Matsuda, Z. (2006) Mutations of conserved glycine residues within the membrane-spanning domain of Human Immunodeficiency Virus type 1 gp41 can inhibit membrane fusion and incorporation of Env onto virions. *Jpn. J. Infect. Dis.* **59**, 77-84
61. Reuven, E. M., Dadon, Y., Viard, M., Manukovsky, N., Blumenthal, R., and Shai, Y. (2012) HIV-1 gp41 Transmembrane Domain Interacts with the Fusion Peptide: Implication in Lipid Mixing and Inhibition of Virus-Cell Fusion. *Biochemistry* **51**, 2867-2878

CHAPTER 4

Construction of HIV-1 Gp41 Pre-Hairpin Intermediate

4.1. Specific Aims

The long-term goals of the proposed research were to elucidate the structure of the intermediate state (Pre-Hairpin Intermediate) of envelope fusion protein gp41 and development of the structure-function model for the whole ectodomain region of the viral fusion protein. This study would help us to understand the mechanism of fusion of viral and host cell membranes and put a step forward towards the development of novel antiviral therapeutics targeting the fusion process.

The specific aims of this project include:

- (1) Mutation of Hairpin region to mimic Pre-Hairpin Intermediate (PHI) state.
- (2) Determination of the secondary structure of the wild type and mutated protein constructs by circular dichroism.

4.2. Introduction

During the fusion process of the viral and host cell membrane, it is proposed that the viral protein gp41 undergoes three states sequentially: (a) Pre-fusion native state, (b) Pre-Hairpin Intermediate state (PHI), and (c) Final state (Hairpin) (1) (Figure 4.1).

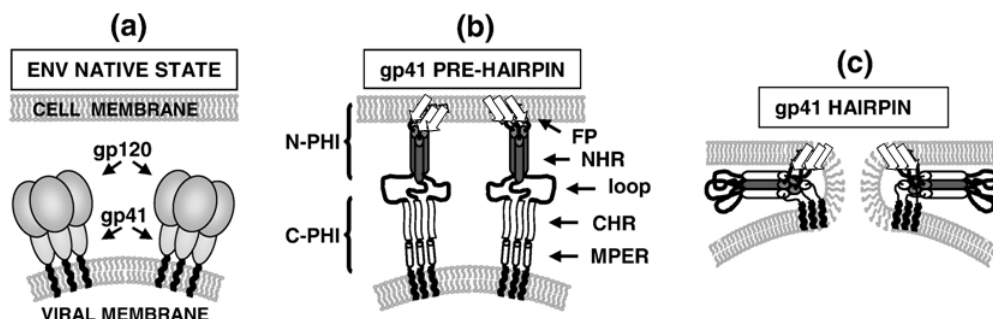


Figure 4.1. HIV fusion model. (a)Native state of envelope protein before fusion, (b) Pre-Hairpin intermediate of gp41, and (c) Final Hairpin state of gp41 (2).

Previously, biophysical studies were done to understand which conformation is involved in the membrane fusion. Lipid mixing assays (assays to measure mixing of membrane lipids) of large unilamellar vesicles was performed to assess the extent of vesicle fusion induced by FP, FP-NHR, FP-Hairpin (FP-HP), and Hairpin. The experimental conditions were low pH 3.0 and physiologic pH 7.4. The vesicles are negatively charged at both pH 3.0 and 7.4. The protein constructs are positively charged at both the pHs, however, the magnitude of charge is greater at pH 3.0 than compared to pH 7.4. It was shown that the FP-HP does not induce significant vesicle fusion, whereas, FP-NHR construct which is a model of N-terminal PHI induces rapid and extensive vesicle fusion. It has also been shown that the FP-HP and Hairpin construct inhibits vesicle fusion when induced by FP-NHR construct (1).

X-ray crystallography studies have been performed to solve the structure of the Hairpin and Hairpin-MPER in its probable final gp41 ectodomain structure during fusion (1,3-5). The trimer of Hairpin structure assembles into a six-helix bundle (SHB). There are recent $\sim 5\text{\AA}$ structures of gp140 trimers that possibly represent the protein state prior to membrane fusion (6-8). The

trimer gp140 is stabilized by mutations and ectodomain cross-linking between gp120-gp41 (9,10). The structure shows loose bundle of three parallel NHR helices and three CHR helices forming a tripod with an angle of 70° between the helices. In this gp140 structure, the FP and MPER are not included. The melting temperature (T_m) of only the NHR-CHR shorter ectodomain construct is ~70 °C whereas the T_m of longer ectodomain constructs is ~110 °C (2,11). This thermal stability leads to propose SHB as the final stable structure of gp41 ectodomain during fusion. PHI state of gp41 ectodomain has been proposed as fully extended (NHR helix/0° -turn/CHR helix) with separate NHR and CHR trimer helical bundles rather than hairpin conformation. To our knowledge, there is only functional evidence for the PHI state of gp41. The presence of this intermediate is supported by inhibition of membrane fusion and HIV infection with NHR and CHR+MPER peptides (12,13). A proposed mechanism of inhibition is tight binding of the peptide to the PHI structure with consequent prevention of formation of the SHB structure.

4.3. Approach

4.3.1. Mutagenesis of CHR

Crystal structure of NHR-CHR peptides shows a trimeric assembly where NHR and CHR region of one monomer are oriented in antiparallel manner. This six-helix bundle formed by the NHR and CHR regions from three monomers is stabilized mainly due to hydrophobic interactions (3). The trimeric NHR assembly is formed by hydrophobic interaction in the “a” and “d” positions as shown in figure 4.2. The residues at “a” and “d” positions of CHR interact with the “e” and “g” residues of the neighboring NHRs (Figure 4.2). W628, W631 of the CHR interacts with W571 of

the NHR. Similarly, I635 of CHR has hydrophobic contact with L565 and L568 of NHR. D632 of CHR forms a conserved salt bridge with K574 of NHR.

Broadly neutralizing antibodies are capable of blocking different strains of virus from infecting target cells. 2F5 and 4E10 are the two broadly neutralizing antibodies that bind to the MPER only in the trimer state of gp41 with MPER fully exposed (14). Another structural study of binding of broadly neutralizing antibodies to MPER show helical structure of peptide MPER as prerequisite for antibody binding (15,16).

It was hypothesized that the PHI state could be obtained by disrupting the interactions between the CHR and NHR. Mutation of W628A and W631R in the CHR might result in the disruption of the hydrophobic interactions. The salt bridge between D632 and K574 can be disrupted by mutating to D632A.

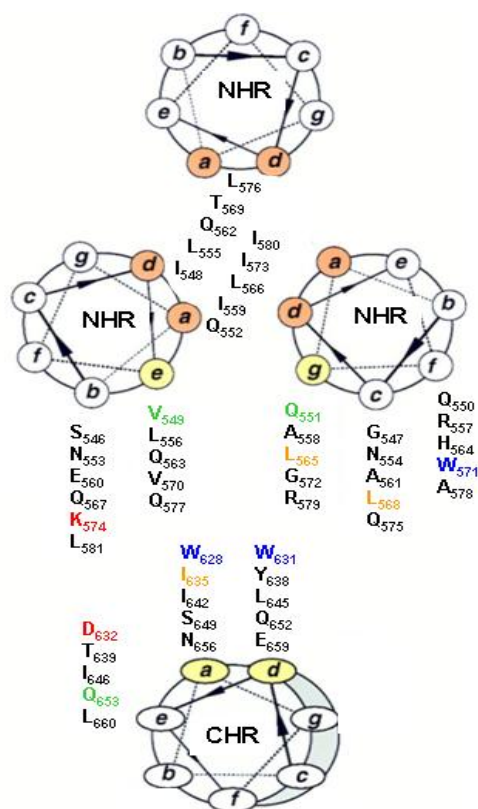


Figure 4.2. Helical wheel representation of NHR and CHR. Three NHR helices and one CHR helix are represented as helical wheel projections. It is an end-on-view of the NHR-CHR complex looking down the axis of the trimer. The NHR helices interact with other NHR helices at the “a” and “d” positions (colored orange). Positions of the NHR and the CHR helices that occupy the interhelical space between two NHR helices and one CHR helix are shaded yellow (3). The residues which interact among themselves are colored accordingly

4.3.2. Verification of Formation of Pre-Hairpin Intermediate

The Hairpin and the FP-HP are highly thermostable with melting temperature at 112°C (2). Circular Dichroism (CD) spectra of Hairpin (HP) in aqueous solution show 100% helicity for the peptide (Figure 4.3). CD is a very useful spectroscopic tool to verify global secondary structure

of large biological molecules. In this technique the difference in absorption of left and right circularly polarized light by chiral molecules is recorded and plotted against the wavelength. Percent helicity can be calculated from the mean residue molar ellipticity (θ_{222}) at wavelength 222nm (minima). θ_{222} of $-33,000 \text{ deg cm}^2 \text{ dmol}^{-1}$ at wavelength 222nm is considered to be 100% (17). CD signal at 222 nm can be measured as a function of temperature and thermal melts can be performed at 2 °C steps between 20 °C – 100 °C.

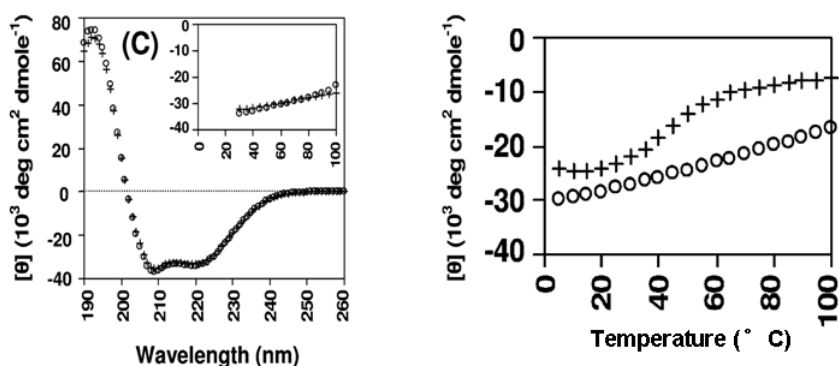


Figure 4.3. (Left) Far-UV CD analysis of Hairpin showing 100% helicity with the mean residue molar ellipticity at 222nm (θ_{222}) is $-33,000 \text{ deg cm}^2 \text{ dmol}^{-1}$. For inset, abscissa is the temperature scale in °C, and the ordinate is θ_{222} , showing the thermal stability of Hairpin with temperature. (Right) Thermal stability of Hairpin (○) and N70 (+). The melting temperature of N70 is 45 °C (1).

Our goal was to have the PHI state such that the NHR region is undisturbed, forming the trimer, and the CHR is disturbed by mutation and unable to interact with NHR. The mean residue molar ellipticity of Hairpin at 222nm ($\theta_{222\text{nm}}$) is $-33,000 \text{ deg cm}^2 \text{ dmol}^{-1}$. $\theta_{222\text{nm}}$ for an unfolded protein is $-5000 \text{ deg cm}^2 \text{ dmol}^{-1}$ (1). There are 47 residues in the NHR, 6 residues in the loop region, and 39 residues in the CHR. Therefore, it is expected that the PHI should have $\theta_{222\text{nm}}$ value as $\sim -$

19,000 deg cm² dmol⁻¹. The melting temperature (T_m) for N70 (FP-NHR) is 45 °C (1). Therefore, the proposed PHI should also have T_m at ~ 45 °C.

4.4. Materials and Method

4.4.1. Construction of Mutant HP and HM

The mutant HP and HM constructs were constructed by site directed mutagenesis. The mutations done were as follows – (a) D632A, (b) W628A, and (c) W631R. All the three mutations were done separately on HP and HM. Mutants M1 comprised of mutation D632A, M2 comprised of mutations D632A and W628A, and M3 comprised of D632A, W628A, and W631R.

Primers were designed based on the required mutations. The template plasmids were extracted from overnight grown cells. PCR was conducted using HP or HM template DNA and designed primers. The DNA sequence after mutation is shown in Figure 4.4.

HP M3

CTLTVQARQL LSGIVQQQNN LLRAIEAQQH LLQLTVWGIK QLQARILSGG
RGGAMERARE INNYTSLIHS LIEESQNQQE KNEQELLELD KWHHHHHH

HP WT DNA Sequence

tgc acg ctg acg gta cag gcc aga caa tta ttg tct ggt ata gtg cag cag cag aac aat ttg ctg agg gct
att gag gcg caa cag cat ctg ttg caa ctc aca gtc tgg ggc atc aag cag ctc cag gca aga atc ctg tct
ggt ggc cgt ggc ggt tgg atg gag tgg gac aga gaa att aac aat tac aca agc tta ata cac tcc tta att
gaa gaa tcg caa aac cag caa gaa aag aat gaa caa gaa tta ttg gaa tta gat aaa tgg cat cac cat cac
cat cac tga

HPM3 DNA Sequence

tgc acg ctg acg gta cag gcc aga caa tta ttg tct ggt ata gtg cag cag cag aac aat ttg ctg agg gct
att gag gcg caa cag cat ctg ttg caa ctc aca gtc tgg ggc atc aag cag ctc cag gca aga atc ctg tct
ggt ggc cgt ggc ggt **gcg** atg gag **cg****g** **gcc** aga gaa att aac aat tac aca agc tta ata cac tcc tta att
gaa gaa tcg caa aac cag caa gaa aag aat gaa caa gaa tta ttg gaa tta gat aaa tgg cat cac cat cac
cat cac tga

Figure 4.4. Amino acid sequence of HP protein construct, DNA sequence of HP WT and HP M3.

4.4.2. Experimental Procedure

The mutated Hairpin (D632A) with His-tag was expressed in BL21 (DE3) competent cells. When proteins of one biological system are expressed in other biological (e.g., bacterial)

system, within the cell they sometimes exist as insoluble aggregates, known as inclusion bodies. Purification of the Hairpin-His (D632A) protein was tried by sonicating the cells and pelleting down the inclusion bodies along with cell debris. The pellet containing the inclusion bodies as well as cell debris was sonicated again in the presence of (a) 8M urea or (b) 6M GuHCl and centrifuged. Both urea and GuHCl could efficiently solubilize HP mutant containing inclusion body. The supernatant containing the solubilized protein was collected. Purification was done by metal affinity chromatography. All the steps in purification were done at pH 7.4 and at room temperature.

Purification of HM mutants was done in similar process as HP mutants, however, only GuHCl solubilized the protein containing inclusion bodies most efficiently. After the purification, the required proteins were dialyzed against sodium formate buffer containing 0.2 mM TCEP and at pH 3.2. All the mutants' i. e. M1, M2 and M3 of both HP and HM were soluble at low pH 3.2.

The secondary structure of the soluble protein constructs were measured by performing circular dichroism (CD). The CD experiments were done at pH 3.2 and 20 °C. The thermal melting of the protein constructs were measured within the temperature range of 20 – 96 °C. All the observations were subtracted by the reading observed by buffer only.

4.5. Results

In the gp41 HP or HM construct, D632 in the CHR forms a salt bridge with K574 in the NHR (Figure 4.5). The mutation of D632A breaks the salt bridge. There is a hydrophobic interaction between W571 in NHR and W628 and W631 in CHR (Figure 4.5). Mutation of W628A and

W631R breaks the hydrophobic interaction. The insertion of charged residue R631 further disturbs the hydrophobic environment.

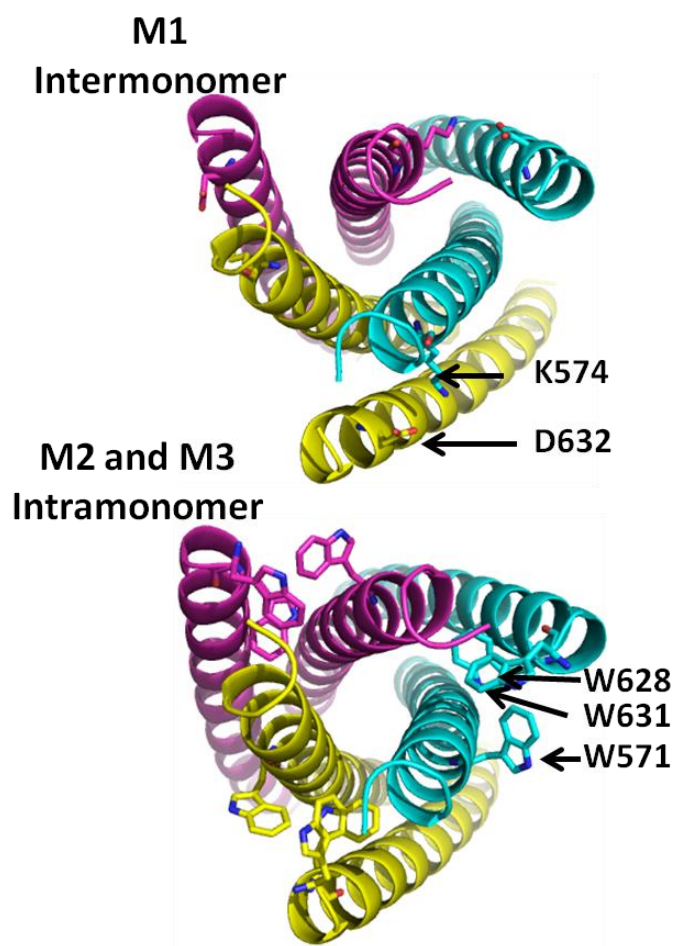


Figure 4.5. The SHB trimer of gp41 comprising of inner core NHR and outer CHR. Each monomer is colored differently. *Top*, residues shown forms intermonomer salt bridge, and *bottom*, shows residues forming hydrophobic interaction between intramonomer strands.

The secondary structure of wild type (WT) HP or HM (Figures 4.6 and 4.8 respectively) showed ~100% helicity. The secondary structure of HP M1 shows similar helicity as HPWT (Figure 4.6).

However, the helicity is reduced by ~50% i.e. θ_{222} is $\sim -20000 \text{ deg cm}^2 \text{ dmol}^{-1}$ with the second mutation (W628A) (Figure 4.7). In case of HP M3 or HM M3, the percentage helicity remains almost same as HP M2 (figure 7, 8, and 6 respectively). To understand the thermal stability of all the three mutants, thermal melt experiments were performed. It was observed that the melting temperature did not change much compared to the WT protein constructs (Figure 4.9). After the unfolding of the protein by thermal melt experiments, the protein sample was allowed to refold at ambient temperature for 2 hours. The extent of refolding was measured by measuring the mean residue molar ellipticity at 20 °C. It was observed that HM WT and HM M3 both refolded after thermal denaturation.

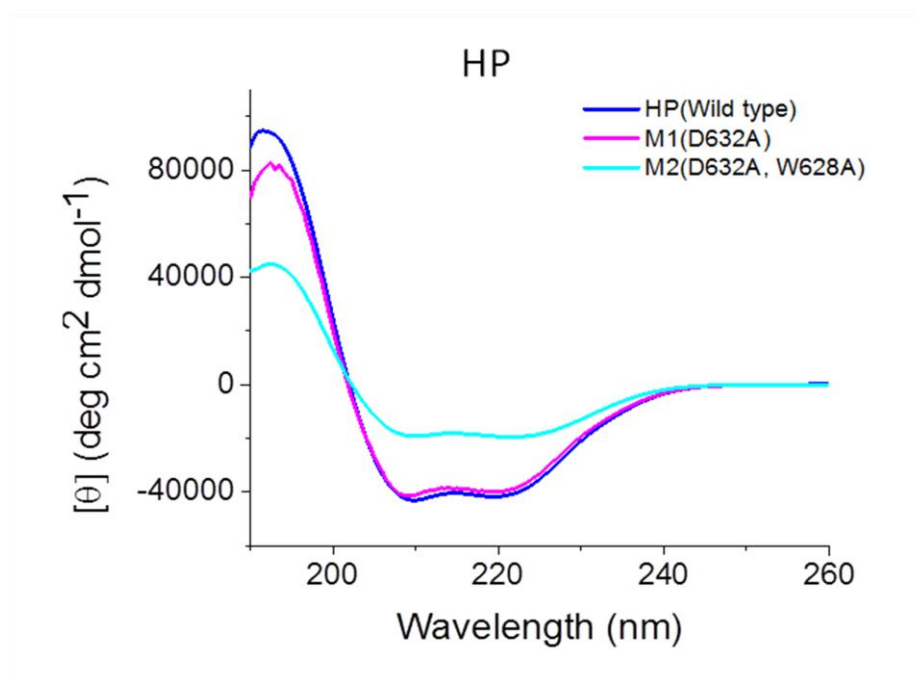


Figure 4.6. CD spectra of WT and mutant HP at pH 3.2. For a SHB trimer, the D632A mutation disrupts the inter-monomer salt bridge with K574 and the W628A mutation disrupts the intra-monomer hydrophobic interaction with W571. The [protein] $\approx 15 \mu\text{M}$ ($\sim 0.2 \text{ mg/mL}$), the pH 3.2 buffers contained 50 mM formate with 0.2 mM TCEP.

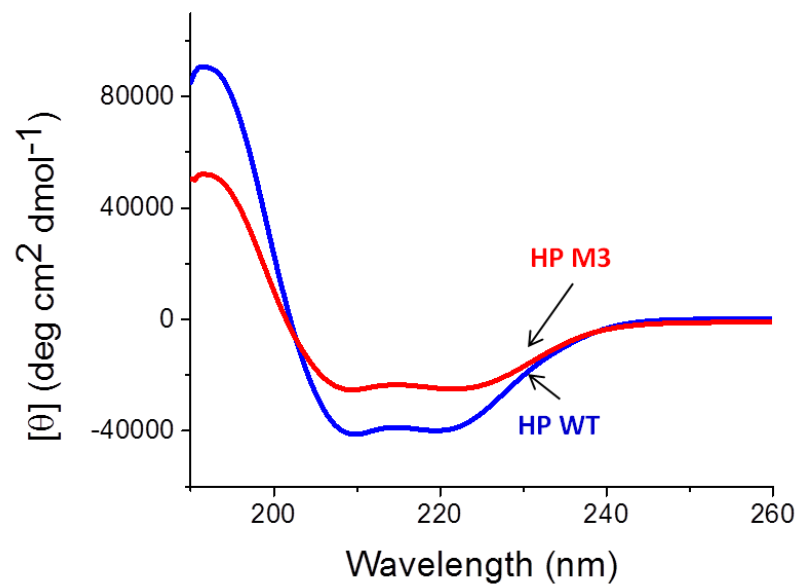


Figure 4.7. CD spectra of HP WT and HP M3 at pH 3.2. W631R mutation disrupt the intra-monomer hydrophobic interaction with W571. The [protein] $\approx 10 \mu\text{M}$ ($\sim 0.2 \text{ mg/mL}$), the pH 3.2 buffers contained 50 mM formate with 0.2 mM TCEP.

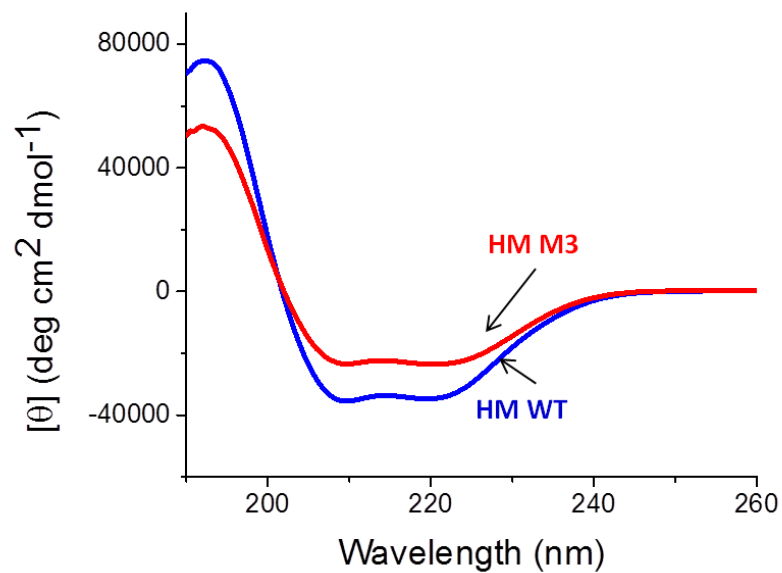


Figure 4.8. CD spectra of HM WT and HM M3 at pH 3.2. For a SHB trimer, W631R mutation along with W628A disrupts the intra-monomer hydrophobic interaction with W571. The [protein] $\approx 10 \mu\text{M}$ ($\sim 0.2 \text{ mg/mL}$), the pH 3.2 buffers contained 50 mM formate with 0.2 mM TCEP.

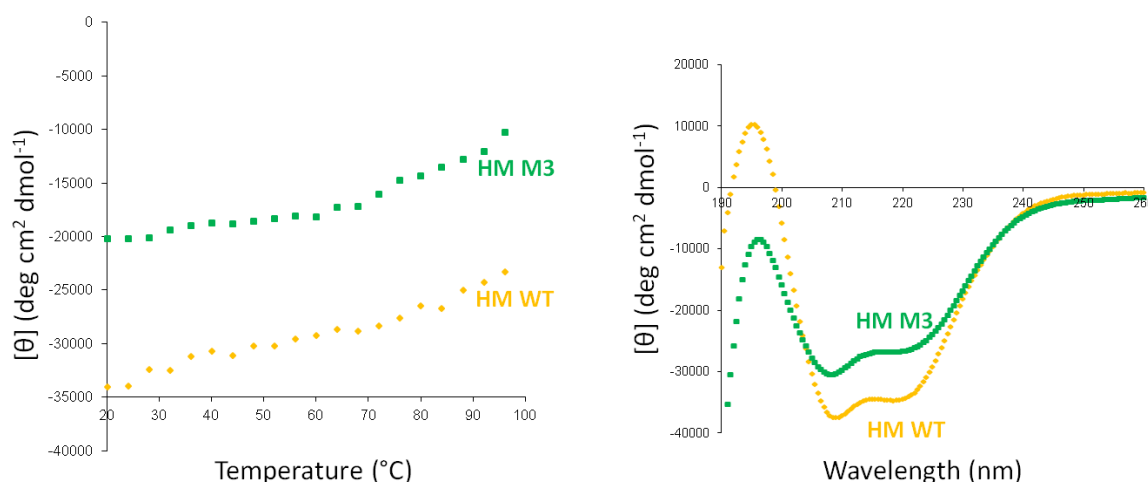


Figure 4.9. Estimation of thermal stability of mutant HM. *Left*, measurement of thermal unfolding of HM WT and HM M3 with mean residue molar ellipticity as a function of temperature at 222nm and pH 3.2, and *right*, CD spectra of thermally denatured protein after refolding.

4.6. Discussion

Based on the fusion model, it is proposed that the gp41 exists as trimer in the PHI state and forms a six-helix bundle structure at the end of the fusion process. The crystal structure of gp41 comprising of the NHR and CHR shows trimer structure (5). However, the mutations we performed to disrupt the NHR-CHR interactions revealed that gp41 NHR-short loop-CHR exists as folded monomer structure. The mutation M1 (D632A) disrupts the inter-monomer salt-bridge with K574 and mutations M2 (W628A and D632A) disrupts the intra-monomer hydrophobic interaction with W571. The D632A (M1) did not change the secondary structure of HP whereas there was decrease in α -helicity in the mutant M2 (Figure 4.6). This indicates that there is a change in secondary structure by disruption of intra-monomer interaction whereas no change by breaking of salt bridge between D632 in CHR and K574 in NHR of two individual

monomers. To add to this, we also verified the T_m of HM M3. It was observed that there was no change in the melting point of M3 as compared with wild type HM. Even though we had achieved the expected θ_{222} value $\sim -20000 \text{ deg cm}^2 \text{ dmol}^{-1}$, we were unable to reach the desired T_m value of the protein constructs. Unfolding of HP or HM (Wild type) to M3 does not contribute to the lowering of melting point.

4.7. Future Work

Our goal was to attain the gp41 ectodomain in the PHI state which would possess low helicity and low T_m . We propose that mutations of few more residues in the CHR would disturb the NHR-CHR interaction. Q653 of CHR has intramolecular hydrogen bond with Q551 of the NHR within the gp41 monomer. Q653 also forms an intermolecular hydrogen bond with the backbone carbonyl oxygen of V549 of NHR (3). The hydrophobic interactions between (a) H564 and Y638 and (b) L565 and I635 are formed between intra-monomer NHR-CHR. These intra-monomer NHR-CHR interactions along with previously mentioned hydrophobic interaction (figure 5 *bottom*) are near the loop region. We propose mutation of Q653A, I635K, and Y638K would disrupt the interaction between NHR and CHR.

REFERENCES

REFERENCES

1. Sackett, K., Nethercott, M. J., Shai, Y., and Weliky, D. P. (2009) Hairpin Folding of HIV gp41 Abrogates Lipid Mixing Function at Physiologic pH and Inhibits Lipid Mixing by Exposed gp41 Constructs†. *Biochemistry* **48**, 2714–2722
2. Sackett, K., Nethercott, M. J., Epand, R. F., Epand, R. M., Kindra, D. R., Shai, Y., and Weliky, D. P. (2010) Comparative Analysis of Membrane-Associated Fusion Peptide Secondary Structure and Lipid Mixing Function of HIV gp41 Constructs that Model the Early Pre-Hairpin Intermediate and Final Hairpin Conformations. *Journal of Molecular Biology* **397**, 301–315
3. Chan, D. C., Fass, D., Berger, J. M., and Kim, P. S. (1997) Core structure of gp41 from the HIV envelope glycoprotein. *Cell* **89**, 263-273
4. Weissenhorn, W., Dessen, A., Harrison, S. C., Skehel, J. J., and Wiley, D. C. (1997) Atomic structure of the ectodomain from HIV-1 gp41. *Nature* **387**, 426-430
5. Shi, W., Bohon, J., Han, D. P., Habte, H., Qin, Y., Cho, M. W., and Chance, M. R. (2010) Structural Characterization of HIV gp41 with the Membrane-proximal External Region*. *Journal of Biological Chemistry* **285**, 24290–24298
6. Bartesaghi, A., Merk, A., Borgnia, M. J., Milne, J. L. S., and Subramaniam, S. (2013) Prefusion structure of trimeric HIV-1 envelope glycoprotein determined by cryo-electron microscopy. *Nat Struct Mol Biol* **20**, 1352-1357
7. Julien, J. P., Cupo, A., Sok, D., Stanfield, R. L., Lyumkis, D., Deller, M. C., Klasse, P. J., Burton, D. R., Sanders, R. W., Moore, J. P., Ward, A. B., and Wilson, I. A. (2013) Crystal Structure of a Soluble Cleaved HIV-1 Envelope Trimer. *Science* **342**, 1477-1483
8. Lyumkis, D., Julien, J. P., de Val, N., Cupo, A., Potter, C. S., Klasse, P. J., Burton, D. R., Sanders, R. W., Moore, J. P., Carragher, B., Wilson, I. A., and Ward, A. B. (2013) Cryo-EM Structure of a Fully Glycosylated Soluble Cleaved HIV-1 Envelope Trimer. *Science* **342**, 1484-1490
9. Schulke, N., Vesanen, M. S., Sanders, R. W., Zhu, P., Lu, M., Anselma, D. J., Villa, A. R., Parren, P. W. H. I., Binley, J. M., Roux, K. H., Maddon, P. J., Moore, J. P., and Olson, W. C. (2002) Oligomeric and conformational properties of a proteolytically mature, disulfide-stabilized human immunodeficiency virus type 1 gp140 envelope glycoprotein. *Journal of Virology* **76**, 7760-7776
10. Sanders, R. W., Vesanen, M., Schuelke, N., Master, A., Schiffner, L., Kalyanaraman, R., Paluch, M., Berkhout, B., Maddon, P. J., Olson, W. C., Lu, M., and Moore, J. P. (2002) Stabilization of the soluble, cleaved, trimeric form of the envelope glycoprotein complex of human immunodeficiency virus type 1. *Journal of Virology* **76**, 8875-8889
11. Lu, M., Ji, H., and Shen, S. (1999) Subdomain folding and biological activity of the core structure from human immunodeficiency virus type 1 gp41: Implications for viral membrane fusion. *Journal of Virology* **73**, 4433-4438

12. Wild, C. T., Shugars, D. C., Greenwell, T. K., Mcdanal, C. B., and Matthews, T. J. (1994) Peptides Corresponding to a Predictive Alpha-Helical Domain of Human-Immunodeficiency-Virus Type-1 Gp41 Are Potent Inhibitors of Virus-Infection. *P Natl Acad Sci USA* **91**, 9770-9774
13. Markosyan, R. M., Cohen, F. S., and Melikyan, G. B. (2003) HIV-1 Envelope Proteins Complete Their Folding into Six-helix Bundles Immediately after Fusion Pore Formation. *Molecular Biology of the Cell* **14**, 926–938
14. Frey, G., Peng, H., Rits-Volloch, S., Morelli, M., Cheng, Y., and Chen, B. (2008) A fusion-intermediate state of HIV-1 gp41 targeted by broadly neutralizing antibodies. *PNAS* **105**, 3739–3744
15. Julien, J. P., Bryson, S., Nieva, J. L., and Pai, E. F. (2008) Structural Details of HIV-1 Recognition by the Broadly Neutralizing Monoclonal Antibody 2F5: Epitope Conformation, Antigen-Recognition Loop Mobility, and Anion-Binding Site. *Journal of Molecular Biology* **384**, 377-392
16. Cardoso, R. M. F., Zwick, M. B., Stanfield, R. L., Kunert, R., Binley, J. M., Katinger, H., Burton, D. R., and Wilson, I. A. (2005) Broadly Neutralizing Anti-HIV Antibody 4E10 Recognizes a Helical Conformation of a Highly Conserved Fusion-Associated Motif in gp41. *Immunity* **22**, 163–173
17. Greenfield, N., and Fasman, G. D. (1969) Computed Circular Dichroism Spectra for Evaluation of Protein Conformation. *Biochemistry* **8**, 4108-4116

CHAPTER 5

Future Work

Previous solid state NMR study of longer ectodomain (FP-HP) of gp41 by our group showed antiparallel beta sheet arrangement of FP domain located in the membrane (1). Our study of gp41 ectodomain reveals existence of the ectodomain in monomeric and hexameric state (explained in Chapter 3). We have also studied the functional significance of the oligomeric gp41 ectodomains. Based on the mutational studies and thermal stability of gp41 ectodomain, it has been proposed that monomer gp41 ectodomain forms a hairpin conformation in aqueous solution at low pH. It is not known whether gp41 ectodomain forms monomer or higher oligomer in membrane environment.

Therefore, in future studies could be done to determine the oligomeric state of whole gp41 ectodomain construct incorporated in lipid membrane (vesicles). A qualitative experiment could be done by 20% ^{13}C labeled and uniformly ^2H labeled protein constructs incorporated in membrane and measuring the dephasing between ^{13}C - ^2H by REDOR solid state NMR experiment. It is expected to observe dephasing if there is any oligomerization and no dephasing for monomeric state. However, by this experiment we could only differentiate between monomer and any oligomer formation. We would not be able to identify between types of oligomers formed by the gp41 ectodomain protein construct incorporated in membrane. After, the determination of the probable oligomeric state of the ectodomain protein construct, solid state NMR studies could be done to investigate the conformation of monomer or higher oligomer structure. This can be done by studying the proximity of FP and MPER in membrane environment. Labeling of ^{13}C F residues in FP and ^{15}N F in MPER can be done. Specific labeling of FP can be done by using labeled amino acid during peptide synthesis. The HM can be labeled by labeling bacterially expressed protein. There is only single F residue

in the whole HM amino acid sequence. Then the whole ectodomain can be constructed by native chemical ligation of labeled FP and HM. The protein construct can then be incorporated in vesicles. REDOR experiments can be done to determine the dephasing between the ^{13}C - ^{15}N F residues due to proximity of the FP and MPER. ^{13}C - ^{15}N REDOR the distance detection limit is approximately $\sim 6\text{-}9\text{ \AA}$. It is expected that the distance between the NMR active nuclei would lie between $7\text{-}9\text{ \AA}$ if the FP and MPER are in close apposition with each other.

However, the FP and MPER might not be close enough to show adequate dephasing. Then the next study that could be done is determination of the membrane insertion depth of MPER included in the longer ectodomain construct. At first a qualitative estimation of membrane insertion can be done by using ^2H labeled lipid molecules with all acyl chain deuteration (D54) and ^{13}C F in the MPER labeled HM protein construct. Then the membrane insertion depth can be found by determining the distance between the specific nuclei of amino acid residue in MPER and in lipid molecule.

Another study can be done is to determine the oligomeric state of FP-HM and gp41 including TM and cytoplasmic endodomain (Endo). Functional activity by lipid mixing assay and antibody binding experiments can be studied by addition of TM and TM+Endo. Also analysis could be done to determine the difference in fusion activity with the addition of TM and TM+Endo. This could be followed by determination of proximity of TM with FP of the whole gp41 protein by solid state NMR after being incorporated in lipid membrane environment. The distance measurement between the FP and TM would help to understand the conformation of gp41 protein at the final stage of membrane fusion.

REFERENCES

REFERENCES

1. Sackett, K., Nethercott, M. J., Zheng, Z. X., and Weliky, D. P. (2014) Solid-State NMR Spectroscopy of the HIV gp41 Membrane Fusion Protein Supports Intermolecular Antiparallel 13 Sheet Fusion Peptide Structure in the Final Six-Helix Bundle State. *Journal of Molecular Biology* **426**, 1077-1094

APPENDICES

APPENDIX I

Designing of FP-HP by Native Chemical Ligation

Expression and Purification of Non-His Tagged Hairpin

The gp41 Hairpin construct without the His-tag was expressed in *E. coli* cells and purified. The cells were lysed in glacial acetic acid and centrifuged. The supernatant contained the protein in soluble form. The supernatant was then dialyzed (using 3500 Da MWCO dialysis tubing) against trifluoroacetic acid: water (1:2000 v/v) in the presence of 150 μ M dithiothreitol (DTT) to prevent disulfide bond formation. After dialysis the protein solution was concentrated using a 10,000 MWCO concentrator. The concentrated protein was then purified by RP-HPLC by using C18 column. A multistep gradient was used between 100% water / 0.1% TFA and 90% acetonitrile / 10% water / 0.1% TFA (1). The purity of the protein was checked with SDS PAGE (Figure AI-1) and MW was also verified by mass spectrometry. The protein concentration was detected by measuring the absorbance at 280 nm. The total amount of protein obtained was 43 mg per 10 g of cells harvested from 1L of culture.

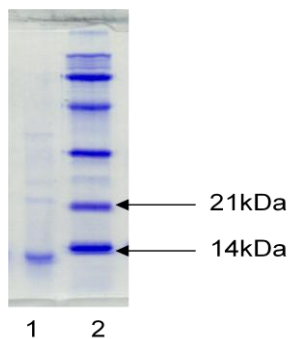


Figure AI-1. SDS-PAGE gel of Hairpin (non His-tag) purified by RP-HPLC. MW of Hairpin (non His-tag) is 10.7 kDa. Lanes: 1- Hairpin; 2- Standard.

The SDS-PAGE gel of the purified Hairpin (Figure AI-1) shows sufficient purity. This protein was used for native chemical ligation with FP.

Native Chemical Ligation

In order to obtain the FP-HP construct, the expressed Hairpin protein (wild type, non His-tag) was ligated with FP. FP had labeled backbone at ^{13}C O of L7 and at ^{15}N amide nitrogen of F8. The labeled FP was supplied by Dr. Kelly Sackett from our group. The labeled FP was dissolved in the ligation buffer (8 M Guanidine chloride, 0.1 M phosphate buffer, and pH 7). The peptide was reacted with 4-mercaptophenyl acetic acid (MPAA) catalyst to modify the thioester linked to the C-terminus of the fusion peptide (2,3) and the mixture was incubated for 30 min. Similarly, HP was dissolved in the ligation buffer including Tris(2-carboxyethyl)-phosphine (TCEP) (2mM) to maintain cysteine residue in reduced state, and the two solutions were mixed together. The reaction was allowed to proceed for 2 days in an inert (Ar) atmosphere. This was followed by purification by RP-HPLC with a C18 column (2). Three peaks comprising of FP23, HP, and FP-HP were collected and MW was verified by mass spectrometry. Acetonitrile was removed from the eluted FP-HP fraction. FP-HP was allowed to refold at 4°C in the refolding buffer (20 mM sodium formate, 200 μM TCEP, and pH-3).

The secondary structure of the whole protein was verified by CD. The signature curve of α -helical structure was observed for FP-HP as shown in the Figure AI-2. The mean residue molar ellipticity was calculated at the minima at 222nm to be 99%.

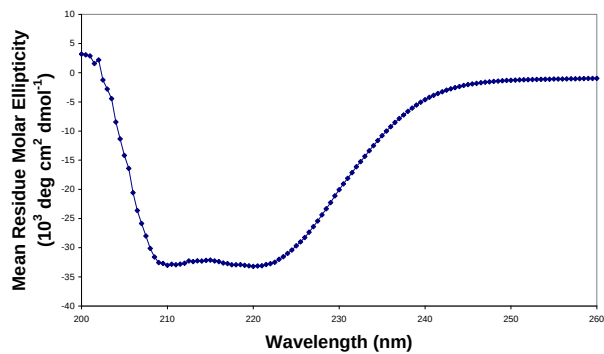


Figure AI-2. CD spectra of the refolded FP-HP.

Purity of FP-HP was verified by SDS-PAGE (Figure AI-3). The total amount of protein obtained was 10 mg i.e., 0.8 μmol from 4 μmol of FP and 4 μmol of Hairpin. This amount of protein was sufficient for preparing a NMR (REDOR) sample. The yield of ligated FP-HP was 20% with respect to the reactant i.e., HP.

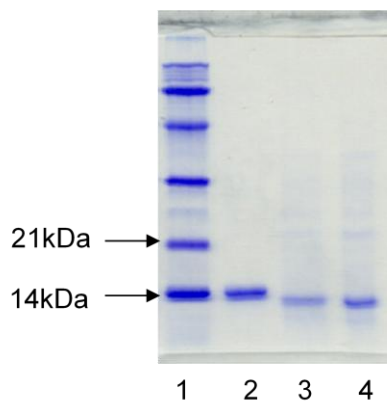


Figure AI-3. SDS-PAGE gel comparing FP-HP and Hairpin. MW of FP-HP is 12.8 kDa, and Hairpin (non His-tag) is 10.7 kDa. Lanes: 1 - Standard, 2 - FP-HP, 3, 4 – Hairpin.

APPENDIX II

Determination of Secondary Structure by Solid State NMR of Membrane Associated FP of Gp41 Ectodomain Construct FP-HP

Protein Reconstitution

The synthesized FP-HP was then incorporated in membrane mimetic environment i.e., in large unilamellar vesicles (LUVs). LUVs were prepared with ether linked lipids 1,2-di-O-tetradecyl-sn-glycero-3-phosphocholine (DTPC) and 1,2-di-O-tetradecyl-sn-glycero-3-phospho-rac-(1-glycerol) sodium salt (DTPG) which lack CO nuclei. Ether linked lipids were used instead of ester linked lipids to avoid observing the carbonyl carbon of the lipids since the chemical shift of carbonyl carbon labeled FP-HP is desired (1,4,5). The molar proportion of lipids used was DTPC: DTPG: Cholesterol at 8:2:5 mol ratios. The molar ratio of Lipid: Protein was 30:1. The LUVs were prepared by dissolving the lipids in chloroform. The mixture was then dried under nitrogen followed by vacuum overnight. Lipid dispersions were prepared by addition of 20mM sodium formate buffer at pH-3 followed by homogenization by alternate cycles of freeze thaws. Extrusion was done through a filter having pores of 100 nm diameter to maintain the size of the LUVs. This was followed by addition of the protein drop wise with continuous stirring. Then it was allowed to stand undisturbed overnight at 4°C. This was followed by centrifugation at 5000rpms for 30min and was repeated to get maximum sample out. The pellet was freeze dried, later hydrated and then packed into a 4mm diameter magic angle spinning NMR rotor.

Solid State NMR

A REDOR of the prepared sample with backbone carbonyl carbon of L7 and amide nitrogen of F8 of FP selectively labeled with ^{13}C and ^{15}N respectively was performed. The labeled FP-HP was reconstituted in a mixture of DTPC: DTPG: Chol (8:2:5) that mimics the lipid composition of host cells. The protein sample was cooled to -50°C to counteract heating with a rotor period of

125 μ s. 4mm magic angle spinning rotor was used and the experiment was set for 80,000 acquisitions. The pulse program used to do the REDOR experiment is explained in section 2.5. FIDs of the S_0 and S_1 signal were acquired. S_0 spectra contain signal of all the ^{13}C without dipolar coupling and S_1 spectra contains signal of all the ^{13}C with dipolar coupling. The difference between the FIDs of S_0 and S_1 gave the chemical shift of CO carbon for isotopically labeled residue L7. The spectra $S_0 - S_1$ are shown in Figure All-1 (right).

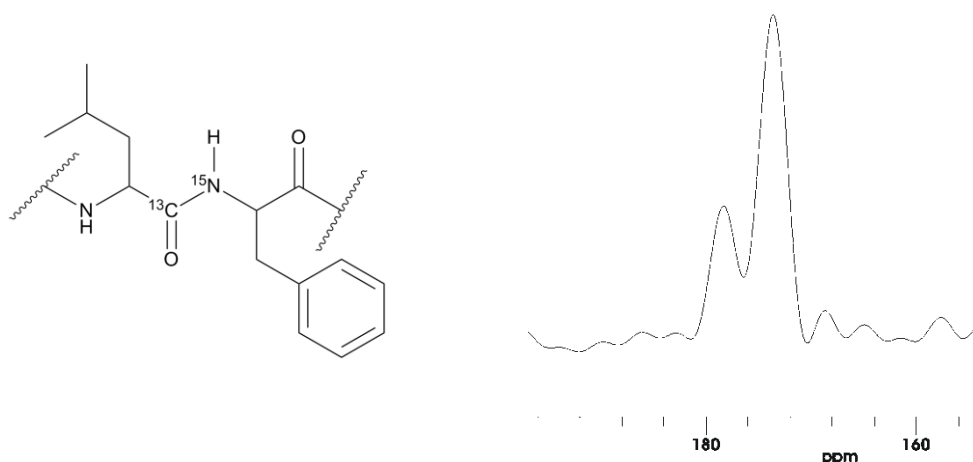


Figure All-1 (Left) The structure of part of the FP showing the positions of the ^{13}C and the ^{15}N labels on Leu (L7) and Phe (F8) respectively. (Right) REDOR subtracted signal for L7F8 in DTPC:DTPG:Chol vesicles. The CO peak of L7 was referenced to the methylene peak of adamantane at 40.5 ppm. The chemical shifts of 178.5 and 173.7 ppm correspond to α -helix and β -sheet respectively.

Conclusion

Peaks at chemical shifts 178.5 and 173.7 represent α -helical and β -sheet structures respectively for Leu residue²⁴. The best-fit deconvoluted peaks were obtained whose integrated areas represent the relative population of β -sheet and α -helical structure. Here it was observed that at pH-3 the structure of FP-HP to be predominantly β -sheet (70%).

APPENDIX III

Study of Binding of Designed HIV Pre-Hairpin Intermediate and Membrane Fusion Inhibitor C34 Peptide

Introduction

The HIV gp41 six helix bundle (SHB) structure consist of inner core formation by three N-terminal helices and the C-terminal helices bind to the conserved grooves on the surface of the NHR trimer core (Figure AIII-1B and C) (6-8). Peptides derived from C-terminal helix region of gp41 also known as C34 are potent HIV membrane fusion inhibitors (9,10). The efficient inhibitory activity of C34 against various strains of HIV is probably due to high conserved sequence of the NHR grooves to which it binds Figure AIII-1D) (6).

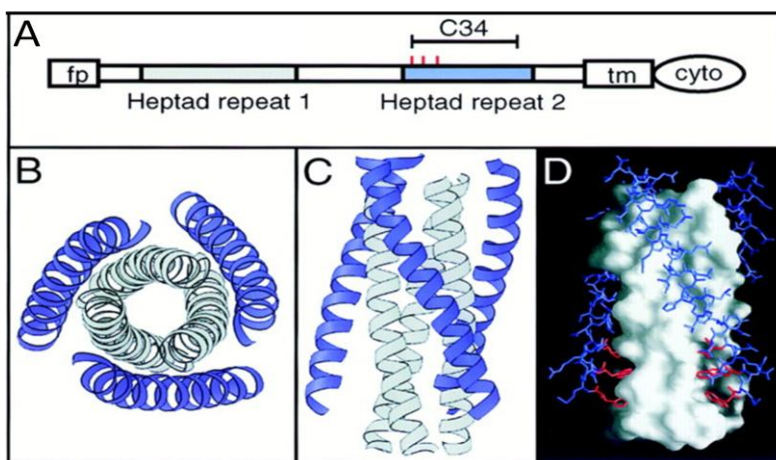


Figure AIII-1. (A) Schematic diagram of gp41 showing different domains – fusion peptide (fp), NHR (Heptad repeat 1), CHR (Heptad repeat 2), transmembrane (tm) and endodomain (cyto). The C34 peptide is shown to compare the sequence length with CHR. The three residues that bind into the groove are shown in red. (B) Down the axis view of SHB structure of gp41 ectodomain, (C) sidewise view of SHB showing the position of CHR on the NHR core grooves. (D) The packing of the C34 on the N36 (NHR) grooves. The N36 is represented as a molecular surface and the three residues of C34 which fit into the binding pocket are shown in red (8).

It has been proposed that in the process of membrane fusion of viral and host cell, at some point the pre-hairpin intermediate (PHI) is an extended structure with the NHRs forming an inner core coiled coil structure. It is possible that C34 which is similar to CHR helix, binds to NHR core and stops the fusion process by inhibiting SHB formation (11). The SHB structure is very thermally stable.

Our aim was to achieve comparatively less stable Hairpin protein with the NHR and CHR not binding to each other and further perform structural studies. Therefore, we had designed mutated Hairpin (HP M3) and mutated Hairpin+MPER (HM M3) as explained in Chapter 4. We have observed that HP M3 and HM M3 are ~50% less folded than respective wild type protein constructs. We then had hypothesized that as the secondary structure of the mutated gp41 ectodomain has the NHR and CHR interaction disturbed, then it is possible that these mutants may bind to the C34 peptides. Based on this hypothesis, we performed some binding studies of our protein constructs by measuring the change in the secondary structure by circular dichroism (CD).

Experimental Conditions

The protein HM M3 was expressed in *E. coli* cells. The protein was purified by inclusion body purification and IMAC. The purified protein was solubilized in 10 mM sodium formate buffer at pH 3.0. The peptide C34 was obtained from NIH Aids Research and Reference Reagent Program (10). The sequence is given below. Equimolar proportion of HM M3 and C34 were mixed and allowed to incubate for some time (2 hours – overnight). After incubation, the secondary

structure of the mixture was measured at 20 °C and at pH 3.0. All the readings were subtracted by buffer only CD spectra.

C34 peptide amino acid sequence (10): WMEWDREINNYTSLIHSLEESQNQQEKNEQELL

Results

The helicity of HM M3 is nearly 50% of the wild type HM with $\theta \sim -20000 \text{ deg cm}^2 \text{ dmol}^{-1}$ (Figure AIII-2A). The CD spectra of C34 (5 μM) only shows an unfolded structure (Figure AIII-2B). After co-mixing of HM M3 and C34 and overnight incubation, the CD spectra (Figure AIII-3) did not show any conspicuous difference. The protein sample was heated and then C34 was added and allowed to incubate overnight. This additional step of heating was performed to ensure that the NHR and CHR have least interaction between each other. However, the CD spectra still did not show any significant change in helicity (Figure AIII-4).

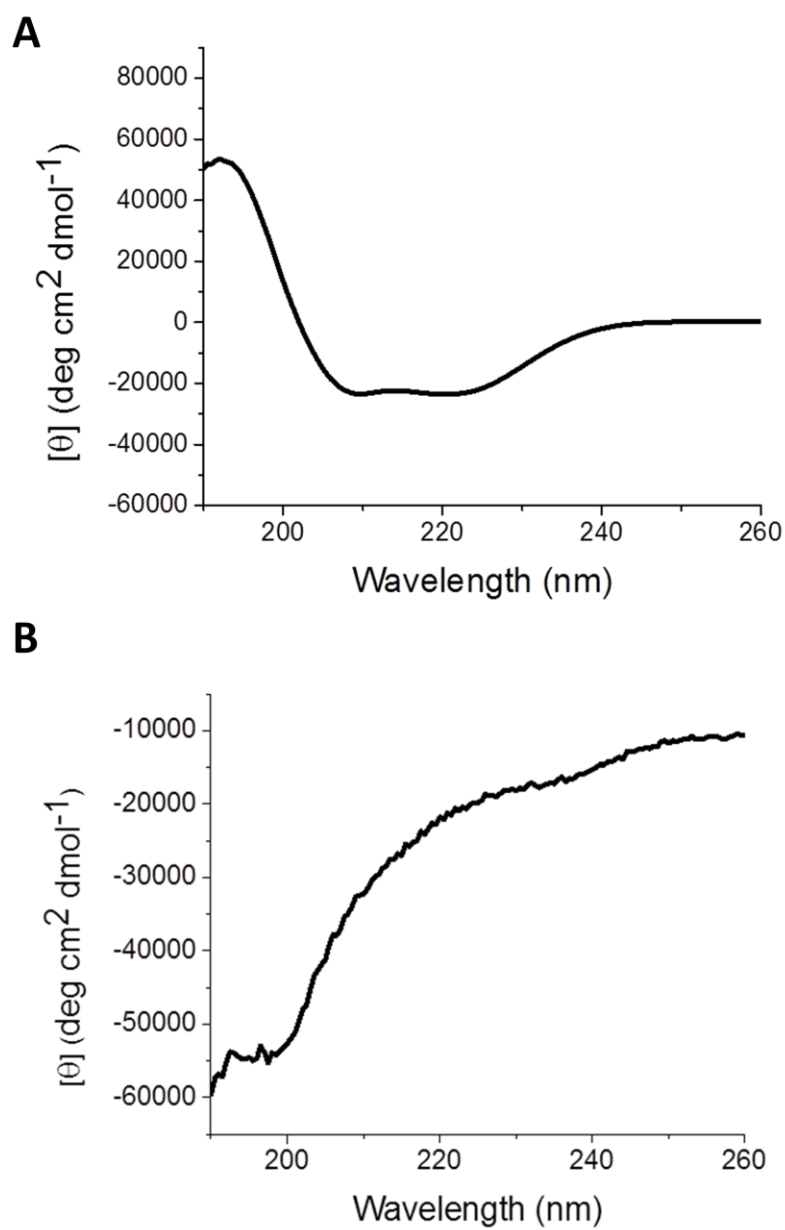


Figure AIII-2. Circular dichroism spectra of (A) 5 μ M of HM M3 and (B) 5 μ M C34 at pH 3.0

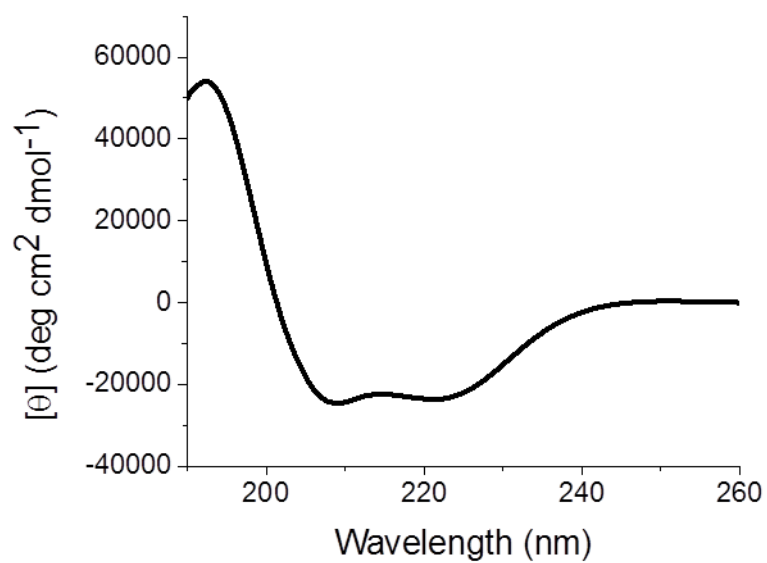


Figure AIII-3. Secondary structure of HM M3 + C34, 5 μ M each. The incubation was performed overnight at 4°C. The CD was performed at pH 3.0.

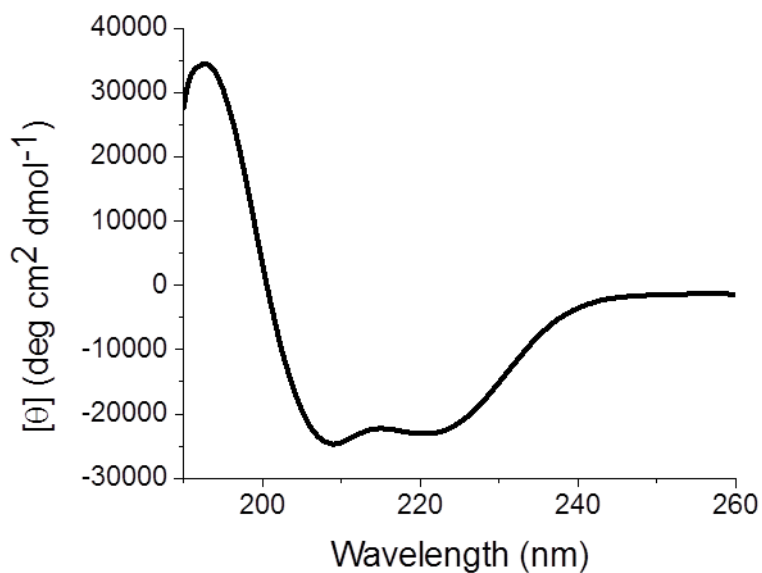


Figure AIII-4. The secondary structure HM M3 + C34, 5 μ M each. The protein was heated for 2 min and the C34 was added and incubated overnight at 4°C.

Discussion

We observed that there was no increase in the helicity of HM M3 even after the addition of C34 as hypothesized. The probable reason could be due to the unbinding of HM M3 with C34 due to repulsive force. HM M3 has a charge of +16 and C34 has a charge of +3 at pH 3.0. As both the protein and the peptide molecules are positively charged, it is likely that there is strong repulsion between the two leading to unbinding.

APPENDIX IV

Sample Preparation of Uniformly Labeled Bacterially Expressed HIV-1 Gp41 Hairpin Ectodomain

The main aim of this study was to determine the oligomeric state of gp41 Hairpin (HP) protein construct in the lipid environment. The oligomeric state of HP has been studied in aqueous solution condition (explained in Chapter 3). We were interested to understand the oligomeric state of the protein molecules when they are incorporated in the vesicles.

Approach

It was planned to label the HP with (1) at least 20% ^{13}C and (2) uniformly ^2H . The labeling of the protein constructs would be optimized by running cross-polarization magic angle spinning (MAS) solid state NMR experiments. With MAS, the polarization of abundant nuclei, i. e. ^1H is transferred to comparatively rare nuclei, in our case ^{13}C , to achieve a very good signal to noise ratio. This experiment would help us to understand whether the protein samples are labeled according to our expectation. According to our hypothesis, after the labeling of the protein samples would be complete, then the differently labeled protein samples will be incorporated in the vesicles, followed by ^{13}C - ^2H rotational echo double resonance solid state NMR experiments.

Materials and Method

Minimal media: For 50 mL media – (1) M9 – 0.564 g and autoclaved, (2) sterile filtered 0.1 M CaCl_2 , (3) sterile filtered 1 M $\text{MgSO}_4 \cdot \text{H}_2\text{O}$, (4) Antibiotic, (5) Vitamin (MEM), (6) 100% labeled (^{13}C or ^2H) glucose.

Cell Growth Conditions

The BL21 (DE3) *E.coli* cells were first grown in LB medium at 37 °C overnight. The cells were then centrifuged, collected and inoculated into already prepared minimal media. After growth for ~3 hours in minimal media, 2 mM IPTG was added for over-expression of the protein of interest and allowed the cells to grow overnight. After the cells were harvested, lysis was done to isolate the inclusion bodies containing labeled HP protein. Controls were done with inclusion body pellet containing unlabeled HP protein.

NMR Experiments

Cross polarization NMR experiments were done with the help of our group member Shuang Liang. The experiments were done at 6 kHz MAS and -50 °C. The ^{13}C O peak of labeled protein was referenced to the methylene peak of adamantane at 40.5 ppm.

Result

According to the cross polarization experiment, a clear ^{13}C O chemical shift for the 100% ^{13}C labeled HP protein in inclusion body sample was observed (Figure AIV-1). Minimal signal was observed for 20% ^{13}C labeled sample. There was no signal for 100% ^2H labeled HP protein in inclusion body sample. This could be due to exchange of ^2H with ^1H during the protein synthesis and sample preparation steps.

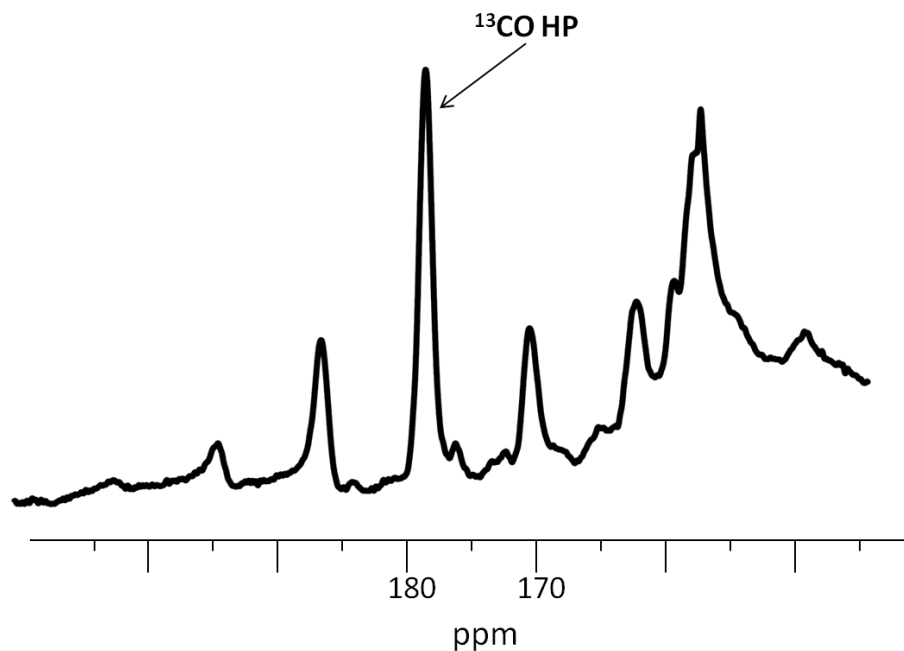


Figure AIV-1. MAS cross polarization signal for ^{13}C labeled inclusion body containing HP protein.

The CO peak of HP was referenced to the methylene peak of adamantane at 40.5 ppm.

Future Work

For labeling of HP protein with ^2H , probably the cells have to be grown in 70-80% D_2O . This could be done by growing the cells in LB in D_2O and the transferring the cells to minimal media in D_2O .

APPENDIX V

Location of Experimental Data Files

The location and file name of experimental data is listed below. There is mention of relevant experiment/conditions at which the experiments were done.

CD

Located in Dr. Lisa Lapidus lab computer

- (a) 3182014, 3192014, 3252014 – HM in DPC at pH 4.0 and 7.4
- (b) Apr2013GelFiltration – HP and HM Oligomers
- (c) 03112013 and 03192013– HM in 3M GuHCl
- (d) 02252013 – HP after Gel Filtration
- (e) 2052013 – FPHM
- (f) 12112012, 12142012 – GuHCl 1-6M
- (g) 12062012 – HM and HMM3 1M GuHCl
- (h) 12012012, 12052012 – M3+C34
- (i) 10192012 – HP and HM
- (j) 5222012 – HMM2 and HMM3
- (k) 02222011 – FP-HP

Fluorimeter

Located in Dr. Lisa Lapidus lab computer

POPC/POPG/Chol Experiments

- (a) 2012013, 2042013, 2072013, 2082013, 02192013, 02202013, 02222013, 02232013,
02282013, 03012013, 03032013, 03042013 – HP, HM, FPHM
- (b) 10262012, 10272012, 10312012, 11012012, 11022012, 11082012 – HP and HM

POPC/DOTAP/Chol Experiments

4142014, 4182014, Apr2_2014, Apr32_14, 4152014, Apr42014 – HP and HM

FPLC

Located in room 14 FPLC computer

- (a) Koyeli/Project for Koyeli/03142013_0.3ml/min_less dilute_2CV
- (b) Koyeli/Project for Koyeli/03142013_0.3ml/min_2CV
- (c) Koyeli/Project for Koyeli/03142013_2.1
- (d) Swetha/Method 3.8/HP(10mg_ml)DPC – 03/12/2014

NMR

- (a) khafre0/mb4b/data/Koyeli_Shuang/cp_13C_ICB_061314_2
- (b) khafre0/mb4b/data/Koyeli_Shuang/cp_unlabeled_P_052014
- (c) khafre0/mb4b/data/Koyeli_Shuang/quecho_2H_ICB_06252014
- (d) khafre0/mb4b/data/koyeli/FPH_L7F8_CPRamp_1 to 14

REFERENCES

REFERENCES

1. Sackett, K., Nethercott, M. J., Epand, R. F., Epand, R. M., Kindra, D. R., Shai, Y., and Weliky, D. P. (2010) Comparative Analysis of Membrane-Associated Fusion Peptide Secondary Structure and Lipid Mixing Function of HIV gp41 Constructs that Model the Early Pre-Hairpin Intermediate and Final Hairpin Conformations. *Journal of Molecular Biology* **397**, 301–315
2. Sackett, K., Nethercott, M. J., Shai, Y., and Weliky, D. P. (2009) Hairpin Folding of HIV gp41 Abrogates Lipid Mixing Function at Physiologic pH and Inhibits Lipid Mixing by Exposed gp41 Constructs. *Biochemistry* **48**, 2714-2722
3. Johnson, E. C. B., and Kent, S. B. H. (2006) Insights into the mechanism and catalysis of the native chemical ligation reaction. *J Am Chem Soc* **128**, 6640-6646
4. Sackett, K., Wexler-Cohen, Y., and Shai, Y. (2006) Characterization of the HIVN-terminal fusion peptide-containing region in context of key gp41 fusion conformations. *Journal of Biological Chemistry* **281**, 21755-21762
5. Qiang, W., Bodner, M. L., and Weliky, D. P. (2008) Solid-state NMR spectroscopy of human immunodeficiency virus fusion peptides associated with host-cell-like membranes: 2D correlation spectra and distance measurements support a fully extended conformation and models for specific antiparallel strand registries. *J Am Chem Soc* **130**, 5459-5471
6. Chan, D. C., Fass, D., Berger, J. M., and Kim, P. S. (1997) Core structure of gp41 from the HIV envelope glycoprotein. *Cell* **89**, 263-273
7. Weissenhorn, W., Dessen, A., Harrison, S. C., Skehel, J. J., and Wiley, D. C. (1997) Atomic structure of the ectodomain from HIV-1 gp41. *Nature* **387**, 426-430
8. Chan, D. C., and Kim, P. S. (1998) HIV Entry and Its Inhibition. *Cell* **93**, 681–684
9. Markosyan, R. M., Cohen, F. S., and Melikyan, G. B. (2003) HIV-1 Envelope Proteins Complete Their Folding into Six-helix Bundles Immediately after Fusion Pore Formation. *Molecular Biology of the Cell* **14**, 926–938
10. Gallo, S. A., Sackett, K., Rawat, S. S., Shai, Y. C., and Blumenthal, R. (2004) The stability of the intact envelope glycoproteins is a major determinant of sensitivity of HIV/SIV to peptidic fusion inhibitors. *Journal of Molecular Biology* **340**, 9-14
11. Chan, D. C., and Kim, P. S. (1998) HIV entry and its inhibition. *Cell* **93**, 681-684