ROLE AND MEMBRANE LOCATION OF HIV GP41 HAIRPIN IN FUSION

By

Md Rokonujjaman

A DISSERTATION

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

Chemistry - Doctor of Philosophy

2024

ABSTRACT

The primary cause of acquired immunodeficiency syndrome (AIDS) is the human immunodeficiency virus (HIV), which infects target white blood cells by connecting (fusion) the membranes of the host cell and HIV, causing the viral nucleocapsid to deposit in the cytoplasm. HIV's Gp41, a single-pass integral viral membrane protein with a 150-residue soluble domain outside the virus, is involved in fusion catalysis. This soluble ectodomain (SE) adopts a hairpin structure that contains N-helix-turn-C-helix, and two helices align themselves antiparallelly and there is van der Waals contact between them. The final fusion structure also consists of three hairpins joining together to form a six-helix bundle. There are competing hypotheses about the catalytic role in HIV/cell fusion of the SE hairpin structure. One hypothesis is that the hairpin is a “post-fusion” structure with no role in catalysis, whereas another hypothesis postulates an important role for the hairpin in catalysis which is associated with binding to the membrane. The latter hypothesis is supported by rapid vesicle fusion induced by the hairpin under some conditions. Further, mixtures of WT and V513E proteins of a large HIV gp41 ectodomain construct with final trimer of-hairpins structure exhibit V513E (gp41 V2E in absence of gp120)-dominant reduction of vesicle fusion that is quantitatively similar to that for gp160 (glycoprotein 160, which is a precursor protein that gets cleaved into two smaller proteins: gp120 and gp41) fusion and infection. Vesicle fusion and helicity were measured for FP_HM (fusion peptide (FP)+hairpin (H)+membrane proximal external region (MPER)) using trimers with different fractions (f’s) of WT and V2E proteins. Correlation between low helicity vs. fusion of V2E was also found in this study. Reductions in FP_HM fusion and helicity vs. f_{V2E} were quantitatively-similar to those for gp160-mediated fusion and infection. Global fitting of all V2E data supports 6 WT gp41 (2 trimers) required for fusion. These data are understood by a model in which the ~25 kcal/mol free energy for initial membrane apposition is compensated by the thermostable hairpin between the Fp in target membrane and Mper/transmembrane domain in virus membrane. The data support a structural model for V2E dominance with a membrane-bound Fp with antiparallel β sheet and interleaved strands from the two trimers. This increased distance is the hypothesized reason for reduced fusion and infection. Role of hairpin and correlation between helicity vs. fusion further studied through characterization and fusion of G10V_FPHM and L9R_FPHM,

respectively. Results established the less helicity in V2E/L9R than WT vs. less fusion in V2E/L9R than WT and similar helicity in WT/G10V vs. similar fusion in WT/G10V. Position of hairpin in membrane could be better understanding of membrane apposition (viral and host membrane) during fusion because deep SE insertion indicates the close membrane apposition and surface insertion predicts far membrane apposition. Moreover, a hairpin contribution to catalysis of membrane fusion is most straightforwardly understood if there is deep insertion as the locations of the surrounding lipid molecules in the membrane would likely be very different than their energetically stable positions in an unperturbed bilayer. We used either ^{13}C-glucose or ^{13}C-glycerol to express a large soluble ectodomain in a minimum medium, with a dilution of the ^{13}C labeling accomplished by mixing with unlabeled glucose or by selectively labeled glycerol. The proximity of protein with ^{13}C-labeled and lipid with ^{2}H labeled positioned either in middle /periphery/ in between of middle and periphery of the membrane were measured using the rotational-echo double-resonance (REDOR) solid-state NMR technique. The production of ^{13}C proteins used a variety of carbon sources, and as a result, various labeled carbon peak assignments on the side chains of amino acids were observed in NMR. There was a significant REDOR dephasing for samples from different carbon sources, which shows that the soluble ectodomain has mostly peripheral location in the membrane.

I dedicate all my PhD efforts to the memory of the martyrs of the July-August 2024 revolution in Bangladesh.

ACKNOWLEDGEMENTS

First and foremost, I would like to express my deepest gratitude to my advisor, Dr. David P. Weliky, for providing me with the opportunity to become a member of this amazing group, for teaching me amazing lessons about life and science, and for being patient and consistently supportive. I would like to express my gratitude to him for all of his advice, strength, and support during my trying times. Over the past six years, he has imparted information and skills related to NMR and biology as well as, most importantly, the ability to learn and solve problems. I am also profoundly thankful to the members of my dissertation committee, Dr. Heedeok Hong, Dr. Jhon McCracken, and Dr. Jian Hu, for their valuable feedback, constructive criticism, and continuous support. Their expertise and suggestions have significantly contributed to the improvement and refinement of this thesis.

I am grateful to the Weliky group's past member Dr. Robert Wolfe for his initial training of my PhD research. I will always remember his tremendous help and support for lab instruction in fundamental biochemistry and molecular biology techniques as well as REDOR solid state NMR technique. I want to express heartfelt thanks to Dr. Daniel Holms and Dr. Xi Lie for their tremendous help and support in learning REDOR solid state NMR for location of HM in membrane. I highly appreciate the help of Dr. Ujjayini Ghosh for initial training of vesicle fusion study in project 1. I want to express my gratitude to the lab's current group member Tahmina Khatun for her immense support and assistance in my PhD journey. I also express my thanks to Forkan Saroar and Noel Chao for their tremendous support in my PhD work. They all contributed to making my research life an enjoyable and memorable experience.

Finally, I would like to express my deepest gratitude to my family. I have not had the opportunity to thank my parents properly. I want to say that everything I have achieved so far is because of you. My brothers have always been a tremendous support throughout my academic journey; thank you so much. A special thanks to my wife, Sabiha Sultana, for her unwavering support throughout my Ph.D. journey. Her tremendous support in taking care of our children and managing our home during the challenging times of my research has been invaluable.

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

ACE2-Angiotensin-converting enzyme 2

AIDS-Acquired immunodeficiency syndrome

AUC-Analytical ultracentrifugation

BCA-Bicinchoninic assay

bNAb-Broadly neutralizing antibody

CCR5-C-C Chemokine receptor type 5

CD-Circular dichroism

CDC-Centers of disease control and prevention

CD4-Cluster of differentiation type 4

CD22-Cluster of differentiation 22

Chol-Cholesterol

CHR-C-terminal helix region

CMC-Critical micelle concentration

CP- Cross-polarization

Cryo-EM Cryo-electron microscopy

CSA-Chemical shift anisotropy

CT- Cytoplasmic tail

CXCR4 C-X-C Chemokine receptor type 4

DCM- Dichloromethane

DM- n-Decyl-β-D-Maltoside

DMF- Dimethylformamide

DNA- Deoxyribonucleic acid

dNTPs- deoxynucleotide triphosphates

DPPC- n-Dodecylphosphocholine

DPPG- 1,2-Dipalmitoyl-sn-glycero-3-phosphocholine

DTT- Dithiothreitol

E. coli- Escherichia coli

EDTA- Ethylenediamine tetra acetic acid

Endo- Endodomain

EPR- Electron paramagnetic resonance

ESR- Electron spin resonance

FP-Fusion peptide

FPHM- FP+HM

FTIR- Fourier-transform infrared spectroscopy

GPCR- G protein coupled receptor

gp160- Glycoprotein 160

gp140- Glycoprotein 140

gp120- Glycoprotein 120

gp4- Glycoprotein 41

GuHCl-Guanidine hydrochloride

GC-Guanine and cytosine

GUVs- giant unilamellar vesicles

G10V- Mutation of the Gly-G to Val-V

1,3-^{3}CGly-1,3-^{13}C labeled Glycerol

2-^{3}CGly-2-^{13}C labeled Glycerol

HA- Hemagglutinin

HA2- Hemagglutinin subunit 2

HP-Hairpin

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

HF- hydrofluoric acid

HFP- HIV Fusion peptide

HIV- Human immune deficiency virus

HM- gp41 NCR+CHR+MPER

HPLC- High-performance liquid chromatography

HRP- horseradish peroxidase

HR1- Heptad repeat 1

HR2- Heptad repeat 2

IDA- Iminodiacetic acid

IMAC- Immobilized metal affinity chromatography

IB- Inclusion body

IPTG- Isopropyl β-D-1-thiogalactopyranoside

kDa-Kilodalton

LB- Luria-Bertani broth

L9R- Mutation of the Leu-L to Arg-R

LUVs- Large unilamellar vesicles

MALDI-MS matrix assisted laser desorption ionization mass spectrometry

MAS- Magic angle spinning

MES- 2-(N-morpholino) ethanesulfonic acid

MDa-Megadaltons MD-Molecular dynamics

MW-Molecular Weight

MPER- membrane-proximal external region

mRNA- messenger RNA

NHR- N-terminal helix region

NMR- Nuclear magnetic resonance

N-NBD-PE N-(7-nitro-2,1,3-benzoxadiazol-4-yl) (ammonium salt) dipalmitoylphosphatidylethanolamine

N-Rh-PE N- (lissamine rhodamine B sulfonyl) (ammonium salt)

NC- Nitro cellulose

NTA- Nitrilotriacetic acid dipalmitoylphosphatidylethanolamine

OD-Optical density

PBS- Phosphate-buffered saline

PCR- polymerase chain reaction

PG- phosphatidyl glycerol

PHI- pre-hairpin intermediate

pI- isoelectric point

POPC- 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine

POPG- 1-palmitoyl-2-oleoyl-sn-glycero-3-[phospho-rac-(1-glycerol)] (sodium salt)

REDOR- Rotational-echo double-resonance

RF-Radio frequency

Rh-PE N-(Lissamine rhodamine b sulfonyl)-1,2-dihexadecanoyl-sn-glycero-3-phosphoethanolamine

RNA- Ribonucleic acid

RP- Recombinant protein

RBS-Ribosome binding site

RP-HPLC Reversed-phase HPLC

Sarkosyl- Sodium lauryl sarcosinate

SDS- Sodium dodecyl sulfate

SDS-PAGE- Sodium dodecyl sulfate polyacrylamide gel electrophoresis

SEC- Size-exclusion chromatography

SE-Soluble ectodomain

SHB- Six helix bundle

SK- Sarkosyl

SSNMR- Solid-state nuclear magnetic resonance

SUVs-Small unilamellar vesicles

SPR-Surface plasmon resonance

TCEP- Tris(2-carboxyethyl) phosphine hydrochloride

THF- Tetrahydrofuran

TM- Transmembrane domain

T_m-Melting temperature

UNAIDS- The Joint United Nations Programme on HIV/AIDS

U-^{3}CGlu-Uniform ^{13}C labeled Glucose

V2E- Mutation of the Val-2 to Glu-2

WB- Western blot

WHO- World health organization

CHAPTER-01: Introduction to Human Immunodeficiency Virus (HIV) Membrane Fusion Protein Gp41.

1.1: Introduction

1.1.1: HIV and AIDS

Human Immunodeficiency Virus (HIV) is the virus that causes the immune system disease known as Acquired Immune Deficiency Syndrome (AIDS). Since the first cases of AIDS were documented more than 35 years ago, over 35 million people have perished from the disease globally, according to The Joint United Nations Programme on HIV/AIDS, also known as UNAIDS[1]. Over 1.7 million new infections and over 38 million people living with the disease worldwide by the end of 2019 led to an estimated 700,000 deaths from the disease[1]. The human immune system can regain its capacity to fight infection with the use of HIV antiretroviral medications, which can stop viral replication in humans. However, just 46% of AIDS patients are taking these antiretroviral medications, and a year's supply costs about $20,000.

There are two kinds of the HIV virus: HIV-1 and HIV-2, however, most recent research has concentrated on HIV-1. The HIV type 1 virus is referred to as HIV in the sections that follow in this dissertation. The structure of the HIV virus, which has a diameter of roughly 100–120 nm, is seen in Figure 1.1[2-4]. The family retroviridae includes the genus Lentivirus, which includes HIV. The virion of the HIV virus is encased in a lipid membrane because it is an enveloped virus. Electron microscopic data showed the viral glycoprotein (gp) of 8-14 trimers on the viral membrane[5-7]. These proteins are external surface protein gp120 and transmembrane protein gp41, respectively are cleaved inside the host cell from a precursor protein called gp160 and non-covalently associated outside the virus surface that promote the fusion of the viral and host cell membranes[4-7]. To maintain the virion's integrity and viral structure, the matrix proteins form a sphere beneath the membrane. The virus' nine genes enclosed by a conical capsid composed of 2,000 copies of the viral protein p24 has been coded by the two copies of positive-sense single-stranded RNA. The nucleocapsid proteins, p7, and enzymes required for the virion's growth, including reverse transcriptase, proteases, ribonuclease, and integrase, are tightly bound to the single-stranded RNA[4,8-11].

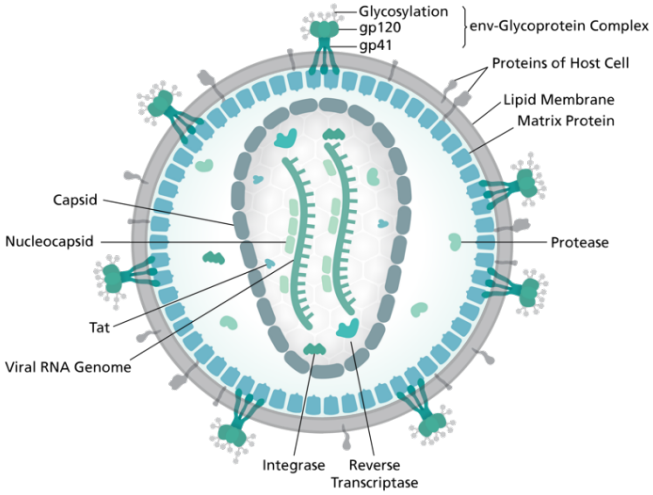


Figure 1.1: An illustration of the HIV virus.

1.1.2: Pathway of HIV viral infection

The HIV infection cycle is depicted in Figure 1.2 panel A. The binding of gp120 to the host cell starts this process. Important immune system components like $CD4^+$ T cells and macrophages are infected by HIV[3, 12, 13]. Normally, gp120 binds to CD4 molecules on the host cell which modifies the structure of gp120. This conformational change makes it easier for the coreceptor to attach to a conserved binding site on gp120, such as CXCR4 and CCR5, which are members of the chemokine receptor family[14, 15]. After the binding, gp120 is dislodged, exposing gp41 as a result. A significant conformational change occurs during the fusion process, which is gp41's primary function in catalyzing the fusion of the viral and host cell membranes[16, 17]. A fusion pore is created at the conclusion of the fusion, allowing the nucleoprotein complexes to enter the host cell. The schematic diagram and images of the fusion process taken using electron microscopy are shown in Figure 1.2, panels B and C[18]. On the other hand, another potential route for HIV entrance is by endocytosis. Results from electron microscopy demonstrated that clathrin-coated pits can endocytose HIV (Figure 1.2 D)[18, 19]. After fusion, reverse transcriptase catalyzes the conversion of the RNA into DNA, allowing the DNA to enter the host cell nucleus and be incorporated into the genome of the host cell. Following the expression of the viral proteins, new viruses can ultimately bud from the host cell membrane and initiate a fresh infection cycle[20]. The design of antiretroviral medications is dependent on how viruses

replicate. Antiretroviral medications are typically taken in combination to prevent drug resistance, prolonging the patient's life expectancy[11].

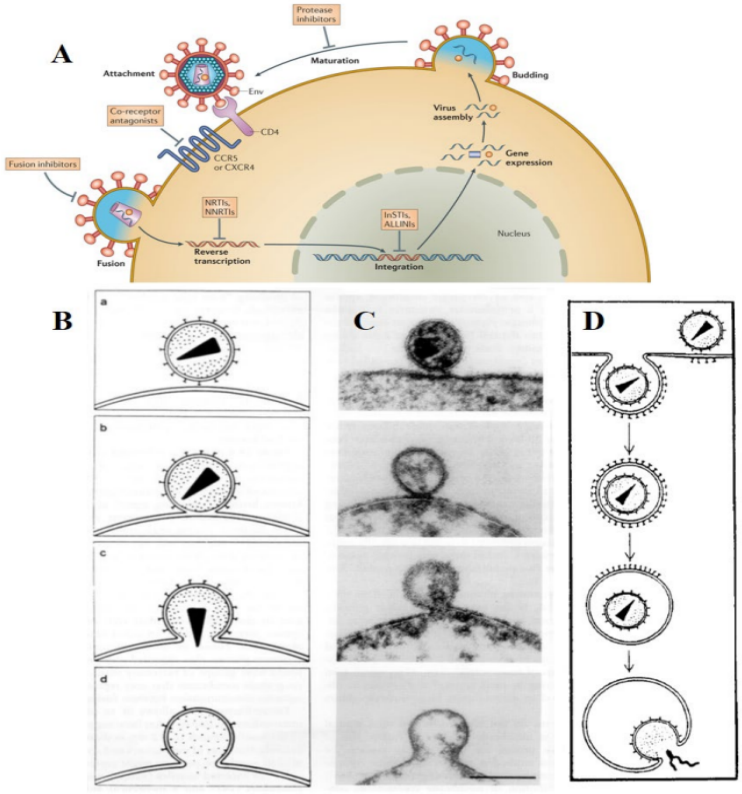


Figure 1.2: (A) The HIV-1 life cycle's stages. The potential mechanism of antiretroviral therapy is depicted in the diagram by the yellow boxes[11]. (B) HIV infection model in diagram form. Virion spikes are made up of the gp120/gp41 complex. The HIV virus binds to the host cell, the viral and host cell membranes partially fuse, pores form, and the viral genetic material completely fuses and enters the host cell. (C) The images from panel B's electron microscopy of the HIV infection process[18]. (D) A schematic infection model showing clathrin-mediated endocytosis as an alternative HIV entry mechanism [18,19].

1.1.3: The gp41's structure

In a noncovalent way, the HIV envelop glycoproteins assemble onto the viral membrane to create a "knob"-shaped complex, with three molecules of gp120 serving as the "cap" and three molecules of gp41 serving as the "stem" (see Figure 1.2 on the surface of the virus)[11,16]. While gp41 joins the viral and host cell membranes to release the RNA into the host cell, gp120 is responsible for attaching to the target cell. The precursor gp160, which

is produced from the viral gene and is cleaved in the host cell, produces the proteins gp120 and gp41[3]. Gp41 has been a target protein in the development of HIV vaccines and antiretroviral medications because it is crucial for membrane fusion. A transmembrane protein containing 345 amino acids is called gp41. Figure 1.3 displays the schematic representations of the entire HIV gp41 protein along with the appropriate amino acid sequence[21]. A significant conformational change in gp41 occurs both before and after fusion, as shown by a wealth of prior research. Big strides have been achieved in revealing the structure of gp41 and its potential fusion process using X-ray crystallography and cryo-electron microscopy[16,17,22-24].

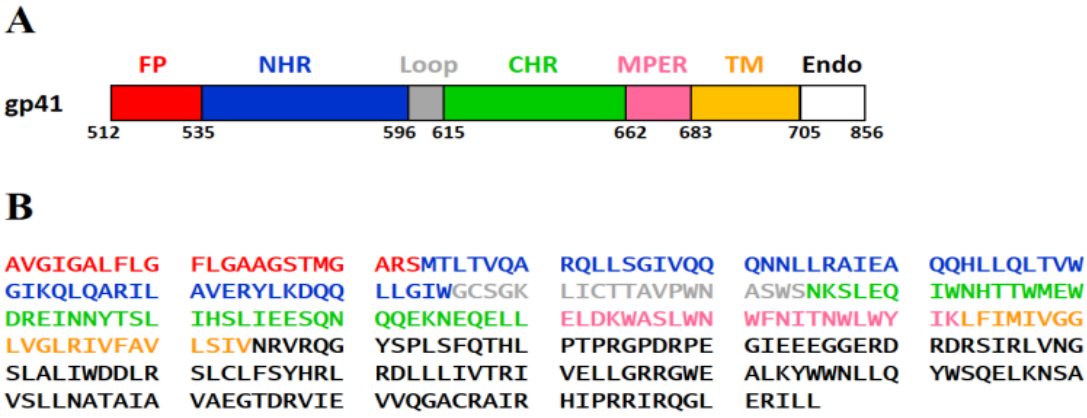


Figure 1.3: (A) Schematic of full-length HIV gp41 and associated color: FP ≡ fusion peptide, red; NHR ≡ N-helix region, blue; Loop, grey; CHR ≡ C-helix region, green; MPER ≡ membrane-proximal external-region, pink; TM ≡ transmembrane domain, orange; and endo ≡ endodomain, white. (B) With the exception of the endo domain in black for amino acid sequences, the segments in panel A are color-matched. The sequence is from the HXB2 laboratory strain of HIV and have the gp160 precursor residue numbering, 1-511 for gp120 and 512-856 for gp41[21].

With a resolution of ~5 Å, X-ray crystallography and cryo-electron microscopy were used to examine the structure of the soluble, cleaved HIV envelope glycoprotein trimer[17,24,25]. These experiments used antibodies to stabilize the HIV virion-derived gp120/gp41 complex, which was in a gp41 prefusion form. Figure 1.5 displays the complex's structure and schematic representation. A ~50 Å-helix trimer bundle of gp41 NHR was produced in this study, and an additional short helix extended but was bent away from the trimer

axis[17]. CHR looped around the bottom of the central trimer at an angle of about 60° to NHR (Figure 1.4 B and C). In contrast to the post-fusion state, where NHR and CHR formed tightly packed SHB, the NHR/CHR interaction in the pre-fusion state was radically different[16]. The results of cryo-electron microscopy revealed similar structures[24, 25]. It is demonstrated how the NHR's coiled-coil structure differs between before and after fusion (Figure 1.5)[25]. Each N-terminal helice's angle changed by 15 degrees between the pre-fusion condition (Figure 1.5 A) and the post-fusion state (Figure 1.5 B). The NHR/CHR SHB was also created in the post-fusion state, along with a tighter packing of the N-terminal helices.

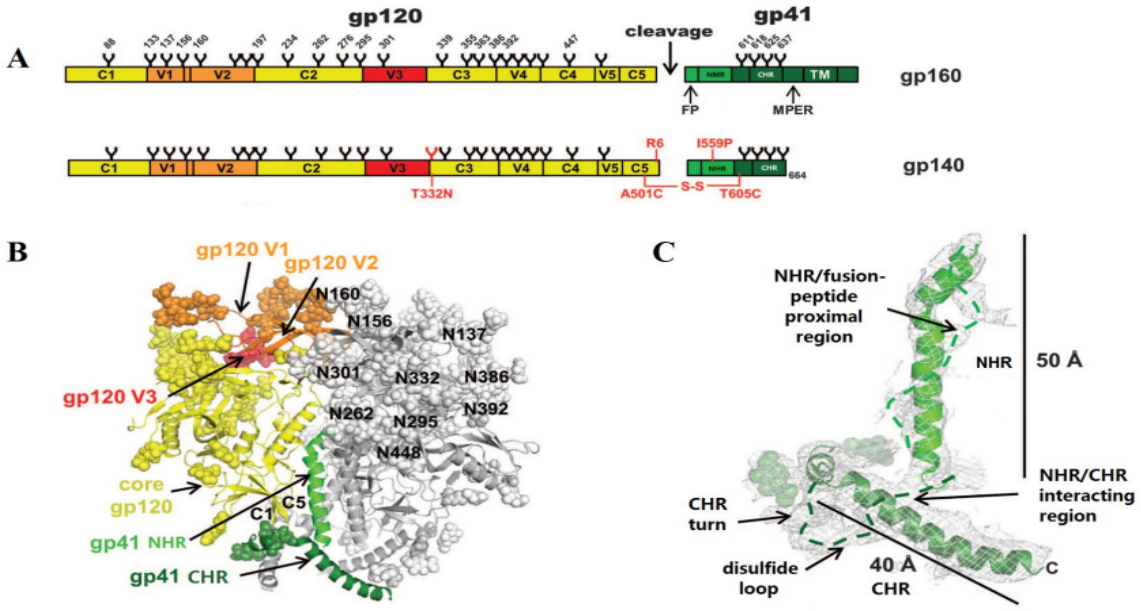


Figure 1.4: (A) Diagram of the HIV gp140 construct compared to full-length gp160 in the study[17]. On each of their corresponding Asn residues, N-linked glycans are displayed and numbered. The transmembrane (TM), endodomain, FP, NHR, CHR, MPER, and endodomain components of gp41 are highlighted. The additional N332 glycan site and the alterations are depicted in red. The color labeling is still present in (B) and (C). (B) The gp140 trimer from the side. Glycans are depicted as spheres, and the main domains are designated to correspond to panel A. Gp41 inside the gp140 complex. The dashed line areas were ambiguous for secondary structure determination.

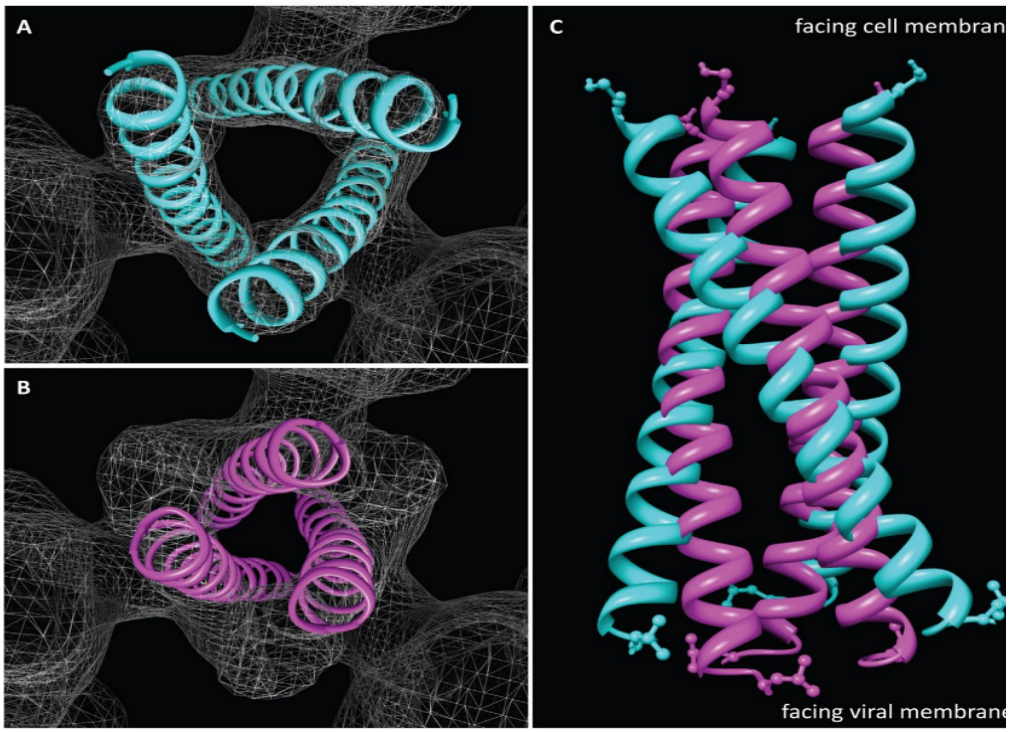


Figure 1.5: (A) NHR coiled-coil trimer comparison in the pre- and post-fusion stages. Cryo-electron microscopy was used to determine the positions of the three N-terminal helices in the pre-fusion condition (B) An X-ray crystallographic top view of the same helices' positions in the post-fusion state. (C) Superposition of the three N-terminal helices' positions in the pre-fusion (cyan) and post-fusion (magenta) conformations.

1.1.3.1: Fusion peptide (FP)

The Glycine (Gly) rich hydrophobic N-terminal domain of gp41 is called the FP region. It has the amino acid sequence AVGIGALFLGFLGAAGSTMGARS, is the first 23 amino acids at the N-terminus of HIV gp41. There are some structural nuclear magnetic resonance (NMR) studies have been conducted using FP in conjunction with detergent micelles[26]. The FP is unstructured in solution, but when detergents such sodium dodecyl sulfate (SDS) or n-dodecyl phosphocholine (DPC) are added, it develops an α-helical structure[27-29]. Conflicting structural studies have proposed a β-sheet structure for the FP in a physiologically relevant membrane environment[30-31], whereas another study concludes that the FP region takes on either α-helical monomeric structure or a β-sheet structure at low and high peptide: lipid concentration, respectively[32]. When cholesterol is

present in the lipid membrane or not, differences in FP conformation have been noted. When cholesterol is present, the FP adopts a β-sheet form; when it is not, it adopts an α-helical structure[31]. There were two membrane places for FP: the main was deeply inserted, while the minor was shallowly put. (Figure 1.6)[29,33]. A positive correlation was observed between the insertion depths and the fusion activities by examining the membrane location and fusogenicity of FP wild type and V2E mutation[34].

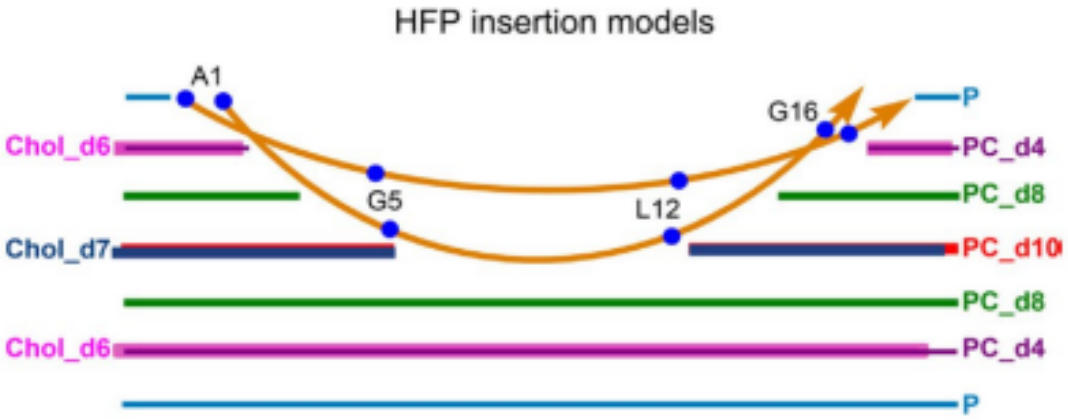


Figure 1.6: Two membrane locations of FP studied by ^{13}C-^{2}H REDOR: majorly deep and minorly shallow insertion. The FP is ^{13}CO labeled on Gly_5, Leu_12, and Gly_16 residues. The labels on both sides of the diagram are the approximate membrane locations of the ^{2}H's and ^{31}P's where P ≡ phosphorous in lipid; Chol_d6, Chol_d7, PC_d4, PC_d8, and PC_d10 are deuterium in cholesterol or phosphocholine lipids[33].

1.1.3.2: N-Helix region (NHR) and C-Helix region (CHR)

Without the loop region, the NHR and CHR's high-resolution crystal structures have a highly helical trimer shape that takes the form of a hairpin conformation[22,35]. Residues 546-581 and 628-661—where numbering starts with the NHR—were the areas under study. The gp41 core region is predicted to have this trimer NHR-CHR structure, also known as the six-helix bundle (SHB) (Figure 1.7). The hydrophobic interactions between the Ile or Leu residues hold the parallel NHR trimers in the SHB's core together. The NHR and CHR both have 4-3 heptad repetitions of hydrophobic residues in their amino acid sequences. The first and fourth hydrophobic residues of a seven-residue helical rotation are repeated to generate heptad repeats. There may be multiple helical turns per heptad

repeat, depending on the size of the protein. The CHR helices are arranged in an antiparallel pattern on the NHR trimer core's outside grooves[16]. A hydrophobic contact occurs primarily between the NHR and CHR. One example of a hydrophobic contact is between W628, W631 of the CHR, and W571 of the NHR, and another is between I635 of the CHR and L565 and L568 of the NHR. Additionally, side chain hydrogen of Q653 in CHR forms an intermolecular hydrogen bond with the backbone CO oxygen of V549 in NHR, as does side chain hydrogen of Q653 of NHR and backbone CO oxygen of Q551 in CHR. There is an inter-molecular salt bridge between the NHR and CHR and between the K574 and D632. The NHR and CHR SHB are thermally stable with melting temperatures of ~80^{o} C for shorter ectodomain constructs with no loop (equimolar mixture of CHR and NHR peptides), whereas melting temperatures up to 110^{o}C have been observed for longer constructs (535-581-SGGRGG-628-662)[36,37]. This thermostability has supported the SHB as the final gp41 structure during fusion. Recent solution NMR structural study of NHR-loop-CHR (538-705) in DPC micelles at pH 4.0 suggests a monomeric structure with much less interaction between the NHR and CHR[26]. Banerjee et al (2014) showed that hairpin ($HP_{535-666}$) is thermostable monomeric structure in 50 mM sodium formate buffer + 150 mM NaCl + 0.2 mM TCEP at pH 3.2[7]. All the NHR-CHR construct's residue ^{13}C chemical shifts were compared to those of the individual peptides' NHR and CHR residues[26]. The study showed that the NHR-CHR protein construct's interaction with detergent compounds by increase of paramagnetic relaxation[38]. Another study showed that the NHR and CHR helices were shown to be amphipathic, with hydrophilic residues located on one side of each helix and hydrophobic residues on the other[39]. This finding leads to the hypothesis that NHR and CHR can both be entrenched at the lipid-water interface (surface representation through ^{1}H- ^{15}N TROSY-HSQC spectra of the NHR and CHR core monomer structure water-exposed in the interface)[39].

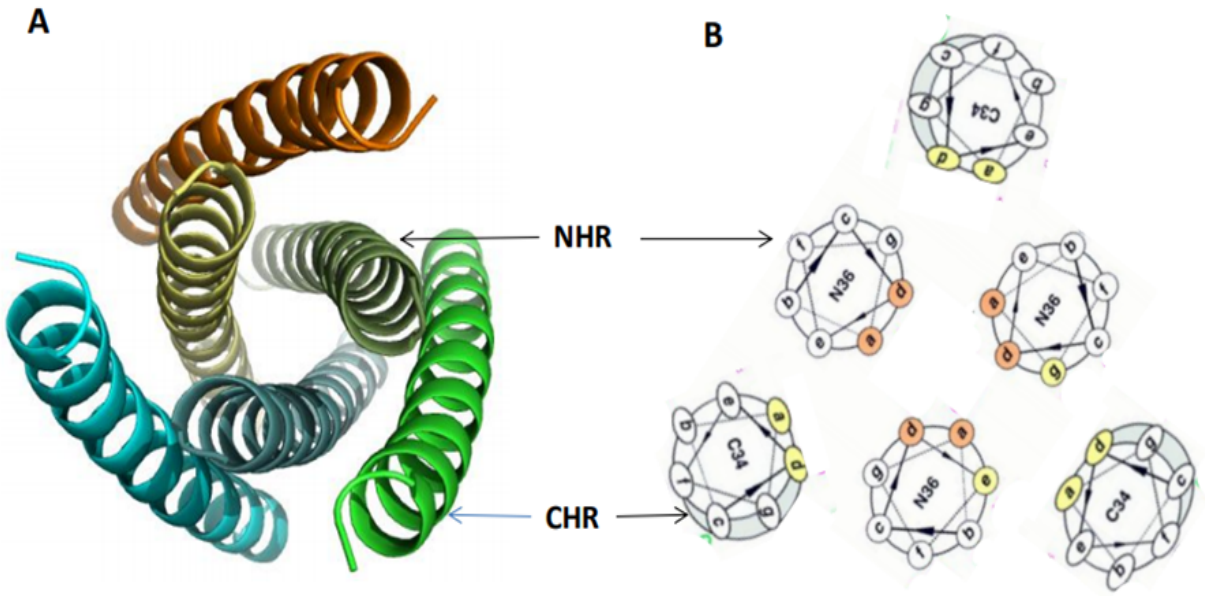


Figure 1.7 (A) NHR and CHR comprise a six-helix bundle of gp41 and (B) Heptad repeat amino acids in NHR and CHR show as yellow and orange, respectively[16]. Residues from 546-581 in NHR are called N36 with 26 residues and residues from 628-661 in CHR are called C34 with 34 residues.

1.1.3.3: Membrane proximal external region (MPER)

The area of the ectodomain that is nearest to the viral membrane is known as the membrane-proximal external region (MPER). The sequence of the 22 amino acid tryptophan rich MPER is ELDKWASLWNWFNITNWLWYIK. The MPER peptide in a DPC micelle was studied using liquid-state NMR, and the results showed a clearly defined -helical peptide (Figure 1.8 A). The MPER peptide also had an amphipathic structure, with the polar residues on one side of the helix and the aromatic residues on the opposite, indicating that the MPER was attached to the viral membrane[40]. MPER was found to have two distinct helical segments with a central hinge, producing an L-shape in a more recent investigation employing liquid-state NMR, electron paramagnetic resonance (EPR), and surface plasmon resonance (SPR) methods (Figure 1.8 B)[40]. A trimeric, coiled coil SHB was seen in the crystal structure of gp41, which also contained the NHR, CHR, and MPER area (Figure 1.8 C). The MPER extended from the CHR in continuation of a slightly bent long helix[41].

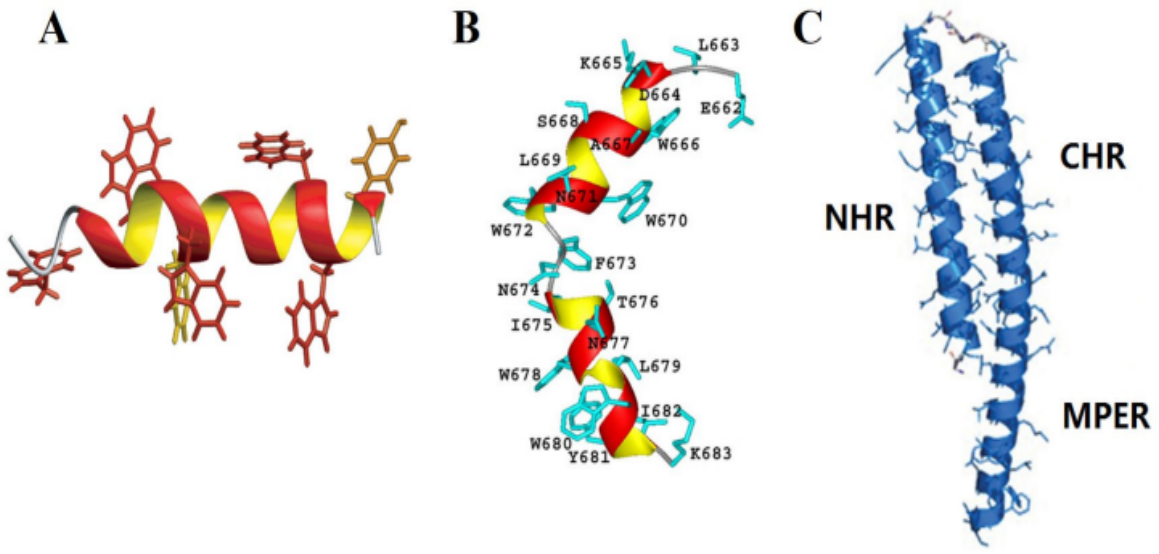


Figure 1.8: Models of MPER (A) exhibiting a straight-helix of $MPER_{665-683}$ in a DPC micelle at pH 3.5[40]**.** (B) An L-shaped kink between two -helices is visible in $MPER_{662-683}$ in DPC micelles at pH 6.6[41]**.** (C) $NHR_{547-575}$-$CHR_{630-662}$-$MPER_{663-675}$ monomer crystallography results[35]. The MPER in this structure terminates at I_{675} and the CHR and MPER form a continuous helix.

1.1.3.4: Transmembrane domain (TM)

The 22-residue transmembrane (TM) domain attaches the gp41 in the viral membrane. It developed a tilted-helical conformation in the membrane and exhibited a tendency to form a trimer, according to computer simulation results (Figure 1.9 A)[42,43]. MPER and TM form an uninterrupted long-helix, as shown by liquid state NMR (Figure 1.9 B)[44]**.** The MPER and TM domains all shown strong helicity, despite the fact that there were some variations in the experimental circumstances, such as pH and construct length. Although the entire gp41 protein does not yet have a high-resolution structure, the isolated ectodomain or membrane segment structures are very informative and aid in future research.

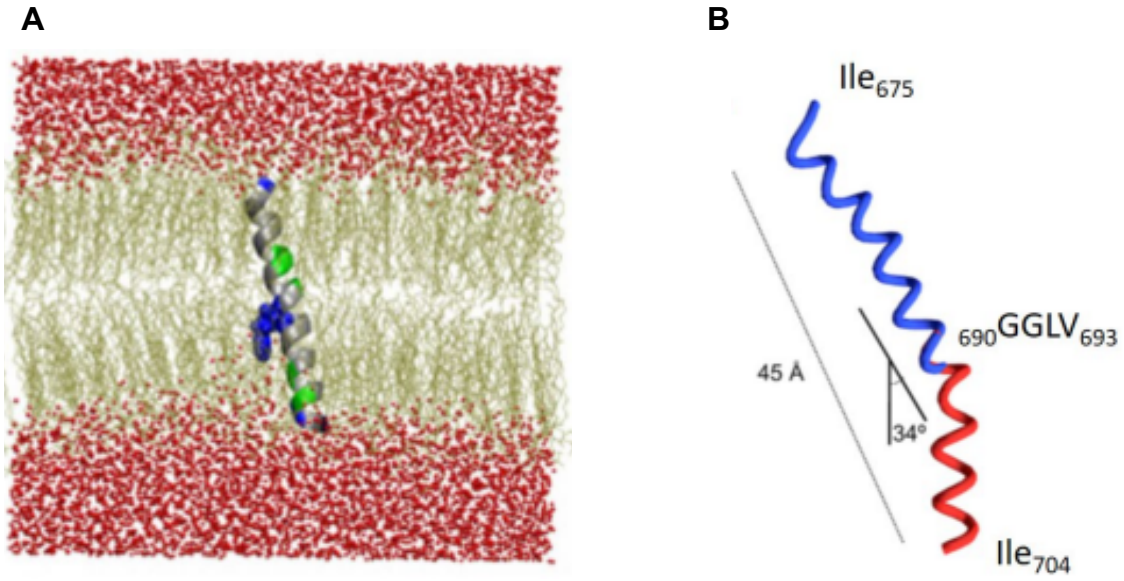


Figure 1.9: $TM_{681-707}$ membrane simulation using molecular dynamics (MD) shows that Arg694 is swimming toward the membrane surface. The TM peptide is depicted in grey, Arg residues are represented in blue, lipid molecules are shown in yellow, and water molecules are indicated in red[42]. Results of the $MPER_{675-683}$-$TM_{684-704}$ peptide's liquid-state NMR. The MPER and TM form a continuous α-helix with a turn at $_{690}GGLV_{693}$ residues[43].

1.1.3.5: Area of the cytoplasmic Endo domain

The region followed by TM domain C-terminal, which contains about 150 amino acids, known as the gp41 cytoplasmic endodomain (Endo). gp41 Endo is a single-pass membrane protein, at least that is what the computationally predicted structure suggests[45]. According to a fluorescent experiment used to analyze HIV-1 particle entrance during virus-cell fusion, juvenile virus particles were less active than mature viral particles[46]. Immature virus particles' inactivity is due to cytoplasmic Endo's connection with matrix protein; only Endo's dissociation from matrix protein and nucleocapsid protein results in gp41-mediated fusion[47]. The cytoplasmic endodomain has been found to be functionally important in the integration of the Env protein into the viral membrane by mutational experiments[48,49]. Mutagenesis of the C-terminal Endo has suggested that the viral matrix protein and cytoplasmic endodomain interact[50,51]. Moreover, the MPER-TM

structure is supported as it spans the membrane by the massive trimeric baseplate that the CT generates around the TM region (Figure 1.10)[51,52].

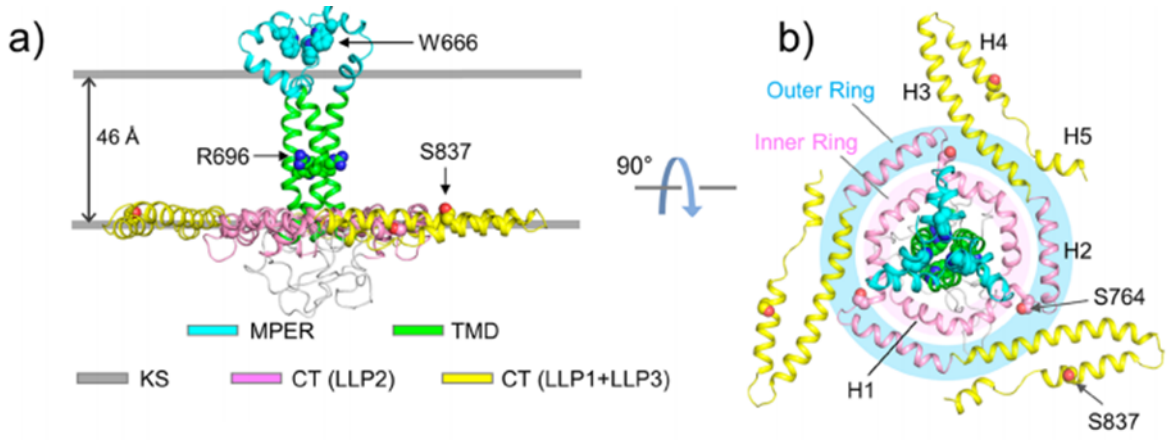


Figure 1.10: Complete membrane region structure of the HIV-1 Env, including CT[52]. (a) Integrated NMR-derived structural model of the MPER-TMD-CT trimer bicelles. (b) The inner and outer rings of the baseplate are coloured in pink and blue, respectively, in this top view (taken from the MPER) of the trimeric complex.

1.1.4: Potential mechanisms of membrane fusion

For a virus to enter host cells, two membranes must first fuse. The steps that membrane fusion takes are as follows: Outer leaflet membrane surface contact, formation of a stalk, hemifusion (where the outer leaflets of two membranes combine to form one membrane first and the distal leaflets remain separate until the opening of a fusion pore), and pore formation are the four processes that take place (Figures 1.11 and 1.12). The entire membrane fusion process takes place isothermally, and according to X-ray observations, the free energy minimum distance between physiologically significant membranes (which have essential roles in maintaining cellular function and overall physiological balance) is between 2 and 3 nm[48]. The repelling force between the two membranes' outer leaflets must be overcome with energy when they are more than 3 nm apart and electrostatic repulsion typically decreases with distance[53]. Water molecules in this membrane environment, tend to bind to the polar head groups of the lipid molecules by hydrogen bonding. Energy is needed to move the bound water molecules out of the way so that membrane fusion between the virus and host cells can take place. The desolvation

energy is the name given to this energy barrier. The activation energy necessary to generate the stalk, which is mostly made up of lipid bilayer, incorporates the desolvation energy. Although the proximal membrane leaflet's lipid molecules rearrange in the following stage, the distal membrane leaflet does not experience lipid mixing during stalk development[54]. Hemifusion occurs after the stalk development stage[53,55]. The structural change from stalk to hemifusion is not significantly exposed to water[54]. Because more energy is ended to pay the desolvation penalty during stalk development compared to the curvature energy (Gaussian curvature elastic energy- is one of the major contributions to the energy of fusion stalks for lipid) of hemifusion formation he activation energy of stalk formation is higher than that of hemifusion formation[53]. Pore formation comes after hemifusion production (Figure 1.12). In one model of membrane fusion, stalk and hemifusion are two intermediate processes with distinct morphologies that take place at local free energy minima. It's possible that the energy released during conformational changes in the viral membrane protein serves as the energy source for all phases of membrane fusion. Although it has been suggested that the folding of the gp41 NHR-CHR ectodomain is an exothermic process that releases about 65 kcal/mol of enthalpy and that enthalpy is used to overcome the activation energy for stalk formation during membrane fusion[37], there is currently no experimental evidence of this energy exchange between membranes and gp41 protein folding. The fact that the already folded hairpin conformation of the gp41 ectodomain is fusogenic contradicts the theory that protein folding causes membrane fusion. The NHR-short loop-CHR (HP) portion of the gp41 ectodomain core sequence forms a stable folded hairpin conformation at pH 3.2 and other pH values i.e. pH 7.4, and the gp41 with hairpin shape induces vesicle fusion[7].

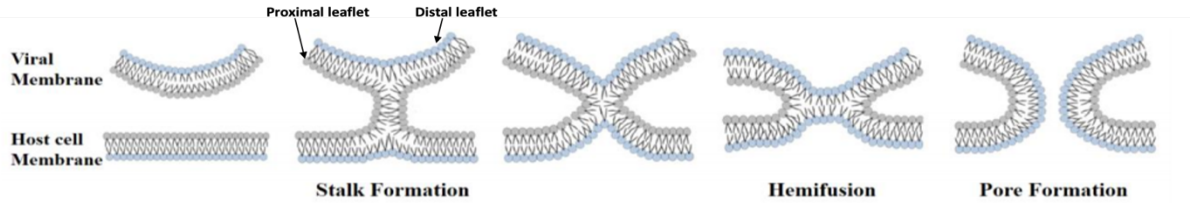


Figure 1.11: Membrane fusion stages.

In order to transmit viral genetic material into the host cell and further employ the host cell's machinery to complete the viral life cycle and replication, the membranes of HIV-1 and the host cell fuse[7]. The gp160 Env glycoprotein is essential for mediating the union of the two membranes. Gp120 and gp41, which are not covalently linked, make up gp160. Gp120 has a high affinity for CXCR4 and CCR5 co-receptors as well as CD4 receptors[56,54]. Following binding, the fusion-efficient gp41 is completely exposed (binding of recombinant soluble CD4 (sCD4) or the purified VI domain of sCD4 to the surface glycoprotein gp120 on virions resulted in rapid dissociation of gp120 from its complex with the transmembrane glycoprotein gp41) when gp120 dissociates[57]. A conformational shift (binding of Env to antibody A12 results in a conformational change that is comparable to changes observed upon its binding to the CD4-binding site antibody, b12) of gp120 still linked to gp41 was seen in a recent cryoelectron tomography analysis of virion particles with Env gp160 coupled to HIV-neutralizing proteins[58]. Another cryoelectron tomography study of conformational change of gp120 upon binding with CD4 receptor and coreceptor ligand showed gp140 (gp120+gp41 ectodomain truncated) intact structure[59]. However, this gp140 structure is stabilized by non-native disulfide bonds between gp120 and gp41 residues. There are two proposed models for the fusion of membranes caused by gp41[54].

1.1.4.1: First Model

Gp41 undergoes a conformational change (figure 1.12-1) to expose the FP toward the host membrane following the dissociation of gp120. The gp41s are a trimer bundle in this extended intermediate stage (Figure 1.12-2)[54]. Following the creation of the NHR and CHR SHB and the subsequent close apposition of the N- and C-terminal TM areas, the two membranes are subsequently brought near to one another (Figure 1.12-3)[54]. After hemifusion, complete fusion, and stalk development between the viral and host cell membranes. After fusion, the gp41 FP and TM are thus located in the same membrane (Figure 1.12-4)[54]. Fusion requires roughly 25 kcal/mol of activation energy, which is satisfied by the folding of gp41 ectodomain N-helix-loop-C-helix from an intermediate pre-hairpin state to final trimer of hairpin as SHB folded form.

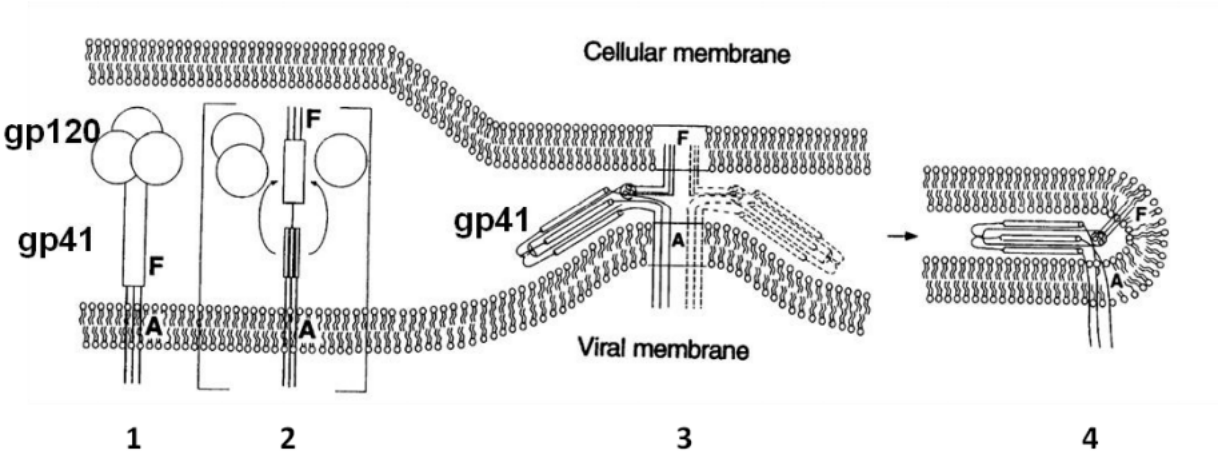


Figure 1.12: Model's schematic diagram (1) gp120 and gp41 trimers in the pre-fusion condition; (2) gp120 and gp41 trimers in the pre-fusion condition; (3) forming a SHB; (4) gp41 in the condition of the SHB after fusing after pore creation. 'A' stands for the transmembrane domain, while 'F' stands for the FP domain[54].

1.1.4.2: Second Model

This model states that the gp41 develops an extended pre-hairpin intermediate (PHI) structure after the gp120 separates from it Figure 1.13-a, and the FP inserts into the host membrane Figure 1.13-b. Following the hemifusion state, fusion pores are formed by folding PHI to SHB Figure 1.13-d[60]. Based on cell/cell fusion studies where the closing of pores after the addition of NHR or CHR peptides was observed, the experimental data indicating the pore development prior to SHB formation[61]**.** Experimental tests of the 512-581SGGRGG628-667 vesicle fusion assay at physiological pH reveal negligible fusion[37,62]. The fusion function of the constructs varied greatly, with FP-Hairpin inducing minimal fusion, FP34 (gp_{41}1-34) causing moderate fusion, and N70 (gp_{41}1-70) causing rapid fusion. The findings support the acceleration of gp41-induced membrane fusion by early PHI conformation and fusion arrest after folding to the final six-helix bundle structure[62]. These models are for vesicle fusion assays with and without the FP present in **Figure 1.3.** These results validate the assembly of membrane-associated gp41 trimers through N-terminal FP interleaving among different trimers. The 70 N-terminal residues of gp41 (N70), a construct that represents a portion of the putative pre-hairpin

intermediate state of gp41, and FP-Hairpin both have similar SSNMR results. Thus, FP assembly might be occurred before fusion[60,62].

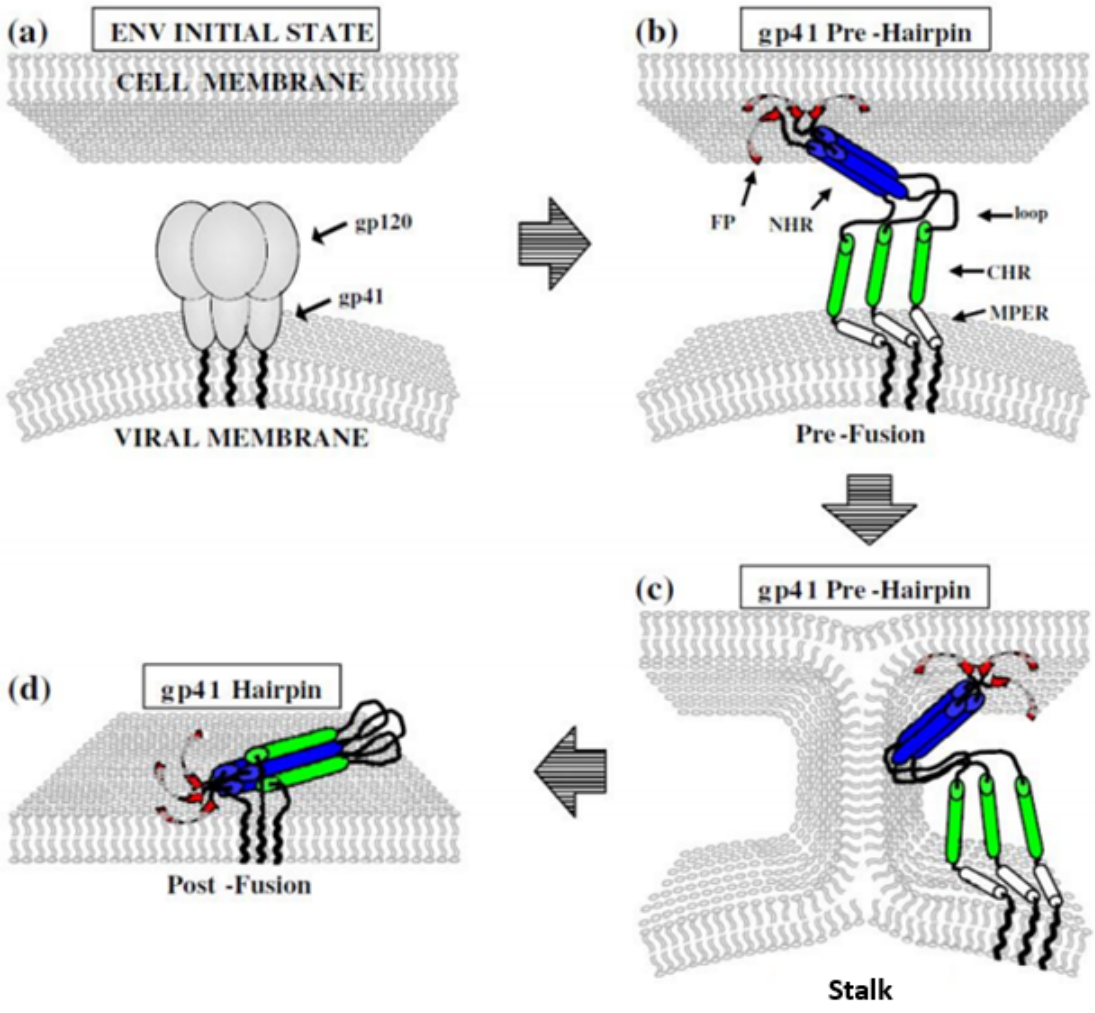


Figure 1.13: Model 2 schematic diagram[60]. (a) Pre-fusion trimer gp120 and gp41; (b) FP insertion into the host cell membrane and PHI production; (c) development of viral and host cell membrane stalks; (d) gp41 in its post-fusion form at the SHB. NHR is displayed in blue, CHR in green, MPER in white, and FP in red.

1.1.4.3: Limitations of Model I & II

The experimental data of fusion inhibition and syncytia formation caused by the binding of NHR or CHR peptides, or CHR+MPER peptide are inconsistent with the sequence of protein folding stages in model I (a temperature jump technique, to determine whether folding of Env into a bundle is complete by the time fusion pores have formed)[61,63]. Model I indicates that the creation of six helix bundle comes first, followed by stalk formation. Model II demonstrates that the development of the stalk occurs before the creation of the SHB and fusion inhibited gp41 is prevented from forming SHB. The PHI has been assumed to be a trimer structure in both models I and II, with NHR serving as the core. No structural evidence exists for gp41's PHI state. However, studying the PHI structure of a related protein is instructive. Recombinant parainfluenza virus 5 (PIV5)'s prefusion stage is characterized by HR2 forming a trimeric coiled coil stalk next to the viral membrane and HR1 lying along the protein surface of coiled coil[64]. While HR2 is thought to remain largely intact and attached close to the virus surface, HR1 is thought to separate from the protein surface and refold into a long, extended three-strand coiled coil projecting towards the target bilayer[54]. The original HR2 bundle separates during the development of the hairpin, and each of its constituent chains binds along the outside of the HR1 trimer, dragging the viral membrane toward the target bilayer until the post-fusion trimer of helical hairpins is entirely formed. This phenomenon was experimentally supported by high thermostability (Tm~110^0C) of hairpin which release required energy of membrane apposition. However, HIV-1 gp41's loop domain (areas 581-628) is typically irrelevant in both the pre- and post-fusion structures because it does not play a central role in the structural rearrangements or functional activities required for viral entry [54]. The loop was truncated and an SGGRGG linker was used in its stead for our purposes. It appears that during these conformational rearrangements, the domain forms a hinged hub around which HR1 and HR2 pivot. The PHI might be a monomeric form of ectodomain gp41. If so, the inhibitors would prevent the SHB from forming by preventing the synthesis of trimers. The gp140 (gp120+gp41; without FP and MPER) structure, as shown by recent x-ray crystal and cryo-EM studies, has loosely packed NHR domains (Figure 1.2). Presumably, this was the state prior to membrane fusion[17].

1.1.4.4: Third Model

Figure 1.14 (1-7) shows several steps of third fusion model as follows (1) gp41 exposure due to gp120 binding with receptor and co-receptor; (2) dissociation of the gp41 ectodomain into monomers, the formation of an extended PHI, and the insertion of FP into the host cell membrane; (3) PHI → hairpin monomer folding that results in the two membranes getting closer to one another; (4) hemifusion; (5) initial pore formation; (6) hairpin monomer → SHB trimer → hexamer ectodomain assembly; and (7) fusion pore expansion[7]. The PHI → SHB transition initiates the subsequent hemifusion and fusion pore development, as seen in Model 3. Hyper-thermostable ectodomain monomer additionally assisted the PHI → SHB monomer folding. Furthermore, the stable hexamer is in line with the hypothesis that HIV infection and membrane fusion need many gp160 trimers[7].

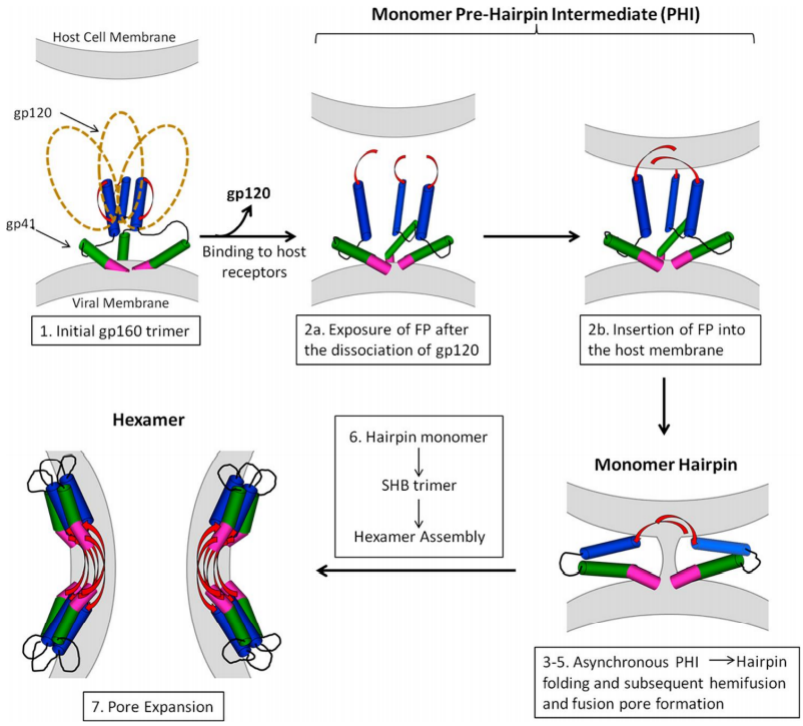


Figure 1.14: Fusion of the membrane Model 3 gp41 ectodomain monomer and hexamer[54]. The TM and endo domains are not depicted, and the other gp41 domains are color-coded similarly to Figure 1.3. In stages 3–5, one of the monomers is not shown. High-resolution structures serve as the foundation for both the original gp41 structure from step 1 and the final SHB structure from step 7[7].

1.1.5: Fusion proteins at the site of membrane fusion: Stoichiometry

Earlier studies and recent fusion model showed that more than one gp160 trimer might be needed for HIV fusion. Mathematical estimation of the number of trimers per virion by Klasse et. al. (2007) and Magnus et. al. (2012) showed that there are 5-8 gp41 trimers are required for HIV viral fusion[65,66]. Cryo-electron tomography research with virions indicates that Env trimers form a claw-like shape when HIV-1 particles engage with CD4+ cells. The "entry claw" of 5-7 gp160 trimers binds to host cells[67]. The FP of gp41 proteins prefer to arrange themselves antiparallel to one another after insertion into lipid bilayers. There is an antiparallel sheet arrangement in the membrane associated FP in the gp41 ectodomain FP-HP, according to recent solid state NMR investigations. This antiparallel arrangement is possible by interleaved FP strands from two gp41 trimers[60].

A proposed gp160-mediated fusion mechanism is also supported if it offers logical explanations of the effects of gp160 mutations on fusion, with special emphasis on mutations that significantly impair fusion but don't affect the initial assembly of gp160 spikes in the HIV membrane. At the N-terminus of the Fp of the gp41 subunit of gp160, V2E is this kind of point mutation (Fig. 1.3). V2E also prevents cell infection with HIV and the fusion of cells expressing CD4 and chemokine receptors with cells expressing gp160[68]. To our knowledge, V2E is the only gp41 mutation that exhibits dominance in mixed trimers of wild-type (WT) and V2E gp160, with f_{V2E} ≡ fraction V2E and $f_{V2E} = 1 - f_{WT}$. As an illustration, compared to $f_{V2E} = 0$, $f_{V2E} = 0.09$ only causes 40% of cell-cell fusion and HIV infection. The number of WT trimers (T) necessary for effective fusion and infection has been determined by mathematical modeling of the dependences of the reductions in fusion and infection on f_{V2E}. The study has been performed by many research groups using the same or comparable data, and the derived values of T largely depend on the model assumptions that were employed in the analysis. The variations in T between the various analyses range from 1 to 19 trimers[65-70]. The large range for T makes it more difficult to distinguish between different possible mechanisms of gp160-mediated fusion. The normal virion only contains about 15 trimers; hence it is possible to predict that the fusion rate and viral infectivity will be inversely associated with T[6,70]. For instance, fusion might be individually catalyzed by each trimer if T = 1, but when T = 15, fusion requires coordinated activity from all of the virion's trimers. On the other hand,

fusion might proceed more quickly when T is higher due to the accumulation of several gp160-related decreases in the energy barriers between fusion intermediates. T might potentially represent the approximate number of trimers undergoing fusion in the specific area of the viral membrane. This theory is consistent with images of virions bound to target cells before fusion captured using electron microscopy. A cluster of gp160 trimer ectodomains linked to cellular receptors has been theorized to be the cause of the inter-membrane electron density in the micrographs[67,70].

1.2: Solid state NMR for membrane protein

The proteins that interact with biological membranes are known as membrane proteins and serve important roles as membrane receptors and enzymes. However, due to the difficulty of growing membrane protein crystals, only a small fraction of membrane proteins (typically in the range of 1-5 mg for 1.6 mm rotor) has high resolution crystal structures[71]. Making enough folded protein for crystallographic tests is a contributing factor in crystallography's difficulties. Larger proteins (>100kDa) typically perform better when studied with cryo-electron microscopy (Cryo-EM) for membrane protein structure. Nuclear magnetic resonance (NMR) in solid states is a potent method for figuring out the structure and function of biomolecules at high resolution. Solid-state NMR is more appropriate for membrane proteins, big proteins, protein aggregates, and nucleic acids that cannot crystallize or are too massive for solution NMR spectroscopy than the more common X-ray crystallography and liquid-state NMR[72]. In solution NMR, a protein's molecular weight is typically limited to about 35 kDa.

1.2.1: Magic angle spinning (MAS)

Effects of anisotropies i.e. chemical shift anisotropy (CSA), dipolar coupling, and quadrupolar are rarely seen in solution NMR spectra. This is because the molecules in a solution rapidly tumble as they average the molecular orientation dependence, resulting in the sharp, narrow peaks that are seen. As opposed to solid state NMR, which lacks this molecular tumbling, broad NMR peaks result from anisotropic interactions rather than being averaged out. MAS was developed to get high-resolution solid-state NMR spectra[73]**.** Because all anisotropic interactions are averaged out under MAS, rotating MAS produces the same outcome as solution for solids. Consequently, only the isotropic chemical shift is observed. However, the rate of MAS must be higher than the anisotropy

of the interaction that is being averaged out in order for MAS to reduce the powder pattern to a single isotropic shift. In addition to the isotropic line, slower spinning results in a collection of spinning sidebands. The isotropic line radiates out into sharp lines called spinning sidebands that are spaced apart by a spinning rate (see picture 1.16). The geometry of MAS is shown in Fig 1.15; the angle between the rotor axis or the spinning axis and the external magnetic field B_0, α, is the magic angle and is equal to 54.7°. Figure 1.8 shows several angles; θ is the angle between the ^{13}C- ^{15}N internuclear vector and B_0, α is the angle between the rotor axis and B_0 and β is the angle between the rotor axis and the ^{13}C – ^{15}N internuclear vector. The angle θ changes over time as the molecule spins with the sample when the sample is spinning at MAS (α = 54.7°). Therefore, each rotor period's average value (<) of $3\cos^2(t)-1$ becomes:

$$<3\cos^2\theta(t)-1>= \tfrac{1}{2}(3\cos^2\alpha-1)(3\cos^2\beta-1)= 0 \qquad \ldots\ldots\ldots\ldots\ldots\ldots\ 1.1$$

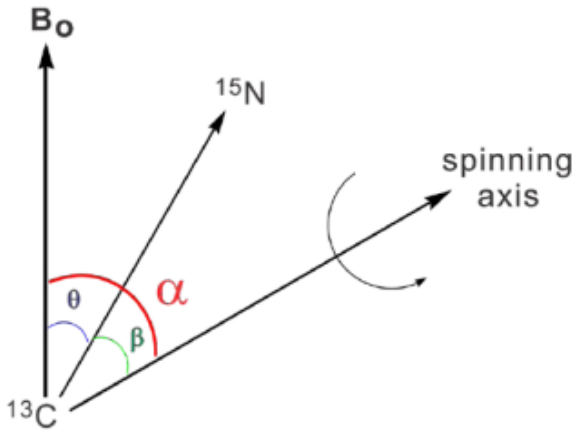


Figure 1.15: The ^{13}C - ^{15}N vector's shape is depicted schematically in the solid-state NMR sample under MAS. A cylindrical rotor rapidly rotates the sample around an axis that is positioned at the magic angle (α= 54.7°) with respect to B_0.

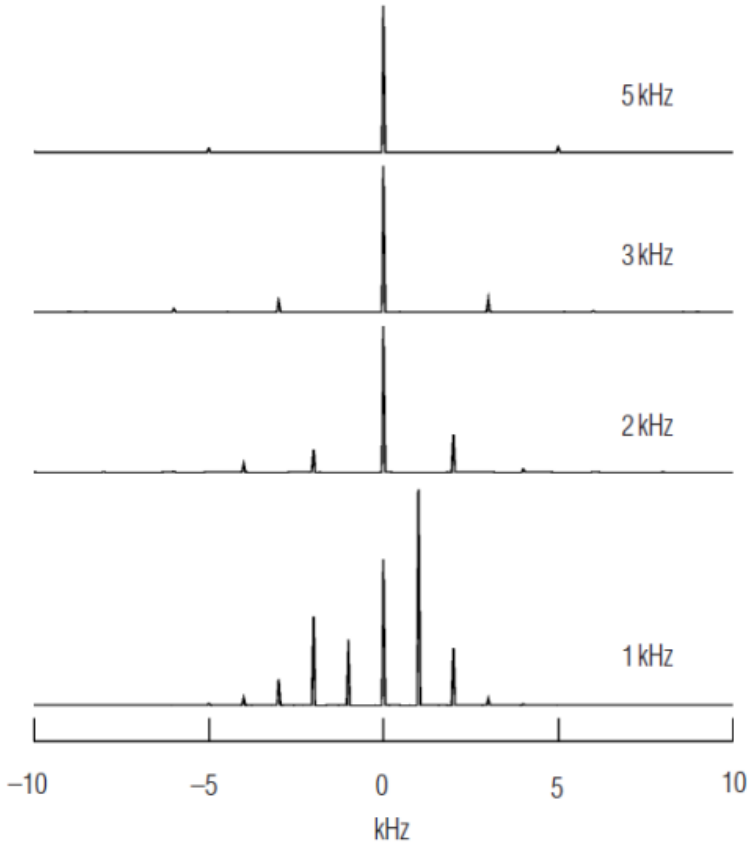


Figure 1.16: The results of MAS's slow rate of powder sample of a single isotropic shift. The isotropic shift causes a set of spinning sidebands to form. At the spinning frequency, the spinning sidebands are spaced[74].

Where figure 1.15 gives a description of the angles θ, α, and β. This technique averages the anisotropy of the interactions (e.g., CSA, heteronuclear dipolar coupling, etc.) that result in a change in the Zeeman energy but do not mix the Zeeman states. But it affects secular interactions, which mixes the Zeeman functions i.e., homonuclear dipolar coupling.

1.2.2: Cross polarization (CP)

Cross polarization is typically employed to facilitate the observation of weak spins, such as ^{13}C. Low sensitivity and long relaxation time, T_1, are the two main disadvantages of monitoring dilute or rare spins. Transferring polarization from the numerous nuclei, often ^{1}H, to the uncommon nuclei, such as ^{13}C, is the most used technique for boosting

sensitivity and shortening experiment duration. The pulse sequence is depicted in figure 1.17, and the transfer is known as cross-polarization (CP).

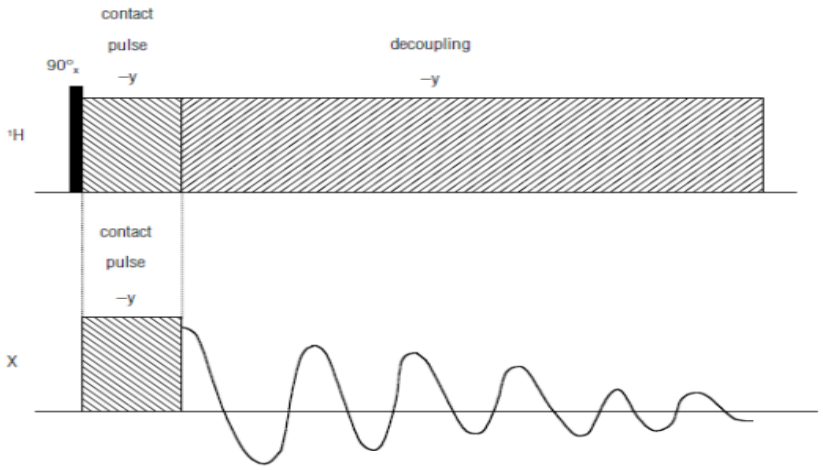


Figure 1.17: Sequence of CP pulse. The heteronuclear dipolar coupling between the ^{1}H and X spins causes the CP pulse sequence to transfer magnetization from the abundant spins (^{1}H) to the rare spins, X (e.g. ^{13}C), as a result.

The CP transfer takes place in the doubly rotating frame, which is composed of the static ^{1}H B_1 field and the static X spin $B_1(X)$ field. Applying a ^{1}H 90° pulse to rotate the 1H magnetization along the -y axis is the first step in CP. To spinlock the ^{1}H magnetization along the -y axis, the ^{1}H contact pulse is then applied along this direction. $B_1(^{1}H)$ represents the spinlock field for the given ^{1}H. At this time, the magnetic field's $B_1(X)$ amplitude and the RF in the X channel are both turned on. The two contact pulses' amplitudes have now been set in the CP experiment to achieve the Hartmann-Hahn matching condition[75,76].

$$\gamma_H B_1 (^{1}H) = \gamma_X B_1 (X) \quad \text{.....(1.2)}$$

1.2.3: Introduction to rotational echo double resonance solid state NMR (REDOR ssNMR)

In order to recouple the heteronuclear dipolar couplings under MAS and determine the internuclear distances in the spin system, REDOR has been widely employed in solid-state NMR. The inventors of this method are Gullion and Schafer[77]. To recouple the heteronuclear dipolar coupling under MAS, REDOR applies a sequence of rotor-synchronized π-pulses to the S-channel. The inverse relationship between r^3 and the heteronuclear dipolar coupling strength means that the REDOR technique is extremely

sensitive to the distance between the connected spins. The ^{13}C-to-^{2}H REDOR pulse sequence is shown in Figure 1.18. A cross-polarization pulse sequence is used at the start of the REDOR pulse sequence to transfer the ^{1}H magnetization to the ^{13}C nucleus. The Hartmann-Hahn condition, which is given by equation 1.2, must be satisfied in order to obtain an efficient cp, as was mentioned in the section before this one. In the absence of a resonance offset, the aforementioned condition is valid. However, a resonance offset field, $B_{r.o}$, exists in actual experiments. The effective magnetic field, B_{eff} at the Hartmann-Hahn condition is reduced in the presence of $B_{r.o}$ to:

$$\gamma_H B_{eff}(^{1}H) = \gamma_C B_{eff}(^{13}C) \quad \text{.....(1.3)}$$

Where, $\boldsymbol{B_{eff} = \sqrt{(B_1 + B_{r.o.})^2}}$**.** The Hartmann-Hahn matching conditions for a static sample are described by the equations 1.3. However, when MAS is present, the aforementioned equation is altered because MAS has an impact on the dipolar coupling of $^{1}H \rightarrow ^{13}C$. For instance, in a peptide, the distance between the ^{13}CO backbone label and ^{1}H is 2 Å. The peptide sample's maximum dipolar coupling is therefore 4 kHz or less. Since the MAS speed in REDOR studies is 9 kHz, the ^{13}C - ^{1}H dipolar coupling should average out at 4 kHz. The ^{1}H - ^{1}H dipolar coupling does not average out under 9 kHz MAS, despite the fact that the ^{13}C - ^{1}H dipolar coupling is meant to. This is due to the ^{1}Hs' network-wide dipolar coupling in peptides, which causes a rapid flip-flop ($\alpha \leftrightarrow \beta$ transition) between the ^{1}Hs via ^{1}H - ^{1}H homonuclear dipolar coupling. The strength of the ^{1}H - ^{1}H dipolar coupling (typically in the range of 10 – 50 kHz) is slower relative to the rate of $\alpha \leftrightarrow \beta$ transition on the ^{1}H spin[78]. The ^{1}H flip-flop disrupts the average of the ^{1}H - ^{13}C dipolar coupling throughout each rotor period under 9 kHz MAS, leading to effective polarization transfer. Consequently, the MAS match condition is provided by[79]:

$$\gamma_H B_{eff}(^{1}H) = \gamma_C B_{eff}(^{13}C) + n\omega_R \quad \text{.....(1.4)}$$

Where n = 0, ±1, ±2 and denotes the nth spinning sideband and ω_R = MAS frequency. In the dipolar coupling, MAS creates time dependency. Hence, $\hat{H}_{HC}$ is no longer constant under MAS condition and oscillates between $\pm\omega_R$, $\pm 2\omega_R$ because $\hat{H}_{HC} \propto [3\cos^2\theta(t) - 1]$ and θ becomes time-dependent (θ(t) varies periodically with the spinning frequency ω_R of the rotor)

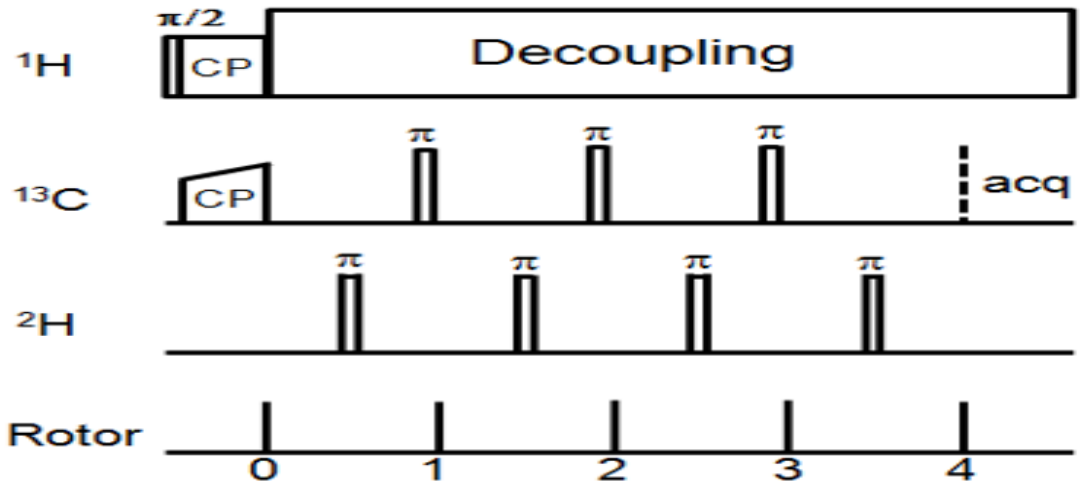


Figure 1.18: Pulse sequence of REDOR NMR. The measured spin in this instance is ^{13}C, and the dephased spin is ^{2}H.

The ^{13}C spinlock field is subjected to a ramp in the REDOR experiment because there is a distribution of Larmor frequency and the resonance offset field, Br.o. Consequently, in a powder sample, molecules with various orientations exhibit various cross-polarization efficiencies. As a result, the efficiency of the magnetization transfer from ^{1}H to ^{13}C under MAS is increased by using a ramping CP.

Following $^{1}H \rightarrow ^{13}C$ CP, REDOR is carried out in two stages: S_1 with rotor-synchronized dephasing pulses and S_0 without them. The heteronuclear dipolar coupling Hamiltonian, as previously discussed, is given by:

$$\hat{H}_{hetero} = \frac{\mu_0 h \gamma_C \gamma_D}{16\pi^2 r^3}(3cos^2\theta - 1)\left(2\hat{C}_z\hat{D}_z\right) \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (1.5)$$

Where C and D, respectively, stand for ^{13}C and ^{2}H. There are three contributors for this dipolar coupling: space part, C spin and N spin. MAS affects the space component, whereas the application of π-pulses modifies the spin component. The π-pulse's purpose is to flip the spins by 180 degrees (for instance, x→-x). This results in the reversal of the sense of the rotation of the observed spins by altering the sign of the dipolar coupling for the observed spins linked to the dephasing spins. The ^{13}C – ^{2}H dipolar coupling in the S_0 experiment is averaged over each rotor period. Furthermore, MAS additionally averages out the ^{13}C CSA. The ^{13}C isotropic chemical shift has been refocused by rotor synchronized π-pulses on ^{13}C channel. Acquisition occurs right after the rotor cycle is

finished. Figure 1.19 shows how MAS averages out the dipolar interactions. The S_1 experiment applies π -pulses in the midst of each ^{2}H-channel rotor period. Figure 1.20 shows how the ^{13}C – ^{2}H dipolar coupling under MAS is recoupled by ^{2}H -pulses. As an oscillating space component, the sign change of the Hamiltonian in the later half of each rotor period is cancelled by the ^{2}H π-pulses at the middle of each rotor period. To make sure that the dipolar coupling is accumulated from one cycle to the next, the π-pulses are placed in the center of each rotor period. The detected echo intensity is decreased because of the impact of the accumulated dipolar coupling. For the S_1 experiment, the density operator is given by:

$$\begin{aligned}\hat{\rho}(t) &= \exp\{-i\tilde{\omega}_{CN} 2C_z N_z t\} C_x \exp\{i\tilde{\omega}_{CN} 2C_z N_z t\} \\ &= C_x \cos(\tilde{\omega}_{CN} t) + 2C_y \mathbf{N}_z \sin(\tilde{\omega}_{CN} t)\end{aligned} \quad \text{..... (1.6)}$$

Where $\tilde{\omega}_{CN}(t) \propto \frac{3\cos^2\theta(t)-1}{r^3}$ is the average dipolar coupling frequency over each rotor period. $C_y^2 = 1$ and tr (Nz) =0 and $C_y$tr (Nz) = 0. There is a distribution of $\tilde{\omega}_{CN}$ due to the distribution of the θ in the powder sample and consequently, there is a decay in the ^{13}C transverse magnetization as a function of dephasing time, τ. Therefore, S_1's integrated ^{13}C signal intensity is lower than S_0's. For the ^{13}C spin for S_0 and S_1 experiments, REDOR calculates the difference in signal intensity as:

$$\frac{\Delta S}{S_0} = \frac{(S_0 - S_1)}{S_0} \quad \text{..... ..... (1.7)}$$

$\hat{H}_{HC}^{D}$= Space part x C spin part

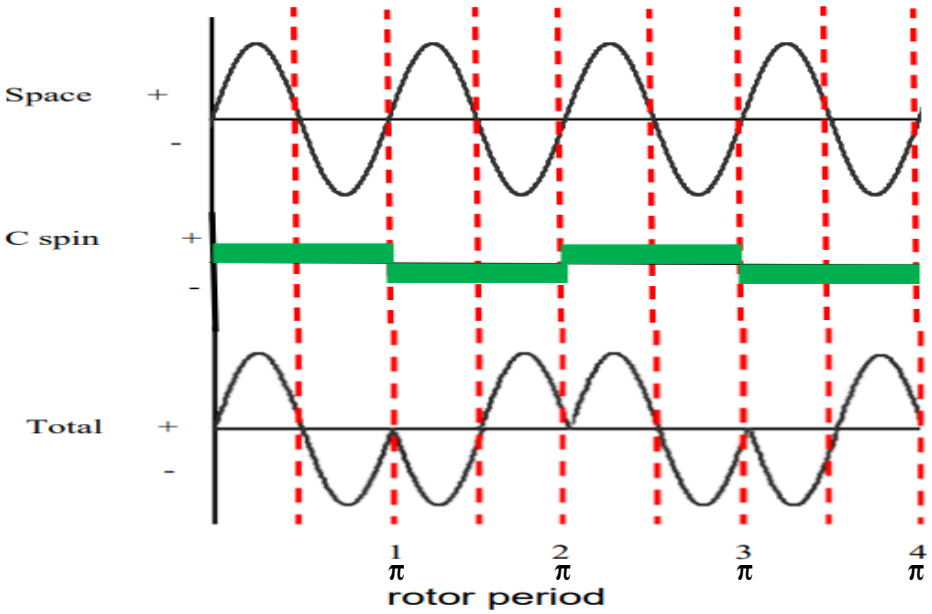


Figure 1.19: Dipolar coupling evolution in the S_0 experiment as a function of rotor period. Rotor-synchronized ^{13}C π-pulses had no effect on the heteronuclear dipolar interaction's MAS averaging.

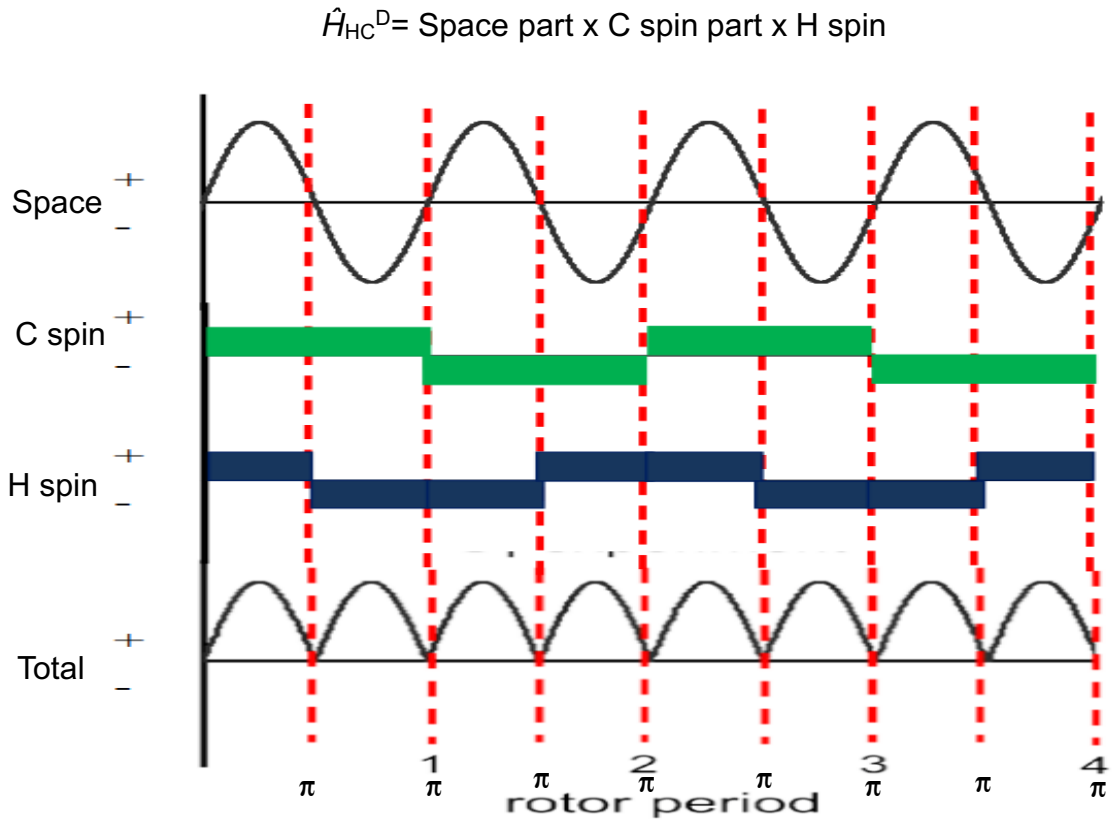


Figure 1.20: Dipolar coupling evolution in the S_1 experiment as a function of rotor period. The heteronuclear dipolar coupling is prevented from MAS averaging by rotor synchronized ^{2}H π-pulses.

To obtain the dipolar coupling, the ($\Delta S/S_0$) buildup at different τ can be fitted using the SIMPSON program or the analytical solution of the dipolar dephasing[80,81]. The internuclear distance can be determined using the formula: d_{CN} (Hz) = 3066 / r^3 (Å) (The $d \propto r^{-3}$ means that the buildup is typically dominated by the closest ^{15}N), once the ^{13}C - ^{15}N dipolar coupling has been determined.

There is a so-called universal dephasing curve of $\Delta S/S_o$ versus λ for REDOR between a spin 1/2 and spin 1/2 pair, such as ^{13}C and ^{15}N, where λ is the product of the dipolar coupling d and dephasing time τ (Figure 1.21a)[82]. According to SIMPSON simulations, there is also a universal dephasing curve of $\Delta S/S_o$ versus λ for REDOR between a spin 1/2 and spin 1 pair, such as ^{13}C and ^{2}H (Figure 1.21b).

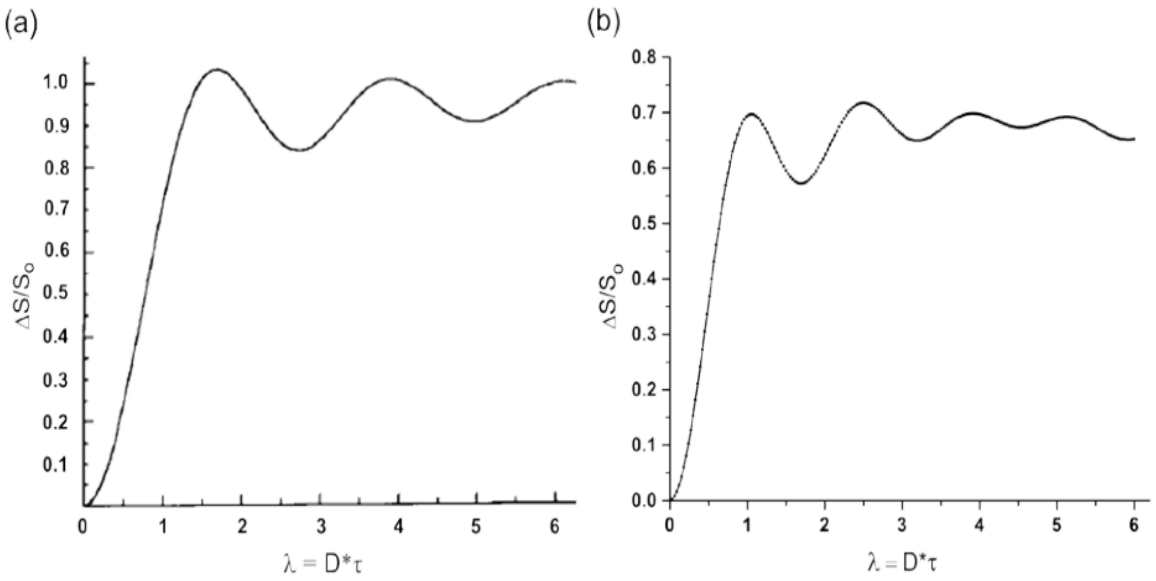


Figure 1.21: For a spin 1/2 - spin 1/2 pair (panel a) and a spin 1/2 - spin 1 pair (panel b), $^{13}CO-^{2}H$ (m = ±1) or $^{13}CO-^{2}H$ (m = 0) over the whole dephasing time in the absence of ^{2}H relaxation. A $^{13}CO-^{2}H$ (m = ±1) pair's dipolar development is fully refocused by MAS during each rotor cycle of the S_0 acquisition, allowing the pair to contribute fully to the ^{13}CO S_0 signal[83]. During a rotor cycle of the S_1 acquisition, dipolar evolution is not refocused, and the pair contributes to the S_1 signal attenuated. There is no evolution for either the S_0 or S_1 acquisition and full and equal contributions to the S_0 and S_1 signals when a $^{13}CO-^{2}H$ (m = 0) pair does not experience coupling. Since only $^{13}CO-^{2}H$ (m = ±1) pairings can cause the $\Delta S/S_0$ accumulation in the absence of relaxation, the resulting $(\Delta S/S_0)^{max}$ is anticipated to be ~(2/3) (Figure 1.22)[83]. This is seen in the SIMPSON calculation (Figure 1.22), which excludes relaxation[83].

The S_1 signal is more affected by relaxation. In certain time segments of the dephasing period, a spin pair is $^{13}CO-^{2}H$ (m = ±1), and in other segments, it is $^{13}CO-^{2}H$ (m = 0). In the $^{13}CO-^{2}H$ (m = ±1) segments, there is dipolar evolution and a corresponding attenuated contribution to the S_1 signal; in the $^{13}CO-^{2}H$ (m = 0) segments, there is no evolution and a corresponding complete contribution to the S_1 signal. Since the m = ±1 ↔ m = 0 transitions are stochastic, as was already mentioned, there is no correlation between the temporal segments of the ^{2}H nuclei. Qualitatively, each pair spends approximately one-third as $^{13}CO-^{2}H$ (m = 0) and two-thirds of the dephasing period as $^{13}CO-^{2}H$ (m = ±1). Therefore, for $^{13}CO-^{2}H$ (m = ±1) pairings, it is anticipated that the

building rate of $\Delta S/S_0$ will be ~(2/3) the dipolar coupling. This agrees with the fitting of SIMPSON computations to the l4 peptide data. Additionally, relaxation influences $(\Delta S/S_0)^{max}$. For various segments of the dephasing period, the majority of pairs are $^{13}CO-^{2}H$ (m = ±1), which means they add to the $\Delta S/S_0$ accumulation. The $\Delta S/S_0$ for large τ values is therefore expected to be >(2/3)[83]. Therefore, it is expected that the ΔS/S0 for large τ values will be >(2/3), which is consistent with the ΔS/S0 shown in Figure 1.22.

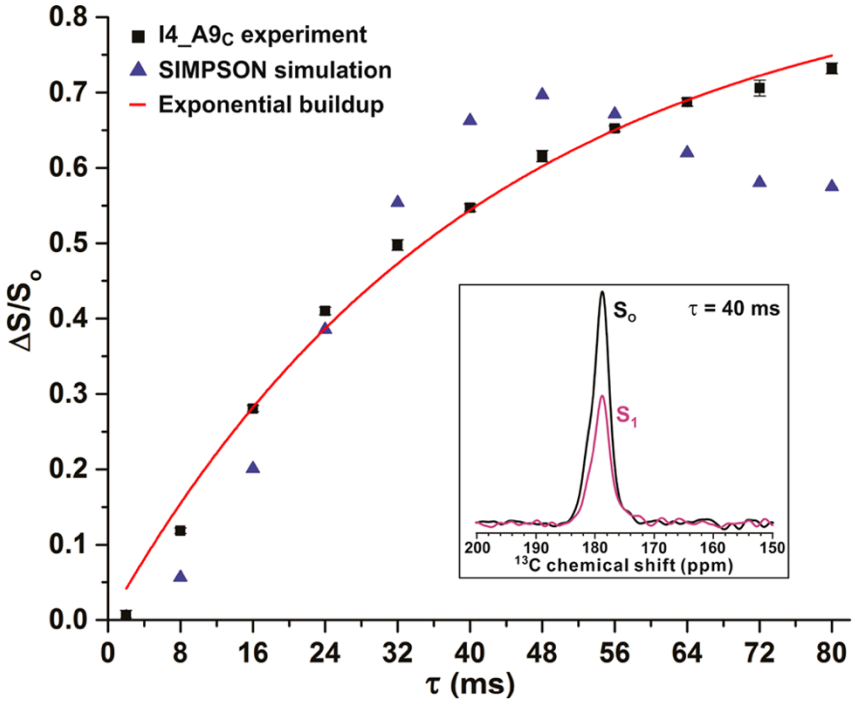


Figure 1.22: $^{13}CO-^{2}H$ REDOR spectra showing $\Delta S/S_0$ vs τ and the I4_A9C peptide with a 40 ms dephasing time. The designated A9 ^{13}CO nuclei's $(\Delta S/S_0)^{lab}$ values, represented by black squares, are numerically rather near to the experimental $(\Delta S/S_0)^{exp}$. Table S2 of the Supporting Information shows that the difference between the $(\Delta S/S_0)^{lab}$ and $(\Delta S/S_0)^{exp}$ values is ≤0.06 for all values of τ. All $(\Delta S/S_0)^{lab}$ data have uncertainties displayed; however, in many cases, the uncertainties are smaller than or equal to the black squares that are depicted. The SIMPSON algorithm is used to calculate $(\Delta S/ S_0)^{sim}$ without ^{2}H relaxation for the blue triangles. The $(\Delta S/S_0)^{sim}$ values with a 22 Hz~$^{13}CO-^{2}H$ dipolar coupling provide the best fit to the $(\Delta S/S0)^{lab}$. The best-fit exponential increase is shown by the red line. A distinct $^{2}H\alpha$ label at A8 and a ^{13}CO label at A9 are used during the peptide's synthesis. The sample has isolated $^{13}CO-^{2}H$ spin pairs with a corresponding dipolar coupling (d) of 37 Hz and an internuclear separation (r) of 5.0 Å, all within a good approximation. About 15 μmol of lyophilized peptide are present in the sample. For τ values between 2 and 48 ms, there are 500 summed S_0 or S_1 scans, and for τ values between 56 and 80 ms, there are 2000[83].

REFERENCES

1. https://www.unaids.org/en
2. Briggs, J. A. G., Simon, M. N., Gross, I., Krausslich, H. G., Fuller, S. D., Vogt, V. M., and Johnson, M. C. (2004) The stoichiometry of Gag protein in HIV-1, Nat. Struct. Mol. Biol. 11, 672-675.
3. Levy, J. A. (1993) PATHOGENESIS OF HUMAN-IMMUNODEFICIENCY-VIRUS INFECTION, Microbiol. Rev. 57, 183-289.
4. McCune, J. M., Rabin, L. B., Feinberg, M. B., Lieberman, M., Kosek, J. C., Reyes, G. R., and Weissman, I. L. (1988) ENDOPROTEOLYTIC CLEAVAGE OF GP160 IS REQUIRED FOR THE ACTIVATION OF HUMAN IMMUNODEFICIENCY VIRUS, Cell 53, 55-67.
5. 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.
6. Zhu, P., Liu, J., Bess, J. et al. Distribution and three-dimensional structure of AIDS virus envelope spikes. Nature 441, 847–852 (2006).
7. Banerjee, K., and Weliky, D. P. (2014) Folded Monomers and Hexamers of the Ectodomain of the HIV gp41 Membrane Fusion Protein: Potential Roles in Fusion and Synergy Between the Fusion Peptide, Hairpin, and Membrane-Proximal External Region, Biochemistry 53, 7184-7198.
8. Geetha Bansalet al. (2008). HIV Sequence Compendium 2008. Retrieved March 31, 2009.
9. Gelderblom, H. R., Ozel, M., Hausmann, E. H. S., Winkel, T., Pauli, G., and Koch, M. A. (1988) FINE-STRUCTURE OF HUMAN IMMUNODEFICIENCY VIRUS (HIV), IMMUNOLOCALIZATION OF STRUCTURAL PROTEINS AND VIRUS-CELL RELATION, Micron and Microscopica Acta 19, 41-60.
10. Gelderblom, H. R., Ozel, M., and Pauli, G. (1989) MORPHOGENESIS AND MORPHOLOGY OF HIV - STRUCTURE-FUNCTION RELATIONS, Arch. Virol. 106, 1-13.
11. Laskey, S. B., and Siliciano, R. F. (2014) A mechanistic theory to explain the efficacy of antiretroviral therapy, Nat. Rev. Microbiol. 12, 772-780.
12. Gottlieb, M. S., Schroff, R., Schanker, H. M., Weisman, J. D., Fan, P. T., Wolf, R. A., and Saxon, A. (1981) PNEUMOCYSTIS-CARINII PNEUMONIA AND MUCOSAL CANDIDIASIS IN PREVIOUSLY HEALTHY HOMOSEXUAL MEN - EVIDENCE OF A NEW ACQUIRED CELLULAR IMMUNODEFICIENCY, N. Engl. J. Med. 305, 1425-1431.
13. Dalgleish, A. G., Beverley, P. C. L., Clapham, P. R., Crawford, D. H., Greaves, M. F., and Weiss, R. A. (1984) THE CD4 (T4) ANTIGEN IS AN ESSENTIAL COMPONENT OF THE RECEPTOR FOR THE AIDS RETROVIRUS, Nature 312, 763-767.

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

15. Broder, C. C., and Dimitrov, D. S. (1996) HIV and the 7-transmembrane domain receptors, Pathobiology 64, 171-179.

16. 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.

17. 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.

18. Grewe, C., Beck, A., and Gelderblom, H. R. (1990) HIV - EARLY VIRUS-CELL INTERACTIONS, J. Acquir. Immune Defic. Syndr. Hum. Retrovirol. 3, 965-974.

19. 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.

20. Frankel, A. D., and Young, J. A. T. (1998) HIV-1: Fifteen proteins and an RNA, Annu. Rev. Biochem. 67, 1-25.

21. S. Liang, P. U. Ratnayake, C. Keinath, L. Jia, R. Wolfe, A. Ranaweera, and D. P. Weliky Efficient Fusion at Neutral pH by Human Immunodeficiency Virus gp41 Trimers Containing the Fusion Peptide and Transmembrane Domains, Biochemistry 2018, 57, 1219-1235.

22. Chan, D. C., and Kim, P. S. (1998) HIV entry and its inhibition, Cell 93, 681-684.

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. 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.

25. Tran, E. E. H., Borgnia, M. J., Kuybeda, O., Schauder, D. M., Bartesaghi, A., Frank, G. A., Sapiro, G., Milne, J. L. S., and Subramaniam, S. (2012) Structural Mechanism of Trimeric HIV-1 Envelope Glycoprotein Activation, PLoS Pathog. 8, 18.

26. Nils-Alexander Lakomek, Joshua D. Kaufman, Stephen J. Stahl, and Paul T. Wingfield. HIV-1 Envelope Protein gp41: An NMR Study of Dodecyl Phosphocholine Embedded gp41 Reveals a Dynamic Prefusion Intermediate Conformation. Structure 22, 1311–1321, September 2, 2014.

27. Jaroniec, 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.

28. 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.

29. 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.

30. 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.

31. 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.

32. Rafalski, M., Lear, J. D., and Degrado, W. F. (1990) Phospholipid Interactions of Synthetic Peptides Representing the N-Terminus of HIV Gp41. Biochemistry 29, 7917-792229.

33. Jia, L. H., Liang, S., Sackett, K., Xie, L., Ghosh, U., and Weliky, D. P. (2015) REDOR solid-state NMR as a probe of the membrane locations of membrane-associated peptides and proteins, J. Magn. Reson. 253, 154-165.

34. Qiang, W., Sun, Y., and Weliky, D. P. (2009) A strong correlation between fusogenicity and membrane insertion depth of the HIV fusion peptide, Proc. Natl. Acad. Sci. U. S. A. 106, 15314-15319.

35. 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

36. 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.

37. 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.

38. Lakomek NA; Kaufman JD; Stahl SJ; Louis JM; Grishaev A; Wingfield PT; Bax A, Internal Dynamics of the Homotrimeric HIV-1 Viral Coat Protein gp41 on Multiple Time Scales. Angew. Chem. Int. Ed. Engl 2013, 52, (14), 3911–3915.

39. 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.

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

41. 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.

42. Gangupomu, V. K., and Abrams, C. F. (2010) All-Atom Models of the Membrane-Spanning Domain of HIV-1 gp41 from Metadynamics, Biophys. J. 99, 3438-3444.

43. 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, Biochim. Biophys. Acta-Biomembr. 1788, 1804-1812.

44. Apellaniz, B., Rujas, E., Serrano, S., Morante, K., Tsumoto, K., Caaveiro, J. M. M., Jimenez, M. A., and Nieva, J. L. (2015) The Atomic Structure of the HIV-1 gp41 Transmembrane Domain and Its Connection to the Immunogenic Membrane-proximal External Region, J. Biol. Chem. 290, 12999-13015.

45. 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

46. 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.

47. 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.

48. 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.

49. 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.

50. 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.

51. 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.

52. Alessandro Piai, and James Chou. (2021) NMR Model of the Entire Membrane-Interacting Region of the HIV-1 Fusion Protein and Its Perturbation of Membrane Morphology, J. Am. Chem. Soc. 2021, 143, 6609−6615.

53. 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.

54. 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.

55. Chernomordik, L. V., Zimmerberg, J., and Kozlov, M. M. (2006) Membranes of the world unite! J Cell Biol 175, 201-20733.

56. Sattentau, Q. J., and Weiss, R. A. (1988) The Cd4 Antigen - Physiological Ligand and Hiv Receptor. Cell 52, 631-633.

57. 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.

58. 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.

59. 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., Mooree, 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.

60. 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, J. Mol. Biol. 426, 1077-1094.

61. 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, 926–93834.

62. 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.

63. 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.

64. DeGrado, William F., Capture and imaging of a prehairpin fusion intermediate of the paramyxovirus PIV5, 20992–20997, PNAS, December 27, 2011,108, 52.

65. 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.

66. Magnus, C., and Regoes, R. R. (2012) Analysis of the Subunit Stoichiometries in Viral Entry. Plos One 7, e33441.

67. 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.

68. E.O. Freed, E.L. Delwart, G.L. Buchschacher Jr., A.T. Panganiban, 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 (1992) 70–74.

69. O.F. Brandenberg, C. Magnus, R.R. Regoes, A. Trkola, The HIV-1 entry process: a stoichiometric view, Trends Microbiol. 23 (2015) 763–774.

70. Md. Rokonujjaman, Abdulrazak Sahyouni, Robert Wolfe, Lihui Jia, Ujjayini Ghosh, and David P. Weliky. A large HIV gp41 construct with final trimer-of-hairpins structure exhibits V2E mutation-dominant attenuation of vesicle fusion and helicity that is quantitatively similar to V2E attenuation of HIV fusion and infection and supports: (1) hairpin stabilization of initial membrane apposition with larger apposition distance for V2E; and (2) V2E dominance mediated by an antiparallel β sheet with interleaved fusion peptide strands from two gp41 trimers. Biophysical Chemistry 293 (2023) 106933.

71. RCSB PDB: Homepage

72. Harper, James K., An Explanation of Magic Angle Spinning NMR Experiments in the Time Domain. Magnetic Resonance. 2009, 34:5 249-263.

73. Andrew, E.R., A. Bradbury, and R.G. Eades, NUCLEAR MAGNETIC RESONANCESPECTRA FROM A CRYSTAL ROTATED AT HIGH SPEED. Nature, 1958. 182(4650): p. 1659-1659.

74. Duer, M.J., Introduction to Solid State NMR. 2008: Blackwell.

75. Hartmann, S.R. and E.L. Hahn, NUCLEAR DOUBLE RESONANCE IN ROTATING FRAME. Physical Review, 1962. 128(5): p. 2042.

76. Pines, A., M.G. Gibby, and J.S. Waugh, PROTON-ENHANCED NMR OF DILUTE SPINS IN SOLIDS. Journal of Chemical Physics, 1973. 59(2): p. 569-590.

77. Gullion, T. and J. Schaefer, ROTATIONAL-ECHO DOUBLE-RESONANCE NMR. Journal of Magnetic Resonance, 1989. 81(1): p. 196-200.

78. Yannoni, C.S., HIGH-RESOLUTION NMR IN SOLIDS - THE CPMAS EXPERIMENT. Accounts of Chemical Research, 1982. 15(7): p. 201-208.

79. Meier, B.H., CROSS POLARIZATION UNDER FAST MAGIC ANGLE SPINNING - THERMODYNAMICAL CONSIDERATIONS. Chemical Physics Letters, 1992. 188(3-4): p. 201-207.

80. Bak, M., J.T. Rasmussen, and N.C. Nielsen, SIMPSON: A General Simulation Program for Solid-State NMR Spectroscopy. Journal of Magnetic Resonance, 2000. 147(2): p. 296-330.

81. Mueller, K.T., ANALYTIC SOLUTIONS FOR THE TIME EVOLUTION OF DIPOLARDEPHASING NMR SIGNALS. Journal of Magnetic Resonance Series A, 1995. 113(1): p. 81-93.

82. Gullion, T., Introduction to rotational-echo, double-resonance NMR. Concepts in Magnetic Resonance, 1998. 10(5): p. 277-289.

83. Li Xie, Lihui Jia, Shuang Liang, and David P. Weliky. Multiple Locations of Peptides in the Hydrocarbon Core of Gel-Phase Membranes Revealed by Peptide 13C to Lipid ^{2}H Rotational-Echo Double-Resonance Solid-State Nuclear Magnetic Resonance. Biochemistry, 2015, 54, 677−684.

CHAPTER-02: Materials and Methods.

2.1: Materials

Orders were placed with GenScript (Piscataway, NJ) for DNA plasmids containing the HM, and FP_HM genes. The Novagen (Gibbstown, NJ) BL21(DE3) strain of Escherichia coli was obtained for protein expression cells. Dot Scientific (Burton, MI) provided the Luria-Bertani broth (LB) medium, Goldbio (St. Louis, MO) provided the isopropyl -D-thiogalactopyranoside (IPTG), and Thermo Scientific (Waltham, MA) provided the cobalt affinity resin. The 1,3- ^{13}C Gly, 2- ^{13}C Gly and U_^{13}C Glu were obtained from Cambridge Isotope Laboratories (Andover, MA). Avanti Polar Lipids (Alabaster, AL) was the vendor to purchase 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine (POPC), 1-palmitoyl-2-oleoyl-sn-glycero-3- [phospho-rac-(1-glycerol)] (sodium salt) (POPG), 1,2-Dipalmitoyl-sn-glycero-3-phosphocholine (DPPG), DPPC-D10, DPPC-D8, cholesterol, and d6-cholesterol. Position of ^{2}H in DPPC-D10, DPPC-D8, DPPC-D4, d6-cholesterol, and d6-cholesterol are shown in Figure 2.1[1]. Sigma-Aldrich (St. Louis, MO) provided all other materials.

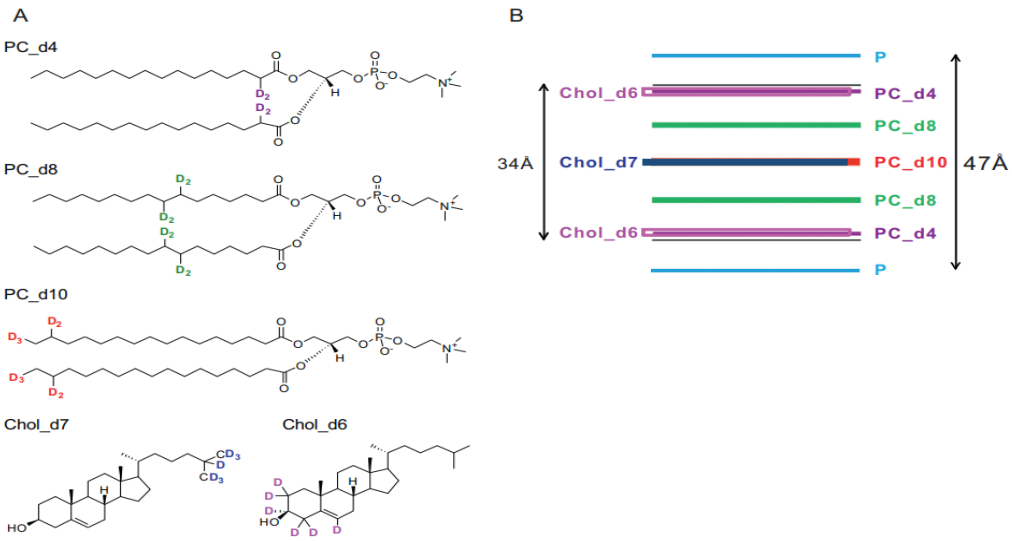


Figure 2.1: (A) Lipid and cholesterol ^{2}H patterns. (B) The ^{2}Hs and the ^{31}Ps (P) in the membrane bilayer without protein are approximately located on the membrane. The cholesterol ^{2}H locations are for the liquid-ordered phase with cholesterol and the lipid ^{2}H and ^{31}P locations are for the membrane gel-phase without cholesterol[1,2]. Subsequent figures employ the same color scheme[1].

2.2: Molecular subcloning

Inserting the specific DNA that codes for the desired protein into the expression vector is necessary to get a protein expressed in E. coli. A gene was purchased from Genscript Inc. in New Jersey. The gene is made available by Genscript in their vector, pUC57-Kan. Cut the gene (insert), which was initially in the pUC57-Kan vector, and ligate it with cut pet24a+ expression vector are the goals of the subcloning procedure being presented here (Figure 2-2). The term "ligation" refers to the connecting of two DNA fragments by the activity of an enzyme called T4 DNA Ligase, in this case the gene (insert) with the cut pet24a+ vector. The restriction sites to be used for the subcloning must be chosen before ordering the gene from the company. We asked Genscript Inc. to add XhoI restriction site to the gene sequence's 3' end and Nde1 restriction site to the gene sequence's 5' end. To avoid the presence of non-native sequences in the produced proteins, Xho1 and Nde1 restriction sites are chosen over other restriction sites.

(1) N-terminal-related restriction site:

The unnecessary T7 tag (MetAlaSerMetThrGlyGlyGlnMetGlyArg) will be present before the N-terminus of the target protein if BamHI is used in place of NdeI. As a result, the expressed protein will have a T7 tag followed by the protein sequence. A tag used in immunoaffinity purification is the T7 tag. (In immunoaffinity purification, a cell lysate is mixed with an antibody that has been covalently linked to a resin. These antibodies are capable of attaching to T7-tagged proteins.) The BamHI site can be utilized in place of the NdeI restriction site if the produced protein is to be purified by immunoaffinity purification.

(2) C-terminal-related restriction site

Depending on where the His-tag is located, the second restriction site, XhoI, is chosen at the 3' end. This restriction site was utilized because it was intended to purify the expressed protein using IMAC, which necessitates the presence of the His tag in the expressed protein (explained briefly under section 2-4(a)). After the protein is expressed, only L (Leucine) and E (Glutamate) amino acids occur before the tag as a result of the use of XhoI. The C-terminal of the protein will develop an unneeded amino acid sequence if Sal1 is used instead of Xho1. When Sal1 is used, the amino acid sequence ArgGlnAlaCysGlyArgThrArgAla will come before the tag at the protein's C-terminus

(Figure 2-4). Since only amino acids L and E are added before the tag, using Xho1 at the 3' end is favored overusing any other restriction site.

As a result, the use of XhoI and NdeI is recommended over the use of any other restriction sites because these two restriction sites prevent the insertion of lengthy non-native sequences into the N- or C-terminal of the produced protein. CATATG and CTCGAG are the nucleotide sequences that correspond to the Nde1 and Xho1, respectively.

2.2.1: The fundamental steps of subcloning

(1) Separate transformations of the gene (insert) contained in the pUC57-Kan expression vector and the empty (pet24a+) expression vector.

2) The empty expression vector (pet24a+) and the plasmid (pUC57-Kan) containing the gene of interest (insert) are extracted from the overnight-grown cell cultures. Miniprep, as this procedure is often known, was carried out using a plasmid extraction kit that is readily accessible for purchase. Plasmid extraction kit from Promega

3) Using the restriction enzymes Nde1 and Xho1, the pUC57-Kan plasmid, which contains the target gene (insert) and the vector (pet24a+), is digested.

Specific DNA sequences can be recognized by restriction enzymes, which then cut the DNA and create cohesive ends. Only the 5' and 3' ends of the gene and vector should have certain restriction sites; the rest of the gene and vector sequence should be free of these sites. If the chosen restriction sites (in this case, Xho1 and Nde1) are present in various locations across the gene sequence and vector sequence, the gene and vector will be digested into various parts.

In vectors, frequently utilized restriction sites are concentrated in a single region called a "multiple cloning site." Multiple cloning sites are found in the pet24 a+ vector between the Nde1 and Xho1 sites. That area is classified as a "multiple cloning site" due to the presence of numerous restriction sites. Users are free to use any of these restriction sites depending on their needs, as was previously described. Restriction sites that are present in several cloning sites are absent from the rest of the vector sequence. The use of restriction sites ensures that the vector won't split into various parts because they are present in the multiple cloning sites.

(4) The chopped pet24a+ vector and the gene (insert) from the pUC57-Kan vector were separated using two preparative DNA gels (agarose gels). The proportion of agarose

needed for the DNA gel preparation varies depending on the size of the vector and insert. Cut pet24a+ vector was separated from uncut pet24a+ vector using a 0.5% agarose gel (approximately 5.3 kbp), and the gene of interest (insert), which is around 500 bp in length, was isolated using a 1% agarose gel (about 2.5 kbp) from the pUC57-Kan vector. Large pores in the gel produced by little amounts of agarose are helpful for isolating constructions with many base pairs. Similar to this, more agarose in the gel causes smaller pores, which makes it possible to separate constructions with few base pairs.

(5) Establishing a ligation reaction. Considering the molar ratio between the empty vector and the insert is necessary while setting up ligation processes. 80–200 ng of empty vector are typically utilized for the ligation reaction. The vector molar ratios of 1:1, 3:1, and 1:3 are typically used to set up three ligation processes. In the beginning, these ratios of the insert and vector were combined. This combination is then mixed with T4 DNA ligase enzyme and its buffer and left overnight at 16°C to allow ligation to occur. Purchased from New England Biolabs Inc. were T4 DNA ligase and its unique buffer (50 mM Tris-HCl, 10 mM MgCl2, 1 mM ATP, 10 mM DTT, pH of buffer is 7.4).

(6) Conversion of the ligation product. On agar plates containing kanamycin, an antibiotic to which the pet24a+ vector exhibits resistance, transformed cells were plated. After being defrosted on ice for 50 L of BL21DE3 competent cells, a volume of 1 to 5 uL of ligation product was added. The bottom of the microcentrifuge tube was gently flicked to combine the ligature product with the competent cells. After that, ice was applied to the mixture of competent cells and ligation resulted for 20–30 minutes. The sample was then exposed to a 45-second heat shock in a water bath at 420°C. A pressure difference between the outside and the inside of the cell is produced when cells are suddenly heated up, which causes pores to form so that supercoiled plasmid DNA may enter[3]. After that, the sample spent an additional two minutes on ice. After LB (Lauria Broth) was used to dilute the competent cell/ligation mixture to 500 L, the sample was incubated at 37^0C for one hour.

(7) Individual colonies were taken from the agar plate once the first colonies had formed and put to different Eppendorf tubes with 1 mL of LB for each sample. Colonies were raised in 1 mL of LB medium supplemented with 1 mL of antibiotic at a 50 mg/mL concentration. Following the transfer, cells were cultured at 37°C for an overnight period.

(8) The plasmid was then removed using a Promega Wizard Plus DNA miniprep kit, and the presence of the required gene in the pet24a+ vector was verified by DNA sequencing.

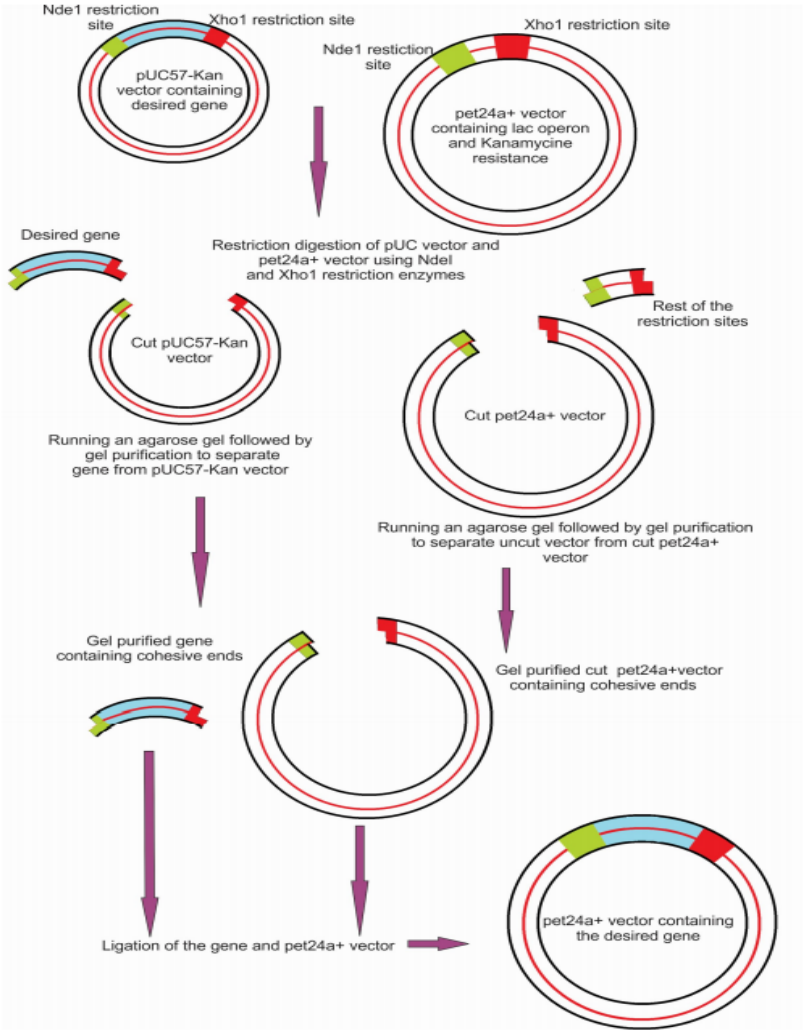


Figure 2.2 The subcloning procedure is depicted in a schematic diagram.

2.2.2: Site-directed mutagenesis

A short primer containing the necessary mutation was used to start the insertion site-directed mutagenesis process. To facilitate primer hybridization with the target gene, primers were created that were complementary to the template DNA sequence close to the insertion site. Anywhere in the DNA sequence might be chosen as the insertion site. The forward and reverse primers were created to complement the DNA double strand of the target gene. DNA polymerase is used to extend the primers that have insertion mutant nucleotides added. Primers to plasmid ratios might range from 3:1 to as high as 7:1. To

optimize the particular polymerase chain reaction (PCR), the ratio should be changed. Through the use of PCR, the necessary DNA was amplified. The multiple PCR stages included:

1. Initialization - Initialization is carried out using a hot start lasting five minutes at 95 °C. The thermally stable Pfu DNA polymerase is thereby activated[4].
2. Denaturation- This quick heating process (95 °C) breaks the H-bonds between the double strand DNA's nucleotides, causing the template DNA to melt[4]. Single strand template DNA is obtained in this stage.
3. Annealing- During this phase, the primers bind to the DNA's single strand. The temperature is decreased, typically to around 5 °C below the melting point of the primers, to promote binding of the primers to the template. To improve binding to the template DNA, primers were created to be guanine and cytosine (GC) rich. Synthesis of the whole DNA begins following the creation of primer-template DNA hybrids[4].
4. Elongation: The DNA polymerase extends the complementary DNA to the template DNA by adding deoxynucleotide triphosphates (dNTPs) after the primer binds to the template DNA. The ideal temperature is typically 34 °C, or 72 °C.
5. Final Elongation - To make sure that all single-stranded DNA has been entirely produced, this stage is likewise carried out at 72 °C. Depending on the DNA's melting point, different annealing temperatures were used during PCR to perform an insertion mutation. Up to three nucleotides could be substituted at once. The completed PCR product was converted, and the DNA was then sent for sequencing to determine whether the desired DNA had been created or not.

2.2.3: Chemical competent cells production

On the first day, BL21DE3 PLysS bacterial cells from a frozen glycerol stock were streaked onto an LB plate (BL21DE3 PLysS was treated with chloramphenicol as an antibiotic). The environment was sterile, where the work was done. At 37°C, plates were grown overnight. A number of solutions, including 1 L LB, 1 L of 100 mM $CaCl_2$, 1 L of 100 mM MgCl2, 100 mL of 85 mM $CaCl_2$ 15% glycerol v/v, centrifuge bottles and caps, and microfuge tubes were autoclaved on day 2. Overnight, the 100 mM $CaCl_2$, 100 mM

$MgCl_2$, 85 mM $CaCl_2$, and 15% v/v glycerol were refrigerated. A single colony of E. coli from the LB plate was used to create a starting culture of cells, which was then used to inoculate a 50 mL starter culture of LB (chloramphenicol). Overnight, culture was cultivated in a shaker at 37°C. On third day, 1 mL of the 50 mL starter culture was added to 300 mL of LB medium, and the mixture was shaken and cultivated at 37 °C. Every hour, then every 15 to 20 minutes until the OD reached 0.2, the OD600 was measured. The cells were put on ice as soon as the OD_{600} reached 0.35–0.4. For 20 to 30 minutes, the culture was refrigerated, occasionally being swirled to achieve equal chilling. At this moment, centrifuge bottles were also put on ice. The OD should not be allowed to increase past 0.4. As the cells multiply exponentially, the OD should be closely watched and checked frequently, especially when it rises above 0.2. When employing a 10 mL starter culture, the typical time to obtain an OD of 0.35 is roughly 3 hours. In order to avoid premature cell proliferation, it is crucial to maintain the cells at 4°C for the duration of the treatment. The cells must be pre-chilled to 4°C, as well as any bottles or solutions they come in contact with. About 50 mL of the culture was added to six ice-cold centrifuge bottles before being divided into six portions. Centrifugation at 3000g for 15 minutes at 4°C was used to harvest the cells. About 20 mL of ice-cold $MgCl_2$ were used to gently resuspension the supernatant after it had been decanted. One centrifuge bottle held all of the suspensions. Centrifugation at 2000g for 15 minutes at 4°C was used to harvest the cells[5]. The pellet resuspended in about 20 mL of ice cold $CaCl_2$ and the supernatant was decanted. For at least 20 minutes, this suspension was put on hold. On ice, 1.5 mL microfuge tubes were cooled. Centrifugation was used to separate the cells for 15 minutes at 4°C at 2000 g. After the supernatant was decanted, the pellet was resuspended in about 50 mL of ice-cold 85 mM $CaCl_2$, 15% glycerol. A 50 mL conical tube was used to transfer the suspension. By centrifuging at 1000g for 15 minutes at 4°C, cells were collected. The pellet was resuspended in 2 mL of ice cold 85 mM $CaCl_2$, 15% glycerol after the supernatant was decanted. A suspension volume of 50 mL was pipetted into sterile 1.5 mL microfuge tubes before being liquid nitrogen-snapped frozen. After that, cells were kept at -80 °C.

2.3 Expression of protein

The most popular and affordable techniques of protein production for structural and functional investigations are bacterial expression systems. An expression vector is modified using molecular cloning, as described above, to incorporate a DNA sequence encoding for the target protein. This vector is then transformed into a host cell (E. coli) and an inducer such as Isopropyl ß-D-1-thiogalactopyranoside (IPTG) was added to protein synthesis. Growth curves are graphs that are used to illustrate the cell density after the E. coli cells have been inoculated into a culture medium. E. coli cell density is calculated using OD_{600} (optical density). It is known that there are around 1 x 10^9 cells per 1 mL of growth medium when the OD_{600} value approaches 1. Lag phase, log phase, stationary phase, and death phase are the four phases that make up E. coli's growth curve (Figure 2.3). There aren't very many cells during the lag period. E. coli cells grow exponentially with time during log phase due to the presence of plenty of nutrients in the culture medium. They are therefore in the best possible condition to make proteins[6]. Due to the consumption of nutrients in the medium during the first two phases, the stationary phase is characterized by an accumulation of metabolic waste products and cell death. The pH of the medium decreases when metabolic byproducts like ethanol, lactate, and acetate build up because E. coli cells thrive best at a pH of neutral. Measuring the OD_{600} over time allows researchers to examine E. coli cell development over time. The timing at which IPTG is to be administered to promote protein expression is decided after monitoring the measured OD_{600} vs. time. Typically, IPTG addition occurs when OD_{600} reaches 0.5–0.8, indicating that E. coli cell growth has entered the log phase.

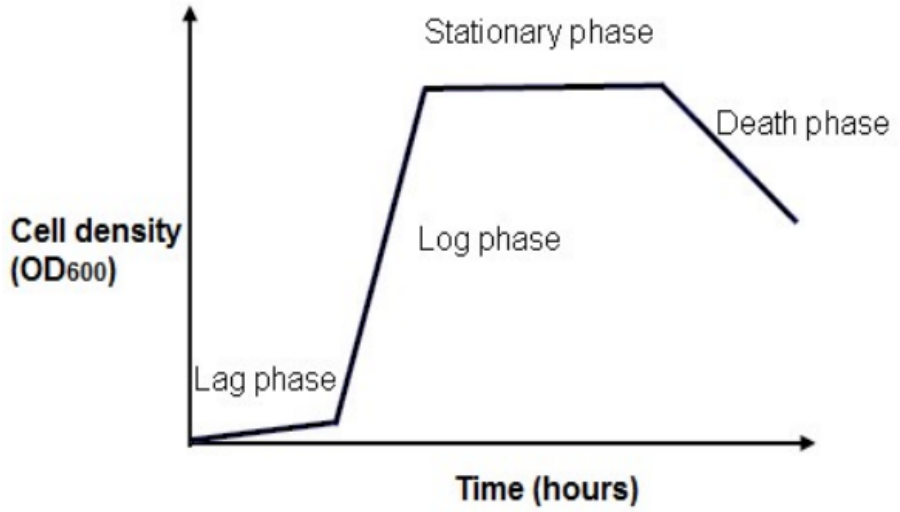


Figure 2.3: The diagram displaying E. coli's four growing phases.

2.3.1: Lac operon

A group of bacterial genes known as an operon are regulated by a single promoter (Figure 2–4). A promoter is a section of DNA that starts a specific gene's transcription. E. coli and other intestinal bacteria naturally include the lac operon, which is used to efficiently transport and digest lactose. The promoter, operator, terminator, and the genes lacZ, lacY, and lacA make up the lac operon (Figure 2–4). The enzyme lactosidase is encoded by lacZ. This enzyme turns lactose into allolactose and breaks down the disaccharide lactose into glucose and galactose[7]. The enzyme lactose permease, which makes it easier for lactose to enter cells, is encoded for by the lacY gene. Galactoside O-acetyltransferase, which is encoded by LacA, is an enzyme that passes an acetyl group from acetyl-CoA to -galactosidase[6,8].

The regulatory gene lacI creates a lac repressor protein that can bind to the operator in the absence of lactose (Figure 2-4 (a)). Repressor protein will stop lacZ, lacY, and lacA mRNA production after it binds to the operator. Lactose is converted to allolactose in the presence of lactose by -galactosidase. The allolactose that is produced binds to the lac repressor protein, changing its shape and preventing it from binding to the operator (Figure 2-4 (b)). As a result, lac genes can be transcribed by RNA polymerase. In order to take use of the lac operon, the gene sequence for the protein that has to be expressed is positioned after the operon in recombinant protein expression. By adding IPTG, an allolactose-like non-hydrolyzable substance, the expression of proteins is induced. IPTG

binds to the repressor protein and prevents it from adhering to the operator, which would otherwise result in the generation of recombinant protein.

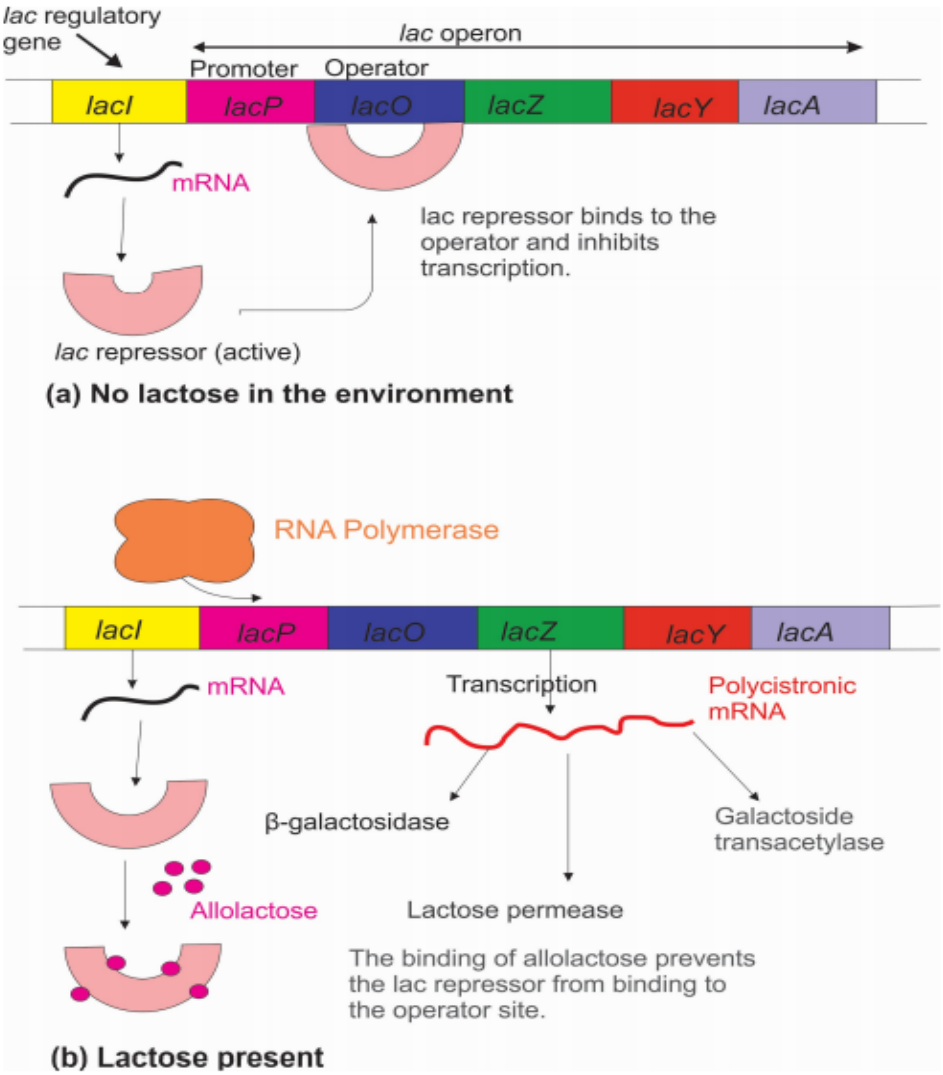
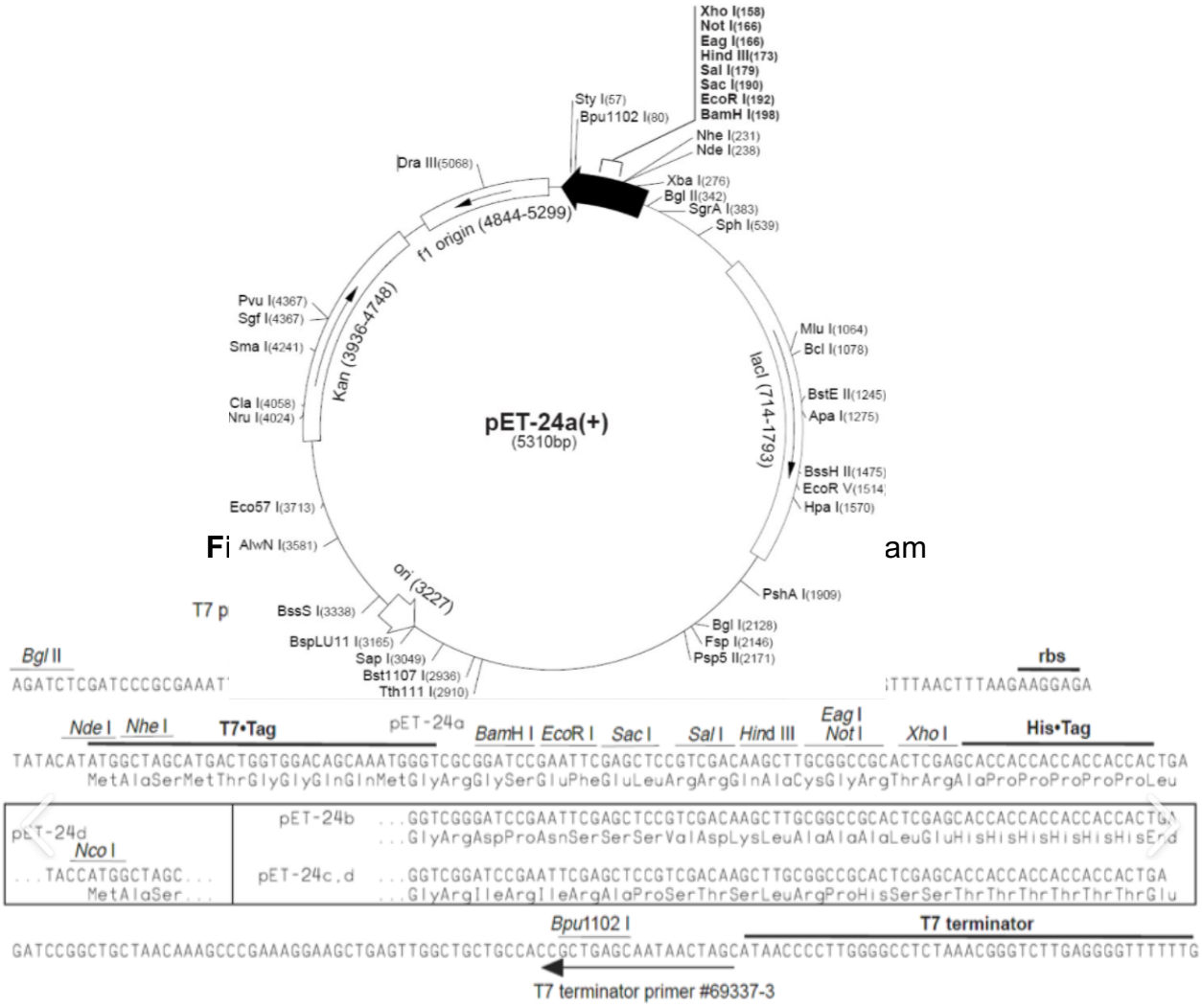


Figure 2.4: Diagram illustrating how the Lac operon functions (a) without lactose (b) with lactose[4].

2.3.2: pET vector system

In this study, the pet24a+ vector (Figure 2.5) was utilized to benefit from the lac operator. The lacI gene, T7 promoter, lac operator, and T7 terminator make up the majority of this vector (Figure 2.6)[9]. The bacteriophage T7 promoter is not a part of the bacterial genome and is solely capable of recognizing T7 RNA polymerase. Since T7 RNA polymerase can transcribe DNA eight times quicker than bacterial RNA polymerase, it is more efficient

than the latter. The BL21 (DE3) competent cells that we have employed in our investigations already have the gene that codes for T7 RNA polymerase, despite the fact that normal bacterial cells cannot make T7 RNA polymerase. The DNA that is downstream of the T7 promoter is transcribed when T7 RNA polymerase specifically binds with the T7 promoter. The gene of interest is not transcribed because T7 RNA polymerase is not produced in absence of the transducer molecule IPTG. When IPTG is added, it causes RNA polymerase to be expressed, which in turn causes the target protein to be produced. The T7 promoter and T7 RNA polymerase tightly regulate the production of a particular protein in pet vectors[9].



pET-24a-d(+) cloning/expression region

Figure 2.6: pET-24a(+) vector various parts are shown in a diagram. Comprising numerous binding sites, the T7 tag, the rbs gene, and the lac operon[9].

2.3.3: Inclusion bodies

E. coli results in the accumulation of proteins in inclusion bodies due to overexpression of many recombinant proteins (Figure 2.7). Along with the overexpressed proteins, the inclusion bodies also contain other E. coli proteins, phospholipids from the E. coli membrane, and RNA. Bacterial inclusion bodies can be spherical, ellipsoidal, cylindrical, or even tear-shaped, and they can range in size from 50 to 700 nm[10]. The weight fraction of inclusion bodies in a cell lysate can vary widely depending on factors such as the expression system, the specific protein being expressed, and the culture conditions.

However, in general, inclusion bodies can constitute a significant fraction of the total protein content in a cell lysate, ranging from 10% to 90% or more in extreme cases. Proteins found in inclusion bodies are either amorphous, which lacks any structural regularity, or paracrystalline, which exhibits short- to medium-term structural ordering[11,12]. This includes methods like X-ray crystallography, nuclear magnetic resonance (NMR) spectroscopy, electron microscopy, and Fourier-transform infrared spectroscopy (FTIR). These techniques can reveal whether the aggregated proteins have amorphous or paracrystalline structures and provide insights into their organization and stability. Paracrystalline proteins are folded, whereas amorphous proteins are unfolded. Transmission electron microscopy makes it simple to determine whether bacterial cells have inclusion bodies. Inclusion bodies can be solubilized using denaturants such as 8M urea, 6M guanidinium hydrochloride or ionic detergents (0.5% sarkosyl, 0.5% SDS)[13,14]. Multiple other detergents have also been used. Protein purified from inclusion bodies for structural and functional studies have some advantages. (1) The amount of protein in inclusion bodies is up to 15-25% of the total cell mass. (2) Those are easy to isolate from cells since their density (about 1.3 mg/mL) is higher than that of cell debris.

Following cell lysis, it is possible to separate inclusion bodies from the less dense cell detritus using high speed centrifugation (15000 rpm, 30 minutes). However, other high density subcellular components like ribosomes will be present in the pellet that has just produced. (3) Protein that is more than 95% pure can be obtained through efficient isolation and purification[15].

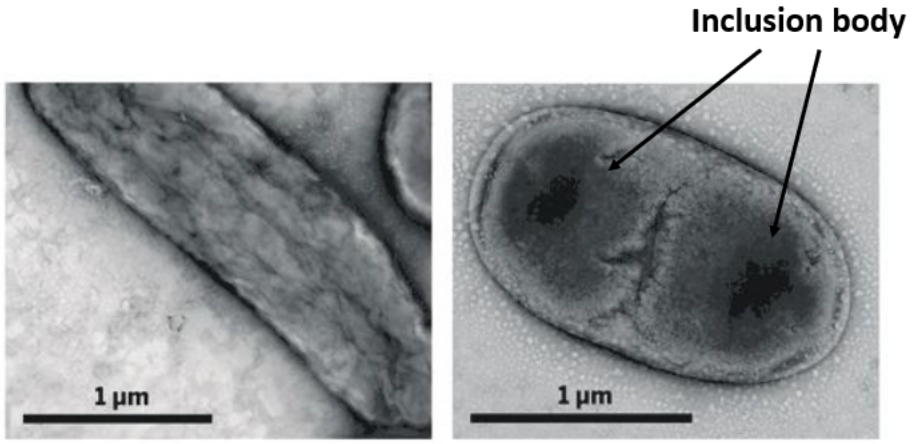


Figure 2.7: Transmission electron microscopy images of bacterial cells without protein expression (Left) and those of bacterial cells with expressed protein (Right).

2.3.4: Isotopically labeled expression proteins

Growing E. coli with a protein construct in isotopically labeled medium, then inducing the expression of the protein in the labeled media, were the steps used to make isotopically labeled protein. The pUC19 vector was also used to express proteins from cells, but this method exhibited leaky expression, which rendered the vector useless for isotopic labeling. Rich medium was first used to cultivate bacteria. No HM protein was produced after a single switch in the medium and IPTG induction. For the expression of our protein, a double media switch procedure was modified[16]. A small portion (20%) of the cell pellet from the rich media culture of E. coli was then transferred to a flask with minimum media and unlabeled glucose for a second growth period (Figure 2.8). To develop E. coli with metabolic systems that enable better use of the scarce nutrients contained in the minimal media, a second growth cycle in the minimal media may be required. Second growth phase's E. coli from the in minimal media was then transferred to a flask containing minimal media and labeled glucose.

D2O was added to the minimum media if non-exchangeable deuterium labeling of the protein was required.

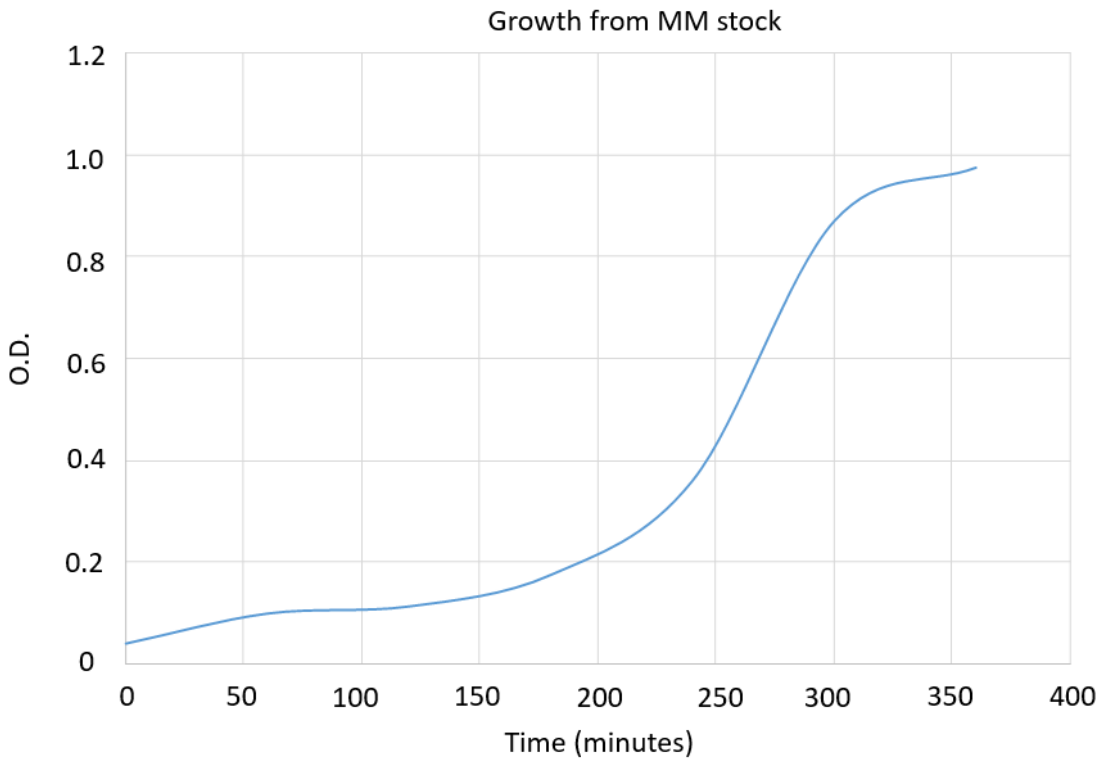


Figure 2.8 E. coli growth graph in minimum media (Robert J. Wolfe Thesis, Weliky Group).

2.4: Protein solubilization and purification

It's important to look in every potential location where protein could be found while developing a protein solubilization and purification technique. Overexpressed proteins tend to assemble as inclusion bodies most of the time[17]. To first break up the cells and release the soluble proteins into the solution, phosphate buffered saline (PBS) was added to the sonication chamber. The lysate was centrifuged at 35000g for 30 minutes at 4^0C at a speed of 15000 rpm. This centrifugation produces a pellet that contains the inclusion body fraction and other heavy cell debris. The cell membrane and soluble proteins are present in the supernatant. Following the filtration of this supernatant, soluble proteins can be extracted. The membrane fraction of the E. coli is isolated as a pellet by ultracentrifugation of this supernatant (speed of 45000 rpm or roughly 300000g for 1 hour at 4^0C). This fraction is more likely to contain membrane proteins[17].

Detergents and chaotropic agents are utilized as solubilizing reagents for membrane protein isolation[18,19,20]. Comparatively to soluble proteins, membrane proteins are more

difficult to solubilize and purify due to a number of characteristics: (1) Membrane proteins have a tendency to build up on cell membranes because they contain hydrophobic domains, therefore cells have a finite amount of membrane surface area to store these membrane proteins in the right folded condition. (2) It is necessary to apply denaturing detergents (such as 0.5% sarkosyl or 0.5% SDS) or denaturing reagents (such as 8M urea or 6M guanidium hydrochloride) in order to dissolve expressed protein, which is frequently stored as an insoluble aggregate known as inclusion bodies. Additionally, before being used in any structural or functional research, proteins isolated from inclusion bodies must be refolded. (3) Various factors can cause downstream protein purification techniques such immobilized metal affinity chromatography (IMAC), ion exchange chromatography, and gel filtration chromatography to fail.

(A) IMAC chromatography requires that the histidine tag be exposed to the metal ion in order to be successful. The hydrophobic nature of membrane proteins frequently results in minimal or no histidine tag exposure, which prevents the targeted protein from bonding with the metal ion and causes the purification to fail. Membrane proteins' hydrophobic domains frequently fold so that they are least exposed to aqueous solvent while the hydrophilic domains are more exposed. If the tag is connected to a hydrophobic domain in such a situation, it can frequently travel with the hydrophobic domain, keeping it out of contact with the solvent. The aforementioned phenomenon is brought on by this lack of exposure, which makes His tag-based membrane protein purification difficult. This problem can be somewhat resolved by adding a longer linker region (such as an extra glycine repeat) between the protein sequence and the His tag sequence[20,21]. To maximize the amount of protein that can be purified by IMAC, our designs have a 6-Glycine repeat in either the N- or C-terminus.

(B) Detergent molecules compete with protein molecules for interaction with the ion exchange resin when both the detergent molecules and the protein molecules are similarly charged, which can cause ion exchange chromatography to fail in the presence of charged detergents.

(C) In the loading solution for the gel filtration columns, some membrane proteins may clump if the detergent concentration is below the critical micelle concentration

(CMC) level of the detergent. To find the detergents where no aggregation takes place, many detergents must be tested. Even though the concentration of the detergent is higher than the CMC, protein aggregation might still happen. Increasing backpressure will result from a clogged gel filtration column from protein aggregation. Protein is also may be combined with other proteins in the void volume or precipitates at the front of the column. To clean the column and solubilize precipitates, denaturing solutions like 8M urea or 6M guanidine hydrochloride can be employed, followed by 0.5M NaOH to remove salts and other contaminants. Before size exclusion chromatography, the proteins utilized in this study were generally solubilized in 6M guanidine hydrochloride.

2.4.1: Immobilized metal affinity chromatography (IMAC)

Histidine residues on the surface of the protein molecule and immobilized metal ions form weak coordinate bonds in immobilized metal affinity chromatography (IMAC) (Figure 2.9). Immobilized metal ions like Ni^{2+}, Cu^{2+}, and Co^{2+} can be chelated using nitrilotriacetic acid (NTA) or iminodiacetic acid (IDA) to form coordinating bonds. Some of the remaining metal coordination sites have water or buffers and are capable of reversible exchange with the histidine sidechain electron donor groups seen in His-tagged protein[21]. Using a strong chelating agent like ethylenediamine tetra acetic acid (EDTA), a low pH buffer or a competitive displacement agent like imidazole can all be used to elute the protein from the affinity matrix.

With a tiny molecule like imidazole, it is simple to elute His-tagged proteins (poly-histidine tags with six histidine residues) that are bound to metal ions since imidazole can exhibit the same interaction as the His tag. As a result, while imidazole binds to the metal, the His-tagged protein separates and elutes from the metal. It is easier to elute the protein from the affinity resin by using 250–300 mM imidazole. The histidine imidazole ring has a pKa of approximately 6.0. Since its lone pairs are given to the H^+ at low pH, histidine becomes protonated at the amine side chain and is unable to establish coordination bonds with metal ions. Therefore, when the eluent has a low pH, a His-tagged protein will elute from the resin. The association between the metal ion and His-tag proteins is also disrupted by powerful chelating agents like EDTA, which causes the protein to elute.

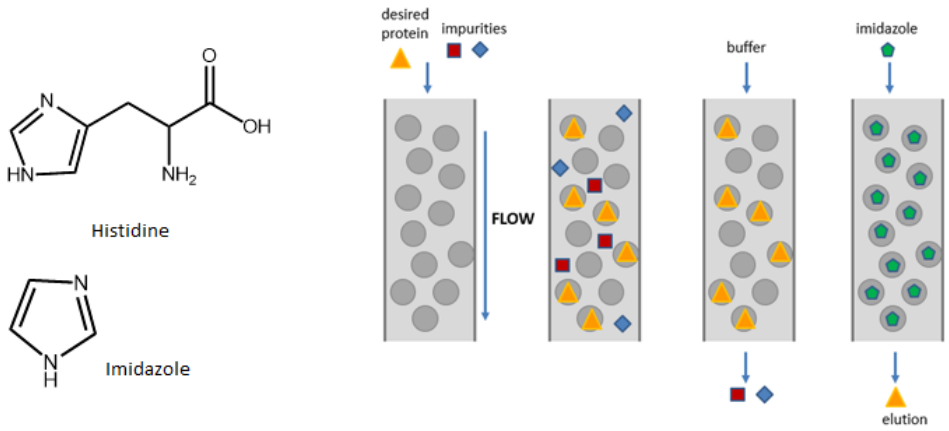


Figure 2.9 Desired proteins and cell lysis contaminants are separated using affinity chromatography. An affinity resin mixture is added to the target protein and cell contaminants. Target protein that binds to affinity resin is cleaned of impurities. A solution containing imidazole is used to elute purified protein from the affinity column.

2.4.2: Ion exchange chromatography

The ionic interaction between protein molecules and the ion exchange resin forms the basis of ion exchange chromatography. Protein molecule attaching to the oppositely charged groups in the ion exchange resin is what causes the separation. A protein's isoelectric point (pI), where its net charge is zero, is determined by the variety of ionizable sidechains of amino acid residues that are normally present at the protein surface. Proteins can be separated by ion exchange because of their net positive or negative charges, which depend on the pH and pI values. With the loading buffer pH values above the protein's pI, anion exchange is carried out utilizing positively charged resin and negatively charged proteins. The protein is positively charged in relation to a negatively charged resin because cation exchange is carried out at pH levels of the loading buffer below the pI value. A salt gradient or adjusting the pH can be utilized to elute the proteins from the ion exchange resin. Proteins that are highly attached to the resin must be eluted at a greater salt concentration than those that are weakly linked to the resin. A protein becomes negatively charged when the mobile phase's pH is increased over its isoelectric point. In order to do cation exchange chromatography, the protein can be eluted off a cation exchange resin using this technique. Protein elutes from the column due to the

cation ion exchange resin and protein becoming similarly charged as a result of the raised pH. Similar to this, in anion exchange chromatography, decreasing the pH of the mobile phase results in protein molecules becoming protonated and positively charged, which causes them to elute out of the positively charged resin.

2.4.3: Sodium dodecyl sulphate poly acrylamide gel electrophoresis (SDS-PAGE)

Protein molecules can be mass-selectively separated using SDS PAGE. Dodecyl aliphatic chain and an anionic sulfate head are characteristics of SDS (Sodium dodecyl sulfate). Protein denaturation happens when SDS detergent covers the proteins. Proteins and SDS bind to one another in a 1:1 ratio (about one SDS molecule for every two amino acids). SDS is bound by hydrophobic and positively charged residues. Proteins develop an intrinsic negative charge as a result of this. Proteins treated with SDS therefore have a comparable charge-mass ratio. Since all protein molecules have the same negative charge-to-mass ratio due to the presence of SDS, SDS-bound protein molecules move through the SDS PAGE gel mostly based on their masses. This enables us to calculate the protein's molecular weight. The crosslinking pore size of the SDS PAGE matrix can be altered by altering the acrylamide to bisacrylamide ratio. The parameters of %T [total polyacrylamide percentage (w/v)] and %C bis [the ratio of bis to monomer (w/w)] are used to calculate the size of pores in polyacrylamide gels. Since the pore size is small, higher % polyacrylamide gels are used to improve the separation resolution of the small proteins. When the gel's pore size is smaller, molecules with larger masses will be more resistant to migrating[4].

2.4.4: Western blots

The method of western blotting can be used to identify a particular protein from a complicated protein mixture. In this method, proteins are initially separated by mass using gel electrophoresis (SDS-PAGE). The proteins are then transferred to a membrane made of nitrocellulose. The membrane is then blocked, which involves adding a milk solution (dry milk dissolved in SDS solution) to the paper to saturate it with protein. The signal is then detected with the addition of an antibody that can specifically bind to the protein of interest. A primary antibody is added first that specifically recognizes the protein of interest, followed by a secondary antibody is added that linked to a reporter enzyme. The primary antibody binds with secondary antibody to produce a specific signal color in

presence of specific substrate. In our laboratory, the majority of produced proteins have a C-terminal histidine tag that facilitates affinity purification. Using an anti-His antibody, these proteins were found. A conjugate of anti-His antibody with the enzyme horseradish peroxidase (HRP) reacts with a chemiluminescence substrate. The His-tagged proteins may be seen thanks to the reaction's brown color.

2.5: Protein characterization

Protein characterization is the process of studying and analyzing proteins to understand their structure, function, and properties. This is a crucial aspect of biochemistry and molecular biology, as proteins play a fundamental role in various biological processes and are essential for the functioning of living organisms. Protein characterization helps researchers gain insights into how proteins work, their interactions with other molecules, and their role in disease and health.

Protein characterization is a multidisciplinary field that combines techniques from biology, chemistry, physics, and informatics to gain a comprehensive understanding of proteins in terms of their structure, function, and roles in biological systems. It is essential for advancing our knowledge of biology, disease mechanisms, drug development, and various other applications in the life sciences. Few of many techniques of protein characterization have been used in our research are discussed as followings:

2.5.1: Gel filtration chromatography: Determination of the oligomeric state of protein

Proteins and oligonucleotides are separated using gel filtration chromatography based on their size. Dextran and agarose, two polysaccharides, were crosslinked to create the matrix of the gel filtration bed. By passing through the stationary phase (column matrix), molecules from the mobile phase can diffuse into the pores. Because the pores in the matrix of the column are too small for larger molecules to enter, they pass through the column more quickly and elute earlier. The pores of the column allow for the entry of small molecules, which travel slowly through the column (Figure 2-10). The exclusion limit in gel filtration chromatography refers to the molecular mass (or size) beyond which molecules start to elute at the column's void volume. The lower limit of a molecule's molecular mass (or size), known as the permeation limit, is what causes all molecules to elute in a single band below that mass. A 1.3 MDa exclusion limit and a 6.5 kDa

permeation limit are features of the superdex 200 column that was employed in our lab. The use of detergent is widespread in research on membrane proteins. Detergents in protein solution will form micelles, and when the protein elutes from the column, its apparent mass is frequently higher than its true molecular mass. Therefore, the measured mass is the result of adding the masses of the detergent and protein molecules[22]. A collection of protein calibration standards with known molecular weights are initially used to calculate the molecular mass of an unknown protein (Figure 2-11). The unidentified protein sample can then be run in the column, and the partition coefficient (K_{av}) can be determined using the equation shown below (Equation 2.1). The term "void volume," or "V_o," describes the volume of a molecule's elution that escapes the pores of the gel filtration matrix. The packed gel filtration bed volume is referred to as the column volume, or V_c.

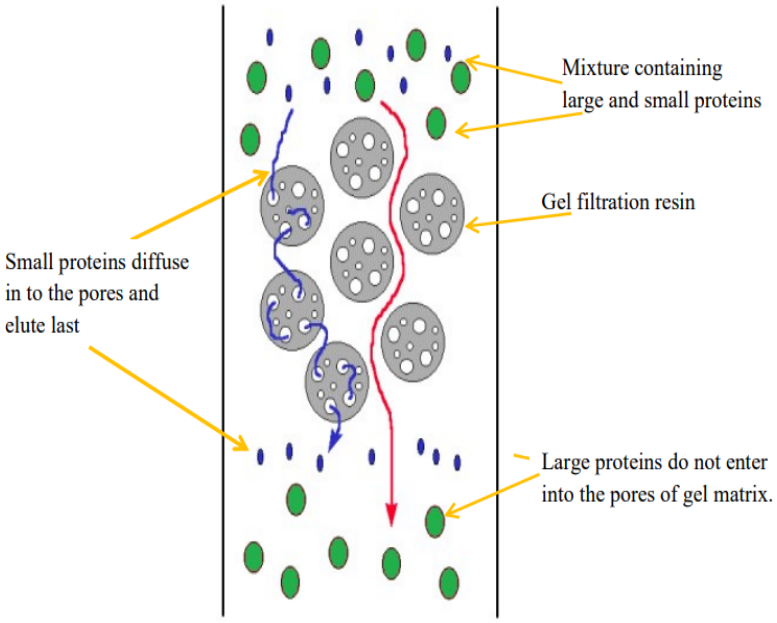


Figure 2.10: Gel filtration's Mechanism.

Partition coefficient, $K_{av} = (V_e-V_o)/(V_c-V_o)$ 2.1

Where V_e, V_c, and V_o, respectively, stand for elution, column, and void volumes. V_c = 24 ml and V_o = 9 ml in the superdex 200 column[23].

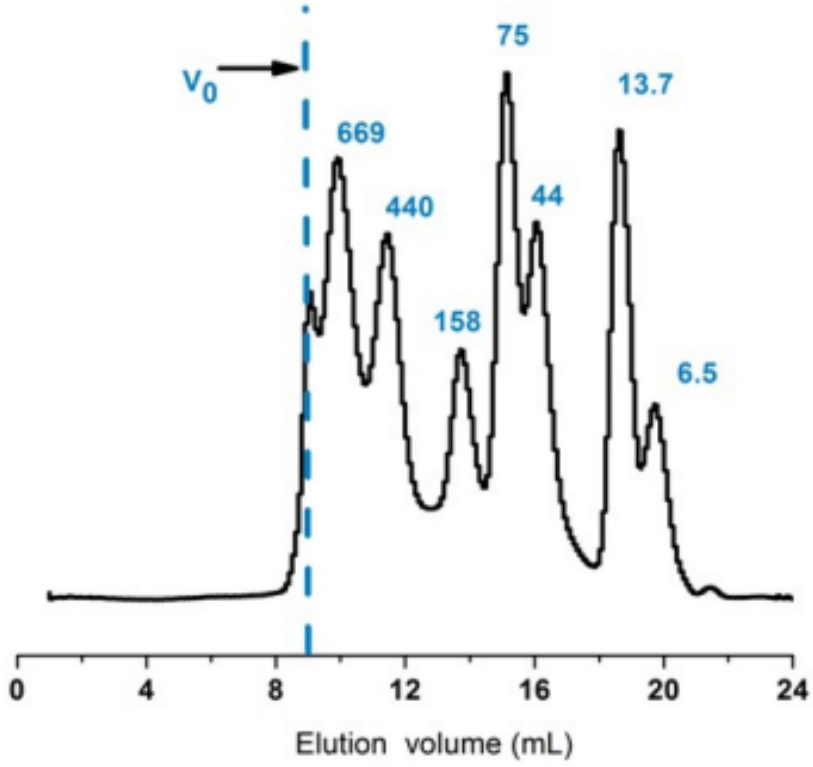


Figure 2.11: A_{280} of MW standards loaded on Superdex 200 columns are shown in the plot. Thyroglobulin (669 kDa), Ferritin (440 kDa), Aldolase (158 kDa), Conalbumin (75 kDa), Ovalbumin (44 kDa), Carbonic Anhydrase (29 kDa), Ribonuclease A (13.7 kDa), and Aprotinin (6.5 kDa) were the protein standards used. The V_0 stands in for the void volume[23].

2.5.2: CD spectroscopy: Determination of the secondary structure of the protein

Protein molecules' global secondary structure can be determined via CD spectroscopy. When a chiral molecule absorbs light with left- and right-circular polarizations, its differential absorption of the two polarizations is measured. The relationship between the mean residue molar ellipticity (θ_{MR}) and wavelength (nm) is represented in a CD plot. In the range of 190-250 nm (Figure 2.12), alpha-helix, beta-sheet, and random coil configurations provide the distinctive shapes of CD spectra. The difference between the absorbance of left-circularly polarized (A_{LCP}) light and right-circularly polarized (A_{RCP}) is used to assess circular dichroism (CD), and it can be written as:

$$\Delta A = A_{LCP} - A_{RCP}$$

According to Beer-Lambert's law, $A = \varepsilon C l$

$$\Delta A = \Delta \varepsilon C l$$

$$\Delta \varepsilon = \varepsilon_{LCP} - \varepsilon_{RCP} = \Delta A/(C \times l)$$

Where; C= molar concentration, and l = path length in centimeters, ε_{LCP} and ε_{RCP} are the molar extinction coefficients for left circularly polarized light (LCP) and right circularly polarized light (RCP), respectively.

The mean residue concentration (CMR) is calculated by multiplying the protein concentration (C) in molar by the number of amino acids (N) in the protein,

$$C_{MR} = C \times N$$

$$\Delta \varepsilon_{MR} = \Delta A/ (C_{MR} \times l)$$

$$\Delta A = \Delta \varepsilon_{MR} (C_{MR} \times l)$$

The apparatus detects differential absorption of the left and right circularly polarized light using a photomultiplier and converts it into ellipticity θ, which has millidegree units. Differential absorption by can be connected to ellipticity by:

$$\Delta A = (\theta/32982)$$

$$\theta_{MR} = (\Delta A) \times 32982$$

$$\theta_{MR} = (\Delta \varepsilon_{MR} (C_{MR} \times l)) \times 32982$$

θ_{MR} = Mean residue molar ellipticity

The mean residue molar ellipticity (θ_{MR}) of -33,000 deg $cm^2 dmol^{-1}$ at wavelength 222 nm is considered as 100% helicity. This value was determined using different reference structures consisting of poly-L- lysine with the secondary structures of completely helical, random coiled and beta-sheet. After scanning over the poly-L-lysine structures' wavelength range of 190 nm–250 nm, the mean molar residue ellipticity was determined. The mean residue molar ellipticity for random coiled and beta-sheet structures at 222 nm was relatively low, whereas the alpha helix had the highest value of -33,000. The percent helicity of a protein molecule can be determined based on that value[24].

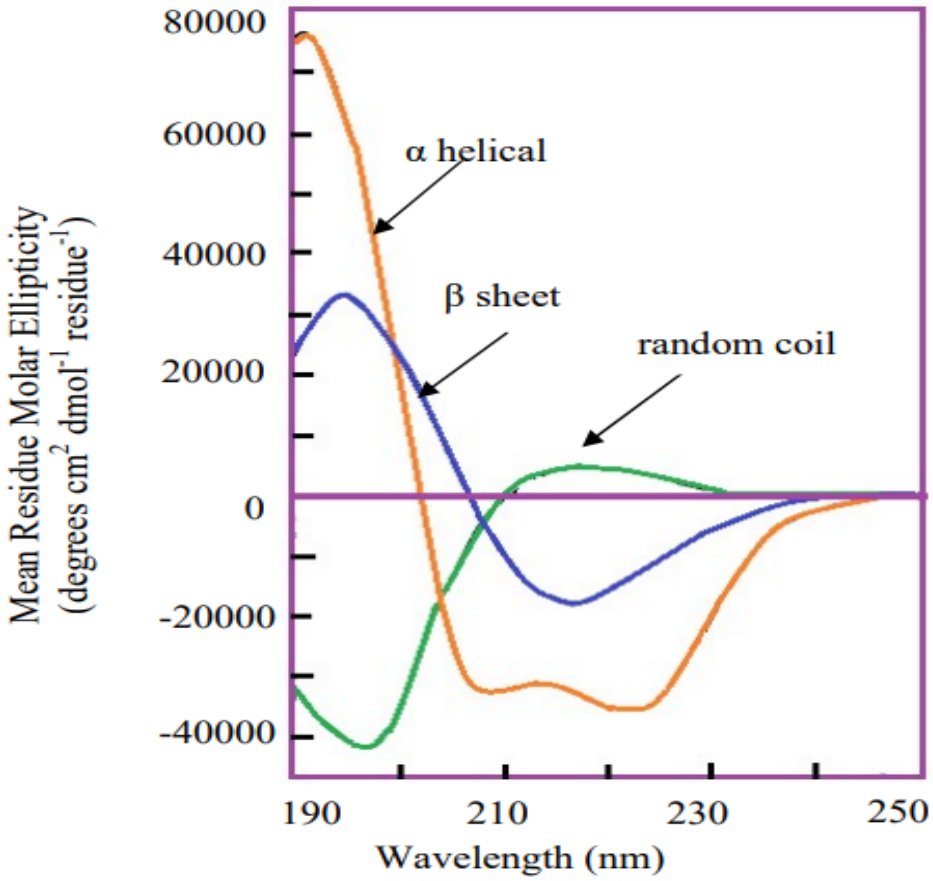


Figure 2.12: Characteristic CD curves for several secondary protein-molecule configurations[24].

2.6: Lipid mixing assay

To examine the fusogenicity of various protein constructions, a fluorescence quenching assay (Figure 2.13) was utilized[25]. In order for this assay to work, the emission band of the donor molecule must overlap with the excitation band of the acceptor molecule, and they must also be located close to one another (Figure 2.14).

The sixth power of the distance between the energy donor and receiver affects how effectively energy is transferred. N-Rh-PE {N-(lissamine rhodamine B sulfonyl) (ammonium salt) dipalmitoylphophatidyl- ethanolamine} is the energy acceptor and N-NBD-PE {N-(7-nitro-2, 1, 3-benzoxadiazol-4-yl) (ammonium salt) dipalmitoylphosphatidylethanolamine is the energy donor[25]. The phosphatidylethanolamine lipid's head group is altered to include a rhodamine quencher and NBD fluorophore. By literature convention, time-dependent percent vesicle fusion is

calculated as $M(t) = 100 \times [F(t) - F_0]/[F_{max} - F0]$ with long-time fusion extent (M_{ext}) calculated using $F(t) = F_{ext}$. The $F \propto \langle 1/ [1 + (R_{For} / R)^6] \rangle$ where R is fluorophore/quencher distance, R_{For} is the Forster distance (~50 Å), and $\langle \ldots \rangle$ is the average over all fluorophores. For the initial F_0 conditions, the estimated $\langle R \rangle \approx R_{For}$ and fusion between two vesicles results in $\langle R \rangle \approx (2)^{1/2} \times R_{For}$ with resulting $M_{ext} \approx 80\%$[30].

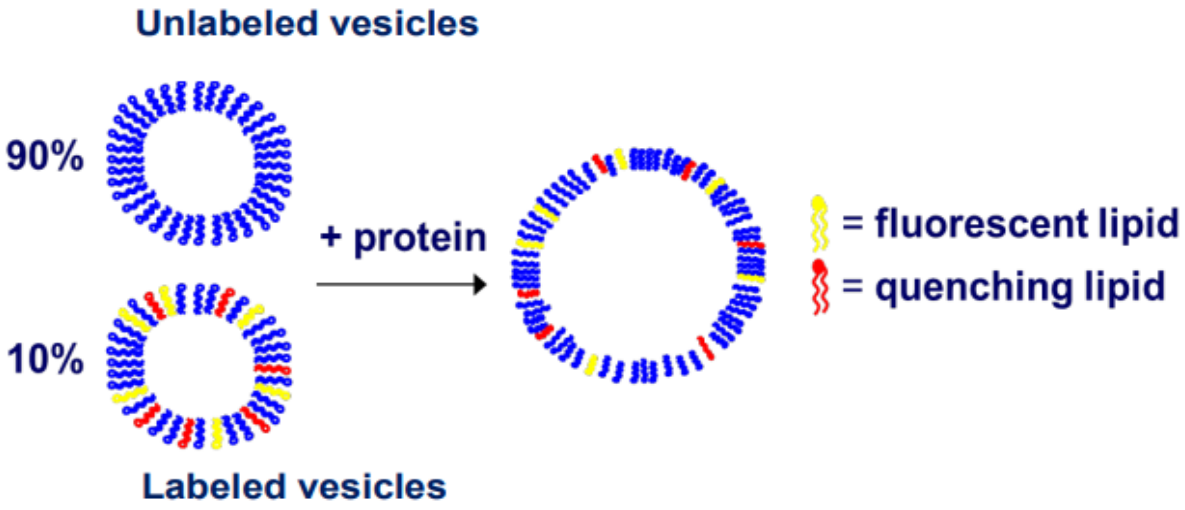


Figure 2.13: The mixing of lipids principle. Unlabeled vesicles are combined with a small portion of labeled vesicles that contain both quenching lipid (acceptor) and fluorescent lipid (donor). The fluorescence signal is amplified as a result of dilution associated with fusion.

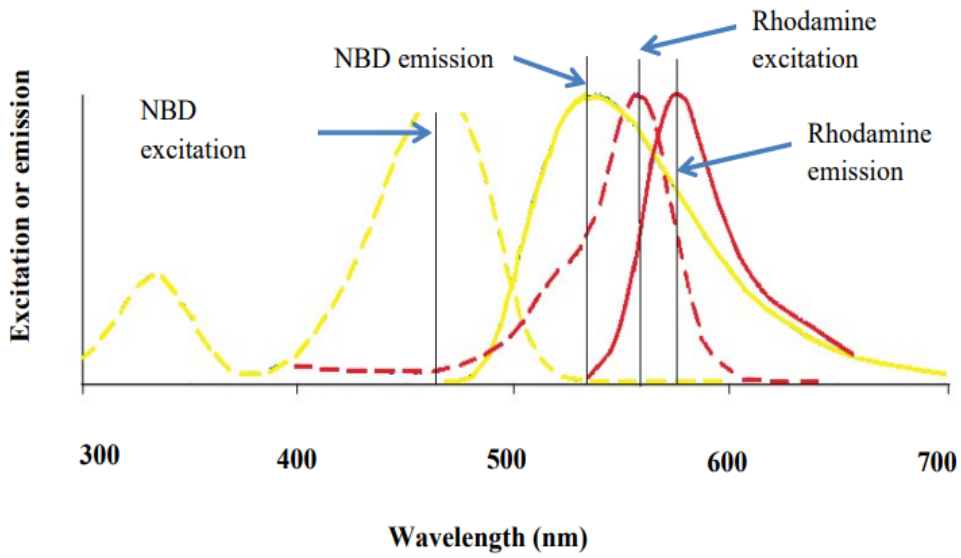


Figure 2.14: Fluorescence quenching assay principle. NBD emits around 530 nm after 467 nm of excitation. When the donor (NBD) and acceptor (Rhodamine) are near together, the acceptor molecule effectively absorbs the donor's emission signal. Protein causes vesicular fusion, which increases the distance between the donor and the acceptor. The acceptor's ability to quench lessens as the distance increases. Consequently, there is an increase in the fluorescence signal.

2.7: Preparation of solid-State NMR samples

The labeled purified proteins after affinity purification were precipitated and centrifuged under vacuum to remove the majority of the water after being dialyzed against distilled, deionized water to remove salts. The materials were then overnight rehydrated after being freeze dried and put into an NMR rotor. The lipid composition was dipalmitoylphosphatidylcholine (DPPC), and 1,2-dipalmitoyl-sn-glycero-3-phospho-(1'-rac-glycerol) (sodium salt) (DPPG) at a 4:1 ratio; and DPPC, DPPG, and cholesterol was at an 8:2:5 mole ratio in lipid containing samples. Both compositions include cholesterol mole fractions that are comparable to those found in the plasma membrane of HIV host cells[18,26,27]. This mixture was chosen because (1) phosphocholine lipids make up a significant portion of the plasma membrane of HIV-1 viral host cells and (2)) the host cell plasma membrane has a charge of approximately -1, which is mimicked by the 4:1 DPPC:DPPG charge[18,26,28]. Lipid (20 μmol) was dissolved in 1 mL of a 9:1 chloroform:methanol solution, and the solvent was then eliminated over the course of an

overnight vacuum pumping procedure and dry nitrogen gas flow. The lipid film was first hydrated using 1 mL of 10 mM 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES) and 5 mM 2-(N-morpholino) ethane sulfonic acid (MES) buffer at pH 5.0 followed by 10 freeze-thaw cycles to make a homogenous suspension of unilamellar vesicles [29]. To obtain large unilamellar vesicles in the retentate, the lipid-buffer suspension was extruded ten times through a polycarbonate membrane with a 100 nm pore size. Large unilamellar vesicles (LUV) are between 100 nm and 1 μm in diameter, gigantic unilamellar vesicles (GUV) are over 1 μm in diameter, and small unilamellar vesicles (SUV) are typically below 100 nm diameter. After dissolving the target protein (0.2 μmol) in 1 mL of HEPES/MES buffer at pH 5.0, the protein was gradually introduced to lipid vesicles drop by drop and incubated for an entire night. The lipid and protein mixture were then diluted to 4 mL with the HEPES/MES buffer and Ultra-centrifuged at 160000g for 4 hours was used to pellet the peptide-lipid complex. Overnight, the pellet of the proteo-liposome complex was lyophilized. Lyophilization aids in minimizing sample loss during NMR rotor sample packing. The sample was placed inside the NMR rotor (phoenix 1.6 mm) and rehydrated over the course of an overnight period at room temperature with 3 μL of the HEPES/MES buffer.

2.8: REDOR Solid state NMR experiment

Oxford NMR AS400 magnet, Bruker Avance Neo console and a Phoenix magic angle spinning (MAS) probe with a 1.6 mm (8 μl) diameter rotor were used for the SSNMR investigations. The probe was simultaneously set to NMR frequencies as ^{1}H (400.1 MHz), ^{13}C (100.6 MHz), and ^{2}H (61.4 MHz). The following pulse sequence was used to capture rotational-echo double-resonance (REDOR) data: (1) ^{1}H ~π/2 pulse; (2) ^{1}H-^{13}C cross polarization (CP); (3) dephasing period of duration and (4) ^{13}C detection. The pulses applied during the dephasing period varied between the S_0 and S_1 REDOR data, which were gathered in an alternate order. Every rotor cycle, with the exception of the last one, ended with a ^{13}C π pulse for S_0 and S_1, and a ^{2}H π pulse was applied in the middle of each rotor period for S_1. (1) 10 kHz MAS frequency and 1.5 ms CP contact time; (2) 50 kHz rf fields for ^{1}H π/2 pulse and CP; (3) 55-66 kHz ^{13}C CP ramp; (4) 60 kHz ^{13}C π pulses, 100 kHz ^{2}H π pulses with XY-8 phase cycling applied to all π pulses; and ~70 kHz two-pulse phase-modulated ^{1}H decoupling during dephasing and acquisition. Recycle delays

of 1 s (= 3.2, 16 ms), 1.5 s (= 24 ms), and 2 s (= 40 ms) were the typical values. External ^{13}C chemical shift reference was carried out utilizing the adamantane methylene peak at 40.5 ppm and subjected to 100 Hz Gaussian line broadening and baseline correction. ^{13}C peak's width from 2-9 ppm at different spectral region were used to compute the S_0 and S_1 intensities. The following is an explanation of the REDOR dephasing:

$$\Delta S/S_0 = (S_0 - S_1)/ S_0$$

S_0 and S_1 ^{13}C intensities determined were used to calculate experimental $(\Delta S/S_0)^{exp} = (S_0 - S_1)/S_0$ values at each τ. The S_0 and S_1 error is based on spectral noise, which is the standard deviation of 5 spectral noise regions with 3 ppm integration window[1] and errors were calculated as follows:

$^{\sigma}S_o$= Stdev. of Noise x Sqrt. (Width of S_o spec. region/ Width of noise region)

and $^{\sigma}S_1$= Stdev. of Noise x Sqrt. (Width of S_1 spec. region/ Width of noise region)

The error of experimental dephasing, $(\Delta S/S_0)$ was then calculated as follows[1]:

$$\sigma_{\Delta S/S_0} = \sqrt{\left(\frac{\sigma_{S_0}}{S_0}\right)^2 + \left(\frac{\sigma_{S_1}}{S_1}\right)^2} \times \frac{S_1}{S_0} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ 2.2$$

The dephasing is a function of the dephasing time, which is the delay between the ^1H to ^{13}C CP and the acquisition of ^{13}C. The dipolar coupling (D) and internuclear distance (r) between ^{13}C and ^2H can be estimated from the dephasing vs. dephasing time (τ) shown against time:

$$D(Hz) = \frac{\mu_0 \gamma_{13C} \gamma_{2H} h}{16\pi^2 r^3} = \frac{4642 Å^3 s^{-1}}{r^3} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ 2.3$$

The I4 peptide AEAAAKEAAAKEAAAKAW with a regular α helical shape was employed as a standard sample (reference compound that is used to calibrate the spectrometer, ensuring that the chemical shifts and other parameters are accurate and reproducible). The reference I4 peptide had ^2H and ^{13}CO tagged at positions A8 and A9, respectively. All of the isolated ^{13}CO-^2H spin pairs had a dipolar coupling D of 37 Hz and a common separation r of 5.0 Å.

REFERENCES

1. Jia, L. H., Liang, S., Sackett, K., Xie, L., Ghosh, U., and Weliky, D. P. (2015) REDOR solid-state NMR as a probe of the membrane locations of membrane-associated peptides and proteins, J. Magn. Reson. 253, 154-165.
2. S. Tristram-Nagle, J.F. Nagle, Lipid bilayers: thermodynamics, structure, fluctuations, and interactions, Chem. Phys. Lipids 127 (2004) 3–14.
3. Iost, I., J. Guillerez, and M. Dreyfus, Bacteriophage-T7 RNA-Polymerase travels far ahead of ribosomes in vivo. Journal of Bacteriology, 1992. 174(2): p. 619-622.
4. Alberts, B., et al., Molecular Biology of the Cell. Third Edition ed. 1994.
5. 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.
6. 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.
7. Juers, D.H., B.W. Matthews, and R.E. Huber, LacZ beta-galactosidase: Structure and function of an enzyme of historical and molecular biological importance. Protein Science, 2012. 21(12): p. 1792-1807.
8. Laskey, S. B., and Siliciano, R. F. (2014) A mechanistic theory to explain the efficacy of antiretroviral therapy, Nat. Rev. Microbiol. 12, 772-780.
9. Schibli, D. J., Montelaro, R. C., and Vogel, H. J. (2001) The membrane-proximal tryptophan-rich region of the HIV glycoprotein, gp41, forms a well-defined helix in dodecylphosphocholine micelles, Biochemistry 40, 9570-9578.
10. Paternal, S. and R. Komel, Active Protein Aggregates Produced in Escherichia coli. International Journal of Molecular Sciences, 2011. 12(11): p. 8275-8287.
11. Taylor, G., et al., Size and density of protein inclusion-bodies. Bio-Technology, 1986. 4(6): p. 553-557.
12. Singh, A., et al., Protein recovery from inclusion bodies of Escherichia coli using mild solubilization process. Microbial Cell Factories, 2015. 14.
13. Kwong, P. D., Wyatt, R., Robinson, J., Sweet, R. W., Sodroski, J., and Hendrickson, W. A. (1998) Structure of an HIV gp120 envelope glycoprotein in complex with the CD4 receptor and a neutralizing human antibody, Nature 393, 648-659.
14. Broder, C. C., and Dimitrov, D. S. (1996) HIV and the 7-transmembrane domain receptors, Pathobiology 64, 171-179.
15. 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.
16. 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.

17. Singh, S.M. and A.K. Panda, Solubilization and refolding of bacterial inclusion body proteins. Journal of Bioscience and Bioengineering, 2005. 99(4): p. 303-310.

18. Banerjee, K. and D.P. Weliky, Folded Monomers and Hexamers of the Ectodomain of the HIV gp41 Membrane Fusion Protein: Potential Roles in Fusion and Synergy Between the Fusion Peptide, Hairpin, and Membrane-Proximal External Region. Biochemistry, 2014. 53(46): p. 7184-7198.

19. Vogel, E.P., et al., Solid state nuclear magnetic resonance (NMR) spectroscopy of human immune deficiency virus gp41 protein that includes the fusion peptide: NMR detection of recombinant Fgp41 in in inclusion bodies in whole bacterial cells and structural characterization of purified and membrane associated Fgp41. Biochemistry 2011. 50 p. 10013-10026.

20. Curtis-Fisk, J., R.M. Spencer, and D.P. Weliky, isotopically labeled expression in E. coli, purification, and refolding of the full ectodomain of the influenza virus membrane fusion protein. Protein expression and purification 2008. 61 p. 212-219.

21. Bornhorst, J.A. and J.J. Falke, Purification of Proteins Using Polyhistidine Affinity Tags. Methods in Enzymology, 2000. 326: p. 245-254.

22. Swalley, S.E., et al., Full-length influenza hemagglutinin HA(2) refolds into the trimeric low-pH-induced conformation. Biochemistry, 2004. 43(19): p. 5902-5911.

23. Punsisi U. Ratnayake, E.A. Prabodha Ekanayaka, Sweta S. Komanduru, David P. Weliky. Full-length trimeric influenza virus hemagglutinin II membrane fusion protein and shorter constructs lacking the fusion peptide or transmembrane domain: Hyperthermostability of the full-length protein and the soluble ectodomain and fusion peptide make significant contributions to fusion of membrane vesicles, Protein Expression and Purification, 117, 2016, 6-16, 1046-5928.

24. Greenfield, N and G.D. Fasman, Computed circular dichroism spectra's for evaluation of protein conformation. Biochemistry, 1969. 8(10): p. 4108-4116.

25. Struck, D.K., D. Hoekstra, and R.E. Pagano, Use of Resonance energy transfer to monitor membrane fusion. Biochemistry, 1981. 20: p. 4093-4099.

26. Sackett, K., A. TerBush, and D.P. Weliky, 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. European Biophysics Journal with Biophysics Letters, 2011. 40(4): p. 489-502.

27. 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.

28. Ratnayake, P.U., et al., 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 gp41with predominant β sheet fusion peptide confirmation. Biochema et Biophysica Acta (BBA) – Biomembranes, 2015, 1848(1, Part B): p. 289-298.

29. Yamaguchi,T. Koga, M. Fujita, Y. Kimoto, E. Effect of pH on Membrane Fluidity of Human erythrocytes. J Biochem. 1982 Apr;91(4): 1299-304.

30. Md. Rokonujjaman, Abdulrazak Sahyouni, Robert Wolfe, Lihui Jia, Ujjayini Ghosh, and David P. Weliky. A large HIV gp41 construct with final trimer-of-hairpins structure exhibits V2E mutation-dominant attenuation of vesicle fusion and helicity that is quantitatively similar to V2E attenuation of HIV fusion and infection and supports: (1) hairpin stabilization of initial membrane apposition with larger apposition distance for V2E; and (2) V2E dominance mediated by an antiparallel β sheet with interleaved fusion peptide strands from two gp41 trimers. Biophysical Chemistry 293 (2023) 106933.

CHAPTER-03 (Project- 01): Quantitative similarity of HIV gp160 V513E-dominant reduction of fusion and infection with fusion by the gp41 ectodomain hairpin supports an important fusion role for the final trimer-of-hairpins structure.

3.1: Introduction

Fusion (joining) of two membranes is important in both physiologically-beneficial and pathogenic processes[1-4]. Fusion is thought to follow a stepwise mechanism with distinct membrane structural intermediates. One common mechanistic model is: (1) close (nm) apposition of the two membranes; (2) formation of a bilayer "stalk" that connects the two membranes and is contiguous with their outer leaflets; (3) topological transformation of the stalk to a planar "hemifusion" bilayer diaphragm that is contiguous with the inner leaflets of the two membranes; and (4) formation and expansion of pores in the hemifusion diaphragm that result in a single bilayer that encompasses the two bodies[3-5]. Steps 2 and 3 are experimentally detected by lipid mixing between the two membranes and step 4 is detected by contents mixing between the two bodies. A variety of computational approaches have been applied to estimate energy barriers for the different steps with typical values of 25 kcal/mol for step 1 and 10 kcal/mol per step for 2, 3, and 4[4].

The present study focuses on fusion relevant to the membrane-enveloped human immunodeficiency virus (HIV) which is the pathogen for AIDS, a global disease. There is currently no cure or vaccine for AIDS, but there are therapeutics and prophylactics that control and prevent HIV infection and are typically taken daily. Fusion between the membranes of HIV and a target cell is an early and required step in infection[3,4]. There is consequent deposition of viral genetic material in the cytoplasm which is necessary for viral replication within the cell. Fusion between HIV and target cells is mediated by the viral envelope glycoprotein 160 kDa (gp160) which had been proteolytically-cleaved into N-terminal gp120 and C-terminal gp41 subunits, that are respectively ~500 and 350 residues. Gp41 has a ~ 180-residue N-terminal ectodomain (Ed) outside the viral membrane which assembles as a trimer and non-covalently associates with three gp120 subunits to form a spike in the viral envelope[6,7]. There is a single transmembrane domain (Tmd) C-terminal of the gp41 Ed followed by a ~ 150-residue endodomain inside the virus. The first step in infection is gp120 binding to the CD4 receptor and then the CXCR4 or CCR5 chemokine receptor of immune system cells like T-cells and macrophages. This is followed by separation of gp120 from the gp41 Ed, and then a large structural change of much of the Ed into a hyperthermostable trimer-of-hairpins structure[8–10]. The ~60-residue

“N-helices” of three gp41 molecules form a parallel coiled-coil within this structure and are followed by 180o turns so that the three ~60-residue “C-helices” are antiparallel and in van der Waals contact with the external grooves of the N-helix core, i. e. each gp41 molecule is a helical hairpin. There are ~25 gp41 residues N-terminal of the N-helix that are denoted the “fusion peptide” (Fp), Fig. 1[11]. The Fp has extended irregular structure in the initial gp41/ gp120 complex, and is the epitope of some broadly-neutralizing antibodies[6,12–14]. Some Fp residues are thought to bind target membrane early in the fusion process. The Fp alone or with C-terminal hairpin often adopts intermolecular antiparallel β-sheet structure in membrane, and can adopt β-sheet or helical monomer structure in detergent-rich media[15–25]. There are ~10 residues between the C-helix and Tmd that are part of the “membrane-proximal external region” (Mper) some of which may also bind the viral membrane with predominantly helical structure[26–32]. In some studies, the Mper includes residues that are part of the C-helix in Fig. 1. These residues may bind the viral membrane in the initial gp41 Ed complex with gp120. The Mper also includes linear epitopes of broadly-neutralizing antibodies which also bind the final trimer-of-hairpins structure of gp41[27,33]. To our knowledge, there aren’t yet clear experimental data that elucidate the relative timings of changes in membrane and gp41 Ed structure during fusion. Models of the mechanism of gp41-induced fusion are instead based on data that include the initial and final structures of the gp41 Ed, fusion inhibitors, and computation. The most common type of model posits that some of the thermodynamic energy released as the gp41 Ed changes from the initial complex with gp120 to the final hairpin is converted into activation energy between membrane intermediates[2,34,35]. These models often hypothesize well-defined structural intermediates of the gp41 Ed that catalyze formation of specific membrane intermediates, e.g. separated N- and C-helices that bind to the initial apposed membranes or to the stalk[36–41]. For this overall group of models, the final Ed hairpin is a “post-fusion” state, i.e. most or all of the membrane changes during fusion happen prior to final hairpin formation. A separate type of model posits that the final gp41 hairpin forms prior to membrane fusion and is the fusion-catalytic structure. Some data supporting this model type are: (1) rapid and extensive vesicle fusion induced by large gp41 constructs with hairpin structure; and (2) vesicle, cell-cell, and computational fusion induced by the related influenza virus hemagglutinin subunit 2 (Ha2)

protein in the final trimer-of-hairpins state[23,25,33,42–47]. For most models, fusion barriers are also reduced by target membrane-binding by the Fp, and by virus membrane-binding by the Mper (Fig. 3.1)[32,48–53]. Similar to enzyme mechanisms, a proposed mechanism of gp160- mediated fusion is also supported if the mechanism provides sensible explanations of the effects of gp160 mutations on fusion, with particular emphasis on mutations that greatly impair fusion but don't affect initial assembly of gp160 spikes in the HIV membrane. V2E is this type of point mutation and is located at the N-terminus of the Fp of the gp41 subunit of gp160 (Fig. 3.1). V2E eliminates fusion between cells expressing gp160 and cells expressing CD4 and chemokine receptors, and also eliminates HIV infection of cells[54]. To our knowledge, V2E is the only gp41 mutation that exhibits dominance in mixed trimers of wild-type (WT) and V2E gp160, with f_{V2E} ≡ fraction V2E and fV2E = 1 – f_{WT}. As one example, relative to f_{V2E} = 0, f_{V2E} = 0.09 results in only 40% cell-cell fusion and HIV infection. The dependences of the reductions in fusion and infection on f_{V2E} have been analyzed with mathematical models to determine a number of WT trimers (T) required for efficient fusion and infection. Different research groups have done the analysis using the same or similar data, and the derived values of T depend strongly on the assumptions of the model used in the analysis, with variation in T between 1 and 19 trimers among the different analyses[55]. The large range for T makes it more difficult to distinguish between different possible mechanisms of gp160-mediated fusion. As one consideration, the typical virion has only ~15 trimers, so the fusion rate and viral infectivity might be predicted to be inversely-correlated with T[56]. As an example, if T = 1, each trimer could independently catalyze fusion, whereas if T =15, fusion requires concerted action of all trimers of the virion. On the other hand, fusion could be faster when T is larger because of summation of multiple gp160-associated reductions in energy barriers between fusion intermediates. T may also be the approximate number of trimers in the local region of the viral membrane undergoing fusion. This hypothesis is correlated with electron micrographs of virions that are bound to target cells prior to fusion. The inter-membrane electron density in the micrographs has been interpreted to be due to a cluster of gp160 trimer ectodomains bound to cellular receptors[57].

Figure 3.1: Amino acid sequence of the Wild-type FP_HM construct with 50 residues per line. The sequence is from the HXB2 laboratory strain of HIV. The DNA sequence is displayed in Fig. 3.2. The construct contains HIV gp41 residues 1–70 and 117–172 separated by SG2RG2 non-native linker, with additional G6LEH6 non-native C-terminal tag for affinity purification. Residues 1–70 with gp41 numbering correspond to 512–581 with gp160 numbering, and 117–172 correspond to 628–683 numbering. Much of FP_HM adopts thermostable hairpin structure like the final state of the soluble ectodomain of gp41 without gp120 (Fig. 3.14). For gp41, the hairpin has a N-helix (residues 24–84/535–595) and C-helix (residues 104–164/615–675) that are in close contact and antiparallel with an intervening turn (Fig. 3.14). There is color coding of the FP_HM segments that are part of N- and C-helices, and different colors for regions that contain the N-terminal fusion peptide and membrane-proximal external region that respectively bind target cell and viral membrane.

It isn't clear why the V2E N-terminal point mutation of gp41 greatly attenuates gp160-mediated fusion and HIV infection, in part because there aren't clear data about the differences in V2E vs. WT structures. Structure-function studies to-date have focused on ~25-residue peptides with the Fp sequence. Peptide-induced vesicle fusion was attenuated for V2E vs. WT, but to our knowledge, V2E dominance has not been observed for WT + V2E peptide mixtures[58–60]. Comparative structural studies have been done for the WT and V2E peptides in detergent and in membrane. For either peptide, there are populations of molecules with either predominant helical or β-sheet structures. The different studies present conflicting data about the helical : β sheet ratio for V2E vs. WT, and include reports of larger, smaller, or no change in ratio[60–64]. There are similarly

conflicting data about whether relative to WT, V2E has shallower or deeper location in the micelle or membrane. The present study addresses unresolved questions about the mechanism of HIV fusion and its V2E-dominant attenuation, and includes a new model for V2E attenuation based on data from the present and earlier studies. The study focuses on the large Ed "FP_HM" construct which includes the Fp followed by the hyperthermostable (T_m > 100 °C) hairpin, i.e. the final gp41 structural state (Fig. 3.1) [25,33,44,65]. Earlier studies showed that vesicle fusion induced by FP_HM has significant similarities with gp160-induced cell/cell fusion, and under some conditions, FP_HM induces vesicle fusion with protein:lipid mole ratio ≈1:5000 similar to the gp160:lipid ratio in virions[33]. The present study investigates whether fusion is highly-attenuated for V2E FP_HM and also whether V2E is dominant in mixed WT/V2E FP_HM trimers. Observation of these effects would support an important catalytic role for the final gp41 hairpin state in HIV fusion and also provide other information about the fusion mechanism.

3.2: Methods and materials

3.2.1: Materials

Commercial sources included: DNA, GenScript (Piscataway, NJ); Escherichia coli (E. coli) Novagen (Gibbstown, NJ); Luria-Bertani (LB) medium, Dot Scientific (Burton, MI); isopropyl β-D-thiogalactopyranoside (IPTG), Goldbio (St. Louis, MO); and Co2+ affinity resin, Thermo Scientific (Waltham, MA). Avanti Polar Lipids (Alabaster, AL) was the source for lipids that included 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine (POPC), 1-palmitoyl-2-oleoyl-sn-glycero-3- [phospho-rac-(1- glycerol)] (sodium salt) (POPG), N-(7-nitro-2,1,3-benzoxadiazol-4-yl) (ammonium salt) Dipalmitoylphosphatidylethanolamine} (N-NBD-DPPE) and N-(lissamine Rhodamine B sulfonyl) (ammonium salt) dipalmitoylphosphatidylethanolamine} (N-Rh-DPPE). Other materials were typically purchased from Sigma-Aldrich (St. Louis, MO).

3.2.2: FP_HM constructs, expression, and purification

FP_HM proteins were produced using E. coli BL21(DE3) cells that were typically incubated at 37 °C in LB medium with kanamycin (50 mg/mL) and shaking at 150 rpm. Fig. 3.2 displays the DNA sequence that coded for the Wild-type (WT) FP_HM sequence and non-native C-terminal G6LEH6.

1
GCCGTGGGTATCGGTGCTCTGTTCCTGGGTTTCCTGGGTGCTGCTGGTTCGACGA
TGGGTGCCCGCTCAATGACG
76
CTGACGGTCCAAGCACGTCAGCTGCTGAGCGGCATTGTGCAGCAACAGAACAATC
TGCTGCGCGCGATCGAAGCC
151
CAACAGCATCTGCTGCAGCTGACCGTTTGGGGTATTAAACAACTGCAGGCTCGTAT
CCTGAGCGGCGGTCGCGGC
226
GGTTGGATGGAATGGGATCGTGAAATTAACAATTATACGAGCCTGATTCACTCTCT
GATCGAAGAAAGTCAAAAC
301
CAACAGGAGAAAAACGAACAGGAACTGCTGGAACTGGACAAATGGGCCTCCCTGT
GGAACTGGTTTAACATTACG
376
AACTGGCTGTGGTACATCAAAGGCGGCGGTGGCGGTGGTCTCGANCACCACCAC
CACCACCAC

Figure 3.2: DNA sequence of Wild-type FP_HM including non-native G6LEH6 C-terminal tag. For the V2E mutant, the T nucleotide at the fifth position is replaced by A.

The WT sequence with G6 was subcloned into a pET-24a (+) plasmid with Lac operon and kanamycin antibiotic resistance and the plasmid with insert was transformed into E. coli cells. After culture growth in LB medium, the plasmid was extracted from a culture aliquot followed by PCR with DNA primers (Fig. 3.3) to produce a plasmid with V2E + G6 DNA that was subsequently transformed into *E. coli* cells. The WT and V2E inserts were confirmed by DNA sequencing of their respective extracted plasmids. WT and V2E culture stocks were separately prepared by streaking cells onto an agar plate, overnight growth, single colony selection, overnight growth in 50 mL LB medium, and mixing 1 mL culture aliquot and 0.5 mL 50% (*v*/v) glycerol with storage at -80 ^{o}C. FP_HM production began with addition of 75 µL bacterial glycerol stock to 50 mL LB medium, ~7 h culture growth, transfer to 1 L fresh LB medium, and ~ 3 h growth. The growth times resulted in final

OD600's of ~0.8. Expression was induced by addition of 2 mmole IPTG followed by incubation at 37 °C for 6 h. The cell pellet was harvested by centrifugation at 9000 *g* for 15 min and stored at -20 °C. For subsequent steps, the PBS pH was 7.4, tip sonication was done with a surrounding ice bath, and centrifugation was done with 48,000 g for 30–45 min at 4 °C. FP_HM has low solubility in PBS without additives and was separated from PBS-soluble molecules by 3 × repetition of sonication of 5 g cell pellet in 30 mL PBS followed by centrifugation and harvesting the new pellet. The final pellet with FP_HM was then solubilized by sonication in PBS with 8 M urea followed by centrifugation and collection of the supernatant. Some dissolved molecules including FP_HM were then precipitated by increasing [NaCl] to ~300 mM. Subsequent steps were: (1) centrifugation, harvesting, and then vortexing the pellet in distilled water; (2) centrifugation, harvesting, and then pellet dissolution in 20 mL PBS with 6 M GuHCl; (3) centrifugation and collection of the supernatant; and (4) Co^{2+} affinity chromatography. The chromatography was done in a fritted column with liquid removal by gravity filtration, and all solutions contained PBS with 6 M GuHCl. There was: (1) filtration of 5 mL of resin suspension; (2) resin washing with 10 mL of 10 mM imidazole; (3) addition of the FP_HM-containing supernatant followed by agitation overnight at 4 °C; and (4) sequential washing/elution in 15 mL without imidazole, 10 mL of 10 mM imidazole, 5 mL of 300 mM imidazole, and 10 mL of 600 mM imidazole. The elutions with 300 and 600 mM imidazole were combined and dialyzed overnight against distilled water with consequent precipitation, and the suspension was centrifuged, followed by harvesting and lyophilization of the pellet, and storage at -20 °C.

Forward Primer Sequence: 5'- CAT ATG GCC GAG GGT ATC GGT G- 3'

Reverse Primer Sequence: 5'- CAC CGA TAC CCT CGG CCA TAT G- 3'

Figure 3.3: Primer nucleotide sequences to produce V2E_FPHM mutant.

3.2.3: Preparation of FP_HM samples including WT/V2E mixtures

Lyophilized FP_HM (either WT or V2E) was typically dissolved in 10 mM Tris buffer at pH 8.0 with 0.17% n-decyl-β-D-maltoside, 2 mM EDTA, and 1 M *L*-arginine. This solution was agitated overnight at 4 °C and then dialyzed against 10 mM Tris buffer at pH 7.4 with 0.2% SDS (Tris/SDS) which is a solution in which FP_HM is predominantly trimeric[25]. The dialysis was for four days with one buffer change. [FP_HM] was then measured using

A280 and sometimes adjusted with dilution or concentration. Preparation of WT/V2E mixed trimers began with separate solutions of WT and V2E in 200 mM Tris buffer at pH 6.8 with 8% SDS and 400 mM dithiothreitol, which is the solution used for SDS-PAGE in which FP_HM is predominantly monomeric. Mixed trimers with a specific fV2E were formed by combining the appropriate volumes of WT and V2E solutions, followed by dialysis against Tris/SDS.

3.2.4: Circular dichroism spectroscopy, Size-exclusion chromatography, and Vesicle fusion

Circular dichroism (CD) spectra of FP_HM in Tris/SDS were obtained with a J-810 spectrometer (Jasco; Easton, MD) with sample temperature controlled by a water bath. The quartz cuvette pathlength was 1 mm, and ultraviolet wavelength was scanned in 0.5 nm steps between 190 and 250 nm. Size-exclusion chromatography (SEC) was done using a DuoFlow Pathfinder 20 instrument (Bio-Rad; Hercules, CA) and HiLoad 16/600 Superdex 75 pg column (Cytiva; Marlborough, MA). FP_HM in Tris/SDS (1 mg/mL) was centrifuged at 9000 g for 30 min and chromatography was done with 100 µL injection of the supernatant, Tris/SDS running buffer with flow rate of 1 mL/min, and A280 detection. Preparation of “unlabeled” POPC:POPG (4:1) unilamellar vesicles for vesicle fusion included freeze-thaw cycles (~10×) followed by extrusion (~10×) through a filter with 100 nm diameter pores. A separate set of “labeled” vesicles was similarly prepared and contained additional *N*-NBD-DPPE fluorescent lipid and *N*-Rh-DPPE quenching lipid, with ~0.02 mol fraction for each labeled lipid. The POPC:POPG (4:1) composition was chosen because: (1) these are common lipids used for earlier studies of viral fusion protein constructs; (2) phosphatidylcholine is a common lipid in membranes of host cells of HIV; and (3) the negatively-charged lipid fraction of host cell membranes is represented by the POPG which is negatively-charged over the pH 3.3–7.4 range of the present study[44,66-69]. The target cell outer membrane normally has little anionic lipid, but there is evidence for scramblase-mediated transport of anionic phosphatidylserine lipid from the inner to outer leaflet prior to gp160-mediated fusion[70]. A vesicle solution was prepared with labeled:unlabeled ratio =1:9, [total lipid] ≈ 150 µM, and typical 25 mM citrate buffer at pH 5.3. The solution was placed in the fluorimeter, warmed to 37 ^{o}C, and fluorescence (F) measured at 530 nm using excitation at 467 nm. Initial fluorescence is denoted F_0. An

aliquot of stock with [FP_HM] ≈ 150 μM in Tris/SDS was added to the vesicle solution at time *t* =0, followed by measurement of F(t) to t ≈ 600 s in 1 s increments, with larger uncertainties for t < 5 s. Because the ratio of labeled:unlabeled vesicles =1:9, a labeled vesicle likely fuses with an unlabeled vesicle, and fluorescence consequently increases with fusion because of the larger separation between fluorescent and quenching lipids. This inter vesicle lipid mixing correlates with lipid mixing that is sometimes assayed and observed in cell-cell fusion and which is a feature of steps of stalk formation and hemifusion for the fusion mechanism described in the Introduction. Specifically, stalk formation results in lipid mixing between the outer leaflets of the viral and target membranes, and hemifusion results in mixing between the inner leaflets of these membranes. FP_HM-induced vesicle fusion was typically nearly complete by t ≈600 s, and the associated fluorescence extent (F_{ext}) was approximately time-invariant. A 35 μL aliquot of 20% Triton X-100 detergent was then added and dissolved the vesicles, with consequent maximum fluorescence (F_{max}) associated with the large increase in separation between fluorescent and quenching lipids. By literature convention, time-dependent percent vesicle fusion is calculated as $M(t) = 100 \times [F_{(t)} - F_0]/[F_{max} - F_0]$ with long-time fusion extent (M_{ext}) calculated using $F(t) = F_{ext}$. The $F \propto \langle 1/ [1 + (R_{For} / R)^6] \rangle$ where R is fluorophore/quencher distance, R_{For} is the Forster distance (~50 Å), and $\langle .. \rangle$ is the average over all fluorophores[71]. For the initial F_0 conditions, the estimated $\langle R \rangle \approx R_{For}$ and fusion between two vesicles results in $\langle R \rangle \approx (2)^{1/2} \times R_{For}$ with resulting Mext ≈ 80%. There is typical 2% variation in M_{ext} for replicate assays using the same vesicle and FP_HM stocks[23]. Assays with different WT:V2E ratios were done on the same day using the same vesicle stock.

3.3: Results

3.3.1: Purified FP_HM

Fig. 3.4 displays SDS-PAGE of (A) WT and (B) V2E FP_HM at different stages of purification. The samples for the gel lanes had been solubilized in PBS and 8 M urea and were: (i) the cell pellet after removing PBS-soluble molecules; (ii) the precipitate formed in 300 mM NaCl; and (iii) the precipitate formed after the combined 300 and 600 mM imidazole elutions from the Co^{2+}-resin had been dialyzed against water.

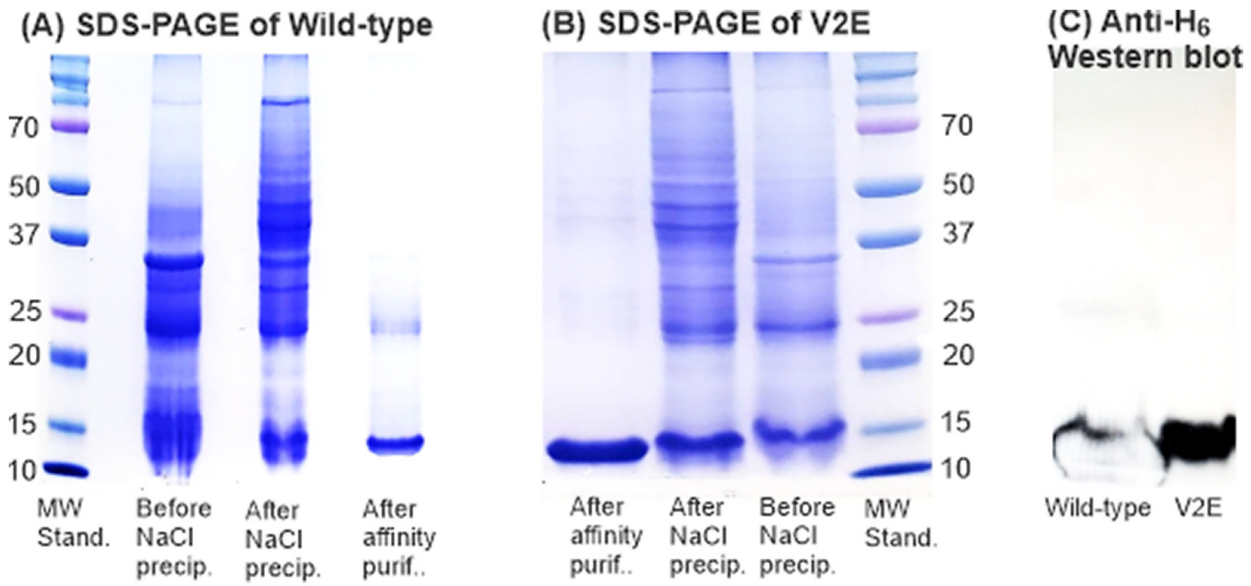


Figure 3.4: SDS-PAGE at different steps of purification of (A) wild-type and (B) V2E FP_HM and (C) anti-H6 western blotting of SDS-PAGE of both proteins after Co^{2+} affinity purification. The samples for the gel lanes had been solubilized in PBS and 8 M urea and were: (i) the cell pellet after removing PBS-soluble molecules; (ii) the precipitate formed in 300 mM NaCl; and (iii) the precipitate formed after the combined 300- and 600-mM imidazole elutions from the Co^{2+}-resin had been dialyzed against water.

The (iii) lanes are the most important because this is the protein used in the subsequent experiments. Both WT and V2E exhibit a prominent band with apparent MW of ~15 kDa which was confirmed to be FP_HM by (C) anti-H6 Western blot and by 88% coverage of the FP_HM sequence by peptides from trypsin digestion of the band (Fig. 3.5). The preparation procedure for the Western blot has been previously described[33]. The 15 kDa band is observable in the (i) cell pellet lane for V2E and in the (ii) NaCl-induced precipitate for WT. When Co^{2+}-affinity purification was done without the NaCl precipitation, the elutions showed more impurities, particularly for WT. The yield after Co^{2+}-affinity purification was ~4 mg/L culture for WT and ~ 8 mg/L for V2E which correlates with a smaller cell pellet for WT vs. V2E. Relative to V2E, WT may be more fusion-active in the E. coli cells and have higher cytotoxicity.

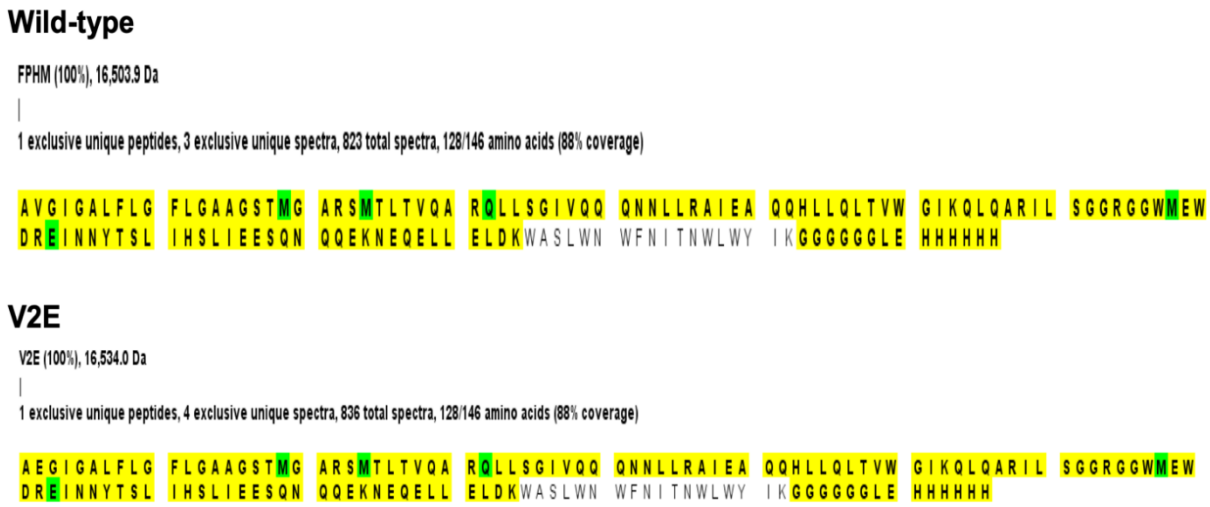


Figure 3.5: Protein sequence coverage from mass spectrometry of peptides formed from trypsin digestion. The highlighted yellow residues were in a peptide. The highlighted green residues were in peptides but had mass changes consistent with oxidation (M), deamination (Q), or dehydration (E).

3.3.2: WT is more helical than V2E and both proteins are hyper thermostable

Fig. 3.6 displays CD spectra of (A) WT and (B) V2E in Tris/SDS that were acquired between 22 and 95 °C. All spectra were acquired on a single day. At all temperatures, the WT and V2E spectra have a double-well shape that is characteristic of predominant helical structure. For WT at 22 °C, the magnitude of the mean-residue-molar-ellipticity at 222 nm ($|\theta_{222}|$) is 2.25×10^4 deg-cm^2-dmol^{-1}, which corresponds to 68% helicity using 3.3×10^4 deg-cm^2-dmol^{-1} for 100% helicity[72]. The 68% value is very similar to the 66% helicity calculated using previous structural data. In particular, residues 23–70 and 117–164 are continuous N- and C-helices in crystal structures of hairpin constructs that are similar to HM[10]. Earlier CD and NMR studies support predominant β -strand conformation for Fp residues, and non-helical structure is expected for the non-native SGGRGG loop and G_6LEH_6 tag[21,23–25]. For V2E at 22 °C, $|\theta_{222}| \approx 1.95 \times 10^4$ deg-cm^2-dmol^{-1}, which corresponds to 59% helicity.

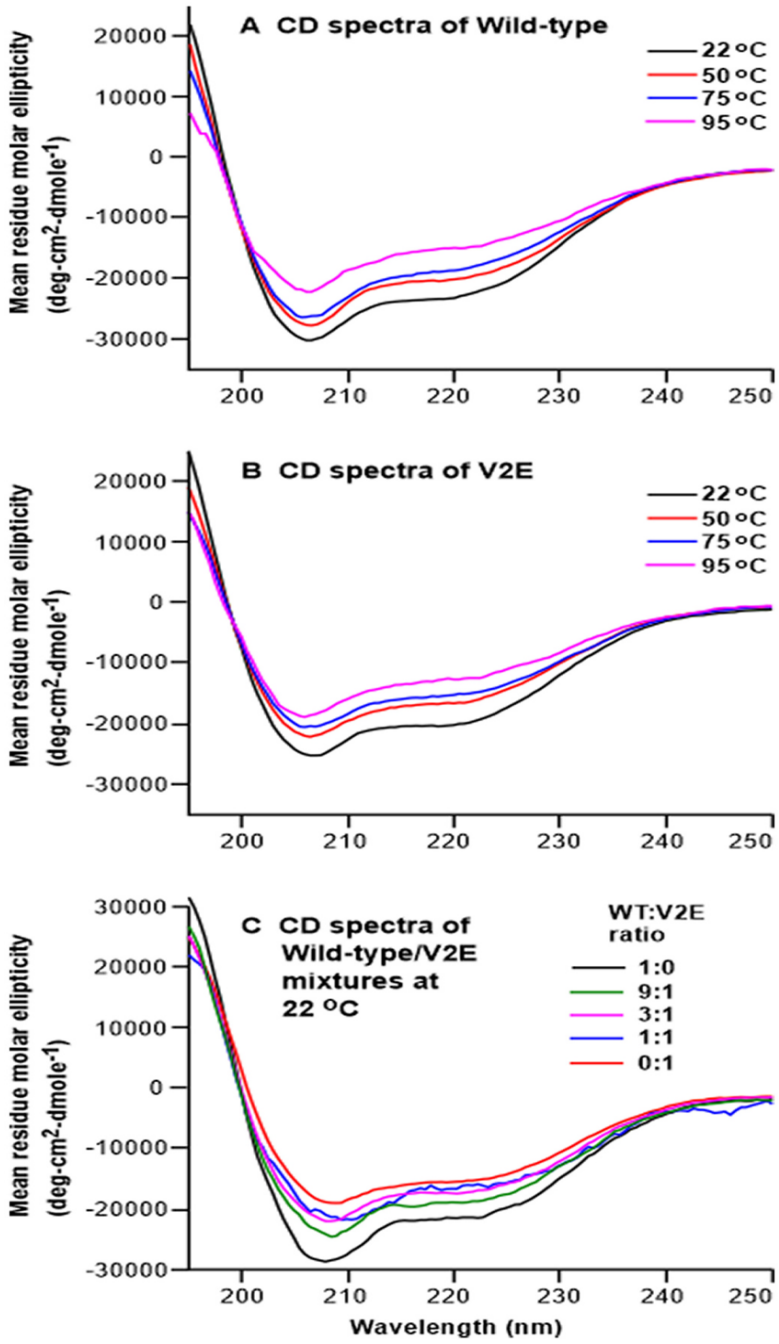


Figure 3.6: Circular dichroism spectra of (A) Wild-type, (B) V2E, and (C) mixtures of Wild-type and V2E FP_HM. A sample typically contained [FP_HM] between 1 and 5 μM in 10 mM Tris at pH 7.4 with 0.2% SDS. The pure Wild-type and V2E samples in panels A and B were at temperatures between 22 and 95 °C. Spectra of replicate samples are displayed in Fig. 7. The mixtures in panel C were at 22 °C and were formed by co-dissolution in 8% SDS in which FP_HM is predominantly monomeric (Fig. 3.4 A, B) followed by dialysis into

Figure 3.6 (cont'd)

0.2% SDS to form mixed trimers (Fig. 3.8). All of the spectra in panels A-C are characteristic of predominant helical structure. The magnitude of the mean residue molar ellipticity at 222 nm ($|\theta_{222}|$) is the parameter typically used to quantitate helicity. At 22 °C, $|\theta_{222}|$ is ~15% smaller for (A) WT vs. (B) V2E and exhibits (C) V2E-dominant reduction in mixed WT/V2E trimers. The $|\theta_{222}|$ decreases with increasing temperature for both (A) WT and (B) V2E, but retains the wavelength-dependent profile that is characteristic of helical structure. These data evidence some helical disordering but not transformation to a random coil structure, and therefore T_m > 95 °C.

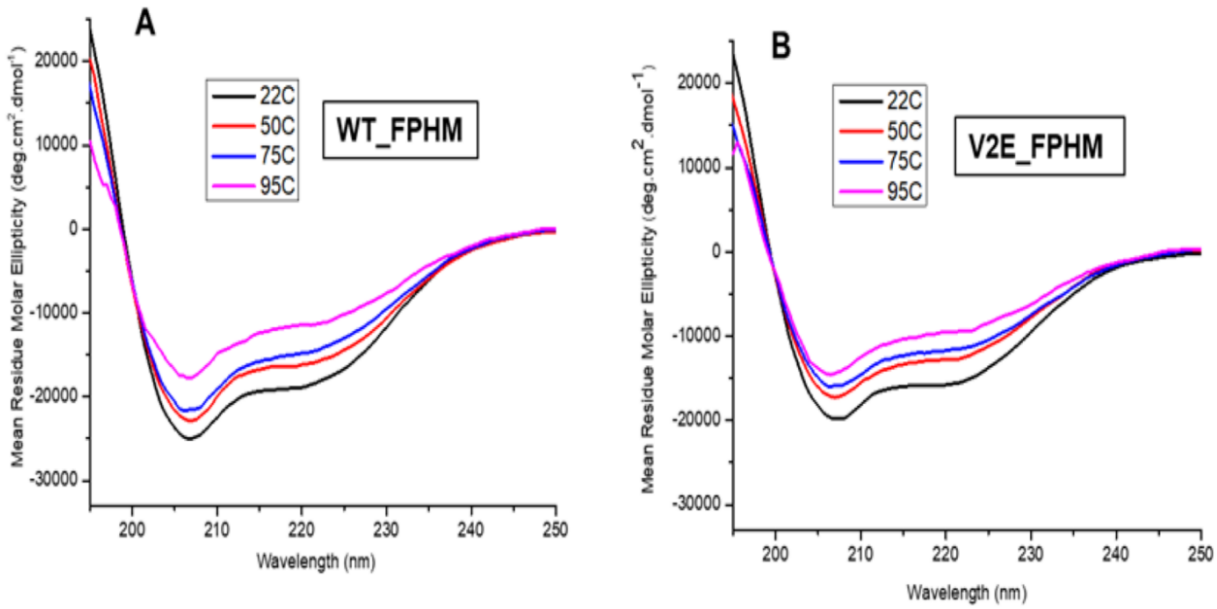


Figure 3.7: CD spectra of (A) Wild-type and (B) V2E mutant FP_HM at temperatures between 22 and 95 ^{0}C. All spectra were recorded on the same day with [protein] ≈ 6 μM in 10 mM Tris buffer at pH 7.4 with 0.2% SDS. The WT and V2E FP_HM are from different preparations than those used for the CD spectra in Fig. 3.6 A,B. The spectra are consistent with helical structure based on shallow minima near 208 and 222 nm. At 22 °C, the θ_{222} values for WT and V2E are 1.82×10^4 and 1.54×10^4 deg-cm^2-dmol^{-1}, respectively, which correspond to 55% and 47% helicity. The CD spectra in Fig. 3.6 A,B in similarly show 9% greater helicity for WT vs. V2E mutant FP_HM.

The reproducibility of greater WT vs. V2E helicity is evidenced by Fig. 3.7 CD spectra of FP_HM proteins from different purifications than in Fig. 3.6. The WT and V2E spectra

between 50 and 95 °C are also consistent with predominant helical structure, but the |θ| values become smaller as temperature increases, with $|\theta_{222}|$95 °C /$|\theta_{222}|$22 °C ≈ 2/3 for spectra of all WT and V2E samples. The CD spectra support T_m > 95 °C for WT and V2E, which is consistent with the previously reported WT T_m ≈ 110 °C from differential scanning calorimetry[43,44].

3.3.3: WT and V2E are predominantly trimeric in Tris/SDS

Earlier studies showed there can be many different oligomeric states of gp41 hairpin constructs in aqueous solution, including monomer, dimer, trimer, hexamer, and dodecamer[22,25,33,37,73,74]. The distribution of oligomeric states depended on the gp41 regions in the construct, concentration of the construct, and additives and pH in the solution. For the present study, the distribution of oligomeric states in Tris/SDS at pH 7.4 was studied because earlier SEC showed predominant trimers for WT in this environment, and hairpin trimers may be the fusion-active state of gp41[25]. Fig. 3.8 displays the SEC profiles for WT and V2E, with down arrows marking the positions of peaks of MW standards. Fig. 3.9 displays the SEC profiles for these standards. The dominant SEC peak for WT and V2E has apparent MW ≈ 90 kDa which was calculated using interpolation between MW standards and $K_{av} \propto \log (MW)$.

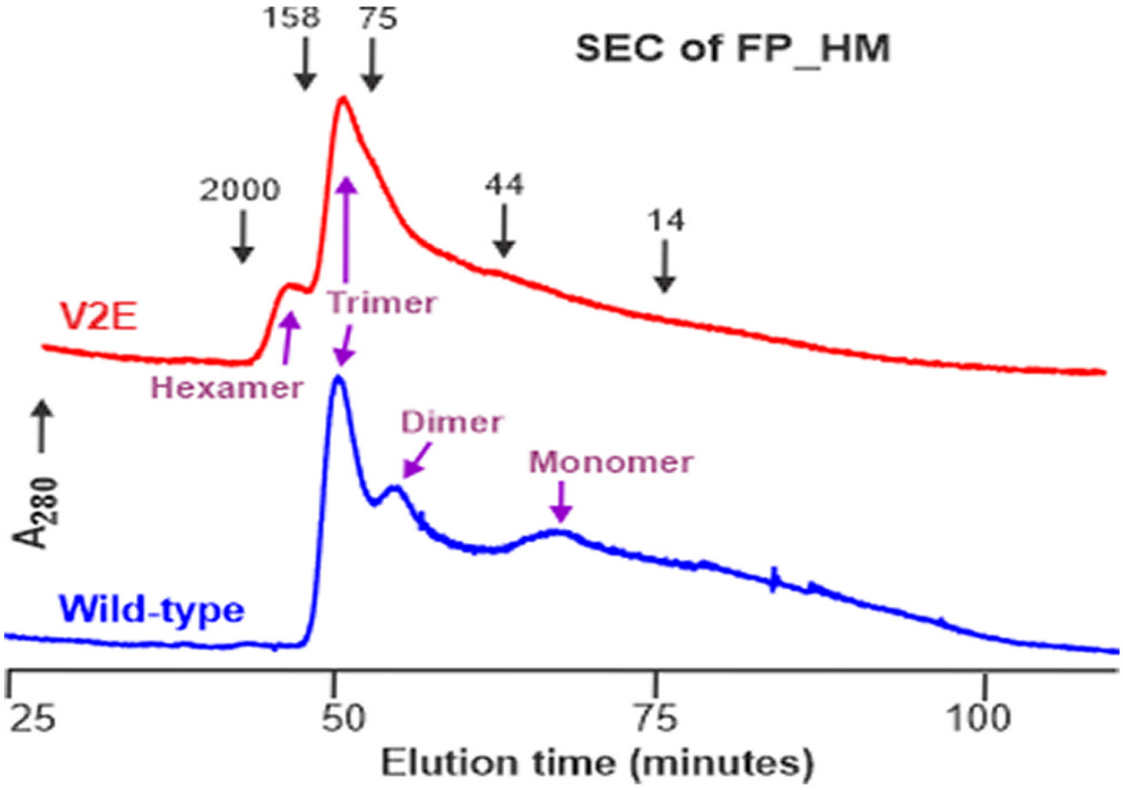


Figure 3.8: Size-exclusion chromatography of FP_HM in 10 mM Tris at pH 7.4 with 0.2% SDS. There are (blue) Wild-type and (red) V2E traces. FP_HM at 1 mg/mL was loaded on a Superdex-75 column with ~10 × dilution in running buffer and final A280 detection. Down-arrows mark the peak elution times of MW standards in kDa (Fig. 3.9). The MW's of peaks in the FP_HM traces were estimated by interpolation between MW standards using log (MW) dependence of elution time. The largest peak for both Wild-type and V2E is at ~90 kDa and assigned to FP_HM trimer with smaller peaks at ~160 kDa (hexamer) for V2E and ~ 70 kDa (dimer) and ~ 30 kDa (monomer) for Wild-type. The chromatograms are similar to earlier data for Wild-type but have broader peaks that may be due to absence of 150 mM NaCl in the running buffer for the present data and its presence for the earlier data. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.

This was also the dominant peak in the earlier WT SEC and is assigned as the trimer[25]. There is a smaller peak for V2E with MW ≈ 165 kDa that is assigned as the hexamer, and smaller peaks for WT with MW ≈ 70 kDa and 35 kDa that are respectively assigned as

dimer and monomer. Peaks with similar masses as the assigned species have been observed in earlier SEC of FP_HM and HM in DPC detergent at pH 4[25]. The monomer, dimer, trimer, and hexamer assignments correlate with SDS mass contributions of ~20, 35, 40, and 65 kDa, respectively. It is likely that most of the SDS molecules bind the apolar Fp and Mper segments of FP_HM. The mass of these segments is ~3 kDa per FP_HM molecule, and the SDS:Fp +Mper mass ratios with these assignments are within the previously-reported range of 3–7 for SDS-bound membrane protein segments[75]. We chose Tris/SDS as the buffer of stock FP_HM for vesicle fusion assays, based on predominant folded helical trimer for WT and V2E (Figs. 3.6 and 3.8), and the likelihood that this is an important catalytic structure of gp41 during fusion. Although some proteins unfold in SDS and others adopt increased helical structure, neither of these effects are apparent in the Figs. 3.6 and 3.8 data. As noted above, the θ_{222} of WT in SDS at 22 ^{o}C matches the fractional helicity calculated using the lengths of helical segments of the crystal structure of a HM-like construct. The crystal was grown in the absence of detergent[10]. Also, the Fig. 3.6 CD spectra of (A) WT and (B) V2E support T_m > 95 ^{o}C, which is consistent with Tm ≈ 110 ^{o}C detected for FP_HM and similar constructs in the absence of detergent[25,44]. In addition, earlier crystal structures and CD studies of gp41 constructs that are smaller than FP_HM show either no change or small increase or decrease in helicity with vs. without detergent, where SDS was sometimes the detergent[24,76].

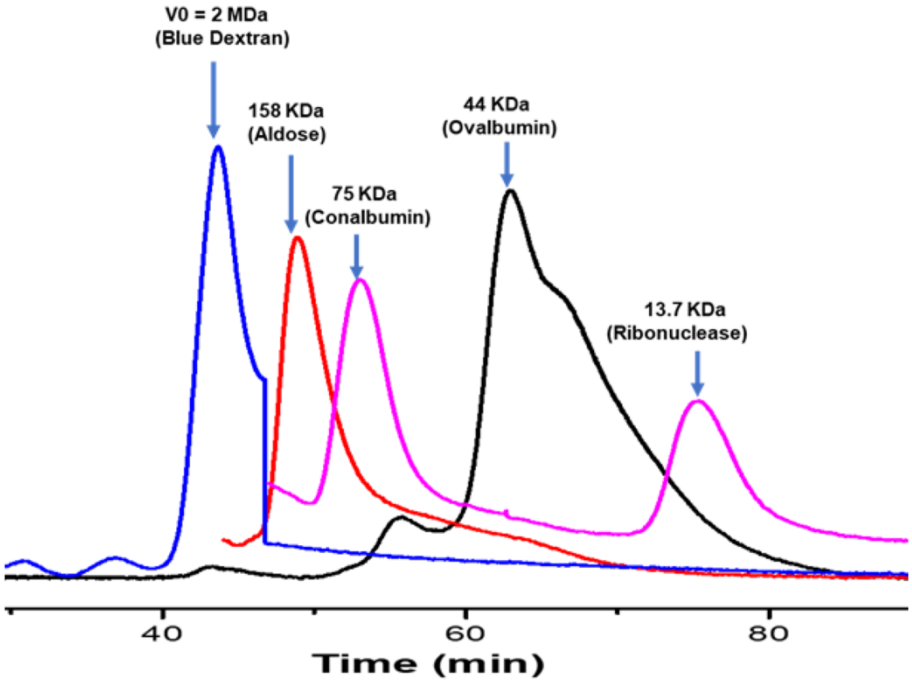


Figure 3.9: Size-exclusion chromatography of molecular weight standards using A280 detection.

3.3.4: FP_HM induces highest vesicle fusion extent near pH 5 with much greater fusion for WT vs. V2E

Fig. 3.10 displays vesicle fusion induced by WT and V2E in the pH range of 3.3–7.4 and with FP_HM:lipid =1:100 mol ratio. For all pH's, WT induces greater fusion extent (M_{ext}) than V2E. For both WT and V2E, Mext exhibits a maximum at pH 5.3 with lower extents at lower and higher pH. This is different than an earlier study for similar gp41 constructs which showed monotonic decrease in fusion extent as pH was raised from 3.5 to 7[45]. These different dependences of fusion extent vs. pH may be related to the initial protein trimer state for the present study vs. predominant monomer state in the earlier study[33]. For the present study, the initial rate of fusion is inversely correlated with pH, which was also observed in an earlier study[23]**.** In addition, the M_{ext}, WT/M_{ext}, V2E ratio becomes larger as pH is increased. This trend is most apparent by comparing pH 4.3 and 6.3 data which both have M_{ext},WT ≈ 30% whereas M_{ext},V2E ≈ 17% at pH 4.3 and 7% at pH 6.3. Subsequent assays were done at pH 5.3 where fusion is maximal.

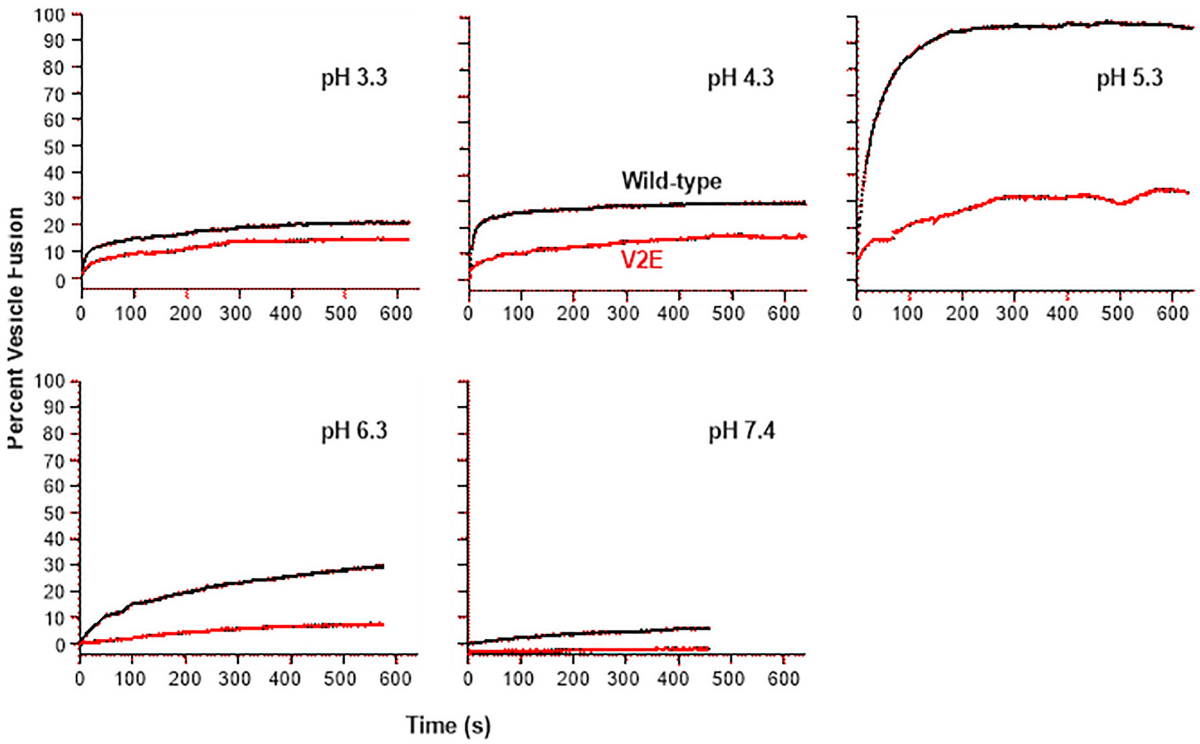


Figure 3.10: Vesicle fusion induced by Wild-type (black) and V2E mutant (red) FPHM at pH's between 3.3 and 7.4. Assays were done at 37 °C using vesicles that had been extruded through 100 nm diameter pores. The vesicles contained POPC:POPG (4:1) with [total lipid] = 150 μM. The vesicle solution contained 25 mM citrate buffer at pH = 3.3, 4.3, 5.3, and 6.3-, and 10-mM Tris buffer at pH = 7.4. At time = 0, an aliquot of FPHM stock solution was added to the vesicle solution to achieve FPHM: lipid with 1:100 M ratio. The stock had 150 μM FPHM at pH 7.4 with 0.2% SDS. The small apparent negative fusion for V2E at pH 7.4 is likely the result of fluorophore dilution because of increased volume of the vesicle solution upon addition of FPHM stock.

3.3.5: Mixed WT/V2E trimers exhibit V2E-dominant reduction of vesicle fusion and helicity

There has been 30-year interest in the striking gp41 V2E-dominant reduction of gp160-induced cell/cell fusion and HIV infection. These data have been analyzed to estimate the number of gp160 trimers required for fusion and infection. We tested whether there are also V2E-dominant reductions of FP_HM-induced vesicle fusion and helicity. These experiments required mixed WT/V2E trimers, so WT and V2E were first co-dissolved in SDS-PAGE buffer in which both proteins are predominantly monomeric (Fig. 3.4) and

then dialyzed into Tris/SDS in which the proteins are predominantly trimeric (Fig. 3.8). Fig. 3.11 displays vesicle fusion with FP_HM:lipid = 1:100 and different WT:V2E ratios and Fig. 3.6C displays CD spectra with different WT:V2E ratios. Both datasets demonstrate that V2E causes dominant reductions of fusion and helicity. The V2E dominance is supported by fusion data obtained with WT and V2E stocks from other purifications and by data with FP_HM: lipid = 1:300 (Fig. 3.12). The fusion extents for all data are presented in table 3.1.

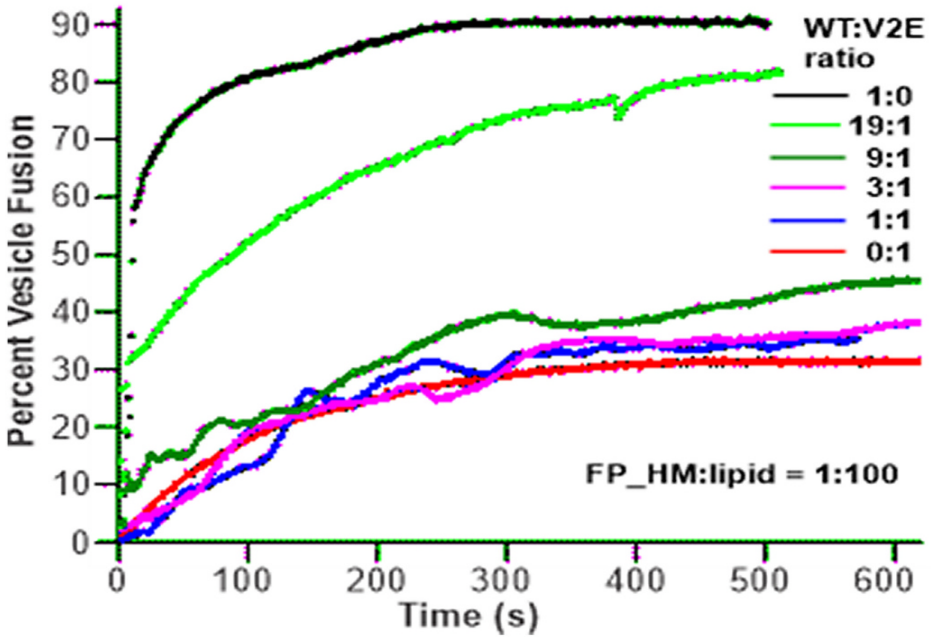


Figure 3.11: Vesicle fusion induced by FP_HM mixed trimers with different Wild-type:V2E ratios. The vesicle solution had 150 μM total lipid, pH 5.3, temperature = 37 °C, and POPC:POPG (4:1) vesicles formed by extrusion through 100 nm diameter pores. At time = 0, an aliquot of FP_HM stock solution was added to the vesicle solution to achieve FP_HM:lipid = 1:100 M ratio. The stock had 150 μM FP_HM at pH 7.4 with 0.2% SDS. Stocks with Wild-type/V2E mixtures were prepared by co-dissolution in 8% SDS with predominantly monomeric FP_HM (Fig. 3.4 A,B) followed by dialysis into 0.2% SDS to form FP_HM mixed trimers (Fig. 3.8). Fig. 3.12 displays vesicle fusion data with different preparations of vesicles and protein stocks, and with FP_HM:lipid either 1:100 or 1:300. The data in Fig. 3.10 and Fig. 3.12 data exhibit similar trends and the long-time fusion

Figure 3.11 (cont'd)

extents are presented in Table 3.1. The fusion extent is <1% when stock buffer without FP_HM is added to the vesicle solution.

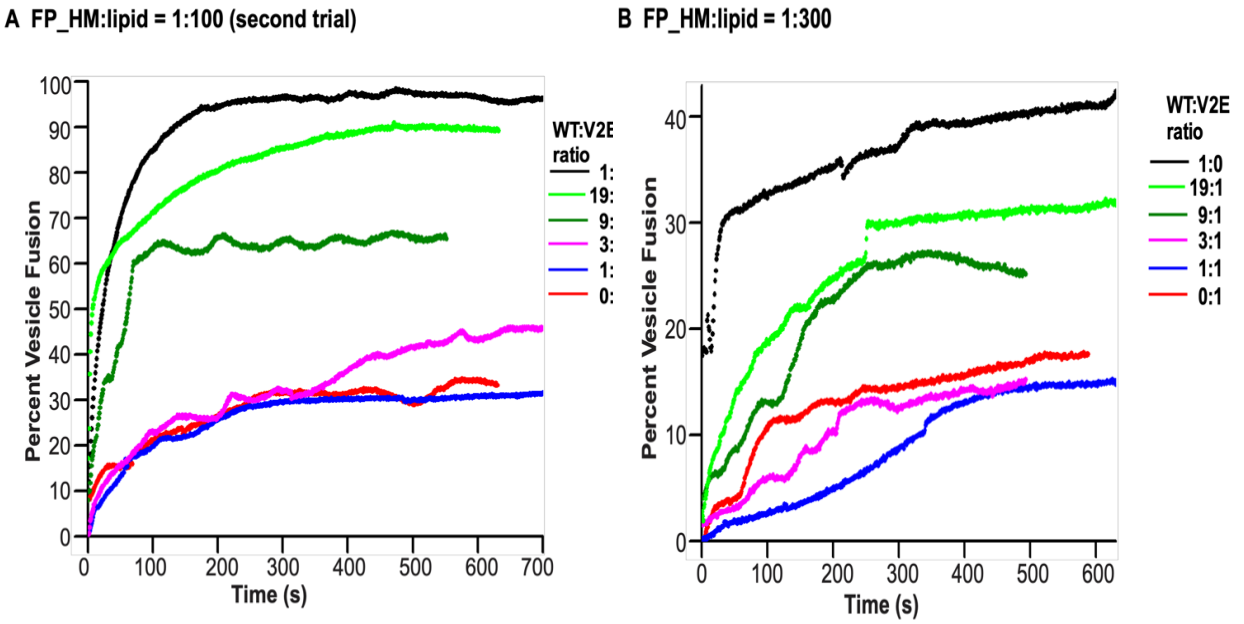


Figure 3.12: Vesicle fusion assays at 37 °C and FP_HM:lipid molar ratios of (A) 1:100 and (B) 1:300. The 1:100 data are a second trial and were obtained with a different batch of vesicles and different FP_HM stocks than Fig. 3.11. At time = 0, an aliquot of FP_HM stock was added to the vesicle solution. The vesicle solution had [total lipid] = 150 μM at pH 5.3 with POPC:POPG (4:1) composition. The FP_HM stock had 150 μM FP_HM in 10 mM Tris buffer at pH 7.4 with 0.2% SDS.

Table 3.1: Percent fusion extent for different V2E mole fractions.

Percent Fusion Extent for different V2E mole fractions.[a]

f_{V2E}	Protein: Lipid = 1:100 (Dataset 1)	Protein: Lipid = 1:100 (Dataset 2)	Protein: Lipid = 1:300
0	90	96	43
0.05	81	86	32
0.10	48	65	25
0.25	38	44	15
0.5	35	31	15
1	32	30	18

[a] Data for each column were obtained using a separate vesicle preparation and a separate protein stock preparation. The $M_{ext} = 100 \times (F_{ext} - F_0)/(F_{max} - F_0)$ where F_0 is the fluorescence prior to protein addition, F_{ext} is the long-time (~600 s) asymptotic fluorescence after protein addition, and F_{max} is the maximum fluorescence after Triton X-100 addition. There is typical ±2% variation in M_{ext} for an assay replicate with the same vesicle preparation and same protein stock preparation.

3.3.6: Quantitative similarity of V2E dominance for FP_HM fusion and helicity, gp160 cell-cell fusion, and HIV infectivity, and global fitting support efficient fusion requiring at least two Wild-type gp41 trimers

It is interesting to understand how quantitatively similar are the V2E dominances for FP_HM-induced vesicle fusion, FP_HM helicity, gp160- induced cell/cell fusion, and HIV infectivity[54]. For each assay type, percent activity (A) is calculated as a function of the fraction gp41 with V2E (f_{V2E}), with A = 100% and 0% when f_{V2E} = 1 and 0, respectively. For cell/cell fusion and infectivity, A = 100 × N (f_{V2E})/N(0), were N = experimental numbers of syncytia and colonies, respectively. For f_{V2E} = 0.5, the A's are ~0, which presumably also holds for f_{V2E} = 1. For FP_HM-induced vesicle fusion, the A (f_{V2E}) = 100 × [M_{ext} (f_{V2E}) – M_{ext} (1)]/[M_{ext} (0) – M_{ext} (1)], and a similar expression is used for FP_HM helicity with θ_{222} substituted for M_{ext}. Table 3.2 displays the A values from the different assays and their associated uncertainties and Fig. 3.13 displays a plot of all A vs. f_{V2E}. A straightforward model for V2E dominance is that activity requires a cluster with number ≡ n WT gp41 molecules, so activity is proportional to the fraction of these clusters with only WT and without V2E gp41 molecules. For this model, A = 100 × $(f_{WT})^n$ = 100 × $(1 - f_{V2E})^n$. Global fitting of all the Fig. 3.13 data yields best fit n = 6.00 ± 0.39, i.e. T = 2 fully WT trimers are required for the greatest cell-cell and vesicle fusion, HIV infection, and helical hairpin structure.

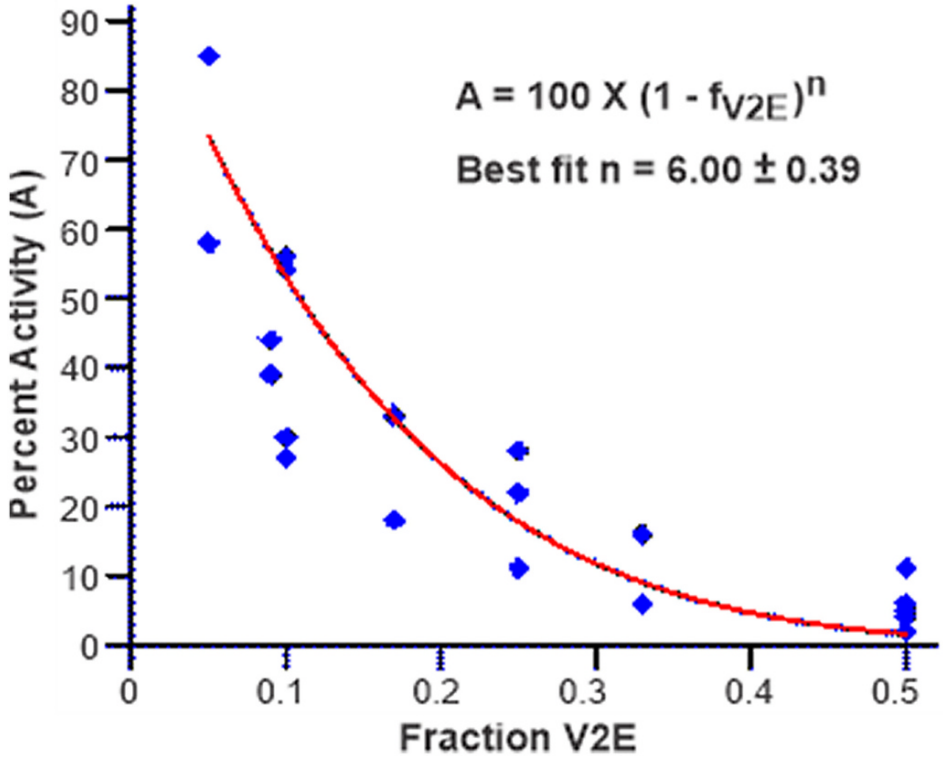


Figure 3.13: Plot of Percent Activity (A) vs. fraction V2E (fV2E) data and fitting to A =100 × $(1 - f_{V2E})^n$. The individual A values and their uncertainties are presented in Table 3.2 and are based on vesicle fusion and circular dichroism data from the present study and earlier cell-cell fusion and HIV infection data. The fitting equation is based on fusion requiring n Wild-type and no V2E molecules. The best-fit n =6.00 ± 0.39, i.e. 6.00 (39). For data subsets from a single approach where f_{V2E} is the only variable parameter (single columns in Table 3.2), best-fit n are: FP_HM-induced vesicle fusion at 1:100 ratio, trial 1, 8.68 (2.36) and trial 2, 5.29 (41); FP_HM-induced vesicle fusion at 1:300 ratio, 11.10 (40); CD-calculated FP_HM helicity, 4.69 (65); gp160-induced cell-cell fusion, 7.14 (40); and HIV infectivity, 6.97(1.49).

Table 3.2: Percent activity for different V2E mole fractions.

Percent activity for different V2E mole fractions.[a]

f_{V2E}	Vesicle fusion P:L = 1:100 (Dataset 1)	Vesicle fusion P:L = 1:100 (Dataset 2)	Vesicle fusion P:L = 1:300	Helicity	Cell-cell fusion	HIV infectivity
0.05	85(6)	85(6)	58(13)			
0.09					39(20)	44(4)
0.10	27(5)	54(5)	30(12)	56(5)		
0.17					18(8)	33(10)
0.25	11(5)	22(4)		28(5)		
0.33					6(1)	16(4)
0.5	6(5)	2(1)		11(5)	4(3)	5(2)

[a] For FP_HM-induced vesicle fusion and FP_HM helicity, Activities (A's) were based on the present study, and for cell-cell fusion and HIV infectivity, on [54]. Uncertainties are in parentheses. For vesicle fusion, $A(f_{V2E}) = 100 \times \{[M_{ext}(f_{V2E}) - M_{ext}(1)]/[M_{ext}(0) - M_{ext}(1)]\}$. The uncertainty of A was calculated using ±2% uncertainty for each M_{ext}. The helicity A was similarly calculated with θ_{222} substituted for M_{ext}. Cell-cell fusion was done with HeLa T4 cells transfected with plasmids coding for the HIV gp160 protein, and f_{V2E} was a fraction of the transfected DNA. Fusion was quantified using the number of cellular syncytia (N), with $A = 100 \times \{N(f_{V2E})/N(0)\}$. The uncertainty of A was the standard deviation from several trials. HIV infectivity was similarly quantified with N = number of colonies after infection and incubation of HeLa T4 cells.

3.4: Discussion

3.4.1: V2E dominance of FP_HM vesicle fusion and helicity supports an important role of the final gp41 hairpin structure in maintaining close membrane apposition prior to fusion

The V2E mutation of gp41 exhibits dominance in both gp160- mediated cell/cell fusion and HIV infectivity[54]. Several different groups have analyzed infectivity data as a function of f_{V2E} with resulting estimates between 1 and 19 gp160 trimers required for infection, with the specific number depending on the assumptions of the model used in the analysis[55]. The major discovery of the present study is V2E dominance for FP_HM in both vesicle fusion and helicity that is quantitatively similar to earlier cellular and viral data (Table 3.2 and Fig. 3.13). FP_HM adopts the hyper thermostable trimer-of-hairpins structure, which is the final state of gp41 without gp120, so this similar dominance supports an important role in viral fusion for the final hairpin structure. These data do not support models in which the hairpin is the "post-fusion" structure that forms after most or all of membrane fusion has occurred. Fig. 3.14 displays a model that correlates the reductions in gp41 helicity and fusion for V2E vs. WT hairpin gp41. Panel A depicts two WT trimers (n =6, T=2) whose hairpins have ~60-residue N- and C-helices, with trimers connected through

an antiparallel β-sheet located in the outer leaflet of the target membrane and composed of six interleaved Fp strands, each with ~17 N-terminal gp41 residues. For clarity, only a single molecule of each trimer is displayed. The hyper thermostability of the hairpin ($T_m \approx$ 110 ^oC), in conjunction with the Fp in target membrane and Mper + Tmd in viral membrane, partially counteracts the ~25 kcal/mol energy for close membrane apposition prior to fusion. The energy for this initial apposition is the highest barrier in many computational studies of membrane fusion[4]. For the Fig. 3.14A model, the final trimer-of-hairpins state is the fusion-catalytic state of gp41 and forms prior to fusion, which is consistent with the results of some earlier experiments and simulations of enveloped virus fusion[46,47]. The N-helix and C-helix in Fig. 3.8A are approximately residues 24–84 and 104–164, respectively, based on high-resolution structures and supported by the WT CD spectra in Fig. 3.6A, as described in the Results section[9,10]. Two trimers for efficient fusion is supported by fitting of the Activity vs. f_{V2E} data (Fig. 3.13 and Table 3.2) and agrees semi-quantitatively with an earlier report of two to three gp160 trimers for efficient infection by HIV primary isolates, i.e. HIV strains obtained from infected patients[55]. One of these isolates was a “transmitted founder” isolate, i.e. an isolate that initiates infection of an uninfected person. In Fig. 3.14, the Fp antiparallel β-sheet structure in membrane is supported by NMR data which also show a distribution of N-terminal strand lengths, with 1 → 17/17 →1 as a major population[18–20]. This registry clusters together the apolar N-terminal Fp residues of the different strands which likely correlates with more favorable free energy of insertion into the membrane hydrophobic core (Fig. 3.14C, right panel, middle and bottom strands). Although FP_HM doesn’t have the Tmd, its C-terminal Mper also binds membrane, and intervesicle lipid mixing could be catalyzed by FP_HM by a similar structure as shown for the Fig. 3.14 A fusion catalyzed by full-length gp41 with Tmd[26,27].

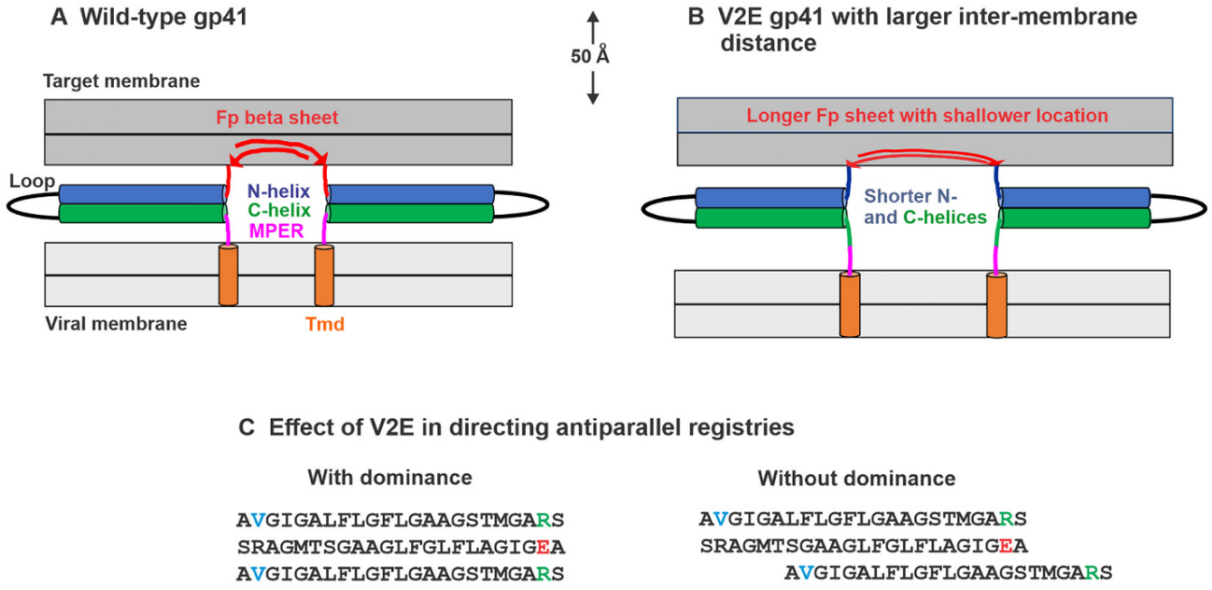


Figure 3.14: (A, B) Model that correlates lower helicity of V2E vs. Wild-type FP_HM with reduced fusion of V2E. The gp41 hairpin reduces the energy of the membrane apposition state that exists just prior to fusion. There are two gp41 trimers with hairpin structure that are joined through an intermolecular Fp β-sheet in the outer leaflet of the target membrane. The Fp strands are antiparallel and interleaved between trimers. For clarity, only one gp41 monomer is shown in each trimer. The model has dominant C-terminal extension of the Fp β _sheet for V2E vs. Wild-Type that results in loss of helicity at the N-terminus of the N-helix and C-terminus of the C-helix. There is consequent larger separation of the viral and target membranes for V2E and higher activation barrier for subsequent stalk formation because of the need for larger-amplitude protrusion of lipid acyl chains into the intermembrane region. Relative to WT, the V2E Fp β-sheet may also have shallower membrane location with consequent lower probability of lipid protrusion. For (A) Wild-type gp41 hairpin, the N- and C-helices are each ~60 residues. There is a distribution of WT Fp antiparallel registries that includes residues 1→17/17→1 which is depicted in panel A and in panel C, right, V2E middle strand/WT bottom strand. This registry clusters together the apolar N-terminal Fp residues of the different strands which likely correlates with more favorable free energy of insertion in the membrane hydrophobic core. A longer 1→23/23→1 registry is depicted for V2E in panel B, and correlates with loss of 6 helical residues in both the N- and C-helices and the

Figure 3.14 (cont'd)

experimentally observed ~9% loss of helicity for V2E vs. WT FP_HM (Fig. 3.6 A,B). This registry is depicted in panel C, left and could correlate with an interstrand E2/R22 salt bridge. The membrane thickness and lengths and widths of the N- and C-helices and Tmd in panels A and B are drawn approximately to the displayed scale. Helices are drawn with ~1.5 Å length per residue, Fp β-strands are drawn with ~3.5 Å length per residue, and other gp41 segments are drawn with ~2.5 Å length per residue. The Mper may also bind the viral membrane, based on existing Mper structures. Panel C displays schematic registry arrangements of three antiparallel Fp strands, WT/V2E/WT either (left) with or (right) without V2E dominance of registries, i.e. 1 →23/23 →1/1 →23 or 1 →23/23 →1/1 →17.

Fig. 3.14B displays a model with the V2E gp41 hairpin, and reduced fusion is predominantly attributed to the larger membrane separation for V2E vs. WT. The Fp β-sheet in V2E is hypothesized to have longer strands, e.g. 1 →23, because of salt bridge formation and/or hydrogen bonding between the E2 sidechain and sidechains of polar residues in the Fp C-terminal region, e.g. R22 (Fig. 3.14C)[20]. The longer and perhaps more stable Fp β-sheet for V2E vs. WT may also be the basis for the observed minor hexamer population for V2E in SDS vs. minor dimer and monomer populations for WT (Fig. 3.8). The β-sheet structure has been previously observed for V2E peptide in membrane, and there are conflicting data about whether there is greater fractional population of helical or β-sheet structure for V2E vs. WT peptides[61,62,64]. NMR data support predominant β-sheet structure for the Fp segment of a membrane-bound construct similar to WT FP_HM[23]. Given that the present study for FP_HM shows lower helicity for V2E vs. WT (Fig. 3.6 A,B and Fig. 3.7 A,B) as well as V2E-dominant reduction in helicity for WT: V2E mixtures (Fig. 3.6 C), our Fig. 3.14 model uses the V2E β-sheet structure. Longer V2E vs. WT β-sheets are supported by some comparative NMR data for V2E vs. WT peptides in membrane[20]. These data also support a distribution of antiparallel registries for V2E as has been observed for WT, but for clarity the following discussion is focused on the 1 →23/23 →1 V2E antiparallel registry. For WT, residues 18–23 are the linker that connects the Fp β-sheet in the target membrane and the hairpin N-helix in the intermembrane region. For V2E, the linker would instead be residues 24–

29 which in WT form the N-terminus of the N-helix. These residues are no longer helical in V2E, and as a consequence, no longer stabilize the helical structure of six residues at the C-terminus of the C-helix. The loss of helicity for 12 hairpin residues for V2E vs. WT approximately matches the experimental ~9% loss in helicity for V2E vs. WT FP_HM (Fig. 3.6 A,B), i.e. 0.09 × 146. The loss of helicity for the C-helix residues results in a ~ 10 Å increase in intermembrane separation with consequent reduction in fusion. Specifically, the stalk intermediate between outer leaflets of the viral and target membranes is hypothesized to form after apposition via transient "protrusion" (excursion) of lipid acyl chains into the aqueous phase[47–50]. For the larger separation with V2E, stalk formation would require larger-amplitude and therefore higher energy protrusion of the chains, and fusion would be hindered. Earlier NMR studies also showed that relative to membrane without Fp, membrane with WT Fp exhibits increased chain protrusion[50]. Other NMR data showed that the V2E Fp β-sheet membrane location is shallower than WT, and this difference may result in attenuated protrusion with V2E gp41[64].

3.4.2: Interleaved fusion peptide strands from two gp41 trimers in an antiparallel β-sheet structure as the basis for V2E dominance

In our view, there hasn't yet been a detailed structural model that explains V2E dominance in HIV fusion and infection, and this intellectual gap is addressed by the Fig. 3.14 model with antiparallel β-sheet with interleaved Fp strands from two trimers. Fig. 3.14C displays schematic registry arrangements of three antiparallel strands, WT/V2E/WT either (left) with or (right) without V2E dominance of registries, i.e. 1 →23/ 23 →1/1 →23 or 1 →23/23 →1/1 →17, respectively. The (left) dominant arrangement might be favored because a R22/E2 salt bridge can form either between the top/middle or bottom/middle strands. If two or three of the six strands in the sheet are V2E, there would be even higher preference for the 1 →23/23 →1 registry. For f_{V2E} =0.5, the maximum number of R22/E2 salt bridges would be achieved when alternating WT and V2E strands only adopt the 1 →23/23 →1 registry. The predicted absence of fusion and infection with only 1 →23/23 →1 registry matches the calculated A =2% when f_{V2E} =0.5 using the best-fit eq. A =100 × $(1 - f_{V2E})^6$, Fig. 3.13. Our model of V2E-dominant reduction in fusion and helicity caused by V2E-dominant longer Fp strand lengths includes the reasonable idea that larger intermembrane separation and perhaps shallowness of membrane location of the Fp β-

sheet are positively correlated with the number of 1 →23/23 →1 registries within the sheet.

3.4.3: Comparison between the gp41 V2E and Ha2 G1E fusion-impairing mutations

The Fig. 3.14 correlation between increased membrane separation and reduced fusion is also evidenced by the similar structural effects of impairing mutations for the influenza Ha2 protein which catalyzes fusion for this virus. Ha2 and gp41 don't share sequence homology but have similar topologies including a viral Tmd and a ~ 180 residue ectodomain that adopts a final-state trimer-of-hairpins structure[77,78]. Like gp41, Ha2 also has a ~ 25-residue N-terminal Fp that binds the target membrane. However, the Ha2 Fp typically adopts helical rather than β-sheet structure in membrane and there isn't close spatial proximity between the helical Fp's of the same or different trimers[69,79–81]. Two extensively studied Ha2 mutations are I173E in the C-terminal region of the hairpin close to the Tmd, and G1E at the N-terminus of the Fp. Both the I173E and G1E mutations greatly reduce Ha-mediated cell/cell fusion, and also greatly reduce cell-cell fusion mediated by the Ha2 ectodomain protein "FHa2" which is analogous to FP_HM[46,82,83]. Vesicle fusion induced by FHa2 is also highly attenuated with these mutations which is similar to the attenuation of vesicle fusion by V2E FP_HM (Fig. 3.10)[81]. Unlike V2E (Fig. 3.6 A,B), both I173E and G1E cause dramatic 35–40 ^{o}C reductions in the T_m of FHa2, so the loss of fusion with either mutation is correlated with less stable hairpin and with larger inter-membrane separation[81]. The reduction in T_m for G1E is surprising because the N-terminus of the WT helical Fp is spatially far from the hairpin. However, H/D exchange data of Ha2 supports a model for reduced hairpin stability in which G1E Fp binding to the C-terminal region of the hairpin displaces the N-helix from the hairpin[78]. G1E may not be a dominant Ha2 mutation, as each Fp adopts independent helical structure[69,80]. To our knowledge, G1E dominance has not been reported for Ha2 fusion.

3.4.4: Greatest fusion between pH 5 and 6 is correlated with retention of FP_HM trimers

Observation of maximum vesicle fusion between pH 5 and 6 for both WT and V2E (Fig. 10) is different from an earlier study which found maximum fusion at pH ≈3 and decreasing fusion as pH is raised, with negligible fusion when pH is in the 5–6 range[45]. The construct of the earlier study was similar to WT FP_HM and similar lipid compositions

were used for vesicles in both studies, with a minor fraction of POPG lipid that has negative headgroup charge over the entire pH range. One difference between studies is the stock protein buffer which was formate at pH 3 in the earlier study vs. Tris/SDS at pH 7.4 in the present study. For both buffers, the stock protein is a thermostable hairpin with T_m > 100 ^oC; however, the protein is predominantly a monomer at pH 3 vs. trimer at pH 7.4[25,33]. This pH-dependent difference is consistent with the calculated charges of ≈+12 at pH 3 vs. -1 at 7.4, and the consequent much larger electrostatic repulsive energy between monomer units at pH 3. At pH 7.4, vesicle fusion is induced by the stock trimer (Fig. 3.10) but is not induced by the stock monomer in the earlier study. We anticipate the stock trimer remains intact because the pH doesn't change. Although the stock monomer could form the lowest free-energy trimer at pH 7.4, a significant fraction of protein likely binds vesicles as monomers and wouldn't adopt the interleaved FP β-sheet structure from two trimers, so there is less compensation of the 25 kcal/mol apposition barrier and consequent attenuated fusion (Fig. 3.14 A). For the present study, fusion was measured in the pH 3.3–7.4 range and the maximum extent of fusion was observed between pH 5 and 6 (Fig. 3.10). Based on the discussion above, the FP_HM oligomeric state may be an important factor in the pH dependence. An earlier study provides insight into the pH dependence of FP_HM oligomerization and relied on analytical ultracentrifugation data in dodecylphosphocholine (DPC) detergent of a construct containing FP_HM plus the transmembrane domain[84]. There was predominant monomer at pH 4, predominant trimer at pH 6, and approximately equal fractions of monomer and trimer at pH 5. Another earlier study showed greater trimer vs. monomer for FP_HM in SDS vs. DPC so we anticipate FP_HM remains a trimer at pH 5.3 where fusion extent is maximal[25]. The more extensive fusion for FP_HM at pH 5.3 vs. higher pH may be due to the +2 charge at pH 5.3 vs. neutral or negative charge at higher pH, with consequent greater FP_HM binding to the negatively-charged membrane at pH 5.3[23,33]. The lower fusion when pH < 5 may be due to FP_HM dissociation into monomers which are probably less fusion-active than trimer. It isn't yet known whether the maximum FP_HM-induced vesicle fusion in the pH 5–6 range is related to HIV fusion within endosomes in this pH range. Fusion within late endosomes is the major infection route for some viruses like influenza, but for HIV there

is disagreement in the literature about whether infection is primarily via fusion with the plasma membrane or within endosomes[85–87].

3.5: Conclusions

The most important result of this study is observation of V2E mutational dominance in reduction of helicity and vesicle fusion induced by FP_HM, which is a hyper thermostable HIV gp41 hairpin construct that includes the Fp and Mper that bind the target and viral membranes, respectively. This mutation had previously been observed to be dominant in reduction of gp160-mediated cell/cell fusion and HIV infection. Both FP_HM and gp160 are predominantly trimers prior to fusion and infection, and mixed trimers with WT and V2E molecules exhibit quantitatively similar dependences of reduced helicity, fusion, and infection on f_{V2E}. Our observation of similar V2E dominance for FP_HM and gp160 supports the final trimer-of-hairpins structure of gp41 as an important state in fusion catalysis. A global fit of the present and earlier V2E data supports efficient fusion and infection requiring 6 WT gp41 molecules, i.e. two gp41 trimers. A model is developed with two gp41 hairpin trimers stabilizing the highest-energy initial apposition of the viral and target membranes prior to fusion. The Mper/Tmd's are in the viral membrane and the Fp's are in the target membrane with intermolecular antiparallel β-sheet structure and interleaved strands from the two trimers. The hyper thermostable hairpins in conjunction with the Fp in target membrane and Mper/Tmd in viral membrane reduce the membrane apposition energy and thereby catalyze fusion. There are longer Fp strands in V2E vs. WT which results in loss of helicity for V2E at the N- and C-termini of the hairpin and consequent larger intermembrane separation. Subsequent formation of the stalk membrane intermediate is impaired with larger separation because of the need for higher-amplitude and therefore higher-energy protrusion of lipid acyl chains into the aqueous region. The V2E dominance of longer strands in the Fp β-sheet is correlated with formation of stabilizing salt bridges and/or hydrogen bonds between E2 and the polar and charged residues near the Fp C-terminus.

REFERENCES

1. R. Blumenthal, M.J. Clague, S.R. Durell, R.M. Epand, Membrane fusion, Chem. Rev. 103 (2003) 53–69.
2. J.M. White, S.E. Delos, M. Brecher, K. Schornberg, Structures and mechanisms of viral membrane fusion proteins: multiple variations on a common theme, Crit. Rev. Biochem. Mol. Biol. 43 (2008) 189–219.
3. R. Blumenthal, S. Durell, M. Viard, HIV entry and envelope glycoprotein-mediated fusion, J. Biol. Chem. 287 (2012) 40841–40849.
4. S. Boonstra, J.S. Blijleven, W.H. Roos, P.R. Onck, E. van der Giessen, A.M. van Oijen, Hemagglutinin-mediated membrane fusion: a biophysical perspective, Ann. Revs. Biophys. 47 (2018) 153–173.
5. L.V. Chernomordik, M.M. Kozlov, Mechanics of membrane fusion, Nat. Struct. Mol. Biol. 15 (2008) 675–683.
6. M. Pancera, T.Q. Zhou, A. Druz, I.S. Georgiev, C. Soto, J. Gorman, J.H. Huang, P. Acharya, G.Y. Chuang, G. Ofek, G.B.E. Stewart-Jones, J. Stuckey, R.T. Bailer, M. G. Joyce, M.K. Louder, N. Tumba, Y.P. Yang, B.S. Zhang, M.S. Cohen, B.F. Haynes, J.R. Mascola, L. Morris, J.B. Munro, S.C. Blanchard, W. Mothes, M. Connors, P. D. Kwong, Structure and immune recognition of trimeric pre-fusion HIV-1 Env, Nature 514 (2014) 455–461.
7. A.B. Ward, I.A. Wilson, The HIV-1 envelope glycoprotein structure: nailing down a moving target, Immunol. Rev. 275 (2017) 21–32.
8. M. Caffrey, M. Cai, J. Kaufman, S.J. Stahl, P.T. Wingfield, D.G. Covell, A. M. Gronenborn, G.M. Clore, Three-dimensional solution structure of the 44 kDa ectodomain of SIV gp41, EMBO J. 17 (1998) 4572–4584.
9. Z.N. Yang, T.C. Mueser, J. Kaufman, S.J. Stahl, P.T. Wingfield, C.C. Hyde, The crystal structure of the SIV gp41 ectodomain at 1.47 A resolution, J. Struct. Biol. 126 (1999) 131–144.
10. V. Buzon, G. Natrajan, D. Schibli, F. Campelo, M.M. Kozlov, W. Weissenhorn, Crystal structure of HIV-1 gp41 including both fusion peptide and membrane proximal external regions, PLoS Pathog. 6 (2010) e1000880.
11. S.R. Durell, I. Martin, J.M. Ruysschaert, Y. Shai, R. Blumenthal, What studies of fusion peptides tell us about viral envelope glycoprotein-mediated membrane fusion, Mol. Membr. Biol. 14 (1997) 97–112.
12. R. Kong, K. Xu, T.Q. Zhou, P. Acharya, T. Lemmin, K. Liu, G. Ozorowski, C. Soto, J. D. Taft, R.T. Bailer, E.M. Cale, L. Chen, C.W. Choi, G.Y. Chuang, N.A. Doria-Rose, A. Druz, I.S. Georgiev, J. Gorman, J.H. Huang, M.G. Joyce, M.K. Louder, X.C. Ma, K. McKee, S. O'Dell, M. Pancera, Y.P. Yang, S.C. Blanchard, W. Mothes, D. R. Burton, W.C. Koff, M. Connors, A.B. Ward, P.D. Kwong, J.R. Mascola, Fusion peptide of HIV-1 as a site of vulnerability to neutralizing antibody, Science 352 (2016) 828–833.

13. M.J. van Gils, T.L.G.M. van den Kerkhof, G. Ozorowski, C.A. Cottrell, D. Sok, M. Pauthner, J. Pallesen, N. de Val, A. Yasmeen, S.W. de Taeye, A. Schorcht, S. Gumbs, I. Johanna, K. Saye-Francisco, C.-H. Liang, E. Landais, X. Nie, L. K. Pritchard, M. Crispin, G. Kelsoe, I.A. Wilson, H. Schuitemaker, P.J. Klasse, J. P. Moore, D.R. Burton, A.B. Ward, R.W. Sanders, An HIV-1 antibody from an elite neutralizer implicates the fusion peptide as a site of vulnerability, Nat. Microbiol. 2 (2017). Art. No. 16199.

14. K. Xu, P. Acharya, R. Kong, C. Cheng, G.-Y. Chuang, K. Liu, M.K. Louder, S. O'Dell, R. Rawi, M. Sastry, C.-H. Shen, B. Zhang, T. Zhou, M. Asokan, R.T. Bailer, M. Chambers, X. Chen, C.W. Choi, V.P. Dandey, N.A. Doria-Rose, A. Druz, E.T. Eng, S.K. Farney, K.E. Foulds, H. Geng, I.S. Georgiev, J. Gorman, K.R. Hill, A.J. Jafari, Y. D. Kwon, Y.-T. Lai, T. Lemmin, K. McKee, T.Y. Ohr, L. Ou, D. Peng, A.P. Rowshan, Z. Sheng, J.-P. Todd, Y. Tsybovsky, E.G. Viox, Y. Wang, H. Wei, Y. Yang, A.F. Zhou, R. Chen, L. Yang, D.G. Scorpio, A.B. McDermott, L. Shapiro, B. Carragher, C. S. Potter, J.R. Mascola, P.D. Kwong, Epitope-based vaccine design yields fusion peptide-directed antibodies that neutralize diverse strains of HIV-1, Nat. Med. 24 (2018) 857–867.

15. C.P. Jaroniec, J.D. Kaufman, S.J. Stahl, M. Viard, R. Blumenthal, P.T. Wingfield, A. Bax, Structure and dynamics of micelle-associated human immunodeficiency virus gp41 fusion domain, Biochemistry 44 (2005) 16167–16180.

16. C.M. Gabrys, D.P. Weliky, Chemical shift assignment and structural plasticity of a HIV fusion peptide derivative in dodecylphosphocholine micelles, Biochim. Biophys. Acta 1768 (2007) 3225–3234.

17. Y.L. Li, L.K. Tamm, Structure and plasticity of the human immunodeficiency virus gp41 fusion domain in lipid micelles and bilayers, Biophys. J. 93 (2007) 876–885.

18. W. Qiang, M.L. Bodner, D.P. Weliky, 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 (2008) 5459–5471.

19. S.D. Schmick, D.P. Weliky, Major antiparallel and minor parallel beta sheet populations detected in the membrane-associated Human Immunodeficiency Virus fusion peptide, Biochemistry 49 (2010) 10623–10635.

20. S.D. Schmick, High Resolution Tertiary Structure of the Membrane-Asssociated HIV Fusion Peptides by Solid State Nuclear Magnetic Resonance, Michigan State University, East Lansing, MI, USA, 2012.

21. K. Sackett, M.J. Nethercott, Z.X. Zheng, D.P. Weliky, 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 (2014) 1077–1094.

22. N.A. Lakomek, J.D. Kaufman, S.J. Stahl, J.M. Louis, A. Grishaev, P.T. Wingfield, A. Bax, Internal dynamics of the homotrimeric HIV-1 viral coat protein gp41 on multiple time scales, Angew. Chem. Int. Ed. 52 (2013) 3911–3915.

23. P.U. Ratnayake, K. Sackett, M.J. Nethercott, D.P. Weliky, 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 1848 (2015) 289–298.

24. K. Sackett, M.J. Nethercott, Y. Shai, D.P. Weliky, Hairpin folding of HIV gp41 abrogates lipid mixing function at physiologic pH and inhibits lipid mixing by exposed gp41 constructs, Biochemistry 48 (2009) 2714–2722.

25. S. Liang, P.U. Ratnayake, C. Keinath, L. Jia, R. Wolfe, A. Ranaweera, D.P. Weliky, Efficient fusion at neutral pH by Human Immunodeficiency Virus gp41 trimers containing the fusion peptide and transmembrane domains, Biochemistry 57 (2018) 1219–1235.

26. T. Suarez, W.R. Gallaher, A. Agirre, F.M. Goni, J.L. Nieva, Membrane interface-interacting sequences within the ectodomain of the human immunodeficiency virus type 1 envelope glycoprotein: putative role during viral fusion, J. Virol. 74 (2000) 8038–8047.

27. M. Montero, N.E. van Houten, X. Wang, J.K. Seott, 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 (2008) 54–84. 28.

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

29. B. Apellaniz, E. Rujas, S. Serrano, K. Morante, K. Tsumoto, J.M.M. Caaveiro, M. Angeles Jimenez, J.L. Nieva, The atomic structure of the HIV-1 gp41 transmembrane domain and its connection to the immunogenic membrane-proximal external region, J. Biol. Chem. 290 (2015) 12999–13015.

30. S.C. Chiliveri, J.M. Louis, R. Ghirlando, J.L. Baber, A. Bax, Tilted, uninterrupted, monomeric HIV-1 gp41 transmembrane helix from residual dipolar couplings, J. Am. Chem. Soc. 140 (2018) 34–37.

31. Q.S. Fu, M.M. Shaik, Y.F. Cai, F. Ghantous, A. Piai, H.Q. Peng, S. Rits-Volloch, Z. J. Liu, S.C. Harrison, M.S. Seaman, B. Chen, J.J. Chou, Structure of the membrane proximal external region of HIV-1 envelope glycoprotein, Proc. Natl. Acad. Sci. U. S. A. 115 (2018). E8892–E8899.

32. B. Kwon, M. Lee, A.J. Waring, M. Hong, Oligomeric structure and three-dimensional fold of the HIV gp41 membrane-proximal external region and transmembrane domain in phospholipid bilayers, J. Am. Chem. Soc. 140 (2018) 8246–8259.

33. K. Banerjee, D.P. Weliky, Folded monomers and hexamers of the ectodomain of the HIV gp41 membrane fusion protein: Potential roles in fusion and synergy between the fusion peptide, hairpin, and membrane-proximal external region, Biochemistry 53 (2014) 7184–7198.

34. M. Kielian, Mechanisms of virus membrane fusion proteins, Annu. Rev. Virol. 1 (2014) 171–189.

35. S.C. Harrison, Viral membrane fusion, Virology 479 (2015) 498–507.

36. S. Shnaper, K. Sackett, S.A. Gallo, R. Blumenthal, Y. Shai, The C- and the N-terminal regions of glycoprotein 41 ectodomain fuse membranes enriched and not enriched with cholesterol, respectively, J. Biol. Chem. 279 (2004) 18526–18534.

37. J. Roche, J.M. Louis, A. Grishaev, J.F. Ying, A. Bax, 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 (2014) 3425–3430.

38. D.J. Benton, S.J. Gamblin, P.B. Rosenthal, J.J. Skehel, Structural transitions in influenza haemagglutinin at membrane fusion pH, Nature 583 (2020) 150–153.

39. C. Aisenbrey, B. Bechinger, Structure, interactions and membrane topology of HIV gp41 ectodomain sequences, Biochim. Biophys. Acta 1862 (2020), 183274.

40. M.W. Eller, H.M.H. Siaw, R.B. Dyer, Stability of HA2 prefusion structure and pH-induced conformational changes in the HA2 domain of H3N2 hemagglutinin, Biochemistry 60 (2021) 2623–2636.

41. M. Lee, C.A. Morgan, M. Hong, Fully hydrophobic HIV gp41 adopts a hemifusion-like conformation in phospholipid bilayers, J. Biol. Chem. 294 (2019) 14732–14744.

42. E. Leikina, D.L. LeDuc, J.C. Macosko, R. Epand, Y.K. Shin, L.V. Chernomordik, The 1-127 HA2 construct of influenza virus hemagglutinin induces cell-cell hemifusion, Biochemistry 40 (2001) 8378–8386.

43. N. Lev, Y. Fridmann-Sirkis, L. Blank, A. Bitler, R.F. Epand, R.M. Epand, Y. Shai, Conformational stability and membrane interaction of the full-length ectodomain of HIV-1 gp41: Implication for mode of action, Biochemistry 48 (2009) 3166–3175.

44. K. Sackett, M.J. Nethercott, R.F. Epand, R.M. Epand, D.R. Kindra, Y. Shai, D. P. Weliky, 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 (2010) 301–315.

45. K. Sackett, A. TerBush, D.P. Weliky, 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 (2011) 489–502.

46. C.S. Kim, R.F. Epand, E. Leikina, R.M. Epand, L.V. Chernomordik, 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 (2011) 13226–13234.

47. A. Pabis, R.J. Rawle, P.M. Kasson, Influenza hemagglutinin drives viral entry via two sequential intramembrane mechanisms, Proc. Natl. Acad. Sci. U. S. A. 117 (2020) 7200–7207.

48. P. Larsson, P.M. Kasson, Lipid tail protrusion in simulations predicts fusogenic activity of influenza fusion peptide mutants and conformational models, PLoS Comp. Biol. 9 (2013) e1002950.

49. U. Ghosh, D.P. Weliky, 2H nuclear magnetic resonance spectroscopy supports larger amplitude fast motion and interference with lipid chain ordering for membrane that contains beta sheet human immunodeficiency virus gp41 fusion peptide or helical hairpin influenza virus hemagglutinin fusion peptide at fusogenic pH, Biochim. Biophys. Acta 1862 (2020), 183404.

50. U. Ghosh, D.P. Weliky, Rapid 2H NMR transverse relaxation of perdeuterated lipid acyl chains of membrane with bound viral fusion peptide supports large-amplitude motions of these chains that can catalyze membrane fusion, Biochemistry 60 (2021) 2637–2651.

51. A.L. Lai, J.H. Freed, HIV gp41 fusion peptide increases membrane ordering in a cholesterol-dependent fashion, Biophys. J. 106 (2014) 172–181.

52. S.T. Smrt, A.W. Draney, J.L. Lorieau, The influenza hemagglutinin fusion domain is an amphipathic helical hairpin that functions by inducing membrane curvature, J. Biol. Chem. 290 (2015) 228–238.

53. G. Meher, H. Chakraborty, The role of fusion peptides in depth-dependent membrane organization and dynamics in promoting membrane fusion, Chem. Phys. Lipids 234 (2021), 105025.

54. E.O. Freed, E.L. Delwart, G.L. Buchschacher Jr., A.T. Panganiban, 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 (1992) 70–74.

55. O.F. Brandenberg, C. Magnus, R.R. Regoes, A. Trkola, The HIV-1 entry process: a stoichiometric view, Trends Microbiol. 23 (2015) 763–774.

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

57. R. Sougrat, A. Bartesaghi, J.D. Lifson, A.E. Bennett, J.W. Bess, D.J. Zabransky, S. Subramaniam, Electron tomography of the contact between T cells and SIV/HIV- 1: implications for viral entry, PLoS Pathog. 3 (2007) e63.

58. F.B. Pereira, F.M. Goni, J.L. Nieva, Liposome destabilization induced by the HIV-1 fusion peptide effect of a single amino acid substitution, FEBS Lett. 362 (1995) 243–246.

59. Y. Kliger, A. Aharoni, D. Rapaport, P. Jones, R. Blumenthal, Y. Shai, Fusion peptides derived from the HIV type 1 glycoprotein 41 associate within phospholipid membranes and inhibit cell-cell fusion. Structure-function study, J. Biol. Chem. 272 (1997) 13496–13505.

60. C.M. Gabrys, W. Qiang, Y. Sun, L. Xie, S.D. Schmick, D.P. Weliky, Solid-state nuclear magnetic resonance measurements of HIV fusion peptide 13CO to lipid 31P proximities support similar partially inserted membrane locations of the α _Helical and β-sheet peptide structures, J. Phys. Chem. A 117 (2013) 9848–9859.

61. F.B. Pereira, F.M. Goni, A. Muga, J.L. Nieva, Permeabilization and fusion of uncharged lipid vesicles induced by the HIV-1 fusion peptide adopting an extended conformation: dose and sequence effects, Biophys. J. 73 (1997) 1977–1986.

62. P.W. Mobley, A.J. Waring, M.A. Sherman, L.M. Gordon, Membrane interactions of the synthetic N-terminal peptide of HIV-1 gp41 and its structural analogs, Biochim. Biophys. Acta 1418 (1999) 1–18.

63. S. Kamath, T.C. Wong, Membrane structure of the human immunodeficiency virus gp41 fusion domain by molecular dynamics simulation, Biophys. J. 83 (2002) 135–143.

64. W. Qiang, Y. Sun, D.P. Weliky, A strong correlation between fusogenicity and membrane insertion depth of the HIV fusion peptide, Proc. Natl. Acad. Sci. U. S. A. 106 (2009) 15314–15319.

65. A. Jacobs, C. Simon, M. Caffrey, Thermostability of the HIV gp41 wild-type and loop mutations, *Protein Pept.* Lett. 13 (2006) 477–480.

66. R. Yang, M. Prorok, F.J. Castellino, D.P. Weliky, A trimeric HIV-1 fusion peptide construct which does not self-associate in aqueous solution and which has 15-fold higher membrane fusion rate, J. Am. Chem. Soc. 126 (2004) 14722–14723.

67. A.L. Lai, A.E. Moorthy, Y.L. Li, L.K. Tamm, Fusion activity of HIV gp41 fusion domain is related to its secondary structure and depth of membrane insertion in a cholesterol-dependent fashion, J. Mol. Biol. 418 (2012) 3–15.

68. M. Lorizate, T. Sachsenheimer, B. Glass, A. Habermann, M.J. Gerl, H.G. Krausslich, B. Brugger, Comparative lipidomics analysis of HIV-1 particles and their producer cell membrane in different cell lines, Cell. Microbiol. 15 (2013) 292–304.

69. J.C. Macosko, C.H. Kim, Y.K. Shin, The membrane topology of the fusion peptide region of influenza hemagglutinin determined by spin-labeling EPR, J. Mol. Biol. 267 (1997) 1139–1148.

70. E. Zaitseva, E. Zaitsev, K. Melikov, A. Arakelyan, M. Marin, R. Villasmil, L. B. Margolis, G.B. Melikyan, L.V. Chernomordik, Fusion stage of HIV-1 entry depends on virus-induced cell surface exposure of phosphatidylserine, Cell Host Microbe 22 (2017) 99–110.

71. R.F.M. de Almeida, L.M.S. Loura, A. Fedorov, M. Prieto, Lipid rafts have different sizes depending on membrane composition: a time-resolved fluorescence resonance energy transfer study, J. Mol. Biol. 346 (2005) 1109–1120.

72. Y.H. Chen, J.T. Yang, K.H. Chau, Determination of helix and beta-form of proteins in aqueous-solution by circular-dichroism, Biochemistry 13 (1974) 3350–3359.

73. G.F. Gao, L. Wieczorek, K.K. Peachman, V.R. Polonis, C.R. Alving, M. Rao, V. B. Rao, Designing a soluble near full-length HIV-1 gp41 trimer, J. Biol. Chem. 288 (2013) 234–246.

74. M. Caffrey, J. Kaufman, S. Stahl, P. Wingfield, A.M. Gronenborn, G.M. Clore, Monomer-trimer equilibrium of the ectodomain of SIV gp41: insight into the mechanism of peptide inhibition of HIV infection, Protein Sci. 8 (1999) 1904–1907.

75. A. Rath, M. Glibowicka, V.G. Nadeau, G. Chen, C.M. Deber, Detergent binding explains anomalous SDS-PAGE migration of membrane proteins, Proc. Natl. Acad. Sci. U. S. A. 106 (2009) 1760–1765.

76. W. Shu, H. Ji, M. Lu, Interactions between HIV-1 gp41 core and detergents and their implications for membrane fusion, J. Biol. Chem. 275 (2000) 1839–1845.

77. J. Chen, J.J. Skehel, D.C. Wiley, N- and C-terminal residues combine in the fusion-pH influenza hemagglutinin HA2 subunit to form an N cap that terminates the triple-stranded coiled coil, Proc. Natl. Acad. Sci. U. S. A. 96 (1999) 8967–8972.

78. A. Ranaweera, P.U. Ratnayake, E.A.P. Ekanayaka, R. Declercq, D.P. Weliky, Hydrogen-deuterium exchange supports independent membrane-interfacial fusion peptide and transmembrane domains in subunit 2 of influenza virus hemagglutinin protein, a structured and aqueous-protected connection between the fusion peptide and soluble ectodomain, and the importance of membrane apposition by the trimer-of-hairpins structure, Biochemistry 58 (2019) 2432–2446.

79. J.L. Lorieau, J.M. Louis, A. Bax, The complete influenza hemagglutinin fusion domain adopts a tight helical hairpin arrangement at the lipid: water interface, Proc. Natl. Acad. Sci. U. S. A. 107 (2010) 11341–11346.

80. U. Ghosh, L. Xie, L.H. Jia, S. Liang, D.P. Weliky, Closed and semiclosed interhelical structures in membrane vs closed and open structures in detergent for the Influenza Virus hemagglutinin fusion peptide and correlation of hydrophobic surface area with fusion catalysis, J. Am. Chem. Soc. 137 (2015) 7548–7551.

81. A. Ranaweera, P.U. Ratnayake, D.P. Weliky, The stabilities of the soluble ectodomain and fusion peptide hairpins of the Influenza virus hemagglutinin subunit II protein are positively correlated with membrane fusion, Biochemistry 57 (2018) 5480–5493.

82. H. Qiao, R.T. Armstrong, G.B. Melikyan, F.S. Cohen, J.M. White, A specific point mutant at position 1 of the influenza hemagglutinin fusion peptide displays a hemifusion phenotype, Mol. Biol. Cell 10 (1999) 2759–2769.

83. E. Borrego-Diaz, M.E. Peeples, R.M. Markosyan, G.B. Melikyan, F.S. Cohen, Completion of trimeric hairpin formation of influenza virus hemagglutinin promotes fusion pore opening and enlargement, Virology 316 (2003) 234–244.

84. Z. Dai, Y.S. Tao, N.N. Liu, M.D. Brenowitz, M.E. Girvin, J.R. Lai, Conditional trimerization and lytic activity of HIV-1 gp41 variants containing the membrane-associated segments, Biochemistry 54 (2015) 1589–1599.

85. C. Grewe, A. Beck, H.R. Gelderblom, HIV: early virus-cell interactions, J. AIDS 3 (1990) 965–974.

86. N. Herold, M. Anders-Osswein, B. Glass, M. Eckhardt, B. Muller, H.G. Krausslich, HIV-1 entry in SupT1-R5, CEM-ss, and primary CD4(+) T Cells occurs at the plasma membrane and does not require endocytosis, J. Virol. 88 (2014) 13956–13970.

87. G.B. Melikyan, HIV entry: a game of hide-and-fuse? Curr. Opin. Virol. 4 (2014) 1–7.

CHAPTER-04 (Extension of Project- 01): Expression, purification, characterization and fusion study of HIV gp41 ectodomain WT, V2E, L9R and G10V to confirm the role of hairpin and correlation between helicity and fusion.

4.1: Introduction

The human immunodeficiency virus type 1 (HIV-1) must fuse its envelope with the host cell's plasma membrane in order to infect cells[1]. The viral envelope glycoprotein (Env) complex, gp120–gp41 is responsible for HIV-1 to fuse with target membranes[2,3,4]. Env spike is a type-I fusion machine that facilitates HIV-1 entry by rearranging from a mature unliganded state (pre-fusion), through receptor and co-receptor bound intermediates (Env activation), to a post-fusion state[5]. In the pre-fusion state, inter-protomer (one gp120 and one gp41 form a protomer) interactions occur through assembled variable loop regions, V1, V2 and V3 of gp120, which comprise the trimer association domain at the membrane-distal portion of the spike, and through gp41, primarily between inter-helical interactions around the trimer axis[5,6]. Env activation occurs after binding of gp120 to receptor (CD4) and co-receptors (CXCR4, CCR5) of host cell followed by gp41 dissociates from gp120 for catalyzing fusion of HIV virus into the host cell[7]. CD4 binding results in a major reorganization of the Env trimer, causing an outward rotation and displacement of each gp120 monomer and exposing V3 loop to bind with co-receptors in the host cell (Fig. 4.1)[8].

Initially synthesized as a gp160 precursor, Env is trimerized and subsequently cleaved by proteases of the furin family into gp120 and gp41 subunits, which associate noncovalently before the native complex reaches the surface of infected cells and is then packaged into virions[9,10]. HIV gp41 has a series of unique α helices joined by loops in the complex. The central α7 helix ($gp41_{572-595}$), out of the four helices, forms an internal parallel trimeric bundle with the C-terminal end likely oriented toward the viral membrane and the N-terminal end pointing toward gp120. The expanded structure of the FP is found in the N-terminal fusion peptide (FP) domain ($gp41_{512-529}$), which is connected to the α6 helix ($gp41_{530-543}$) by a loop[11]. However, as shown in Figure 4.1a, a loop from the α7 C-terminus links to the α8 helix ($gp41_{619-623}$) and subsequently the α9 helix ($gp41_{628-664}$). The α6, α8, and α9 helices appear to be on three sides of the bundle, outside, and roughly perpendicular to the α7 bundle. Finally, large-scale expansions and topological rearrangements of helical segments are revealed when comparing this original gp41 ectodomain structure in association with gp120 to the subsequent gp41 structure without gp120 (post fusion state)[11]. In the latter structure, each hairpin is made by an N-helix

($gp41_{536-596}$), loop, and C-helix ($gp41_{616-675}$). This structure is similar to a "trimer of hairpins." According to Fig. 4.1 b&c, this "trimer of hairpins" is arranged as a six-helix bundle (SHB), with three N-helices ($gp41_{536-596}$) forming an inside bundle, loops, and then C-helices ($gp41_{616-675}$) running antiparallel in the external grooves of the bundle [11,12-15]. A gradual left-handed superhelix is formed by three N-helices forming a parallel coiled coil in the six-helix bundle (SHB), and a left-handed antiparallel wrap of three C-helices around the outside of the center coiled coil trimer of N-helices (Figure 4.1c)[12].

Fusion peptide (FP) is the N-terminal construct of HIV gp41 responsible for fusion (initial binding with host membrane) consists of ≈23 amino acids residues from 512 to 535 (Figure 4.1b) of gp41 in the env trimer becomes buried into a new pocket inside the trimer distant from the host membrane during CD4 binding and further triggering is required for gp41 fusion into host cell membrane[7]. This triggering is done by co-receptor engagement with gp120 and consequently, gp41 detaches from gp120 to fusion. The salient feature of fusion peptide amino residues with an average high hydrophobicity among which has 11 glycine (H as side chain) and alanine ($-CH_3$ as side chain) in total 23 residues[16]. Different isolates of HIV-1 have extreme conserved amino acids sequence of fusion peptide and other immunodeficiency viruses such as HIV-2 and SIV also have high homologous amino acids sequence[16,17,18].

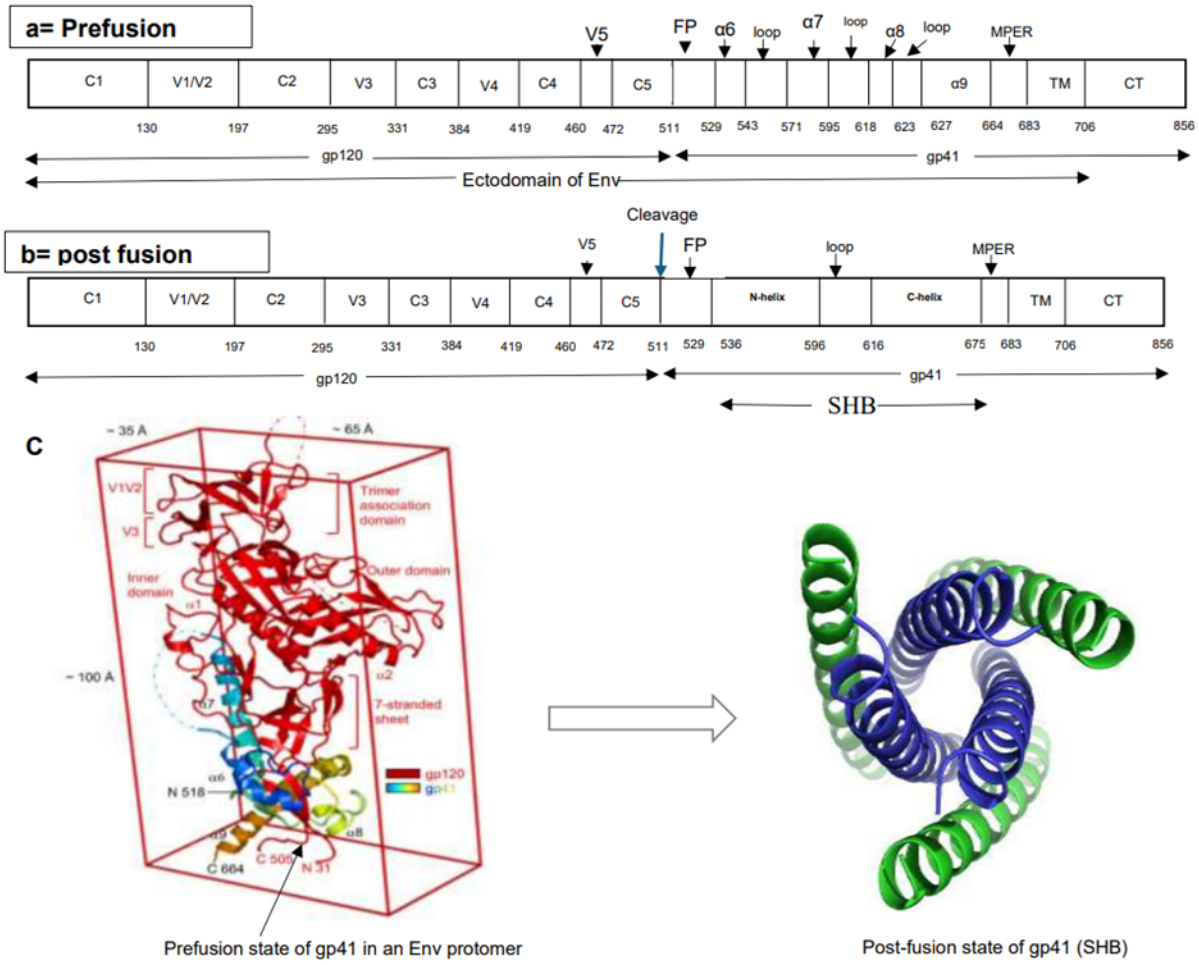


Figure 4.1: (a) Schematic representation of one Env protomer (gp120 + gp41) in prefusion state (b) schematic representation of Env protomer in post-fusion state in which gp120 cleaves around amino acid residue 511 after binding with receptor or co-receptor. On the other hand, four helices (α6, α7, α8 and α9) of gp41 in prefusion state rearrange from amino acid residues 536 to 675 to form six helix bundle (SHB) in post-fusion state. [In both Fig. a, b & C= constant region, V= variable loop region, FP= fusion peptide, MPER= membrane proximal external region, TM= transmembrane region, CT= cytoplasmic region] (c) structural rearrangement of gp41 in prefusion vs. post-fusion state. Here, prefusion state is one Env protomer and post-fusion state is SHB of gp41 trimer (one Env protomer has one gp120+gp41, Env trimer has three gp120+gp41). Gp41 trimer of SHB in the post-fusion state dissociated from gp120 in Env trimer.

Several mutational analyses for fusion on HIV-1 gp41 FP region and other gp160 has been conducted[16,19,20]. Freed et al. (1992) co-transfected the $CD4^+$ HeLa cell line HeLa T4 with identical numbers of wild-type and mutant envelope expression constructs to ascertain whether mutant HIV-1 envelope glycoproteins might dominantly interfere with syncytium formation produced by wild-type HIV-1 envelope glycoproteins[19,21]. The findings showed that syncytium formation induced by wild-type envelope was less than twice affected by mutations at gp120 amino acids 318, 319, and 320 (within the V3 loop), and gp41 amino acids 9, 15, 26, and 29 (within the amino terminus of gp41). Previous research by Freed et al (1989,90,91), also demonstrated that mutations at these locations significantly decreased or eliminated syncytium formation[22-24]. On the other hand, the syncytium formation caused by the wild-type envelope glycoprotein was reduced by a factor of 25 when it was co-expressed with an envelope glycoprotein that had a polar substitution (Val→Glu) at gp41 amino acid 2 (also known as 513E/V2E mutation)[19]. Polar substitution at gp41 amino acid 2 (the 41.2 mutation) also results in an envelope glycoprotein that dominantly interferes with both syncytium formation and infection mediated by the wild-type HIV-1envelope glycoprotein[19]. Delahunty et al., (1996) mutagenized each glycine residue in the fusion peptide individually to a valine residue in the fusion peptide region of HIV gp41 to assess the importance of the glycine residues for the fusogenic activity of the envelope glycoprotein complex[16]. They tested the mutant envelope constructs for their capacity to mediate infection (virus/cell fusion) and produce syncytia (cell/cell fusion) in $CD4^+$ cells. Findings indicate that, while other glycine residues (G3, G5, and G20) are more tolerant of substitutions, two glycine residues (G10 and G13) in the middle of the fusion peptide sequence, FLGFLG, are essential for the formation of syncytium's and the establishment of a productive infection. Although less so than mutation of G10 and G13, fusion was also reduced by changing each of the two phenylalanine's (F8 and F11) of the FLGFLG sequence to valine. Rokonujjaman et al. (2023, First project) showed quantitative similarity of HIV gp160 V513E-dominant reduction of fusion and infection with fusion by the gp41 ectodomain hairpin supports an important fusion role for the final trimer-of-hairpins structure[20]. There is a large V2E attenuation and dominance were observed for vesicle fusion induced by FPHM, a large gp41 ectodomain construct with Fp followed by hyper thermostable hairpin with N- and

C-helices, and membrane-proximal external region (Mper). The gp41 ectodomain also exhibits V513E-dominant reduction of hairpin helicity and correlation of reductions in gp41 helicity and fusion for V2E vs. WT hairpin gp41[20].

Current project focuses on making few point mutants in fusion peptide region of ectodomain (FPHM) of HIV gp41. We express, purify, characterize and fusion study of WT_FPHM, V2E_FPHM, L9R_FPHM and G10V_FPHM. Though mutational analysis in fusion peptide region in gp160 (gp120 + gp41) to identify fusiogenic region and role if specific residues in fusion peptide in HIV fusion, we wanted to address how single mutation in fusion peptide region of gp41 (no gp120) affect in fusion and helicity. This study also identified the role HIV gp41 hairpin in fusion since point mutation in fusion peptide region interfere with fusion of haipin in absence of gp120. Moreover, we wanted to compare our vesicle fusion with single mutant in gp41 fusion peptide (no gp120) region with the earlier cell-cell fusion by gp160 single mutant in fusion peptide region. We also studied the correlation between helicity and fusion with mutant i.e. L9R, G10V since our first project showed the low helicity vs. low fusion of V2E.

4.2: Methods and materials

4.2.1: Materials

Most of the materials of this study were purchased from Sigma-Aldrich (St. Louis, MO). Other materials were purchased from the following commercial sources: DNA, GenScript (Piscataway, NJ); Escherichia coli (E. coli) Novagen (Gibbstown, NJ); Luria-Bertani (LB) medium, Dot Scientific (Burton, MI); isopropyl β-D-thiogalactopyrano-side (IPTG), Goldbio (St. Louis, MO); and Co^{2+} affinity resin, Thermo Scientific (Waltham, MA). Membrane lipid vesicle materials that included 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine (POPC), 1-palmitoyl-2-oleoyl-sn-glycero-3- [phospho-rac-(1- glycerol)] (sodium salt) (POPG), N-(7-nitro-2,1,3-benzoxadiazol-4-yl) (ammonium salt) Dipalmitoylphosphatidylethanolamine} (N-NBD-DPPE) and N-(lissamine Rhodamine B sulfonyl) (ammonium salt) dipalmitoylphosphatidylethanolamine} (N-Rh-DPPE) were purchased from Avanti Polar Lipids (Alabaster, AL).

4.2.2: Protein constructs, expression, and purification

We prepared, expressed and purified four gp41 ectodomain constructs i.e. wild-type (WT) FPHM, V2E_FPHM (valine is replaced by glutamic acid in WT), L9R_FPHM (leucine is

replaced by arginine in WT) and G10V_FPHM (glycine is replaced by valine in WT). E. coli BL21(DE3) cells were used to produce all the constructs of this study that were typically incubated at 37 °C in LB medium with kanamycin (50 mg/mL) and shaking at 150 rpm. DNA sequences of all constructs were coded with non-native C-terminal G6LEH6 (Fig. 4.2). pET-24a (+) plasmid with Lac operon and kanamycin antibiotic resistance was used to subclone wild-type (WT) FPHM and G10V_FPHM DNA sequences, respectively with G6LE spacer in the C-terminal. This spacer will aid purification because it will expose the affinity tag during the process[11]. The plasmid with insert was then transformed into E. coli cells. pET24a (+) [plasmid] containing DNA sequence of WT protein was extracted from previously subcloned E. Coli cell cultures followed by PCR reactions using DNA primers (Fig. 4.3) for V2E/L9R mutation. After that, the plasmid containing mutated DNA was purified followed by transformed into BL21(DE3) competent cells that were then streaked on an agar plate. A single colony was produced after overnight growth of BL21(DE3) competent cells in agar and one of the single colonies was transferred to 50 mL of LB medium in the presence of 50 mg/L kanamycin antibiotic[25]. The cell cultures were then grown overnight at 150 rpm and 37 °C. These overnighted cell cultures were used for preparing stock aliquots for future growth by mixing 1 mL of culture and 0.5 mL of 50% (v/v) glycerol and stored at −80°C[25]. This V2E/L9R mutation in WT is confirmed by DNA sequencing (Fig. 4.2) and proteomics (Fig. 4.4). On the other hand, the DNA (Ordered from Gene Scripts) of G10V was transformed into BL21(DE3) competent cells by the same way as V2E/L9R except for PCR reactions followed by preparing glycerol stock. Production of all four constructs separately began with the addition of 80 µL bacterial glycerol stock to 50 mL LB medium, ~7 h culture growth, transfer to 1 L fresh LB medium, and ~ 3 h growth till final OD_{600}'s of ~0.8. Induction of these cells was started by adding 2 mmoles IPTG to express proteins for 8 hours at 37 °C. Centrifugation was performed at 9000 g for 15 min to harvest the expressed cell pellets (~5g) followed by storage at -20 °C. Protein purification started with phosphate buffer sulfate (PBS; 30 ml for ~5g cell pellets) washing at pH 7.4 by tip sonication with 50% amplitude at 0.6 pulses on and 0.4 pulses off for 1 minute with a surrounding ice bath, and centrifugation was done with 48,000 g for 1 hr at 4 °C. This process was repeated 2 times as all the constructs have low solubility in PBS without

additives and was separated from PBS soluble molecules. For subsequent steps, WT and G10V cell plates were solubilized by sonication in PBS with 6 M GuHCl followed by centrifugation, and collection of the supernatant. Dialysis with 10 KDa MW cutoff membrane was then done against Millipore water to precipitate some dissolved molecules including WT and G10V, respectively followed by centrifugation. The resultant solids were solubilized by sonication in 8M urea with PBS+0.5% SDS+ 0.8% SK followed by centrifugation, and collection of supernatants. Finally, Co^{2+} affinity chromatography was done in a fritted column with liquid removal by gravity filtration, and all solutions contained 8M urea with PBS+0.5% SDS+ 0.8% SK. The process began with (1) filtration of 5 mL of resin suspension; (2) resin washing with 10 mL of 10 mM imidazole; (3) addition of the WT/G10V containing supernatant followed by agitation overnight at 4 °C; and (4) sequential washing/elution in 15 mL without imidazole, 10 mL of 10 mM imidazole, 5 mL of 300 mM imidazole, and 10 mL of 600 mM imidazole. The elutions with 300- and 600-mM imidazole were combined followed by mixing (1:1) with refolding buffer (10 mM Tris-HCl (pH 8.0), 0.17% n-decyl-β-D-maltoside, 2 mM EDTA, and 1 M L-arginine), dialysis against 10 mM Tris+ 0.2% SDS at pH 7.4 to CD/fusion study.

1

GCCGTGGGTATCGGTGCTCTGTTCCTGGGGTTCCTGGGTGCTGCTGGTTCGACGA
TGGGTGCCCGCTCAATGACG

76

CTGACGGTCCAAGCACGTCAGCTGCTGAGCGGCATTGTGCAGCAACAGAACAATC
TGCTGCGCGCGATCGAAGCC

151

CAACAGCATCTGCTGCAGCTGACCGTTTGGGGTATTAAACAACTGCAGGCTCGTAT
CCTGAGCGGCGGTCGCGGC

226

GGTTGGATGGAATGGGATCGTGAAATTAACAATTATACGAGCCTGATTCACTCTCT
GATCGAAGAAAGTCAAAAC

301

CAACAGGAGAAAAACGAACAGGAACTGCTGGAACTGGACAAATGGGCCTCCCTGT
GGAACTGGTTTAACATTACG

376

AACTGGCTGTGGTACATCAAAGGCGGCGGTGGCGGTGGTCTCGANCACCACCAC
CACCACCAC

Figure 4.2: DNA sequence of Wild-type FP_HM including non-native G6LEH6 C-terminal tag in the plasmid. T nucleotide at the 5th position is replaced by A for V2E mutation, T nucleotide at the 26th position is replaced by C for L9R mutation and G nucleotide at the 26th position is replaced by C for L9R mutation, and G nucleotide at the 30th position is replaced by T for G10V mutation.

Primer nucleotide sequences to produce V2E_FPHM mutant.

Forward Primer Sequence: 5'- CAT ATG GCC GAG GGT ATC GGT G- 3'

Reverse Primer Sequence: 5'- CAC CGA TAC CCT CGG CCA TAT G- 3'

and primer nucleotide sequences to produce L9R_FPHM mutant.

Forward Primer Sequence: 5'- ATC GGT GCT CTG TTC CGC GGT TTC CTG- 3'

Reverse Primer Sequence: 5'- CAG GAA ACC GCG GAA CAG AGC ACC GAT- 3'

Figure 4.3: Primer sequence of V2E_FPHM and L9R_FPHM.

On the other hand, V2E and L9R cell pellets were solubilized after PBS without additives by sonication in 8 M Urea with PBS followed by centrifugation, and collection of the supernatant. These constructs and including other dissolved molecules were then precipitated by increasing [NaCl] to ~300 mM followed by centrifugation, harvesting, and then vertexing the pellet in distilled water. The insoluble pellets in water were then collected by centrifugation followed by harvesting, pellet dissolution in 20 mL PBS with 6M GuHCl centrifugation, and collection of the supernatant. Co^{2+} affinity chromatography was done then in a fritted column with liquid removal by gravity filtration, and all solutions contained 6M GuHCl in PBS by subsequently following the same steps as WT/G10V. The reverse order of Co^{2+} resin mixing with WT/G10V vs. L9R/V2E was found by trial-and-error methods since more hydrophobic WT/G10V purified proteins were collected after elution with 8M urea with PBS+0.5% SDS+ 0.8% SK and more polar L9R/V2E shows more purified proteins after elution with 6M GuHCl. Finally, dialysis with 10 KDa MW cutoff membrane was done overnight against distilled water to all the elutions (both 300- and 600 mM imidazole were combined) with consequent precipitation, and the suspension was centrifuged, followed by harvesting and lyophilization of the pellet, and storage at − 20 ˚C.

Figure 4.4: Mass spectrometry of peptides formed from trypsin digestion was used to cover protein sequence. Yellow residues that are highlighted here were present in our peptide of the SDS-PAGE gel. Green residues that are highlighted here were in peptides but had mass changes consistent with oxidation (M), deamination (Q), or dehydration (E).

4.2.3: Western Blots (WB) of Protein SDS-PAGE Gel

SDS-PAGE gel of purified proteins was transferred to a nitrocellulose membrane and incubation in a 20mL solution of milk buffer (2.5 g of non-fat dry milk in 50 mL of 25 mL 10^{x}TBS buffer+ 225 mL DI water+0.5 mL 50% Tween 20). Milk buffer was discarded by washing with 20 mL of 25 mL 10^{x} TBS buffer+ 225 mL DI water+0.5 mL 50% Tween for 5 mins and subsequent incubation with 10 µL penta -His-HRP conjugate (antibody)+ 10 mL milk buffer for 1 hour. Finally, nitrocellulose membrane was developed with SuperSignal West Pico chemiluminescent substrate to get image of protein bands with His tag.

4.2.4: Circular dichroism spectroscopy (CD) and Vesicle fusion

CD sample preparation began with combining the proteins after elutions into refolding buffer (10 mM Tris-HCl (pH 8.0), 0.17% n-decyl-β-D-maltoside, 2 mM EDTA, and 1 M L-arginine at pH 8.0 at 1:1 ratios followed by agitation overnight at 4 °C[20]. This solution was

dialyzed two times against 10 mM Tris+ 0.2% SDS at pH 7.4 to remove arginine, EDTA, imidazole and maltoside. The protein concentration in buffer was ≈5μM. All the CD spectra were buffer subtracted with a J-810 instrument (Jasco) equipped with a water circulation bath, a quartz cuvette with a 1 mm path length and a 190−260 nm wavelength range scanned in 0.5 nm steps. Moreover, all the CD spectra were acquired over a 22−95 °C range with ≈25 °C temperature increments. Vesicle fusion started with preparing "unlabeled" POPC:POPG (4:1) unilamellar vesicles included freeze-thaw cycles (~10×) followed by extrusion (~10×) through a filter with 100 nm diameter pores. A similar preparation protocol was taken for "labeled" vesicles that was prepared in a separate set and contained additional NBD-DPPE fluorescent lipid and N-Rh-DPPE quenching lipid, with ~0.02 mol fraction for each labeled lipid. POPC:POPG (4:1) are the common lipids used for earlier studies of viral fusion protein constructs and

phosphatidylcholine is a common lipid in membranes of host cells of HIV[20]. In addition, POPG represents a negatively charged lipid fraction of host cell membranes over a pH range 3.0-8.5. Though the target cell's outer membrane has little anionic lipid, there is evidence that gp160-mediated fusion happens after scramblase-mediated transport of anionic phosphatidylserine lipid from the inner to the outer leaflet[20]. Typically, 25-mM citrate buffer was used for vesicle solutions for pH's~3.12, 4.23, 5.13 & 6.3- and 10-mM Tris buffer was used for pH's~7.53 & 8.40. The total membrane lipid vesicle concentration was ≈150 μM in which 10% labeled lipids were mixed with 90% unlabelled lipids (total fluorescence cuvette volume ~1.4 ml). Fluorescence (F) of solution was measured at 530 nm using excitation at 467 nm in a fluorimeter that was warmed to 37°C. At t=0, an aliquot of stock with [FP_HM] ≈ 110 μM in Tris/SDS was added to the vesicle solution as this denoted initial fluorescence F_0 followed by F(t) is measured to t ≈ 700 s in 1 s increments. During fusion fluorescence consequently increases with fusion because of the larger separation between fluorescent and quenching lipids as the labeled vesicle likely fuses with an unlabeled vesicle (labeled:unlabeled vesicles = 1:9). The protein-induced vesicle fusion was typically nearly complete (time-invariant fusion extent, M_{ext}) by t ≈ 700 s followed by a 50 μL aliquot of 20% Triton X-100 detergent was then added and dissolved the vesicles, with consequent maximum fluorescence (Fmax) associated with the large increase in separation between fluorescent and quenching lipids. Fluorescence

measurement is set by literature convention and Time-dependent vesicle fusion is calculated as $M(t) = 100 \times [F(t) - F_0]/[F_{max} - F_0]$ with long-time fusion extent (M_{ext}) calculated using $F(t) = F_{ext}$. The $F \propto \langle 1/ [1 + (R_{For} / R)^6 \rangle$ where R is fluorophore/quencher distance, R_{For} is the Forster distance (~50 Å), and $\langle \ldots \rangle$ is the average overall fluorophores[20]. Estimated $\langle R \rangle \approx R_{For}$ at initial F_0 (initially no significant separation due to fusion between fluorophore and quencher) and $R \rangle \approx (2)^{1/2} \times R_{For}$ at the condition of fusion between two vesicles with resulting $M_{ext} \approx 80\%$.

4.3: Results

4.3.1: Protein purification

SDS-PAGE of all four constructs in this study are displayed in Fig. 4.5. Gels of WT/G10V constructs were taken from the samples after combining elutions with 300- and 600-mM imidazole from Co^{2+} resin affinity chromatography. Gels of V2E/L9R constructs were taken from the samples after combining elutions with 300- and 600-mM imidazole from Co^{2+} resin affinity chromatography followed by dialysis against Millipore water, centrifugation and dissolved in SDS-PAGE sample buffer. Gels of G10V, WT, V2E and L9R in Fig. 4.5 show a prominent band with an apparent molecular weight ~15 KDa. These bands were confirmed to be a monomer of WT/V2E/L9R 88% coverage by Fig. 4.4 of the FP_HM sequence by peptides from trypsin digestion of the band and anti-H6 Western blot (Fig. 4.5 D & E). Moreover, V2E and L9R gels show three non-prominent bands between 45 and 55 KDa which could be trimer of proteins or other proteins as impurities in the sample. However, these bands were not detected by anti-H6 Western blot. G10V monomer band was confirmed by 80% coverage of G10V sequence (Fig. 4.4 & 4.5). Among all four constructs, V2E had a high purified yield (~ 6.0 mg/L) after Co^{2+} affinity purification. L9R, WT and G10V's yields were ~ 2.5 mg/L, ~ 1.7 mg/L and ~ 1.1 mg/L, respectively. Low yields of WT and G10V could be hypothesized due to less cytotoxicity (relative to V2E/L9R, WT/G10V may be more fusion-active in the E. coli cells and have higher cytotoxicity) than V2E/L9R[20].

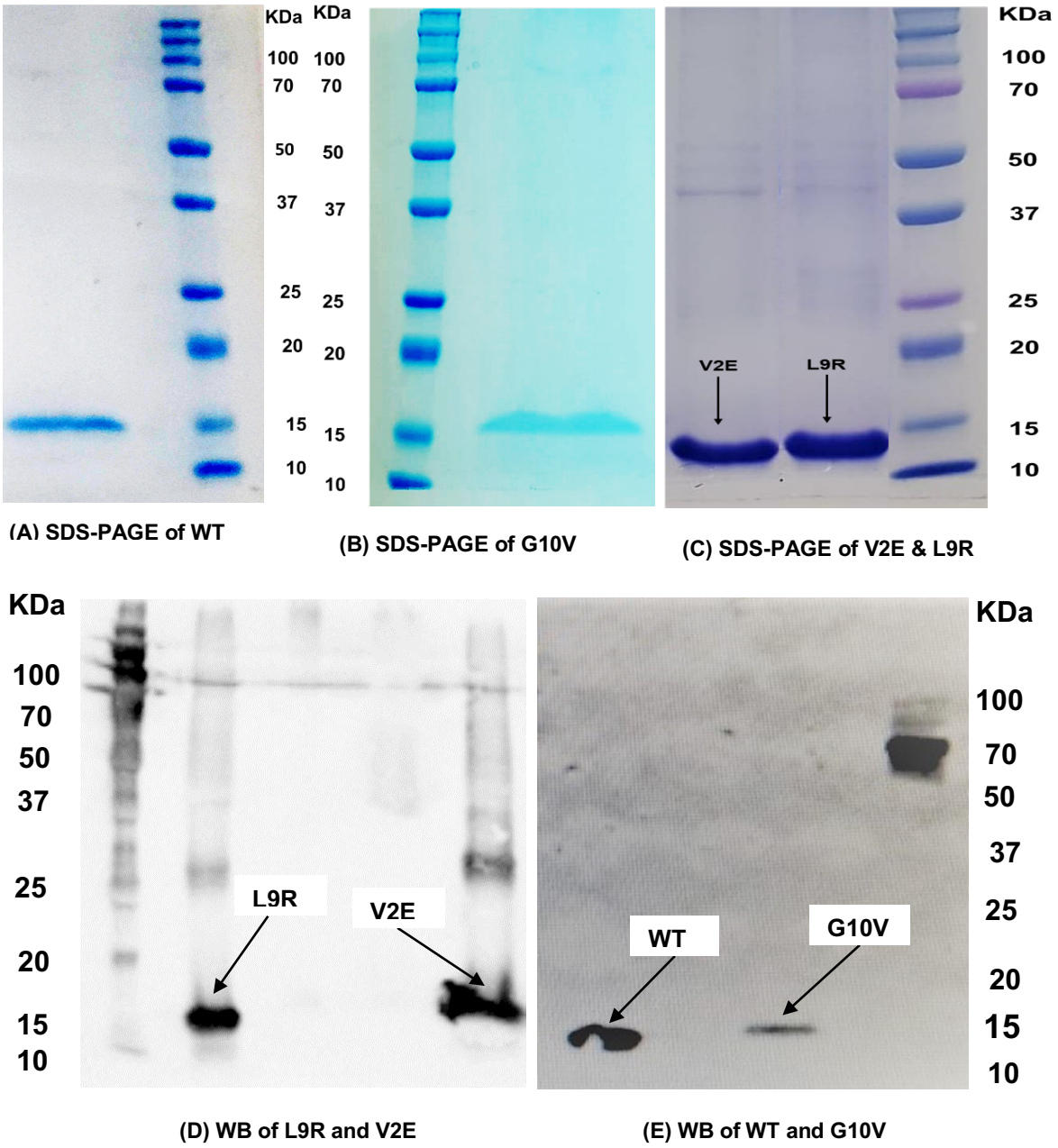


Figure 4.5: SDS-PAGE of combined 300- and 600-mM imidazole elutions from the Co^{2+} resin affinity chromatography of (A) Wild type FPHM (B) G10V_FPHM (C) V2E_FPHM and L9R_FPHM (D) WB of L9R and V2E after SDS-PAGE of different preparation. The L9R and V2E lanes have lighter bands at ~30 KDa; which could be a dimer of protein and band ~100 KDa could be due to contamination from the standard lane during transferring of SDS-PAGE gel to nitrocellulose paper (E) WB of WT and G10V after SDS-PAGE of different preparation. The band near ~70 KDa in (E) is for standard molecular marker and

Figure 4.5 (cont'd)

other position of molecular marker was determined after poking the gel before transferring to the nitrocellulose (NC) paper since all the standards band were visible in the gels before transferring to NC and only ~70 KDa band transferred to NC paper because antibody binds to that standard protein of MW~70 KDa.

4.3.2: Hyper thermostable proteins, WT and G10V are more helical than V2E & L9R

All the spectra displayed in Fig. 4.6 were acquired on the same day in Tris/SDS between 22 and 97°C. All spectra in panel A-E are characteristics of the predominant helical structure. $|\theta_{222}|$ (magnitude of mean residue molar ellipticity at 222 nm) is typically used to quantitate helicity. WT shows ~72% helicity using $|\theta_{222}|=3.3 \times 10^4$ deg.cm^2.dmol^{-1} for 100% helicity[26]. Earlier CD studies of WT had 68% helicity whereas 66% helicity was calculated using previous structural data from residues 23–70 and 117–164 (continuous N-helices and C-helices) in crystal structures of hairpin constructs that are similar to HM[20,27]. This difference in helicity between the earlier and present study of FPHM is negligible because experimental samples were from different expressions and purifications (can lead to differences in protein folding) and the experiment was not done on the same day (there could be there could be temperature, concentration variation). FP residues in WT FPHM have predominant β strand conformation supported by earlier CD and NMR studies and non-native SGGRGG loop and G6LEH6 tag have non-helical structure [21,23–25]. G10V shows 70% helicity at 22°C, V2E shows ~62% at 22°C, and L9R shows ~51% helicity at 22°C. WT and G10V show almost similar helicity (only ~2% difference) at 22°C whereas V2E show ~10% less helicity and L9R shows 20% less helicity, respectively than WT.

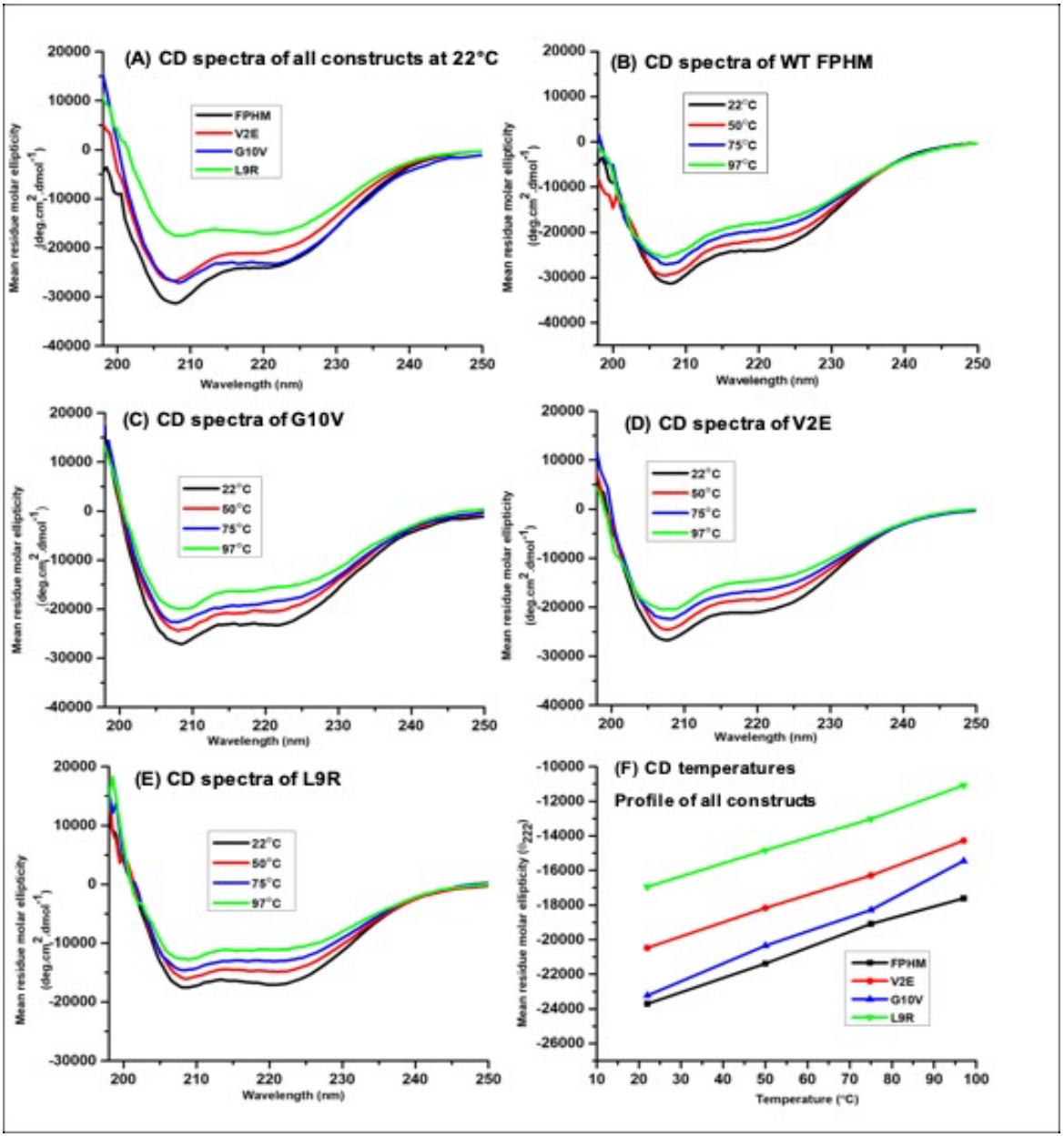


Figure 4.6: Circular dichroism spectra (CD) of (A) All constructs at 22°C (B) WT_FPHM (C) G10V (D) V2E (E) L9R (F) θ_{222} Vs. Temperature (°C) for all constructs. All the spectra in panel A-E were recorded on the same day. Typical protein concentrations of this study were 4.5-7.5 µM. In panel B-E all the spectra were at temperatures between 22 and 97°C and panel A shows the spectra of all constructs at 22°C. All spectra in panel A-E are characteristics of the predominant helical structure. $|\theta_{222}|$ (magnitude of mean residue molar ellipticity at 222 nm) is typically used to quantitate helicity.

Figure 4.6 (cont'd)

We also reproduced the similar WT vs. G10V helicity and greater WT vs. V2E /L9R helicity in Fig. 4.7. Reproducible CD spectra in Fig. 4.7 of WT show ~74% helicity, G10V show ~75% helicity, V2E show ~55% helicity, and L9R show ~48% helicity. Panel 4.6F shows all the constructs have the wavelength-dependent profile of characteristics helical structure to 97°C and evidence of some helical disordering but not transformation to a random coil structure, therefore Tm> 97 °C (since CD spectra is predominantly helical till 97°C) for all the constructs which is consistent with the previously reported WT Tm ≈ 110 °C from differential scanning calorimetry[28,29].

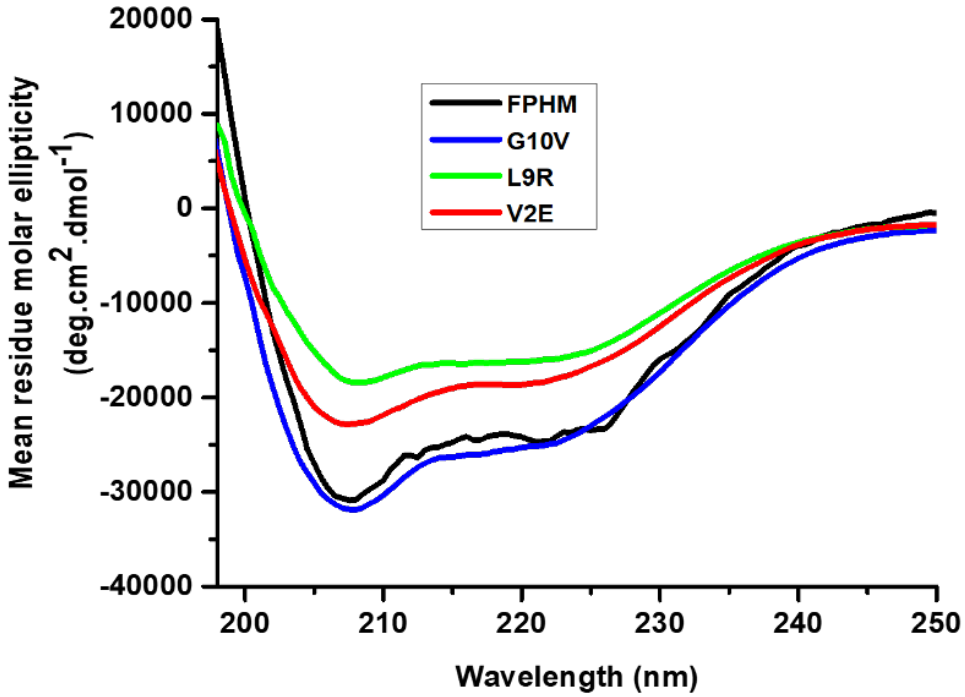


Figure 4.7: CD spectra of all four constructs at 22°C from different preparations than those used for the CD spectra in Fig. 4.6A. All spectra show characteristics of alpha-helical structures based on shallow minima near 208 and 222 nm. ($|\theta_{222}|$) of WT is 2.46 × 10^4 deg.cm^2.dmol^{-1}, which corresponds to ~74% helicity, ($|\theta_{222}|$) of G10V is 2.50 × 10^4 deg.cm^2.dmol^{-1} , which corresponds to ~75% helicity, ($|\theta_{222}|$) of V2E is 1.82 × 10^4 deg.cm^2.dmol^{-1} , which corresponds to ~55% helicity and ($|\theta_{222}|$) of L9R is 1.59 × 10^4 deg.cm^2.dmol^{-1} , which corresponds to ~48% helicity.

4.3.3: Protein-induced vesicle fusion at high, physiological, and low pH

Figure 4.8 displays lipid vesicle fusion induced by WT and other three different single gp41 ectodomain mutant constructs between pH's 3.12 and 8.40 at 37° C with lipid vesicle prepared from POPC:POPG at 4:1 and with lipid:protein (100:1)**.** All the constructs in this study show significant fusion at all pH's except L9R at pH 8.4 At pH 8.40, L9R shows a small apparent negative fusion which might be due to fluorophore dilution because of an increased volume of the vesicle solution upon the addition of protein stock[20]. For all pH's, G10V shows a maximum long time fusion extent (M_{ext}~600 s) than all other constructs. pH's 3.12-5.13, G10V shows ~90% fusion extent (M_{ext}), at pH's 6.30-7.45, G10V shows M_{ext} ~80% fand at pH 8.40 G10V shows M_{ext}~28%. pH's 3.12- 6.30, WT FPHM shows M_{ext} ~80%, at pH 7.45, G10V shows M_{ext} ~70% and at pH 8.40, WT shows M_{ext}~12%. Moreover, at pH 6.30, both G10V and WT show almost the same M_{ext}~80%. Here, M_{ext} of G10V shows the maximum fusion up to pH 5.12 and then decreases at pH 6.30 to higher pH's. On the other hand, M_{ext} of WT has the highest fusion up to pH 6.30 (or little more than that) and fusion decreases at pH 7.45 (M_{ext} ~70 %) to pH 8.40 (M_{ext} ~10-12 %). Both V2E and L9R show low fusion extent (M_{ext}) than G10V and WT at all pH's. M_{ext} of V2E ~50-52 % observes at pH 5.13, M_{ext} then decreases to second highest M_{ext}~36-38 % at pH 6.30 and M_{ext} ~8-10 % observes at pH's 7.45 -8.40. L9R shows highest M_{ext} ~38-40 % at pH's 4.23- 6.30, second highest M_{ext}~10-12 % at pH 3.12 & 7.45, respectively and little apparent negative fusion at pH 8.40. This negative fusion could be due to fluorophore dilution because of an increased volume of the vesicle solution upon the addition of protein stock[20].

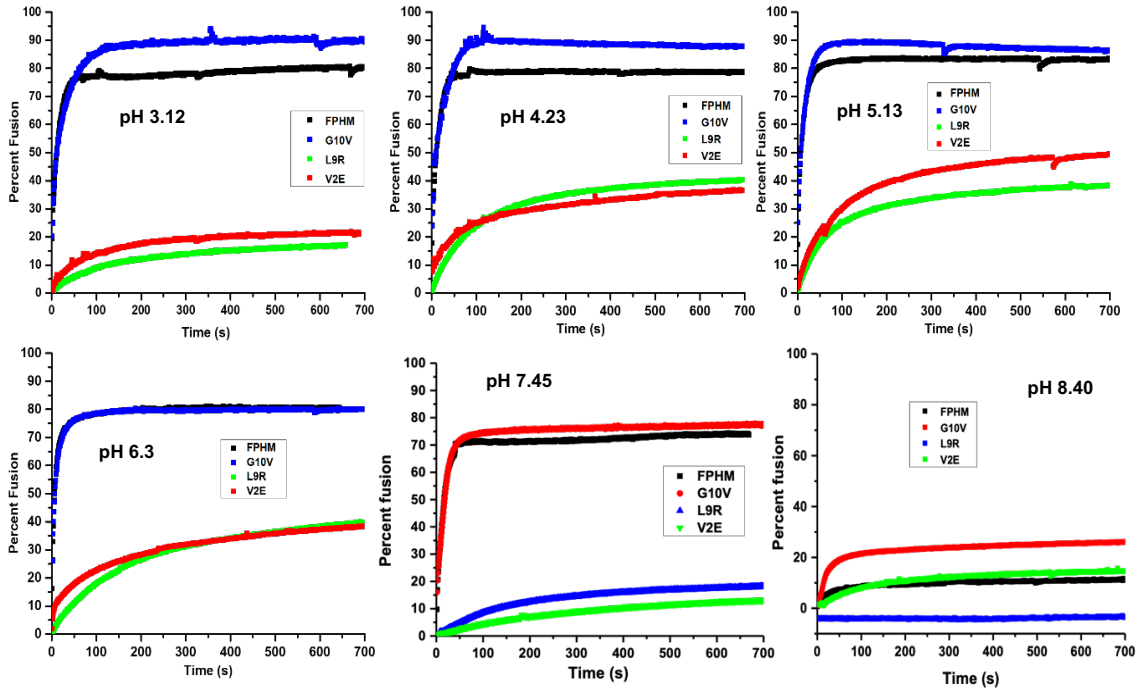


Figure 4.8: Respective wild type (WT), G10V, V2E and L9R induced vesicle fusion at pH's between 3.12 and 8.40 were done at 37 ◦C using vesicles that had been extruded through 100 nm diameter pores. The membrane lipid vesicle was prepared with unlabeled lipids and labeled lipids at 9:1 ratio. The unlabeled lipids were prepared with POPC and POPG at 4:1 ratio with a total 4.81 µmol. The labeled lipids were prepared from one aliquot of unlabeled lipid (4.81 µmol) to dilute it into ten aliquots (0.481 µmol). Each aliquot of labeled lipid also had 2% N-NBD-DPPE and 2% N-Rh-DPPE, respectively of total lipid. The total lipid concentration was = 150 µM and the vesicle solution contained 25 mM citrate buffer at pH = 3.12, 4.23, 5.13, and 6.30-, and 10-mM Tris buffer at pH = 7.40 and pH=8.40. The stock solution of proteins was 150 µM at pH 7.40 with 0.2% SDS and at time = 0, an aliquot of protein stock solution was added to the vesicle solution to achieve protein: lipid with 1:100 ratio. L9R shows a small apparent negative fusion at pH 8.40 due to fluorophore dilution because of an increased volume of the vesicle solution upon the addition of protein stock[20].

Lipid vesicle fusion from different samples (express & purified in different day, using same reagents for fusion with lipid: protein (600:1) is also shown in **Fig. 4.9.** The samples of different preparations also show a similar trend of M_{ext} as lipid: protein (600:1) for G10V

and WT at all pH's. However, V2E and L9R show very low M_{ext}~2-3 % at all pH's with lipid: protein (600:1). The reason behind the low fusion of L9R and V2E could be due to very low volume of proteins (~0.8 μl) was added to fusion cuvette with Lipid: Protein (600:1). Moreover, V2E/L9R also show very less fusion even at lipid: protein (100:1) where volume of proteins (~3 μl) in fusions cuvette.

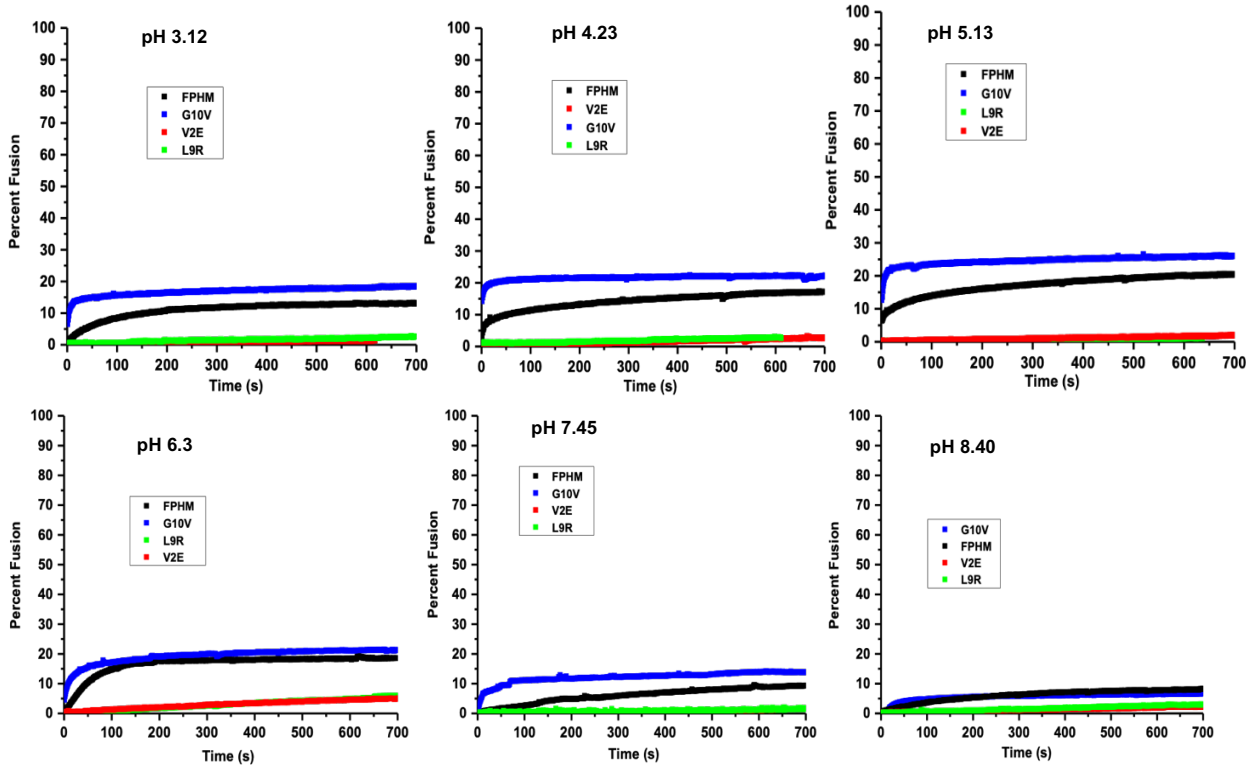


Figure 4.9: Membrane vesicle fusion assays of different batches of membrane lipid vesicles and different protein stocks with protein: lipid molar ratios of 1:600 at 37 °C. The vesicle solution had [total lipid] = 150 μM at pH's between 3.12 and 8.40 with POPC: POPG (4:1) composition and at time t= 0, an aliquot of protein stock (150 μM protein in 10 mM Tris buffer at pH 7.4 with 0.2% SDS) was added to the vesicle solution.

Earlier vesicle fusion study by Rokonujjaman et. al (2023) [Chapter-3, Figure 3.10] showed that WT_FPHM has the highest M_{ext}≈ 80% and V2E_FPHM has the highest M_{ext}≈ 30% at pH 5.13, respectively[20]. At pH 3.3, 4.3, 6.3, WT_FPHM showed M_{ext}<35% and at 7.4 it showed the lowest fusion (M_{ext}<10%). Besides, at pH 3.3, 4.3, 6.3, V2E_FPHM showed M_{ext}<10% and at 7.4 it showed the lowest fusion (M_{ext}<2%). Fusion results of

current study showed different results (more M_{ext} at all pH's) from earlier study by Rokonujjaman et. al (2023) though it had same lipid and protein composition and ratios. Current study showed WT_FPHM has M_{ext}>70% from pH's 3.12-7.45 and fusion decrease to M_{ext}≈10% at pH 8.40. Similarly, V2E_FPHM also showed higher values of M_{ext} at pH's (3.12-8.40) than earlier study by Rokonujjaman et. al (2023). This differences of results in fusion between current study and earlier study [Chapter 3, Rokonujjaman et. al (2023)] is unknown to us; however, better lipid mixing with protein (using constant slow stirring during lipid mixing assay; earlier studied had not have constant stirring of solution after addition of protein lipid solution)[20], some instrumental variabilities such as temperature control (using a thermostat during assay at 37^0C), light interference (room light can interfere with fluorescence instrument during lipid mixing and addition of protein to lipid etc. can make the difference in fusion.

4.3.4: Rate analysis of vesicle fusion

The time course of fusion of POPC:POPG (anionic vesicle) by four different HIV gp41 ectodomain constructs were well fitted with either the sum of fast and slow exponential buildup[40]:

$$M(t)= M_f\ (1-e^{-k_f(t+c)}) + M_s\ (1-e^{-k_s(t+c)})$$

or a single slow exponential buildup[40]:

$$M(t)= M_s\ (1-e^{-k_s(t+c)})$$

M_s and M_f are the long-time fusion extents of the slow and fast processes, respectively. K_s and k_f are the fusion rate constant for the slow and fast processes, respectively. c is a constant that shifts the time variable. This could account for a delay, or an initial time offset in the fusion process. Table. Apx.2 shows all the fitted values (M_s, M_f, K_s , k_f & c) through single slow process and double slow and fast process for all protein constructs at Lipid:Protein ratios 100:1 and 600:1, respectively. At Lipid:Protein 100:1, WT shows highest rate at the single slow process (k_s) at all pH's except 8.40 whereas V2E shows highest k_s at 8.40. However, long time fusion extent (M_{ext}) at the single slow process is the highest for G10V for all pH's (Fig. 4.10 & Table. Apx. 2). This similar kind of trend also observed at Lipid:Protein 600:1 at the single slow process in Table. Apx. 2 and Fig. Apx. 2.2. L9R shows negative fusion at pH 8.4 and the curve was not fitted well either by single or double exponential buildup (Fig. 4.8 & 4.10). At pH 6.3, WT and G10V show similar K_s

and M_s and L9R and V2E also show similar K_s and M_s (Fig. 4.10). This kind of trend also observed for percent vesicle fusion in Fig. 4.8. WT and G10V show maximum rate (k_s≈88 ms^{-1}) at pH 6.3 and second maximum at pH 5.13 for WT (k_s≈74 ms^{-1}) and for G10V (ks≈68 ms^{-1}) at single slow process (Fig. 4.11 and Table Apx. 2). V2E shows a maximum k_s≈ 45 ms^{-1} at pH 8.4 (Fig. 4.11 and Table Apx. 2). WT shows rate at the single slow process, k_s≈ 62-88 ms^{-1} for pH's 3.12-7.45 and k_s≈ 9 ms^{-1} at pH 8.4. G10V shows rate at the single slow process, k_s≈ 30-88ms^{-1} for pH's 3.12-6.3, k_s≈ 52 ms^{-1} at pH 7.45 and k_s≈ 17 ms^{-1} at pH 8.4. V2E shows rate at the slow process, k_s≈ 3-8 ms^{-1} for pH's 3.12-7.45 and k_s≈ 45 ms^{-1} at pH 8.4.

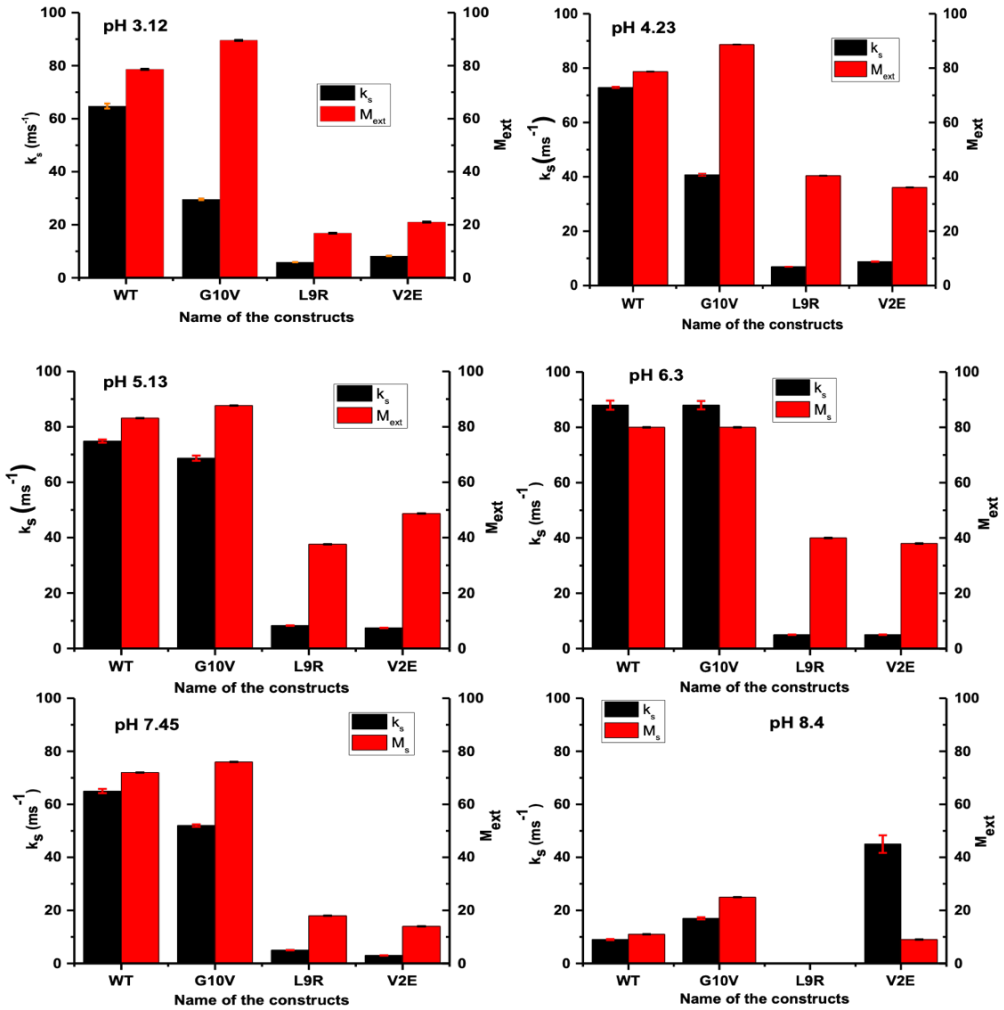


Figure 4.10: Comparison of k_s and M_s with error bars of four gp41 constructs at pH's 3.12-8.4 for lipid:protein 100:1. Though long time fusion extent at slow process (M_s) at pH 8.4

Figure 4.10 (cont'd)

was exceeded >100 for L9R and V2E, respectively all the plots are made vertical scale 0-100. k_s of L9R and V2E at pH 7.45 and 8.4, respectively are very negligible which does not visual at the bar graphs.

L9R shows rate at the slow process, $k_s \approx$ 4-8 ms^{-1} for pH's 3.12-7.45. k_s for the single slow process between V2E and L9R are 3-8 ms^{-1} at pH's 3.12-7.45 except $k_s \approx$ 45 ms^{-1} for V2E at pH 8.4.

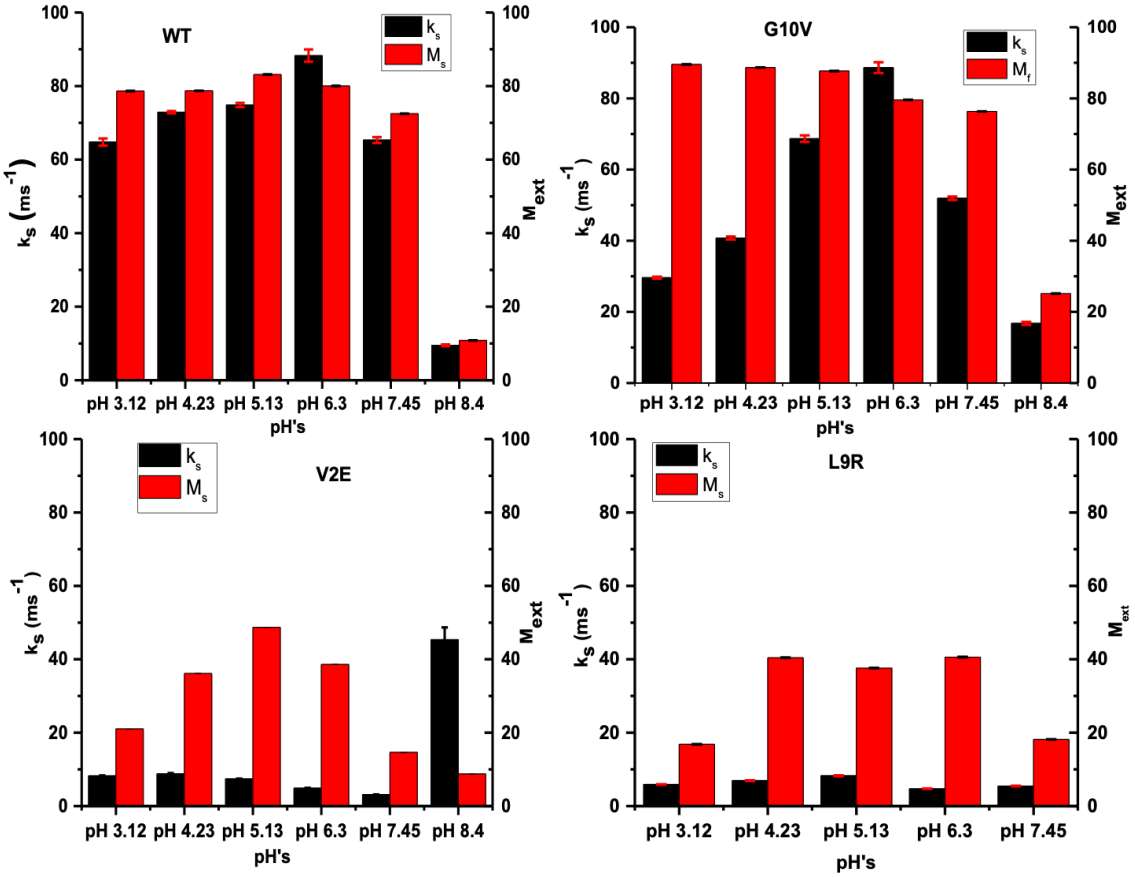


Figure 4.11: Comparison of k_s and M_s through single exponential buildup with error bars of four different gp41 constructs at different pH's (3.12, 4.23, 5.13, 6.3, 7.45 and 8.4) for lipid:protein 100:1. Some error in the bars of the plots are too small to observe. However, some errors of k_s and M_s in the plots are not determined after fitting with the single slow exponential buildup. A similar set of plots for kf for double buildup are shown in appendix 2.

The fast rate (k_f) of double exponential buildup of WT is lowest ($k_s \approx 47\ ms^{-1}$) at pH 8.4 and highest ($k_f \approx 793\ ms^{-1}$) at pH 5.13. The slow rate (k_s) of double exponential buildup of WT is lowest ($k_s \approx 3\ ms^{-1}$) at pH 8.4 and highest ($k_s \approx 73\ ms^{-1}$) at pH 4.23. G10V shows highest double exponential buildup, $k_f \approx 447\ ms^{-1}$ at pH 6.3 and lowest $k_f \approx 40\ ms^{-1}$at pH 4.23. The k_s of G10V for single exponential buildup is highest as $69\ ms^{-1}$at pH 5.13 and lowest as $1.0\ ms^{-1}$ [1]at pH 7.45. V2E shows highest $k_f \approx 45\ ms^{-1}$ and $k_s \approx 45\ ms^{-1}$ at pH 8.4 and lowest $k_f \approx 3.0\ ms^{-1}$at pH 7.45 and $k_s = 2\ ms^{-1}$at pH 4.23. L9R shows highest $k_f \approx 21\ ms^{-1}$ at pH 5.13 and $k_s \approx 5\ ms^{-1}$ at pH 3.12 and lowest $k_f \approx 8.0\ ms^{-1}$ and $k_s = 0.6\ ms^{-1}$at pH 6.3, respectively.

The two exponential buildups fitted well (R^2=0.93-0.99) for all the constructs at all pH's except V2E at pH 8.4 (R^2=0.55). k_f and k_s either double or single exponential buildup are same for WT at pH's 4.23 and 7.45, for G10V at pH's 4.23 and 5.13, for V2E at pH's 7.45 and 8.4. Either from single or two exponential buildup mutants (V2E/L9R) has always has low rate (k_s/k_f) compare to WT/G10V except V2E has highest values at pH 8.45.

All the k_s from slow single exponential buildup and k_f from two double exponential buildup are similar (close values) for WT at all pH's except pH 5.13, 6.3 and 8.4. k_f of WT at pH 5.13 has the values $793\ ms^{-1}$, at pH 6.3 $k_f \sim 160\ ms^{-1}$ and at pH 8.4 $k_f \sim 47\ ms^{-1}$. G10V shows all the k_s from slow single exponential buildup and k_f from two double exponential buildup are similar (close values) except values except pH 3.12, 6.3 and 8.4. k_f of G10V at pH 3.12 has the values $202\ ms^{-1}$, at pH 6.3 $k_f \sim 447\ ms^{-1}$ and at pH 8.4 $k_f \sim 48\ ms^{-1}$. WT and G10V have always higher k_s of slow single exponential buildup and k_s and k_f of two double exponential buildup from V2E/L9R except k_s of slow single process has higher value. Either k_s/k_f values in single/double exponential buildup process of V2E or L9R are quite small compared to WT/G10V. Therefore, mutation of valine (V) by glutamic acid (E) in V2E and Leucine (L) by arginine (R) in L9R has great impact on fusion in the pH's 3-8.5. Fast process might be favorable in pH's 5-6.5 because the FPHM oligomeric state (fusion mainly happens in trimeric state of FPHM; trimeric state is mostly favorable in this pH range) may be an important factor in the pH dependence[20]. However, slow process might be favorable at physiological pH 7.4 since all the fast and slow process rates (k_s and k_f) are same for WT.

4.4: Discussion

G10V and WT show a M_{ext}>70% at physiological pH (7.4) and WT shows M_{ext} ≈10 % and G10V shows M_{ext} ≈28 % fusion at pH (8.4). Moreover, V2E and L9R show M_{ext} ≈10 % and M_{ext} ≈15 % at pH 7.45, respectively. V2E shows M_{ext}≈7 % at pH 8.40 and L9R shows at negative M_{ext} at pH 8.40 (fluorophore dilution because of an increased volume of the vesicle solution upon the addition of protein stock). Vesicle membrane fusion at physiological pH is important because HIV viral fusion occurs at this pH and electrostatic attraction does not play a role in fusion (fusion mainly occurs between anionic lipids and negatively charged proteins) as it does for lower pH's. Therefore, hydrophobicity of WT and G10V, respectively plays an important role in fusion at pH 7.4. Liang et. al. (2018) showed large ectodomain construct containing FP and/or TM segments significantly triggered fusion of both neutral and anionic vesicles at neutral pH, reflecting expected physiologic circumstances of HIV/host cell fusion[11,38]. Fusion under physiologic conditions in their study explained due to inclusion of more hydrophobic segments, and also stock solutions with predominant trimer in the study versus monomer in previous studies[11, 39,40]. The hydrophobic effect's impact is demonstrated by the maximum fusion extent for FP_HM_TM, which has the highly conserved FP and TM regions[11,41,42]. The greater local concentration of FP and TM may be the cause of the trimer's fusion efficiency relative to monomer[11]. This is also correlated with a considerably higher vesicle fusion extent caused by FPs that have C-terminally cross-linked topologies, resembling hairpin trimer structure[11].

In current study, we see more vesicle fusion for WT_FPHM and G10V_FPHM at physiological pH compared to previous studies[11,20]. Moreover, V2E_FPHM and L9R_FPHM also show little fusion at physiological pH. Though we did not run size exclusion chromatography for proteins in this study, all samples were in tris buffer in fusion after using same expression, purification and dialysis protocol as Rokonujjaman et al. (2023) where protein was predominantly trimer. Therefore, our proteins were also predicted to be in trimeric condition in fusion. The reason behind more fusion by WT and mutants' protein at physiological pH in our study could be due to more purified proteins, large lipid to protein ratio (100:1; Liang et al. (2018) has 1:450)[11].

Despite not having similar sequences, Ha2 and gp41 have similar topologies, such as a viral Tmd and an ectodomain with about ~180 residues that adopts a trimer-of-hairpins structure in its final state[30,31]. Ha2 also has ~ 25-residue N-terminal Fp like gp41, which binds to the target membrane. Nonetheless, the Ha2 Fp typically adopts helical rather than β sheet structure in membrane and there isn't close spatial proximity between the helical Fp's of the same or different trimers[25,32-34]. I173E, located in the C-terminal region of the hairpin near the Tmd, and G1E, located at the N-terminus of the Fp, are two widely researched Ha2 mutations. Both the I173E and G1E mutations significantly lessen cell-to-cell fusion mediated by Ha and significantly lessen cell-to-cell fusion mediated by the Ha2 ectodomain protein "FHa2," which is comparable to FPHM[28,35-37]**.** Similar to the suppression of vesicle fusion by V2E/L9R FP_HM (Fig. 8), vesicle fusion triggered by Ha2 is likewise greatly attenuated with these mutations (figure 8a in reference 25)[25]. Moreover, vesicle fusion triggered by FHa2 is also attenuated with G1E mutation (figure 2a in reference 35)[35]**.** The melting point (T_m) of FHa2 is dramatically reduced by 35–40 ^oC by both I173E and G1E, in contrast to V2E/L9R (Fig. 6), so the loss of fusion caused by either mutation is associated with a less stable hairpin and a wider inter-membrane separation[25]. Given that the N-terminus of the WT helical Fp is spatially distant from the hairpin, the decrease in T_m for G1E is unexpected.

L9R shows a similar fusion decrease trend from WT as V2E. Rokonujjaman et. al. studied the trans-dominance effect through membrane vesicle fusion of V513E to wild-type ectodomain of gp41 showing the quantitative similarity of HIV gp160 V513E-dominant reduction of fusion and infection with fusion by the gp41 ectodomain hairpin supports an important fusion role for the final trimer-of-hairpins structure[19,20]. Though no studies so far showed dominance effect by L9R, attenuation of fusion as V513E by point mutation R to L in L520R further establishes the important role of the final trimer-of-hairpin structure.

G10V shows a similar/little higher membrane vesicle fusion trend as WT. However, Delahunty et. al. (1996) showed G10 is a critical element of HIV fusion peptide and G10V mutation has a great reduction of gp160 cell-cell fusion from the WT. Though this article did not explain the reason of reducing fusion due to G10V mutation; it could be asserted from the article that expression of G10V gp160 spikes is quite different from WT (number of syncytium formation for WT≈ 516 and for G10V≈ 0) and production of number of spikes

is different for G10V from WT (number of envelope expression for WT≈ 16890 and for G10V≈ 6756). Therefore, inability of syncytium formation/less envelope expression of G10V reduce the cell-cell fusion. On the other hand, membrane vesicle fusion by G10V is similar/a little higher than WT because more hydrophobicity in G10V (G is replaced by V increase hydrophobicity) helps in fusion[11,16].

Helicity influences fusion. A recent study by Md Rokonujjaman et. al. (2023) showed a decrease in V2E helicity influences a decrease in the fusion of V2E from WT. The model predicted extended β-sheet in V2E_FPHM form salt bridge or H-bond that loses few residues in N-helix in helix. Consequently, few residues in C-helix also lose their helicity, and membrane apposition distance (viral and host) increases and less fusion than WT. In this study, V2E and L9R show a decrease in helicity from WT and show a significant decrease in fusion from WT. Moreover, G10V and WT show similar helicity, and their fusion is also similar. These results will be established previous study findings since less helicity in V2E/L9R from WT vs. less fusion in V2E/L9R from WT and similar helicity in WT/G10V vs. similar fusion in WT/G10V.

A previous study, which used analytical ultracentrifugation data in dodecylphosphocholine (DPC) detergent of a construct including FPHM (WT) plus the transmembrane domain, shed light on the pH dependence of FPHM oligomerization[43]. At pH 5, there were roughly equal amounts of trimer and monomer, while at pH 4, there was predominant trimer than monomer. At pH 5.3, where the fusion rate is greatest, we expect FPHM to stay a trimer because a previous study shown higher trimer vs. monomer for FPHM in SDS compared to DPC[11]. The +2 charge at pH 5.3 compared to the neutral or negative charge at higher pH may be the cause of the more extensive fusion (rapid process) for FPHM at pH 5.3 vs. higher pH, with increased FPHM binding to the negatively-charged membrane at pH 5.3 as a result[39,40]. When pH is less than 5, FPHM may dissociate into monomers, which are most likely less fusion-active than trimers, which could explain the decreased fusion (less rate).

Initial viral and cell membrane hemifusion, or joining, is correlated with the lipid mixing assay of ectodomain-induced vesicle fusion. At low pH levels, the protein is monomeric in solution and is attracted to anionic membranes through electrostatic forces. At neutral pH, lipid and protein interactions are repulsive in nature and repulsive electrostatic forces

prevent lipid-protein binding. Moreover, protein aggregation might occur at neutral pH due inability of binding through electrostatic repulsion. In contrast to vesicle fusion at neutral pH, HIV/cell fusion is likely to occur at neutral pH with anionic membranes, where viral protein attached to the cell membrane without the help of electrostatic attraction between lipid and protein.

It is likely that the slow rather than the fast component of vesicle fusion better models' virus/cell hemifusion. The attractive protein/membrane electrostatic energy, which is substantially positively associated with the fast process, is probably negligible or even repulsive for viral fusion at neutral pH. Furthermore, the host cell membrane's |charge/lipid| is ~0.07 and HIV's is ~0.15[44]. Vesicle fusion is maximally slow and minimally fast for this regime. Although the rate of viral hemifusion has not been determined, the average total duration for viral fusion is approximately 30 minutes[45]. The slow process favors at pH 7.45 in ~65 ms^{-1} is close to the duration of viral fusion in comparison to the fast process in ~793 ms^{-1} at pH 5.13. G10V also shows similar kind of trend of slow/fast processes at pH's 7.45 and pH 5.13, respectively as WT since this mutation increase the hydrophobicity in FP region (maintain the originality of FP as all residues are hydrophobic in nature). However, V2E/L9R mutation decreases the rates Vs. WT/G10V at all pH's which indicates mutations not only has impact on great reduction of HIV/vesicle fusion but also rate in fusion[40].

REFERENCES

1. Stein, B. S., Gowda, S. D., Lifson, J. D., Penhallow, R. C., Bensch, K. G., and Engleman, E. G. (1987). pH-independent HIV entry into CD4- positive T cells via virus envelope fusion to the plasma membrane. Cell 49, 659–668.
2. Lifson, J. D., Feinberg, M. B., Reyes, G. R., Rabin, L., Banapour, B., Chakrabarti, S., Moss, B., Wong-Staal, F., Steimer, K. S., and Engleman, E. G. (1986). Induction of CD4 dependent cell fusion by the HTLV-III/LAV envelope glycoprotein. Nature 323, 725–728.
3. Sodroski, J., Goh, W. C., Rosen, C., Campbell, K., and Haseltine, W. A. (1986). Role of the HTLV-III/LAV envelope in syncytium formation and cytopathicity. Nature 322, 470–474.
4. Freed, E. O., and Martin, M. A. (1995). The role of human immunodefi-ciency virus type 1 envelope glycoproteins in virus infection. J. Biol. Chem. 270, 23883–23886.
5. M. Pancera et al. Structure and immune recognition of trimeric pre-fusion HIV-1 Env. Nature 2014, 514, 455-461.
6. D. Lyumkis et al. Cryo-EM Structure of a Fully Glycosylated Soluble Cleaved HIV-1 Envelope Trimer. Science 2013, 342, 6165, 1484-1490.
7. G. Ozorowski et al. Open and closed structures reveal allostery and pliability in the HIV-1 envelope spike. Nature 2017, 547, 560-363.
8. J. Liu et al. Molecular architecture of native HIV-1 gp120 trimers. Nature 2008, 455, 109-113.
9. J.P. Julien et al. Crystal Structure of a Soluble Cleaved HIV-1 Envelope Trimer. Science 2013, 342, 6165, 1477-1483.
10. McCune, J. M., Rabin, L. B., Feinberg, M. B., Lieberman, M., Kosek, J. C., Reyes, G. R., and Weissman, I. L. (1988). Endoproteolytic cleavage of gp160 is required for the activation of human immunodeficiency virus. Cell 53, 55–67.
11. S. Liang et al. Efficient Fusion at Neutral pH by Human Immunodeficiency Virus gp41 Trimers Containing the Fusion Peptide and Transmembrane Domains. Biochemistry 2018, 57, 1219−1235.
12. D. C. Chan et al. Core Structure of gp41 from the HIV Envelope Glycoprotein. Cell 1997, 89, 263-273.
13. V. Buzon et al. Crystal Structure of HIV-1 gp41 Including Both Fusion Peptide and Membrane Proximal External Regions. PLoS Pathogens 2010, 6 (5), e1000880.
14. M. Caffrey et al. Three-dimensional solution structure of the 44 kDa ectodomain of SIV gp4. EMBO Journal 1998,17, 16, 4572–4584.
15. Yang, Z. N et al., The crystal structure of the SIV gp41 ectodomain at 1.47 A resolution. J. Struct. Biol., 1999, 126, 131−144.

16. M. D. Delahunty, I. Rhee, E. O. Freed, and J. S. Bonifacino. Mutational Analysis of the Fusion Peptide of the Human Immunodeficiency Virus Type 1: Identification of Critical Glycine Residues. VIROLOGY 218, 94–102 (1996).

17. Myers, G. (1988). In ‘‘Human Retroviruses and AIDS 1988’’ (G. Myers et al., Eds.), p. 68.

18. Bosch, M. L., Earl, P. L., Fargnoli, K., Picciafuoco, S., Giombini, F., Wong, S. F., and Franchini, G. (1989). Identification of the fusion peptide of primate immunodeficiency viruses. Science 244, 694–697.

19. E.O. Freed, E.L. Delwart, G.L. Buchschacher Jr., A.T. Panganiban, 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 (1992) 70–74.

20. Md. Rokonujjaman, Abdulrazak Sahyouni, Robert Wolfe, Lihui Jia, Ujjayini Ghosh, and David P. Weliky. A large HIV gp41 construct with final trimer-of-hairpins structure exhibits V2E mutation-dominant attenuation of vesicle fusion and helicity that is quantitatively similar to V2E attenuation of HIV fusion and infection and supports: (1) hairpin stabilization of initial membrane apposition with larger apposition distance for V2E; and (2) V2E dominance mediated by an antiparallel β sheet with interleaved fusion peptide strands from two gp41 trimers. Biophysical Chemistry 293 (2023) 106933.

21. Maddon, P. J., Dalgleish, A. G., McDougal, J. S., Clapham, P. R., Weiss, R. A. & Axel, R. (1986) Cell 47, 333-348.

22. Freed, E. O., Myers, D. J. & Risser, R. (1989) J. Virol. 63, 4670-4675.

23. Freed, E. O., Myers, D. J. & Risser, R. (1990) Proc. Natl. Acad. Sci. USA 87, 4650-4654.

24. Freed, E. O., Myers, D. J. & Risser, R. (1991) J. Virol. 65, 190-194.

25. A. Ranaweera, P.U. Ratnayake, D.P. Weliky, The stabilities of the soluble ectodomain and fusion peptide hairpins of the Influenza virus hemagglutinin subunit II protein are positively correlated with membrane fusion, Biochemistry 57 (2018) 5480–5493.

26. Y. H. Chen, J.T. Yang, K.H. Chau, Determination of helix and beta-form of proteins in aqueous-solution by circular-dichroism, Biochemistry 13 (1974) 3350–3359.

27. V. Buzon, G. Natrajan, D. Schibli, F. Campelo, M.M. Kozlov, W. Weissenhorn, Crystal structure of HIV-1 gp41 including both fusion peptide and membrane proximal external regions, PLoS Pathog. 6 (2010) e1000880.

28. N. Lev, Y. Fridmann-Sirkis, L. Blank, A. Bitler, R.F. Epand, R.M. Epand, Y. Shai, Conformational stability and membrane interaction of the full-length ectodomain of HIV-1 gp41: Implication for mode of action, Biochemistry 48 (2009) 3166–3175.

29. K. Sackett, M.J. Nethercott, R.F. Epand, R.M. Epand, D.R. Kindra, Y. Shai, D. P. Weliky, 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 (2010) 301–315.

30. J. Chen, J.J. Skehel, D.C. Wiley, N- and C-terminal residues combine in the fusion-pH influenza hemagglutinin HA2 subunit to form an N cap that terminates the triple-stranded coiled coil, Proc. Natl. Acad. Sci. U. S. A. 96 (1999) 8967–8972.

31. A. Ranaweera, P.U. Ratnayake, E.A.P. Ekanayaka, R. Declercq, D.P. Weliky, Hydrogen-deuterium exchange supports independent membrane-interfacial fusion peptide and transmembrane domains in subunit 2 of influenza virus hemagglutinin protein, a structured and aqueous-protected connection between the fusion peptide and soluble ectodomain, and the importance of membrane apposition by the trimer-of-hairpins structure, Biochemistry 58 (2019) 2432–2446.

32. J.C. Macosko, C.H. Kim, Y.K. Shin, The membrane topology of the fusion peptide region of influenza hemagglutinin determined by spin-labeling EPR, J. Mol. Biol. 267 (1997) 1139–1148.

33. J.L. Lorieau, J.M. Louis, A. Bax, The complete influenza hemagglutinin fusion domain adopts a tight helical hairpin arrangement at the lipid: water interface, Proc. Natl. Acad. Sci. U. S. A. 107 (2010) 11341–11346.

34. U. Ghosh, L. Xie, L.H. Jia, S. Liang, D.P. Weliky, Closed and semiclosed interhelical structures in membrane vs closed and open structures in detergent for the Influenza Virus hemagglutinin fusion peptide and correlation of hydrophobic surface area with fusion catalysis, J. Am. Chem. Soc. 137 (2015) 7548–7551.

35. C.S. Kim, R.F. Epand, E. Leikina, R.M. Epand, L.V. Chernomordik, The final confirmation of the complete ectodomain of the HA2 subunit of Influenza Hemagglutinin can by itself drive low pH-dependent fusion, J. Biol. Chem. 286 (2011) 13226–13234.

36. H. Qiao, R.T. Armstrong, G.B. Melikyan, F.S. Cohen, J.M. White, A specific point mutant at position 1 of the influenza hemagglutinin fusion peptide displays a hemifusion phenotype, Mol. Biol. Cell 10 (1999) 2759–2769.

37. E. Borrego-Diaz, M.E. Peeples, R.M. Markosyan, G.B. Melikyan, F.S. Cohen, Completion of trimeric hairpin formation of influenza virus hemagglutinin promotes fusion pore opening and enlargement, Virology 316 (2003) 234–244.

38. Zaitseva, E., Zaitsev, E., Melikov, K., Arakelyan, A., Marin, M., Villasmil, R., Margolis, L. B., Melikyan, G. B., and Chernomordik, L. V. (2017) Fusion stage of HIV-1 entry depends on virus-induced cell surface exposure of phosphatidylserine. Cell Host Microbe 22, 99−110.

39. Banerjee, K., and Weliky, D. P. (2014) Folded monomers and hexamers of the ectodomain of the HIV gp41 membrane fusion protein: Potential roles in fusion and synergy between the fusion peptide, hairpin, and membrane-proximal external region. Biochemistry 53, 7184−7198.

40. Ratnayake, P. U., Sackett, K., Nethercott, M. J., and Weliky, D. P. (2015). 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, Biomembr. 1848, 289−298.

41. Durell, S. R., Martin, I., Ruysschaert, 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.

42. Long, Y., Meng, F., Kondo, N., Iwamoto, A., and Matsuda, Z. (2011) Conserved arginine residue in the membrane-spanning domain of HIV-1 gp41 is required for efficient membrane fusion. Protein Cell 2, 369−376.

43. Z. Dai, Y.S. Tao, N.N. Liu, M.D. Brenowitz, M.E. Girvin, J.R. Lai, Conditional trimerization and lytic activity of HIV-1 gp41 variants containing the membrane-associated segments, Biochemistry 54 (2015) 1589–1599.

44. B. Brugger, B. Glass, P. Haberkant, I. Leibrecht, F.T. Wieland, H.G. Krasslich, The HIV lipidome: a raft with an unusual composition, Proc. Natl. Acad. Sci. U. S. A. 103 (2006) 2641–2646.

45. S.A. Gallo, C.M. Finnegan, M. Viard, Y. Raviv, A. Dimitrov, S.S. Rawat, A. Puri, S. Durell, R. Blumenthal, The HIV Env-mediated fusion reaction, Biochim. Biophys. Acta 1614 (2003) 36–50.

CHAPTER-05 (Project- 02): ^{13}C-^{2}H REDOR solid state NMR for the location of hairpin of HIV gp41 in membrane.

5.1: Introduction

The HIV viral envelope (Env) spike is the trimer of protomers (each protomer consists of a gp120 and a gp41) that dominantly contributes for HIV-1 fusion (the process by which HIV viral membrane fuses with the host CD4 cell membrane) into host cell[1]. Env spike facilitates HIV-1 entry by rearranging from a mature unliganded state (pre-fusion), through receptor and co-receptor bound intermediates (Env activation), gp41 catalysis to a final fusion state[1]. Initially synthesized as a gp160 precursor, Env is trimerized and subsequently cleaved by proteases of the furin family into gp120 and gp41 subunits, which associate noncovalently before the native complex reaches the surface of infected cells and is then packaged into virions[2]**.** During fusion, gp120 binds to receptor (CD4) and co-receptors (CXCR4, CCR5) of host cell followed by gp41 dissociates from gp120 for catalyzing fusion of HIV virus into the host cell[3]. The ectodomain part of gp41 has ~170-amino residues located outside the virus which is important for fusion into host cell, and it is subdivided into several domains i.e., fusion peptide (FP) ($gp41_{513-535}$), N-helix ($gp41_{536-596}$), loop, C-helix ($gp41_{616-675}$), and membrane proximal external region (MPER) ($gp41_{676-683}$). Among these ectodomain amino acid residues, ~150 residues (Hairpin (HP= N-helix- loop- C- helix) + MPER= HM) are considered as large soluble ectodomain (SE) of gp41[4]. To compare the initial gp41 ectodomain structure in complex with gp120 to the later gp41 structure without gp120 (final fusion state) shows large-scale extensions and topological rearrangements of helical segments[2]. The latter structure is like a “trimer of hairpins” in which each hairpin (HP) is formed by N-helix- loop- C-helix. This “trimer of hairpins” arranges like six helix bundle (SHB) in which three N-helices ($gp41_{536-596}$) that form an interior bundle, followed by loops, and then C-helices ($gp41_{616-675}$) that run antiparallel in the exterior grooves of the bundle[4,5,6,7,8]. In the six-helix bundle (SHB), three N-helices form parallel coiled coil wrapped in a gradual left-handed superhelix and three C-helices wrapped antiparallel to the N-helices in a left-handed direction around the outside of the center coiled coil trimer[5]**.**

^{13}C-labeled glucose or glycerol used as the sole carbon source for labeling the various carbons of amino acids in protein. Using uniform labeled glucose, or 25% uniform labeled glucose as the sole carbon source in M9 (minimum) media, allows for the uniform labeling of the carbon of amino acids in protein. Similarly, the sole carbon supply in M9 medium

for the growth of selectively labeled protein can be [1,3-^{13}C] glycerol, [2-^{13}C] glycerol, [1-^{13}C] or [2-^{13}C] glucose. Growing a protein on [1,3-^{13}C] glycerol, [2-^{13}C] glycerol, [1-^{13}C], or [2-^{13}C] glucose as the sole carbon source known as isotopic dilution and which simplifies the labeling patterns of amino acids and offers several practical advantages[9-13]**.** In the isotopic dilution method, we can label certain carbon sites with ^{13}C at high levels while the other sites remain completely unlabeled by supplying specifically ^{13}C-labeled glucose or glycerol. Some carbons of amino acid are labelled by either [1,3-^{13}C] glycerol or [2-^{13}C] glycerol and rest is unlabeled and consequently, spin is diluted in amino acids in a protein. Through the removal of most one bond ^{13}C–^{13}C J-coupling, the selective labeling approach improves carbon resolution of NMR spectra (spectrum without J-coupling are sharper and clearer, which enhances the resolution and makes it easier to distinguish closely spaced signals). Spectroscopic assignments of aromatic side chains have been demonstrated to be facilitated by selective labeling, which also increases resolution and simplifies the spectrum in the aromatic region[14]. Selective labeling can target specific parts of the molecule, allowing for detailed study of the aromatic region without interference from other parts of the molecule. The selective labeling would likely result in a spectrum where the aromatic side chains produce distinct, well-resolved peaks, making it easier to assign these signals to specific atoms or groups within the aromatic region. Furthermore, reduction of dipolar truncation effects (the recoupling of a small dipolar coupling between a distant spin pair is severely attenuated in the presence of a third spin strongly coupled to one of the first two spins)[15], simplification of polarization transfer (transfer of nuclear spin polarization from spins with large Boltzmann population differences to nuclear spins of interest with low Boltzmann population differences)[16] pathways, and facilitation of measuring long-range distance restraints, which are critical for structural determination is helped through dilution of ^{13}C spins by selective labeling in a protein.

Rotational-echo double-resonance (REDOR) solid-state NMR is one of the best tools for probing the membrane locations of specific residues of membrane proteins. REDOR helps to enable unambiguous assignment (In complex spectra where multiple signals overlap, REDOR can help isolate interactions between specific nuclei, making it easier to

assign the signals to specific atoms or residues in the molecule) and ^{2}H labeling covers a small region of the lipid or cholesterol molecule, and couplings are measured between protein ^{13}C nuclei and ^{2}H nuclei[17]. The robustness of $(\Delta S/S_0)^{exp}$ with respect to rf fields, such as ^{1}H decoupling, rf inhomogeneity, resonance offsets, quadrupolar and chemical shift anisotropies, and magic angle spinning (MAS) frequency, are one of the few advantages of REDOR approach's[18]. There was coherent magnetization transfer between spins and data analysis could be validated using protein containing isolated spin pairs with a single dipolar coupling d and internuclear distance r with $d \infty r^{-3}$. The experimental $(\Delta S/S_0)^{exp}$ buildup with dephasing time τ is fitted to a buildup rate and long-time extent that are connected with the fractional population with a specific d. ^{13}C is coupled to several ^{2}H's or ^{31}P's with various pairwise r's and d's in a membrane protein sample. The extent and accumulation rate are used to semi-quantitatively calculate the fractional population and effective d and r. The $d \infty r^{-3}$ indicates that the closest ^{2}H or ^{31}P usually dominates the buildup[18]. The dipolar coupling between two nuclei decreases rapidly with distance, following a r^{-3} dependency. Therefore, nuclei that are closest to the observed nucleus will have much stronger dipolar interactions compared to those that are farther away.

For an enveloped virus fusion protein, there is an initial structure that is very different from the hairpin and the preponderance of the literature denotes the hairpin as the final “post-fusion” state, i.e. a state that forms after fusion is complete and which doesn’t have a functional role in fusion. Data from our and other groups support a different model that the hairpin is formed prior to the start of fusion, i.e. when both the viral envelope membrane and cell membrane are both still intact[19]. For this latter model, the thermostability of the hairpin is required to maintain the apposition of the two membranes through dehydration prior to fusion. In the absence of catalyst, such apposition has a calculated energy 25 kcal/mole higher than separated membranes (this also known as highest energy step during fusion). Hairpin thermostability is evidenced by unusually high melting temperatures of fusion protein hairpin structures, e.g. 110 ^{o}C for the HM protein construct from the HIV gp41 ectodomain[4,19,20,21]. The formation of the 6-HB is highly energetically favorable. The energy released during this transition helps to overcome the

activation energy barrier of approximately 25 kcal/mole. The energy released from the hairpin formation drives the two membranes (viral and host) into close proximity, effectively overcoming the repulsive forces and energy barrier[22]. There aren't yet clear data on hairpin binding and there aren't any data about membrane location of the hairpin. There are at least three possibilities: (1) hairpin is not membrane-bound; (2) hairpin is bound and surface-associated; and (3) hairpin is bound and deeply inserted[17] (Fig. 2.1). A hairpin contribution to catalysis of membrane fusion is most straightforwardly understood if there is deep insertion as the locations of the surrounding lipid molecules in the membrane would likely be very different than their energetically stable positions in an unperturbed bilayer. Furthermore, a hairpin position in the periphery of lipid bilayer of membrane could help better understanding of membrane apposition through dehydration in fusion (Fig.5. 1)[22].

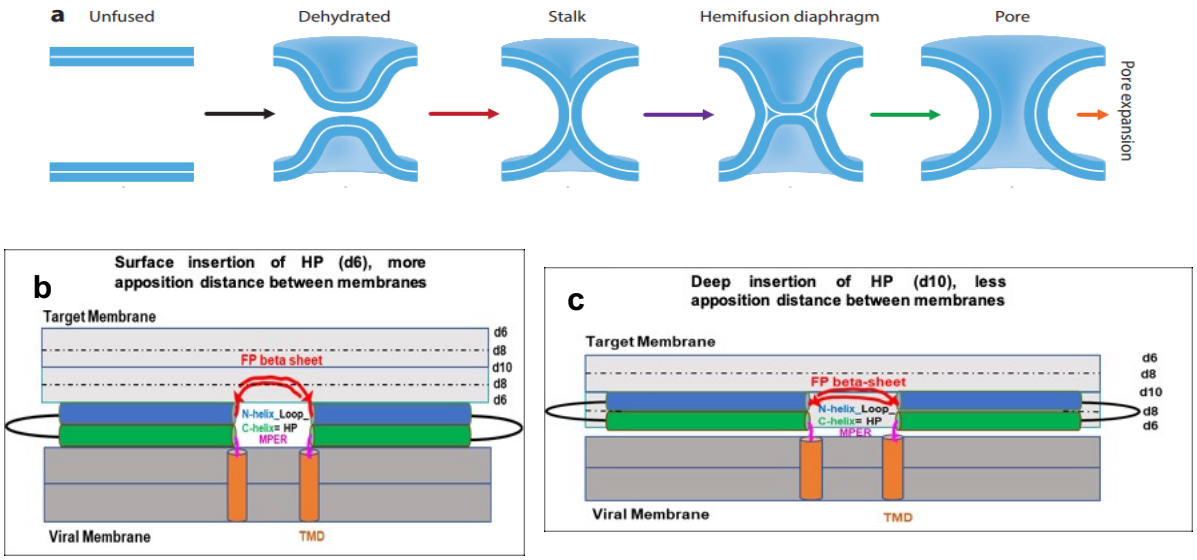


Figure 5.1: (a) Schematic representation of intermediates in the canonical membrane fusion pathway[22]. (b) approximate hairpin (HP) location the periphery of membrane; more apposition distance between host and viral membrane in fusion (c) approximate hairpin (HP) location in the middle of the lipid bilayer in membrane; less apposition distance between host and viral membrane in fusion.

Computational analysis showed apposition of membranes through dehydration is the most energy demanding (~25 Kcal) step in the fusion process[22]. By knowing the apposition distance could further help to know details about the dehydration process.

Schematic representation of hairpin in the periphery of membrane in Fig. 5.1 (b) has more apposition distance compare to less apposition distance in Fig. 5.1 (c) whereas hairpin in the middle of the lipid bilayer in the membrane. Therefore, hairpin position in the membrane perhaps help to know the apposition distance between membranes; then help to know the dehydration process in the fusion.

For this project, we produced "^{13}CHM" gp41 soluble ectodomain construct (contains hairpin) in bacteria using either uniformed ^{13}C-glucose or selective labeled ^{13}C-glycerol and then used REDOR ssNMR to measure proximity between the ^{13}CHM and ^{2}H labels in lipids in the membrane. Proximity of ^{13}CHM to $^{2}HD_{10}$ lipid (Fig. 2.1) lead to the middle of the lipid bilayer in the membrane, proximity of ^{13}CHM to $^{2}HD_{8}$ lipid (Fig. 2.1) lead to the middle of the leaflet of the lipid bilayer in the membrane and proximity of ^{13}CHM to $^{2}HD_{6}$Chol/D_{4}Lipid (Fig. 2.1) lead to periphery of the lipid bilayer in the membrane.

5.2: Materials and Methods

5.2.1: Materials

Avanti Polar Lipids provides the lipids. Typically, the phosphatidylglycerol headgroup lipid (PG) was 1,2-dipalmitoyl-sn-glycero-3-phospho-(10-rac-glycerol) (sodium salt), and the phosphatidylcholine headgroup lipid (PC) was 1,2-dipalmitoyl-sn-glycero-3-phosphocholine. Avanti custom-synthesized PC_d10, and PC_d8, utilizing deuterated palmitic acids derived from CDN isotopes (Fig. 2.1). Other reagents, such as cholesterols labeled with ^{2}H (d6-Chol), were usually purchased from Sigma-Aldrich. The following companies provided the other materials: GenScript (Picataway, NJ) for DNA; Novagen (Gibbstown, NJ) for the Escherichia coli BL21(DE3) strain; Dot Scientific (Burton, MI) for Luria-Bertani (LB) medium; Goldio (St. Louis, MO) for isopropyl β-D-thiogalactopyranoside (IPTG) and tris-(carboxyethyl) phosphine (TCEP); Thermo Scientific (Waltham, MA) for co^{2+}-resin. The 1,3- ^{13}C Gly, 2- ^{13}C Gly and U_^{13}C Glu were obtained from Cambridge Isotope Laboratories (Andover, MA). The majority of the additional materials came from St. Louis, MO's Sigma-Aldrich.

5.2.2: Protein region of interest, expression and purification

Based on the HXB2 laboratory strain of HIV, the gp41 amino acid sequence starts from 512 to 856 and ectodomain part (FP+HM) is from 512-683. HM (Amino acids sequence: CTLTVQARQLLSGIVQQQNNLLRAIEAQQHLLQLTVWGIKQLQARILSGGRGGWMEDR

EINNYTSLIHSLIEESQNQQEKNEQELLELDKWASLWNWFNITNWLWYIKGGGGGGLE HHHHHH) is the soluble ectodomain (SE) of gp41 which develops a hyper thermostable helical hairpin shape with a melting temperature of T_m > 100^{o}C. Though a cysteine residue present in the beginning of N-terminal (C534) of HM with histag/without histag there is very less probability of forming a disulfide bond between inter/intra chain HM. Buzon et al. (2010) showed that C534 are crowed with other amino acids residues except cysteine in the crystal structure of residues (gp 528-581 & 628-683)[6]. Therefore, C534 has very low probability to form inter/intra disulfide linkage in the HM structure. Liang et al. (2018) also showed by SEC (Fig. S5C in reference 4) that HM is predominantly trimer before and after treating with Dithiothreitol (DTT; is a reducing agent which beaks disulfide bonds) which indicates cysteine residue in HM does not form disulfide bond[4]. 1.0 mL glycerol stock was then grown in 50 ml LB medium for 3 hours followed by harvesting the cell pellet to begin production of ^{13}C labeled HM protein. The cell pellet was suspended in 10 mL minimal medium (It was prepared by mixing autoclaved aqueous solutions that included: (1) 50 mL with M9 salts (0.34 g Na_2HPO_4, 0.15 g KH_2PO_4, 0.03 g NaCl, and 0.05 g NH_4Cl); (2) 1 mL with 0.1 M $CaCl_2$; and (3) 1 mL with 1 M $MgSO_4$, as well as 0.5 mL MEM vitamin, and 200 mg glucose), and a 2 mL aliquot of this suspension was then added to the remaining minimal medium, followed by 4 h growth ($OD_{600} \approx 0.6$). Individual 5 mL aliquots of the suspension were then added to four separate flasks that each contained 50 mL fresh minimal medium with 200 mg 25% uniform ^{13}C labeled glucose (for uniform labelling of amino acids in protein), followed by 3h growth, induction of expression by adding solid (0.03g) of isopropyl β-D-1-thiogalactopyranoside (IPTG) and growing the cell overnight to express protein. After that, the cell pellet was harvested by centrifugation (9000g for 10 min) followed by PBS washing with tip sonication. Protein with selective ^{13}C-labeling by 1,3-^{13}C glycerol (99%) and 2-^{13}C glycerol (99%) was done by following the same procedure as ^{13}C glucose labelling of protein.

Centrifugation at 9000 g for 15 min was used to harvest the cell pellet, which was then kept at -20^{0}C. In the following stages, the pH of the PBS was 7.4, the tip was sonicated in an ice bath, and the sample was centrifuged at 4 ◦C for 30 to 45 minutes using 48,000 g. Because HM is not very soluble in PBS, it was isolated from PBS-soluble molecules by sonicating a 2 g cell pellet in 15 mL of PBS, centrifuging the mixture, and then collecting

the fresh pellet. This process repeats two times. The pellet after PBS wash was solubilized by sonication in PBS containing 8 M urea. After that, certain dissolved compounds, including HM, precipitated when [NaCl] was raised to about ~300 mM. The subsequent procedures were followed: (1) centrifugation, harvesting, and vortexing the pellet in distilled water; (2) centrifugation, harvesting, and pellet dissolving in 10 mL PBS with 6 M GuHCl; (3) centrifugation and collecting the supernatant; and (4) Co^{2+}-affinity chromatography of the pellet. The chromatography was carried out in a fritted column using gravity filtration to remove the liquid, and each solution comprised PBS with 6 M GuHCl. The final procedure involved the following steps: (1) filtering 2 mL of resin suspension; (2) washing the resin with 4 mL of 10 mM imidazole; (3) adding the HM-containing supernatant and agitating it for an overnight period at 4°C; and (4) washing/elution in batches of 6 mL without imidazole, 4mL of 10 mM imidazole, 2 mL of 300 mM imidazole, and 6 mL of 600 mM imidazole. After combining the elutions containing 300- and 600-mM imidazole, they were dialyzed overnight against distilled water, causing precipitation. The suspension was then centrifuged, the pellet was harvested and lyophilized, and it was stored at -20 ^{0}C.

On the other hand, ^{13}CHM without His-tag (obtained from uniform 25% labelled ^{13}C glucose in minimal medium) was purified with (1) 2 times PBS lysis at pH 7.4 followed by centrifugation (2) 6M GuHCl in PBS lysis at pH 7.4 followed by centrifugation and (3) dialysis against Millipore water followed by centrifugation, freeze drying. Finally, Sodium dodecyl sulfate−polyacrylamide gel electrophoresis (SDS−PAGE) and MALDI mass spectrometry was performed to check the purity of expressed protein, respectively.

5.2.3: Mass spectrometry-based protein characterization

A Kratos Analytical Axima-CFR Plus device was used for MALDI-TOF mass spectrometry. One miL of 98% formic acid was used to vortex the protein (~0.1 mg), and a 2- µL aliquot was subsequently pipetted with four microliters of a solution containing α-cyano-4-hydrocinnamic acid (10 mg/mL) in a 3:1 acetonitrile:0.1% TFA ratio. The mixed solution was placed in a 2 µL aliquot on the MALDI plate, dried, and then run through MALDI-TOF in linear positive mode. M^{+} and M^{2+} signals were frequently present, and $(m/z)M^{+} \approx 2 \times (m/z)M^{2+}$ was used to attribute each signal to a specific chemical species.

5.2.4: Preparation of solid-state NMR sample

Dipalmitoylphosphatidylcholine (DPPC) and 1,2-Dipalmitoyl-sn-glycero-3-phospho-(1'-rac-glycerol) (Sodium Salt)) (DPPG) had a 4:1 ratio in the lipid composition, while the mole ratio for DPPC, DPPG, and cholesterol was 8:2:5. Both compositions have a cholesterol mole fraction that is comparable to the plasma membrane of HIV host cells. Because the host cell plasma membrane of the HIV-1 virus is negatively charged and has a charge similar to the 4:1 DPPC:DPPG charge, this composition was selected. Moreover, the PC lipids make up a significant portion of the host cell plasma membrane. A 2 mL solution of chloroform and methanol with a 9:1 volume ratio was used to dissolve 20 μmol of lipids. The solvent was then extracted using dry nitrogen gas flow and vacuum pumping for a whole night. The lipid film was hydrated using 3 mL of 10 mM 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES) and 5 mM 2-(N-morpholino) ethane sulfonic acid (MES) buffer at pH 5.0. This was followed by 10 cycles of freeze-thaw to create a homogenous suspension of unilamellar vesicles. To obtain big unilamellar vesicles, the lipid-buffer suspension was extruded ten times through a polycarbonate membrane with a pore size of 100 nm. Large unilamellar vesicles (LUV) are between 100 nm and 1 μm, gigantic unilamellar vesicles (GUV) are above 1 μm, and small unilamellar vesicles (SUV) are usually below 100 nm. After dissolving the target protein (0.2 μmol) in 1 mL of HEPES/MES buffer, the protein was gradually introduced to lipid vesicles drop by drop and incubated for an entire night. In order to pelletize the peptide-lipid combination, ultra-centrifugation was used for four hours at 45000 g. Overnight, the proteo-liposome complex pellet underwent lyophilization. Lyophilization aids in minimizing sample loss during the NMR rotor sample packing process. The material was placed inside the NMR rotor and allowed to rehydrate for an entire night at room temperature using 4μL of the HEPES/MES buffer.

5.2.5: Solid state NMR spectroscopy

Oxford NMR AS400 magnet, Bruker Avance Neo console and a Phoenix magic angle spinning (MAS) probe with a 1.6 mm diameter rotor were used for the SSNMR investigations. The probe was simultaneously set to ^{1}H, ^{13}C, and ^{2}H NMR frequencies. The following pulse sequence was used to capture rotational-echo double-resonance (REDOR) data: (1) ^{1}H ~$\pi/2$ pulse; (2) ^{1}H-^{13}C cross polarization (CP); (3) dephasing period

of duration and (4) ^{13}C detection. The pulses applied during the dephasing period varied between the S_0 and S_1 REDOR data, which were gathered in an alternate order. Every rotor cycle, with the exception of the last one, ended with a ^{13}C π pulse for S_0 and S_1, and a ^{2}H π pulse was applied in the middle of each rotor period for S_1. (1) 10 kHz MAS frequency and 1.5 ms CP contact time; (2) 50 kHz rf fields for ^{1}H $\pi/2$ pulse and CP; (3) 55-66 kHz ^{13}C CP ramp; (4) 60 kHz ^{13}C π pulses, 100 kHz ^{2}H π pulses with XY-8 phase cycling applied to all π pulses; and ~70 kHz two-pulse phase-modulated ^{1}H decoupling during dephasing and acquisition. Typical values of recycle delays of 1 s for 3.2 ms & 16 ms dephasing, 1.5 s for 24 ms dephasing, and 2 s for 40 ms dephasing. External ^{13}C chemical shift reference was carried out utilizing the adamantane methylene peak at 40.5 ppm.

5.3: Results

5.3.1: Protein purification and characterization

In this study, we expressed uniform labelled ^{13}CHM using uniform 25% labelled ^{13}C glucose and selective labelled ^{13}CHM from 1,3-^{13}C labelled glycerol and 2-^{13}C labelled glycerol as carbon source in minimal medium. All the proteins had with C-terminal His tag except the ^{13}CHM using uniform 25% labelled ^{13}C glucose. The reason behind using the different protein than Histag-HM was to check if there any difference purification without histag. After expressing from minimal medium, recombinant protein (RP) contains ^{13}CHM firstly purified by cell lysis with tip sonication in PBS at pH 7.4 followed by centrifugation at 48000g for 20 mins and harvesting the new pellet. This step is repeated for two times. Secondly, cell pellet remaining after PBS lysis was again lysed by tip sonication with 6 M GnHCl in PBS at pH 7.4, followed by centrifugation 48000g for 20 mins. As a result of this process, ^{13}CHM with other little impurities will be dissolved in GnHCl. Further, Co^{2+} affinity chromatography of soluble protein (contains ^{13}CHM with His-tag) in GnHCl was done by eluting with 300 mM+600 mM imidazole followed by dialysis against Millipore water, centrifugation and freeze drying the precipitate. Finally, Sodium dodecyl sulfate−polyacrylamide gel electrophoresis (SDS−PAGE) and MALDI mass spectrometry was performed to check the purity of expressed protein.

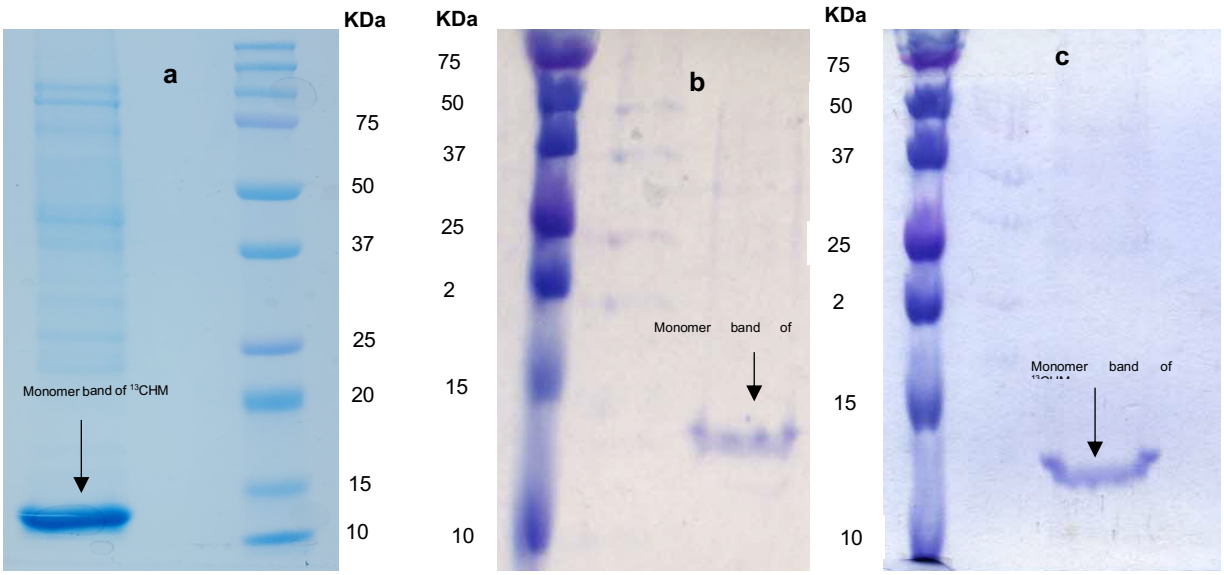


Figure 5.2: SDS-PAGE of the purified (a) ^{13}CHM from U-^{13}C labeled glucose as carbon source (b) ^{13}CHM from 1,3-^{13}C glycerol as carbon source (c) ^{13}CHM from 2-^{13}C glycerol as carbon source.

Fig. 5.2 shows, SDS-PAGE of the purified ^{13}C HM expressed from minimal medium using different ^{13}C sources (a) ^{13}CHM using 25% uniformed labeled ^{13}C glucose as carbon source (b) ^{13}CHM using 1,3-^{13}C glycerol as carbon source (c) ^{13}CHM using 2-^{13}C glycerol as carbon source. All the gels show clear HM monomer band ~13.5 KDa. Gel of uniformed labeled glucose also has some other bands which are not as intense as monomer band of HM. However, these bands could be impurity bands, and this can be observed by MALDI spectrum in Fig. 5.3 (a).

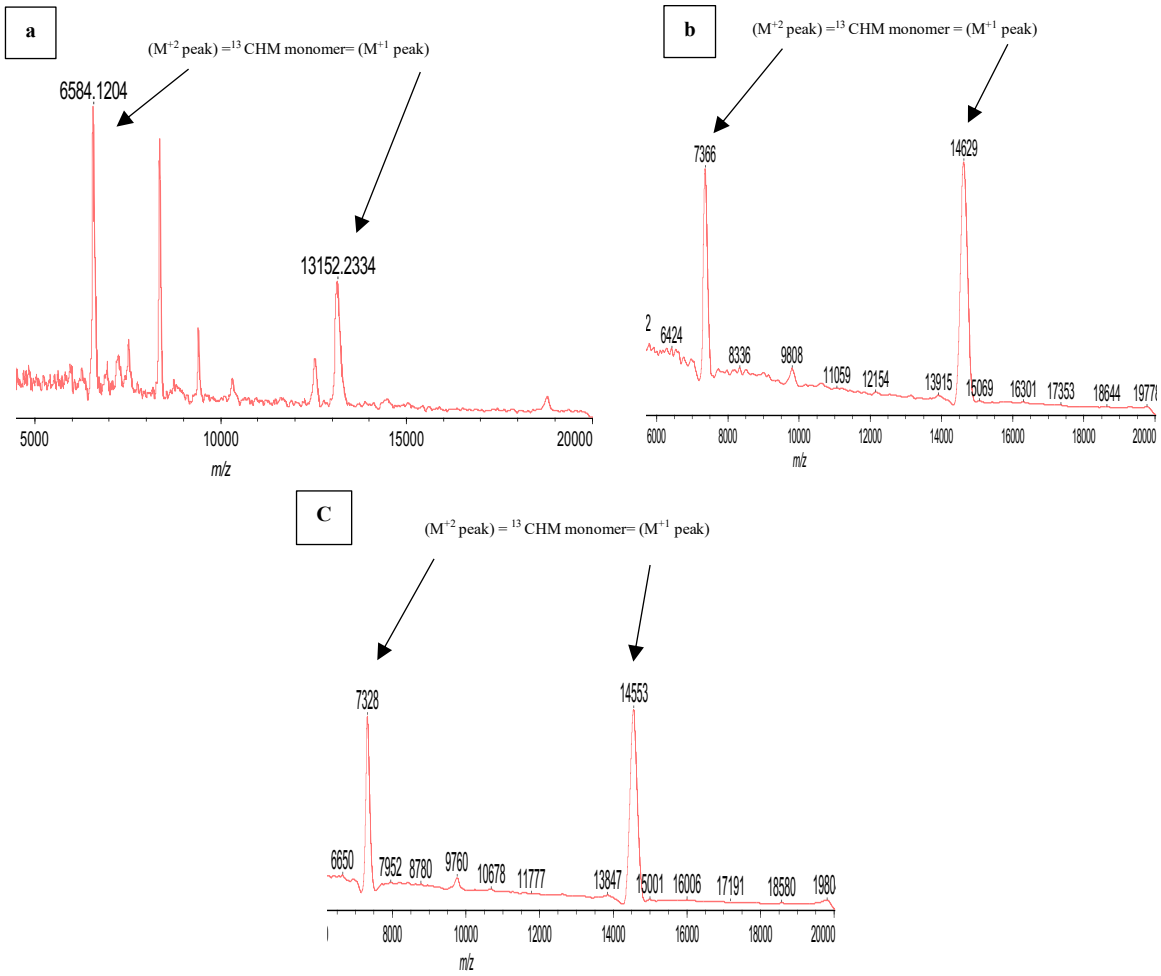


Figure 5.3: MALDI-MS spectrum of purified ^{13}C HM expressed from minimal medium using carbon source as (a) 25% uniformed labeled ^{13}C glucose (b) 99% 1,3-^{13}C glycerol (c) 99% 2-^{13}C glycerol.

In Fig. 5.3 (a), there is ^{13}CHM monomer (expressed from minimal medium with 25% uniformed labeled ^{13}C glucose) M^{+1} peak at M.W. 13152 which is 191 higher than calibrated molecular weight (MW) of HM without His-tag (M.W.12961). This additional mass (191unit) with ^{13}CHM without His-tag is due to labeling of HM from 25% uniformed labeled ^{13}C glucose (act as carbon source for amino acids) in minimal medium. Calibration of MW of HM without His-tag was done by running MALDI with unlabeled HM without His-tag with the same condition as MALDI run with labeled HM without His-tag (uniformed labeled ^{13}C glucose as carbon source). Moreover, this additional mass indicates HM without his-tag expressed from minimal medium with 25% uniformed

labeled ^{13}C glucose is 32.74% labeled (% Carbon Labeling=(Number of ^{13}C in HM/ Total number of Carbon in HM)×100%; equation used in this study of % Carbon Labeling= Mass difference between HM in MALDI spectrum and HM with His tag / Total number of carbons (632) in HM with His tag) × 100% and % Carbon Labeling= Mass difference between HM in MALDI spectrum and HM without His tag / Total number of carbons (585) in HM without His tag)×100%). On the hand, if we look at the Fig. 5.3 (a), there are few peaks of impurities between ^{13}CHM M^{+1} (m/z 13152) and M^{2+} (m/z 6584) peaks. These impurities peaks indicate that PBS urea/GnHCl purification protocol for purifying HM without His-tag did not completely purify the protein. In Fig. 5.3 (b), there is ^{13}CHM MALDI mass spectrum with His-tag (expressed from minimal medium with 1,3-^{13}C glycerol) M^{+1} peak at m/z 14629 and M^{+2} peak at m/z 7366 and in between there is no peaks for impurities. This indicates purification protocol using Co^{2+} affinity chromatography of ^{13}CHM with His-tag works great than only PBS urea/GnHCl protocol of HM without His-tag. This M^{+1} (M.W. 14629) peak of HM is 265 mass unit higher than calibrated calculated mass (M.W. 14364) of HM with His-tag. This extra mass indicates labeling of amino acids of HM by 1,3-^{13}C glycerol in minimal medium and percent labeling of ^{13}CHM is ~42%. Calibration of MW of HM with His-tag was done by running MALDI with unlabeled HM with His-tag with the same condition as MALDI run with labeled HM with His-tag (1,3/2-^{13}C glycerol as carbon source). Fig. 5.3 (c) shows MALDI mass spectrometry M^{+1} peak of ^{13}CHM with His-tag (expressed from minimal medium with 2-^{13}C glycerol) at m/z 14553 and M^{+2} peak at m/z 7328 and in between there is no peaks for impurities. This also indicates purification protocol using Co^{2+} affinity chromatography of ^{13}CHM with His-tag works great than only PBS urea/GnHCl protocol of HM without His-tag. HM with His-tag is ~30% ^{13}C labeled from minimal medium using 2-^{13}C glycerol as carbon source of HM. Moreover, MALDI confirms HM is labelled by ^{13}C either using uniform $^{13}CGlu$/1,3-$^{13}CGly$/2-$^{13}CGly$ is a significant extent.

5.3.2: ^{13}C- ^{2}H REDOR for membrane location of HM

This study used rotational-echo double-resonance (REDOR) solid-state NMR for probing the membrane locations of residues of HM incorporated into host membrane lipid. We expressed ^{13}CHM from minimal medium using 25% uniform labeled ^{13}C glucose (U-$^{13}CGlu$) and selective labeled glycerol (1,3-^{13}C Gly and 2-^{13}C Gly) as carbon source of

amino acids. This expressed protein was purified by various washing and purification protocols i.e. Co^{2+} affinity chromatography, dialysis, PBS washing, urea washing, GnHCl etc. This labeled protein was then incorporated into lipid mixture of DPPC + DPPG + cholesterol with D_{10} (^2H)/ D_8 (^2H)/ D_6 Chol (^2H) and proteoliposome pellet was made through ultracentrifugation followed by lyophilization. REDOR was applied to probe the location of ^{13}C in various amino acids in HM into various position in lipid through measuring dephasing during magnetic dipolar coupling between ^{13}C and ^2H. The buildup of experimental dephasing, $[(\Delta S/S_0)^{exp}= (S_0-S_1)/S_0]$ vs. dephasing time (τ) of a sample provides the basis for assessing the protein labeled (lab) ^{13}C-to-lipid ^2H proximity and r (internuclear distance) values[17,18,23]. S_0 and S_1 ^{13}C intensities determined were used to calculate experimental $(\Delta S/S_0)^{exp}= (S_0-S_1)/S_0$ values at each τ. The S_0 and S_1 error is based on spectral noise, which is the standard deviation of 5 spectral noise regions with 3 ppm integration window[17,18, 24].

$$\Delta_{S/S0} = \sqrt{\left(\frac{s_{S_0}}{S_0}\right)^2 + \left(\frac{s_{S_1}}{S_1}\right)^2} \cdot \frac{S_1}{S_0} \qquad \text{..... E5.1}$$

REDOR buildups have typically been fitted to yield precise distances. For example, the ^{13}CO-detect/^{15}N-dephase buildup of a lab $^{13}CO_{res.i}$, $^{15}N_{res.\ i+4}$ α helical peptide is well-fitted with a single d = 45 Hz and corresponding r = 4.1 Å that support a helical structure in all molecules[17, 24]. However, there are various reasons why the fits of the protein lab ^{13}C–membrane ^2H builds varied. (1) Even though all of the lab ^{13}C's have the same membrane insertion depth, the distance between each lab ^{13}C and a specific ^2H will differ amongst protein molecules due to the locally non-crystalline nature of the membrane environment (2) Multiple ^2H's are present in the lipid, or Chol, and the $(\Delta S/S_0)^{exp}$ of a given protein represents the relative angles and r's of multiple ^{13}C–^2H internuclear vectors. Each protein molecule will have a unique multi-spin geometry. The effect is mitigated for each spin pair by $d \propto r^{-3}$, so that the d associated with the closest ^2H should dominate the $(\Delta S/S_0)^{exp}$ accumulation for a given molecule (the dipolar coupling between two nuclei decreases rapidly with distance, following a r^{-3} dependency. Therefore, nuclei that are closest to the observed nucleus will have much stronger dipolar interactions compared to those that are farther away) (3) During the dephasing period, there are $m = 0 \leftrightarrow m = \pm 1$ ^2H transitions since the ^2H T_1 relaxation ~ 50 ms[17,18]. Nothing is building up for a lab ^{13}C at the m = 0

times. Each ^{2}H is in the m=0 state for approximately 1/3 of the dephasing period so that the observed buildup rate $\gamma \approx$ 2d/3; where d is dephasing constant. A sample including isolated ^{13}C–^{2}H intra-peptide spin pairs with a single r was reported to exhibit this relationship[18]. Overall, these peptide ^{13}CO–lipid ^{2}H REDOR considerations suggest that fitting parameters will relate to membrane location semi-quantitatively rather than quantitatively.

Three types of labeled ^{13}CHM (uniform ^{13}CHM labeled from 25% uniform labeled ^{13}C glucose, selective ^{13}CHM labeled from either 1,3-^{13}C glycerol or 2-^{13}C glycerol) have been expressed in this study. Each labeled ^{13}CHM was incorporated either into D_{10}DPPC+ DPPG+ Chol/ D_8DPPC+ DPPG+ Chol/ DPPC+ DPPG+ D_6Chol at 8:2:5 molar ratio. Finally, REDOR NMR was performed for each of ^{2}HLipid + ^{13}CHM sample.

5.3.3: Calculation of dephasing by ^{13}CHM only in ^{2}HLipid + ^{13}CHM samples

$$\boldsymbol{S_0^{Total}(\tau)=} S_0^{Lipid}(\tau) + S_0^{Protein}(\tau) = 1$$

$$\Rightarrow \boldsymbol{S_0^{Protein}(\tau)=} 1 - S_0^{Lipid}(\tau)$$

$$\boldsymbol{S_1^{Total}(\tau)=} S_1^{Lipid}(\tau) + S_1^{Protein}(\tau)$$

$$\Rightarrow \boldsymbol{S_1^{Total}(\tau)/ S_0^{Total}(\tau)=} [\{S_1^{Lipid}(\tau)/S_0^{Lipid}(\tau)\} \times S_0^{Lipid}(\tau) + \{S_1^{Protein}(\tau)/S_0^{Protein}(\tau)\} \times S_0^{Protein}(\tau)]/ S_0^{Total}(\tau)$$

$$\Rightarrow \boldsymbol{\gamma^{Total}(\tau)=} [\gamma^{Lipid}(\tau) \times S_0^{Lipid}(\tau) + \gamma^{Protein}(\tau) \times S_0^{Protein}(\tau)]$$

$$\Rightarrow \boldsymbol{\gamma^{Protein}(\tau)=} \{\gamma^{Total}(\tau) - \gamma^{Lipid}(\tau) \times S_0^{Lipid}\}/S_0^{Protein}$$

Therefore, **Dephasing by ^{13}CHM**, $\Delta S^{Protein}(\tau) / S_0^{Protein}(\tau) = 1 - S_1^{Protein}(\tau)/S_0^{Protein}(\tau)$

$$= 1 - \boldsymbol{\gamma^{Protein}(\tau)} \text{ .....E5.2}$$

In order to calculate $\gamma^{Protein}(\tau)$, we calculated $\gamma^{Total}(\tau)$ from Lipid+Protein sample at particular dephasing time (τ) and $\gamma^{Lipid}(\tau)$ from lipid sample at particular dephasing time (τ). $S_0^{Protein}(\tau)$ calculate from $\boldsymbol{S_0^{Lipid}(\tau)=} 1 - S_0^{Protein}(\tau)$, whereas $S_0^{Protein}(\tau)$ calculate from following equation:

$$\boldsymbol{S_0^{Protein}(\tau)=} \gamma^{Total}(\tau) - \gamma^{Lipid}(\tau)\}/ \gamma^{Protein}(3.2\text{ ms}) - \gamma^{Lipid}(\tau)$$

We assume that $\gamma^{Protein}(3.2\text{ ms}) = 1$, since at lower dephasing time both $S_0^{Protein}$ and $S_1^{Protein}$ has same value, hence no dephasing. We use $\gamma^{Protein}(3.2\text{ ms}) = 1$ for calculating $S_0^{Protein}(\tau)$ for other higher dephasing times i.e. 16 ms, 24 ms and 40 ms at particular peak region, because spin contribution will be same for all dephasing times.

5.3.4: Calculation of propagation of error only for ^{13}CHM in ^{2}HLipid + ^{13}CHM samples

$$S_0^{Total}(\tau) = S_0^{Lipid}(\tau) + S_0^{Protein}(\tau) = 1$$

$$\gamma^{Lipid}(\tau) = \frac{S_1^{Lipid}(\tau)}{S_0^{Lipid}(\tau)}$$

$$\gamma^{Total}(\tau) = \frac{S_1^{Total}(\tau)}{S_0^{Total}(\tau)}$$

$$\sigma_{\gamma}^{Lipid}(\tau) = \sqrt{\left(\frac{\sigma_{S_0}(Lipid)}{S_0(Lipid)}\right)^2 + \left(\frac{\sigma_{S_1}(Lipid)}{S_1(Lipid)}\right)^2} \cdot \frac{S_1(Lipid)}{S_0(Lipid)}$$

$$\sigma_{\gamma}^{Total}(\tau) = \sqrt{\left(\frac{\sigma_{S_0}(Total)}{S_0(Total)}\right)^2 + \left(\frac{\sigma_{S_1}(Total)}{S_1(Total)}\right)^2} \cdot \frac{S_1(Total)}{S_0(Total)}$$

$$\sigma^{\gamma^{Total}(\tau) - \gamma^{Lipid}(\tau)} = \sqrt{\left(\sigma_{\gamma}^{Total}\right)^2 + \left(\sigma_{\gamma}^{Lipid}\right)^2}$$

$$\sigma^{1-\gamma^{Lipid}(\tau)} = \sqrt{(0)^2 + \left(\sigma_{\gamma}^{Lipid}\right)^2}$$

$$S_0^{Protein}(\tau) = \gamma^{Total}(\tau) - \gamma^{Lipid}(\tau) / 1 - \gamma^{Lipid}(\tau)$$

$$\sigma^{S_0{}^{Protein}(\tau)} = \sqrt{\left(\frac{\sigma^{\gamma^{Total}(\tau) - \gamma^{Lipid}(\tau)}}{\gamma^{Total}(\tau) - \gamma^{Lipid}(\tau)}\right)^2 + \left(\frac{\sigma^{1-\gamma^{Lipid}(\tau)}}{1-\gamma^{Lipid}(\tau)}\right)^2} \cdot \frac{\gamma^{Total}(\tau) - \gamma^{Lipid}(\tau)}{1-\gamma^{Lipid}(\tau)}$$

$$S_0^{Lipid} = 1 - S_0^{Protein}$$

$$\sigma^{S_0{}^{Lipid}(\tau)} = \sqrt{(0)^2 + \left(\sigma^{S_0{}^{Protein}(\tau)}\right)^2}$$

$$\sigma^{\gamma^{Lipid}(\tau) \times S_0{}^{Lipid}(\tau)} = \sqrt{\left(\frac{\sigma^{\gamma^{Lipid}(\tau)}}{\gamma^{Lipid}(\tau)}\right)^2 + \left(\frac{\sigma^{S_0{}^{Lipid}(\tau)}}{S_0{}^{Lipid}(\tau)}\right)^2} \cdot \left(\gamma^{Lipid}(\tau) \times S_0{}^{Lipid}(\tau)\right)$$

$$\sigma^{\gamma^{Total}(\tau) - (\gamma^{Lipid}(\tau) \times S_0{}^{Lipid}(\tau))} = \sqrt{\left(\frac{\sigma^{\gamma^{Total}(\tau)}}{\gamma^{Total}(\tau)}\right)^2 + \left(\frac{\sigma^{\gamma^{Lipid}(\tau) \times S_0{}^{Lipid}(\tau)}}{(\gamma^{Lipid}(\tau) \times S_0{}^{Lipid}(\tau))}\right)^2}$$

$$\sigma^{\gamma^{Protein}(\tau)} = \sqrt{\left(\frac{\sigma^{\gamma^{Total}(\tau) - (\gamma^{Lipid}(\tau) \times S_0{}^{Lipid}(\tau))}}{\gamma^{Total}(\tau) - (\gamma^{Lipid}(\tau) \times S_0{}^{Lipid}(\tau))}\right)^2 + \left(\frac{\sigma^{Protein}(\tau)}{S_0{}^{Protein}(\tau)}\right)^2} \cdot \frac{\gamma^{Total}(\tau) - (\gamma^{Lipid}(\tau) \times S_0{}^{Lipid}(\tau))}{S_0{}^{Protein}(\tau)}$$

$$\sigma^{\frac{\Delta S^{Protein}}{S_0}}(\tau) = \sqrt{(0)^2 + \left(\sigma^{\gamma^{Protein}(\tau)}\right)^2} \quad \text{..... (E5.3)}$$

5.3.5: Calculation of errors of internuclear distance, r (^{13}C-^{2}H)

The uncertainty of average internuclear distance r (^{13}C-^{2}H)= $(4642/d)^{1/3}$ is calculated by

$$\sigma(\mathbf{r}) = 4642^{1/3} \times \sigma(d^{-1/3})$$
$$= 4642^{1/3} \times (-1/3) \times d^{-4/3} \times \sigma(d)$$
$$= (-1/3) \times (4642/d)^{1/3} \times d^{-1} \times \sigma(d)$$
$$= (-r/3) \times \sigma(d)/d \text{.....} \textbf{(E5.3)}$$

Where the -sign can be eliminated because the uncertainty is often given as a positive number (magnitude), and σ(d) is the uncertainty of the dipolar coupling d. The generic equation to calculate the uncertainty of A^n is the basis for deriving σ(r):

$$\sigma(A^n) = n \text{ x } A^{n-1} \text{ x } \sigma(A) \text{ .....} \textbf{(E5.4)}$$

where σ(A) is the uncertainty of A. According to equation **E5.4**, the relative uncertainty of r is only 1/3 of that of d, i.e. σ(r)/r = (1/3) x σ(d)/d.

5.3.6: ^{13}CHM - to- $D_{10}{}^{2}$HPC REDOR NMR

REDOR spectra are shown for D_{10}Lipid in Fig. 5.4, Uniform (U)_^{13}CGlucose_^{13}CHM + D_{10}Lipid in Fig. 5.5, 1,3_^{13}CGlycerol_^{13}CHM + D_{10}Lipid in Fig. 5.6 and 2_^{13}CGlycerol_^{13}CHM + D_{10}Lipid in Fig. 5.7 at four dephasing times (3.2, 16, 24 and 40 ms). Peak at 30-38 ppm spectral region for all spectra is assigned for lipid -CH_2 and also for protein $^{\beta}$C-Val/ $^{\beta}$C-Ile/ $^{\beta}$C-Phe/ $^{\gamma}$C-Met/ $^{\delta}$C-Lys. Peak at 22-28 ppm spectral region assigned for lipid -CH_2 (-CH_3) and also for protein $^{\gamma}$C-Ile/$^{\gamma}$C-Ile/$^{\gamma}$C-Lys/ $^{\gamma}$CThr/ $^{\gamma}$C-Arg. Peak at 12-18 ppm spectral region is the peak of lipid -CH_3 and also for protein $^{\beta}$C-Ala, $^{\gamma}$C-Val, $^{\delta}$C-Ile, $^{\epsilon}$C-Met. Spectral region at 172-180 ppm refers for both lipid and protein >C=O peak. Table Apx. 3.41 shows ^{13}C chemical shifts of different lipids. U-^{13}CGlu_^{13}CHM + D_{10}Lipid sample does not show >C=O peak at 172-180 ppm region at 24 and 40 ms, respectively because there could be faster T2 relaxation of the sample. There is a peak at 52-62 ppm region for 2-^{13}CGly_^{13}CHM + D_{10}Lipid which refers to $^{\alpha}$C-Ala/$^{\alpha}$C-Cys/$^{\alpha}$C-Thr/$^{\alpha}$C-Leu/$^{\alpha}$C-Val/$^{\alpha}$C-Lys/$^{\alpha}$C-Ile/$^{\alpha}$C-Met/$^{\alpha}$C-Phe/$^{\alpha}$C-Tyr/$^{\alpha}$C-Ser. D_{10}Lipid does not show this peak which indicates only protein contributes for this spectral region.

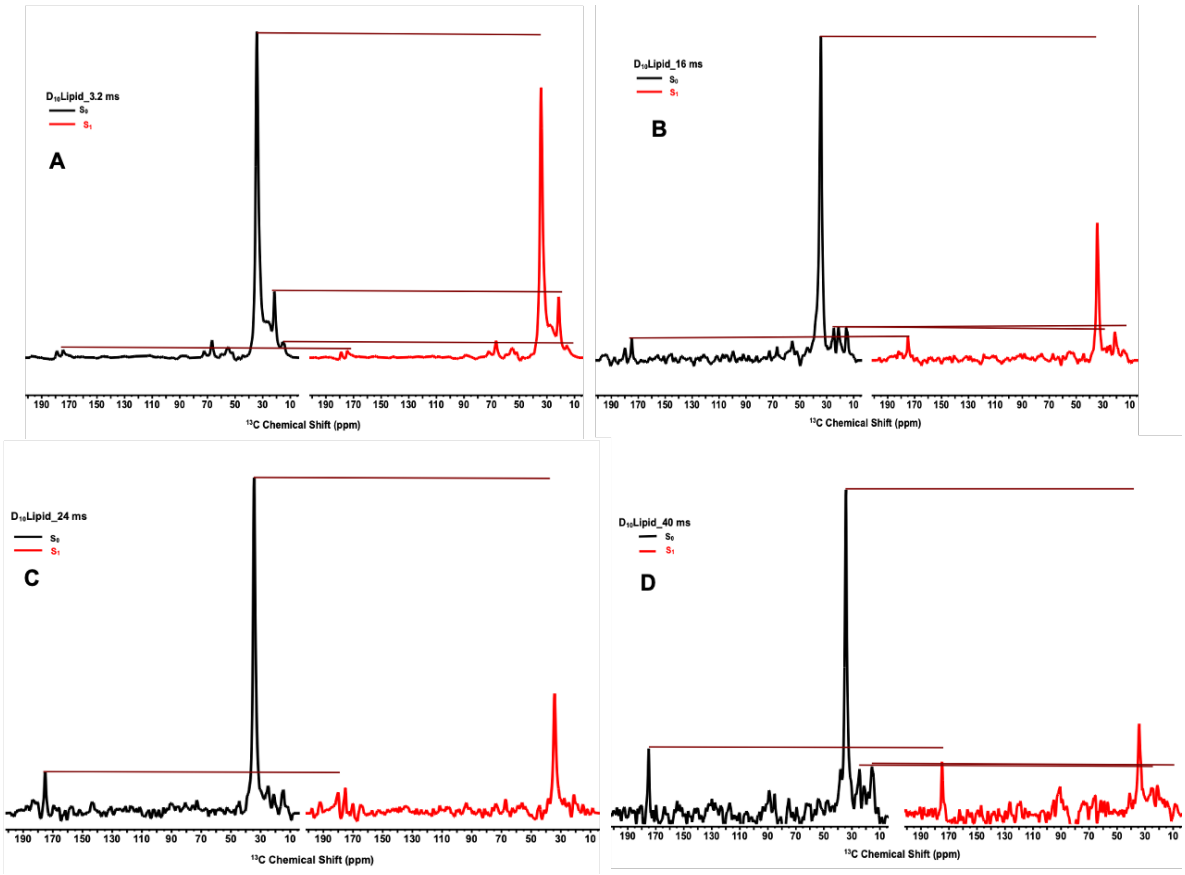


Figure 5.4: $^{13}C–^{2}H$ REDOR spectra of D_{10} Lipid with a dephasing time at (A) 3.2 ms (B) 16 ms (C) 24 ms and (D) 40 ms. The black lines in Fig. (A-D) denoted S_0 spectrum and red lines are denoted S_1 spectrum. The dark brown line over S_0 to S_1 at 172-178 ppm, 30-38 ppm, 22-28 ppm and 12-18, respectively demonstrates the difference between S_0 and S_1 by which dephasing was calculated as $\Delta S/S_0 = S_0\text{-}S_1/S_0$.

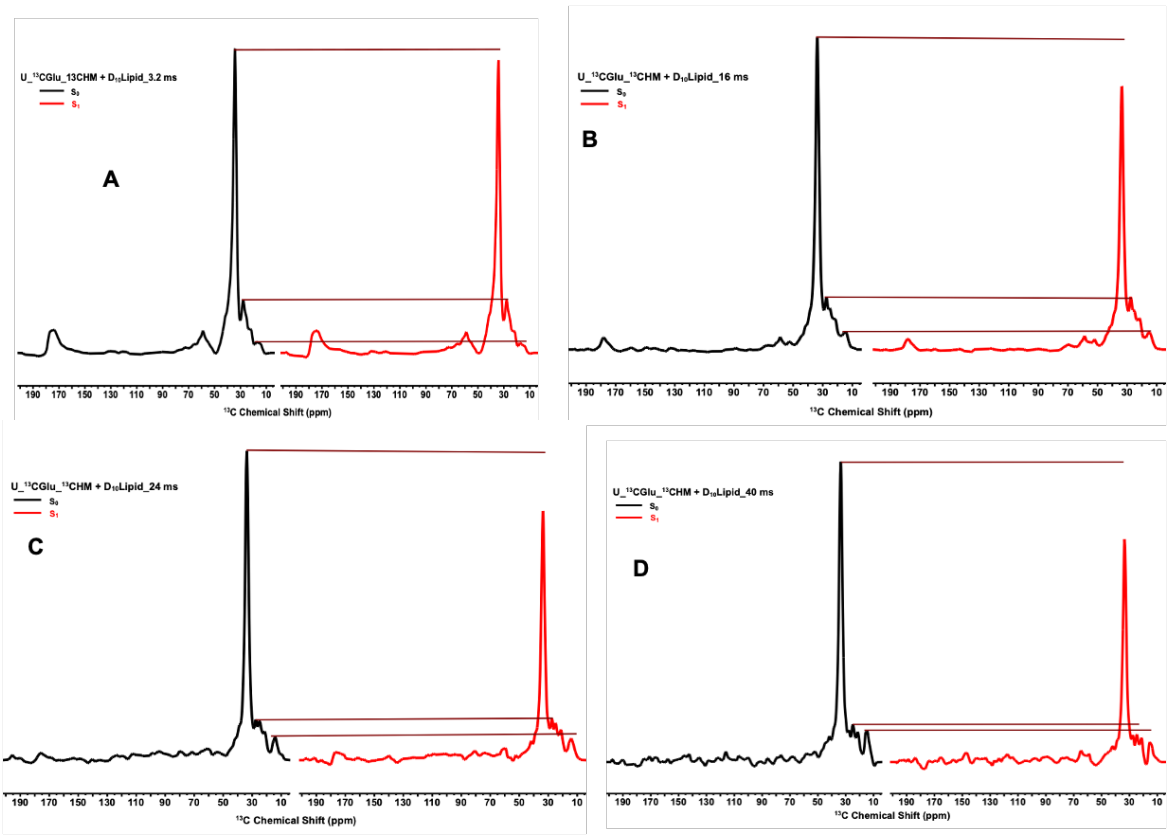


Figure 5.5: $^{13}C–^{2}H$ REDOR spectra of 25% U_^{13}CGlu_^{13}CHM + D_{10} Lipid (U_^{13}CGlu as carbon source for ^{13}CHM) with a dephasing time at (A) 3.2 ms (B) 16 ms (C) 24 ms and (D) 40 ms. The black lines in Fig. (A-D) denoted S_0 spectrum and red lines are denoted S_1 spectrum. The dark brown line over S_0 to S_1 at 30-38 ppm, 22-28 ppm and 12-18, respectively demonstrates the difference between S_0 and S_1 by which dephasing was calculated as $\Delta S/S_0 = S_0\text{-}S_1/S_0$.

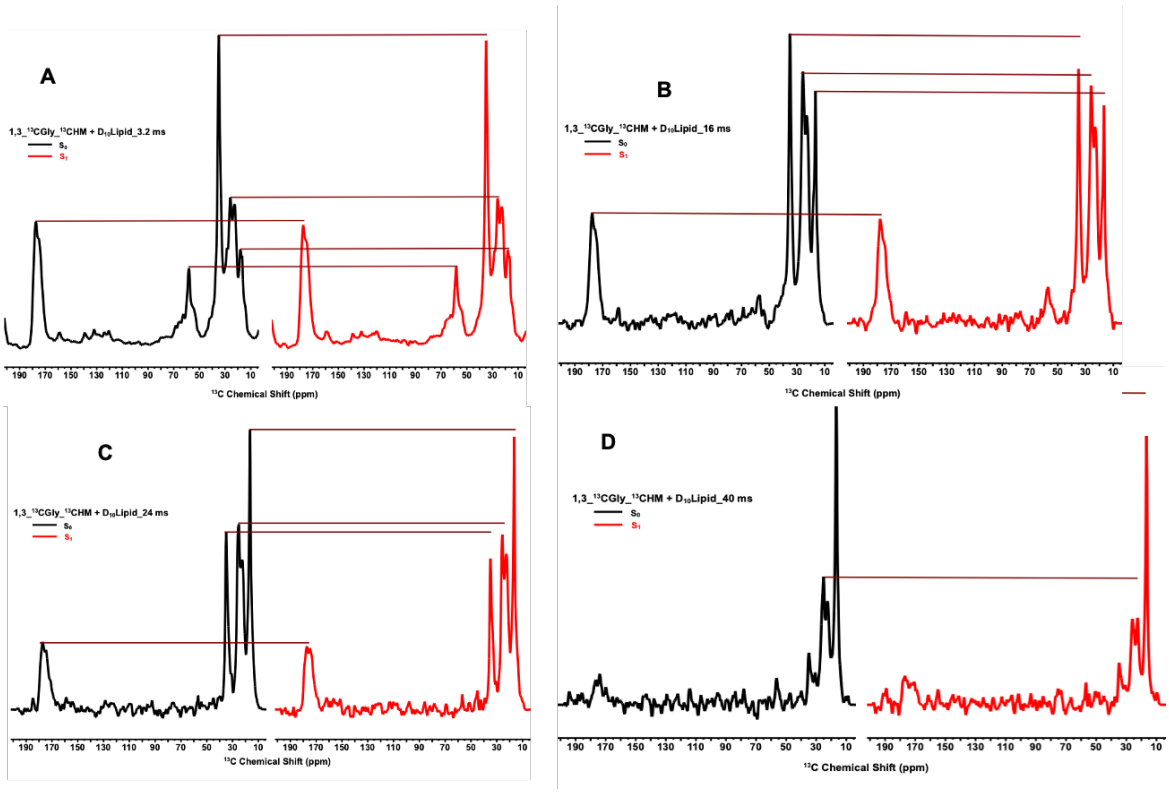


Figure 5.6: $^{13}C-^{2}H$ REDOR spectra of 1,3_^{13}CGly_^{13}CHM + D_{10} Lipid (1,3_^{13}CGly as carbon source for ^{13}CHM) with a dephasing time at (A) 3.2 ms (B) 16 ms (C) 24 ms and (D) 40 ms. The black lines in Fig. (A-D) denoted S_0 spectrum and red lines are denoted S_1 spectrum. The dark brown line over S_0 to S_1 at 172-180 ppm, 30-38 ppm, 22-28 ppm and 12-18, respectively demonstrates the difference between S_0 and S_1 by which dephasing was calculated as $\Delta S/S_0 = S_0\text{-}S_1/S_0$.

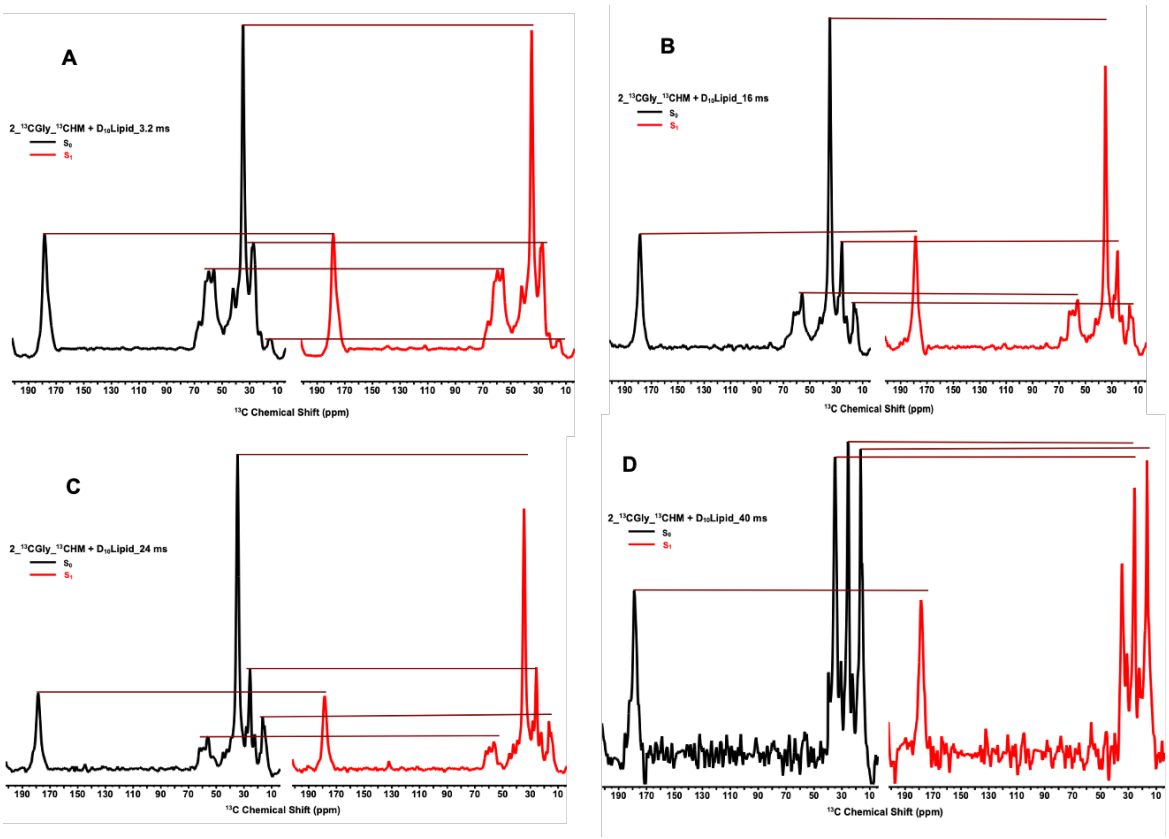


Figure 5.7: $^{13}C-^{2}H$ REDOR spectra of 2_^{13}CGly_^{13}CHM + D_{10} Lipid (2_^{13}CGly as carbon source for ^{13}CHM) with a dephasing time at (A) 3.2 ms (B) 16 ms (C) 24 ms and (D) 40 ms. The black lines in Fig. (A-D) denoted S_0 spectrum and red lines are denoted S_1 spectrum. The dark brown line over S_0 to S_1 at 172-180 ppm, 52-60 ppm, 30-38 ppm, 22-28 ppm and 12-18, respectively demonstrates the difference between S_0 and S_1 by which dephasing was calculated as $\Delta S/S_0 = S_0-S_1/S_0$.

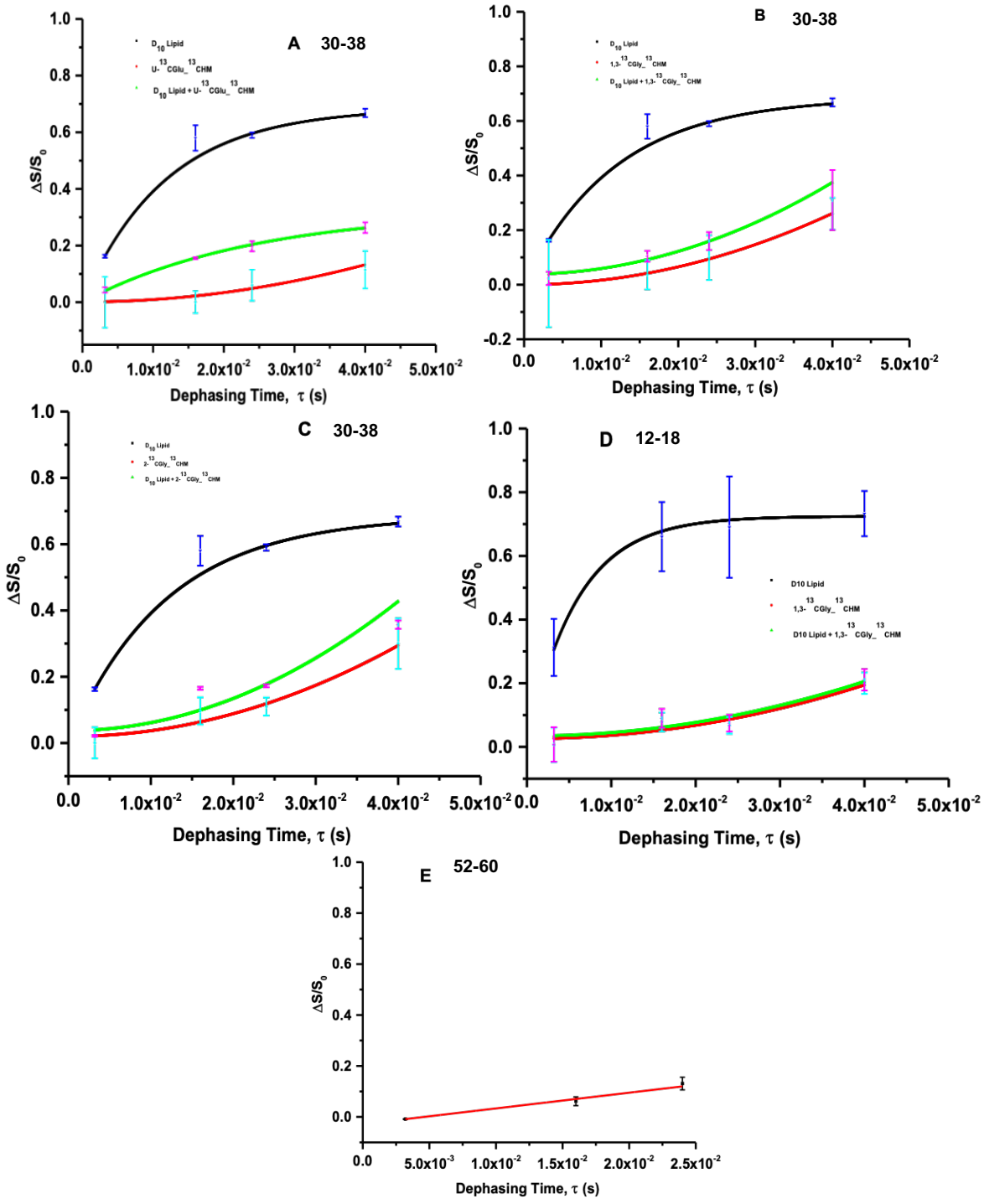


Figure 5.8: Fig. (A-C & D) shows REDOR buildup curve ($\Delta S/S_0$) vs. dephasing time (τ) at 30-38 ppm and 12-18 ppm spectral region, respectively. In (A-D), black fitted line is for D_{10}Lipid with blue error bars at 4 dephasing times, green fitted line is for ^{13}CHM + D_{10}Lipid

Figure 5.8 (cont'd)

with pink error bars at 4 dephasing times, and red fitted line is only for ^{13}CHM with turquoise blue error bars at 4 dephasing times. Dephasing at several τ's for the buildup curve only for ^{13}CHM calculated using $\Delta S^{Protein}(\tau) / S_0^{Protein}(\tau) = 1 - S_1^{Protein}(\tau)/S_0^{Protein}(\tau) = 1 - \gamma^{Protein}(\tau)$ [E5.2]. Fig. E shows REDOR buildup curve for 2-^{13}CHM_^{13}CHM + D_{10}Lipid at 52-62 ppm spectral region. No spectral peak has been observed at 52-60 ppm region for D_{10}Lipid. The $(\Delta S/S_0)^{exp}$ of D_{10}Lipid in (A-D) are fitted to a single exponential buildup $A\times(1-e^{-\gamma\tau})$ with A and γ are fitting parameters. The $(\Delta S/S_0)^{exp}$ of ^{13}CHM + D_{10}Lipid in (A) is also fitted to a single exponential buildup $A\times(1-e^{-\gamma\tau})$. The $(\Delta S/S_0)^{exp}$ of ^{13}CHM + D_{10}Lipid and $(\Delta S/S_0)^{Cal\ (calculated)}$ only ^{3}CHM in (A-D) are fitted to a parabolic buildup $M + \gamma^2 \times \tau^2$ with M and γ are fitting parameters. The $(\Delta S/S_0)^{exp}$ of ^{13}CHM + D_{10}Lipid in (E) are fitted to a linear buildup $M + \gamma \times \tau$ with M and γ are fitting parameters.

(A-C) in Figure 5.8 shows the fitted curve of experimental D_{10}Lipid, ^{13}CProtein (from three different ^{13}C sources) + D_{10}Lipid and calculated ^{13}CProtein (lipid contribution has been omitted after calculation using E5.2) at 30-38 ppm spectral region. $(\Delta S/S_0)^{exp}$ vs. τ of D_{10}Lipid in (Fig. 5.8A-D) are fitted to a single exponential buildup $A\times(1-e^{-\gamma\tau})$, where A and γ are fitting parameters. Due to T_1 relaxation, A is assigned as the approximate fraction of protein with $d = 3\gamma/2$ which is based on equal time spent in the three ^{2}H m ($m=\pm1,0$) states during τ. The fraction of protein with d=0 is (1-A)[17]. The best fit buildup of this curve (Fig. 5.8A-C) at 30-38 ppm spectral region gives buildup rate γ=84 s^{-1}, dephasing constant d=126 Hz and average internuclear distance r (^{13}C-^{2}H)= $(4642/d)^{1/3}$=3.32 Å and A=0.69. Moreover, The best fit buildup of $(\Delta S/S_0)^{exp}$ vs. τ (Fig 5.8D) for D_{10}Lipid at 12-18 ppm spectral region gives buildup rate γ= 171 s^{-1}, dephasing constant d= 256 Hz and average internuclear distance r (^{13}C-^{2}H)= $(4642/d)^{1/3}$= 2.62 Å and A=0.72. ^{2}H's in D_{10}DPPC is in the terminal -CH_3 of acyl chain (six ^{2}H replace six ^{1}H in two -CH_3 group) and -CH_2 right after the terminal -CH_3 of acyl chain (four ^{2}H replace four ^{1}H in two -CH_2 group) (Fig. 2.1). Though carbon in DPPC and DPPG in the lipid sample are not chemically ^{13}Clabeled there are 70% of lipid contribution to the signal by natural abundant ^{13}C (1.108% of ^{12}C) in the lipid could be close enough of Lipid ^{2}H for dephasing. Due to proximity of natural abundance ^{13}C and ^{2}H in lipid sample there is magnetic dipolar interaction between two

nuclei and subsequent net dephasing. -CH_2 peak of lipid at 30-38 ppm spectrum region show more dephasing than lipid + protein samples (Fig. 5.8A-C). When ^{13}CHM (U-^{13}CGlu/1,3-^{13}CGly/2-^{13}CGly as carbon source) was added with D_{10}lipid at 30-38 ppm spectral region S_1 signal are also reduced for all the Lipid + protein samples (Fig. 5.8A-C & 5.5-5.7). However, more interestingly S_1 signal reduced less in ^{13}CHM + D_{10}Lipid samples compare to D_{10}Lipid sample. This indicates ^{13}C protein has proximity to the ^{2}HLipid since it contributes to dephasing for Lipid+Protein sample. Consequently, ^{13}Cprotein itself causes dephasing due to magnetic dipolar coupling of proximal ^{13}CHM and ^{2}HLipid. However, dephasing only by ^{13}CHM in ^{13}CHM + D_{10}Lipid has been calculated using E5.2 at particular dephasing time, τ and best fit curve has shown in Fig. 5.8A-D. Error of dephasing of D_{10}Lipid and ^{13}CProtein + D_{10}Lipid samples at 4 dephasing times (Fig. 5.8A-D) were calculated using E5.1. Errors of dephasing at 4 dephasing times for ^{13}CProtein only was calculated using E5.3. The $(\Delta S/S_0)^{exp}$ vs. τ of U-^{13}CGlu_^{13}CHM + D_{10}Lipid in (5.8A) at 30-38 ppm spectral peak region is also fitted to a single exponential buildup $A\times(1-e^{-\gamma\tau})$ with A=0.33, γ=41s-^{1}, d=61 Hz and average internuclear distance r (^{13}C-^{2}H)= $(4642/d)^{1/3}$=4.23 Å . The $(\Delta S/S_0)^{exp}$ vs. τ of ^{13}CHM + D_{10}Lipid and $(\Delta S/S_0)^{Cal}$ vs. τ only ^{3}CHM in (Fig. 5.8A-D) are fitted to a parabolic buildup M + $\gamma^2 \times \tau^2$ with M (y- intercept) and γ are fitting parameters. The universal buildup curve for REDOR dephasing is sigmoidal in nature; however, the buildup curve can be approximated by a parabolic function. When the dipolar interaction is weak or the evolution time is short, the REDOR signal intensity S(x) can be approximated by a parabolic (quadratic) function: $S(x)\approx 1-x^2$ where: $x=N\delta/2\pi$, N is the number of rotor cycles, δ is the dipolar coupling constant. This parabolic approximation simplifies the analysis and fitting of REDOR data in the regime where x is small. The $(\Delta S/S_0)^{exp}$ vs. τ of ^{13}CHM + D_{10}Lipid in Fig.5.8E are fitted to a linear buildup M (y- intercept) + $\gamma \times \tau$ with M (y- intercept) and γ are fitting parameters. The fitted curve of 1,3-^{13}CGly_^{13}CHM + D_{10}Lipid in 5.8B at 30-38 ppm spectral peak region gives γ=15 s-^{1}, d=22 Hz and average internuclear distance r (^{13}C-^{2}H)= $(4642/d)^{1/3}$=5.97 Å and 2-^{13}CGly_^{13}CHM + D_{10}Lipid in Fig. 5.8C at 30-38 ppm spectral peak region gives γ=16 s-^{1}, d=23 Hz and average internuclear distance r (^{13}C-^{2}H)= $(4642/d)^{1/3}$=5.83 Å. Moreover, fitting of 1,3-^{13}CGly_^{13}CHM + D_{10}Lipid buildup curve in Fig. 5.8D at 12-18 ppm spectral peak

region gives γ=12 s^{-1}, d=18 Hz and average internuclear distance r (^{13}C-^{2}H)= (4642/d)$^{1/3}$=6.41 Å. All the r (^{13}C-^{2}H) values with uncertainties from lipid, lipid+ ^{13}Cprotein and ^{13}Cprotein are shown in table 5.1. Uncertain of r's in table 5.1 are calculated using supplementary E5.4.

The $(\Delta S/S_0)^{Cal}$ of ^{13}CHM at 30-38 ppm and 12-18 ppm, respectively for all samples from three different carbon sources fitted parabolically whereas $(\Delta S/S_0)^{Exp}$ of D_{10}lipid at 30-38 ppm and 12-18 ppm, respectively fitted exponentially. $(\Delta S/S_0)^{exp}$ of D_{10}Lipid at 30-38 ppm spectrum region built up to ≈0.70 (Fig. 5.8A-C) at 40 ms with average internuclear distance r (^{13}C-^{2}H)= 3.32 Å and at 12-18 ppm spectrum region built up to ≈0.75 (Fig. 5.8D) at 40 ms with average internuclear distance r (^{13}C-^{2}H)= 2.62 Å. On the other hand, $(\Delta S/S_0)^{Cal}$ at 30-38 ppm spectrum region for U-^{13}CGlu_^{13}CHM built up to ≈0.12 (Fig. 5.8A) at 40 ms with average internuclear distance r (^{13}C-^{2}H)= 6.99 Å, 1,3-^{13}CGly_^{13}CHM built up to ≈0.20 (Fig. 5.8B) at 40 ms with average internuclear distance r (^{13}C-^{2}H)= 5.97 Å and 2-^{13}CGly_^{13}CHM built up to ≈0.21 (Fig. 5.8C) at 40 ms with average internuclear distance r (^{13}C-^{2}H)= 6.18 Å. Moreover, $(\Delta S/S_0)^{Cal}$ at 12-18 ppm spectrum region for 1,3-^{13}CGly_^{13}CHM built up to ≈0.17 (Fig. 5.8D) at 40 ms with average internuclear distance r (^{13}C-^{2}H)= 6.41 Å. Further, $(\Delta S/S_0)^{Exp}$ at 52-62 ppm spectrum region for 2-^{13}CGly_^{13}CHM + D_{10}Lipid linearly built up to ≈0.15 (Fig. 5.8E) at 24 ms with average internuclear distance r (^{13}C-^{2}H)= 7.94 Å. However, D_{10}Lipid does not show any peak at 52-62 ppm spectrum region (Fig. 5.4); therefore, peak contribution for this region only from 2-^{13}CGly_^{13}CHM. Though all the D_{10}Lipid and ^{13}CProtein + D_{10}Lipid show dephasing at all dephasing times (Figure 5.8A-D & 5.4-5.7) at 22-28 ppm spectral region, $(\Delta S/S_0)^{Cal}$ values for only ^{13}CHM from all carbon sources found very negligible after calculation through E5.2 (Fig. 5.9).

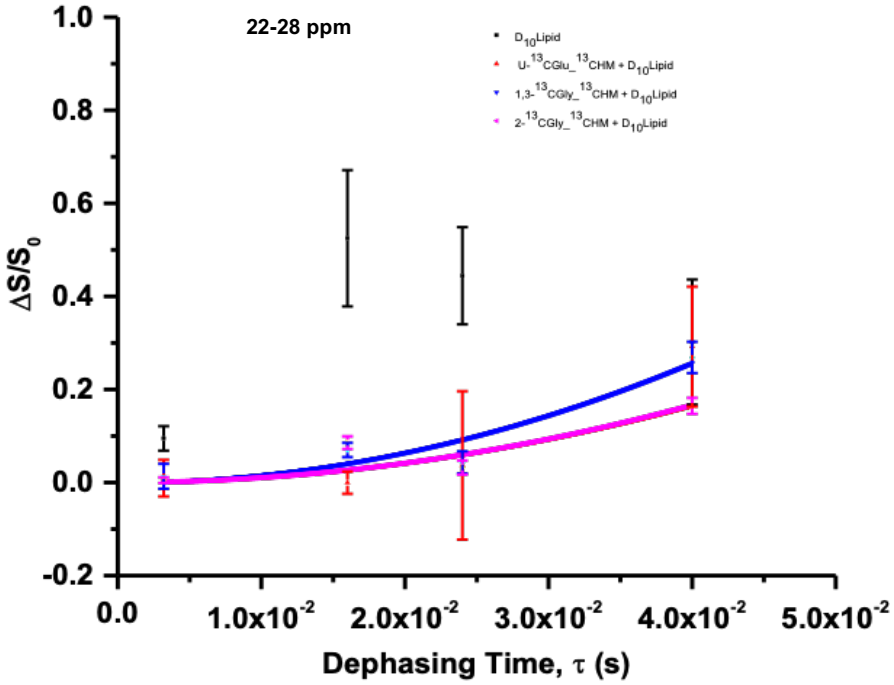


Figure 5.9: $^{13}C-^{2}H$ REDOR buildup curve $(\Delta S/S_0)^{exp}$ vs. dephasing time (τ) at 22-28 ppm spectral region. Black scatter dots with error bars represent buildup of D_{10}Lipid at 4 dephasing times. These dots are not fitted well exponentially/parabolically/linearly though $(\Delta S/S_0)^{exp}$ are ≈.09 at 3.2 ms, ≈.57 at 16 ms, ≈.47 at 24 ms and ≈.3 at 40 ms. $(\Delta S/S_0)^{exp}$ should larger at higher dephasing times; however at 16ms has highest $(\Delta S/S_0)^{exp}$ at 30-38 ppm spectral region and $(\Delta S/S_0)^{exp}$ decrease to 40 ms. This is unusual trend of decreasing $(\Delta S/S_0)^{exp}$ and however, we don't know the explanation for that. Pink fitted line is for 2-^{13}CGly-^{13}CHM + D_{10}Lipid with pink error bars, blue fitted line is for 1,3-^{13}CGly-^{13}CHM + D_{10}Lipid with blue error bars and red fitted line is for U-^{13}CGlu-^{13}CHM + D_{10}Lipid with red error bars at 22-28 ppm spectral region. The $(\Delta S/S_0)^{exp}$ vs. τ of ^{13}CHM (from three different ^{13}C sources) + D_{10}Lipid are fitted to a parabolic buildup $M + \gamma^2 \times \tau^2$ with M and γ are fitting parameters.

Table 5.1: Internuclear distances, r (^{13}C-^{2}H) in different lipid regions. The fitting method for each of the samples in different spectra region (ppm) is denoted as "a" for parabolic buildup, "b" for linear buildup and "c" for exponential buildup.

Samples Name	γ (s^{-1}) 180-188 (ppm)	d= 3γ/2 (Hz) 180-188 (ppm)	r(Å) 180-188 (ppm)	γ (s^{-1}) 172-178 (ppm)	d= 3γ/2 (Hz) 172-178 (ppm)	r(Å) 172-178 (ppm)	γ (s^{-1}) 52-60 (ppm)	d= 3γ/2 (Hz) 52-60 (ppm)	r(Å) 52-60 (ppm)	γ (s^{-1}) 30-38 (ppm)	d= 3γ/2 (Hz) 30-38 (ppm)	r(Å) 30-38 (ppm)	γ (s^{-1}) 22-28 (ppm)	d= 3γ/2 (Hz) 22-28 (ppm)	R (Å) 22-28 (ppm)	γ (s^{-1}) 12-18 (ppm)	γ (s^{-1}) 12-18 (ppm)	r(Å) 12-18 (ppm)
D_{10}Lipid										84.47±5.40	126.7±8.1	3.32±.07 (c)				171±13	257±19.5	2.63±.06 (c)
U-^{13}CGlu_^{13}CHM +D_{10}Lipid										41±3	62±4.5	4.23±.09 (c)						
U-^{13}CGlu_^{13}CHM (Calcu.)										9±1	13.5±1.5	6.99±.35 (a)						
1,3-^{13}CGly_^{13}CHM +D_{10}Lipid										15±1.5	22.5±2.25	6.22±.20 (a)				12±1.3	18±2	6.36±.15 (a)
1,3-^{13}CGly_^{13}CHM (Calcu.)										13±.1	19.5±.15	6.19±.02 (a)				10±1.2	15±1.8	6.76±.27 (a)
2-^{13}CGly_^{13}CHM +D_{10}Lipid							6±.62	9±0.93	8.01±.26 (b)	16±2.5	24±3.8	5.78±.30 (a)						
2-^{13}CGly_^{13}CHM (Calcu.)										13±1.4	19.5±2	6.18±.22 (a)	10±2	15±3	6.85±.43			
D_8Lipid				27±22	40.5±33	4.86±1.35 (c)				115±8	172.5±12	3.0±.07 (c)	15±2.3	22.5±3.5	5.90±.30 (b)	17±2.3	25.5±3.5	5.64±.25 (b)
U-^{13}CGlu_^{13}CHM +D_8Lipid										81±17	121.5±25.5	3.37±.26 (c)	21±29	32±43.5	5.24±2.35 (c)	8.4±1.1	12.6±1.7	7.16±.33 (a)
U-^{13}CGlu_^{13}CHM (Calcu.)										66±61	99±91.5	3.60±1.09 (c)	5±18	7.5±27	8.52±10 (c)	11±2	16.5±3	6.55±.41 (a)
1,3-^{13}CGly_^{13}CHM +D_8Lipid																		
1,3-^{13}CGly_^{13}CHM (Calcu.)																		
2-^{13}CGly_^{13}CHM +D_8Lipid				17±1.5	25.5±2.3	5.66±.17 (a)												
2-^{13}CGly_^{13}CHM (Calcu.)				15±1.1	22.5±1.7	5.90±.15 (a)												
D_6Chol										46±1.5	69±2.3	4.1±0.04 (b)	7.18±.38	10.8±.57	7.55±0.13 (c)	27±.65	40.5±.98	4.85±0.04 (a)
U-^{13}CGlu_^{13}CHM +D_6Chol	19±.68	28.5±1.02	5.46±.06 (a)				145±13	217.5±19.5	2.77±.08 (c)	396±19	594±28.5	1.98±0.03 (c)	16±1.1	24±1.6	5.78±0.1 (a)	6.5±9.6	9.75±14.4	7.80±3.85
U-^{13}CGlu_^{13}CHM (cal)										11±16	16.5±24	6.55±3.35 (c)	59±48	88.5±72	3.75±1.02 (c)	27±2	40.5±3	4.85±.12 (a)

5.3.7: ^{13}CHM - to- D_8 ^{2}HPC REDOR NMR

REDOR spectra are shown for D_8Lipid in Fig. 5.10, Uniform (U)_^{13}CGlucose_^{13}CHM + D_{10}Lipid in Fig. 5.11, 1,3_^{13}CGlycerol_^{13}CHM + D_{10}Lipid in Fig. 5.12 and 2_^{13}CGlycerol_^{13}CHM + D_{10}Lipid in Fig. 5.13 at four dephasing times (3.2, 16, 24 and 40 ms). Peak at 30-38 ppm spectral region for all spectra is assigned for lipid -CH_2 and also for protein $^{\beta}$C-Val/ $^{\beta}$C-Ile/ $^{\beta}$C-Phe/ $^{\gamma}$C-Met/ $^{\delta}$C-Lys. Peak at 22-28 ppm spectral region assigned for lipid -CH_2 (-CH_3) and also for protein $^{\gamma}$C-Ile/$^{\gamma}$C-Ile/$^{\gamma}$C-Lys/ $^{\gamma}$CThr/ $^{\gamma}$C-Arg. Peak at 12-18 ppm spectral region is the peak of lipid -CH_3 and also for protein $^{\beta}$C-Ala, $^{\gamma}$C-Val, $^{\delta}$C-Ile, $^{\epsilon}$C-Met. Spectral region at 172-180 ppm refers for both lipid and protein >C=O peak. U-^{13}CGlu_^{13}CHM + D_8Lipid and 1,3-^{13}CGly_^{13}CHM + D_8Lipid sample does not show >C=O peak at 172-180 ppm region at 16, 24 and 40 ms, respectively because there could be faster T2 relaxation of the sample.

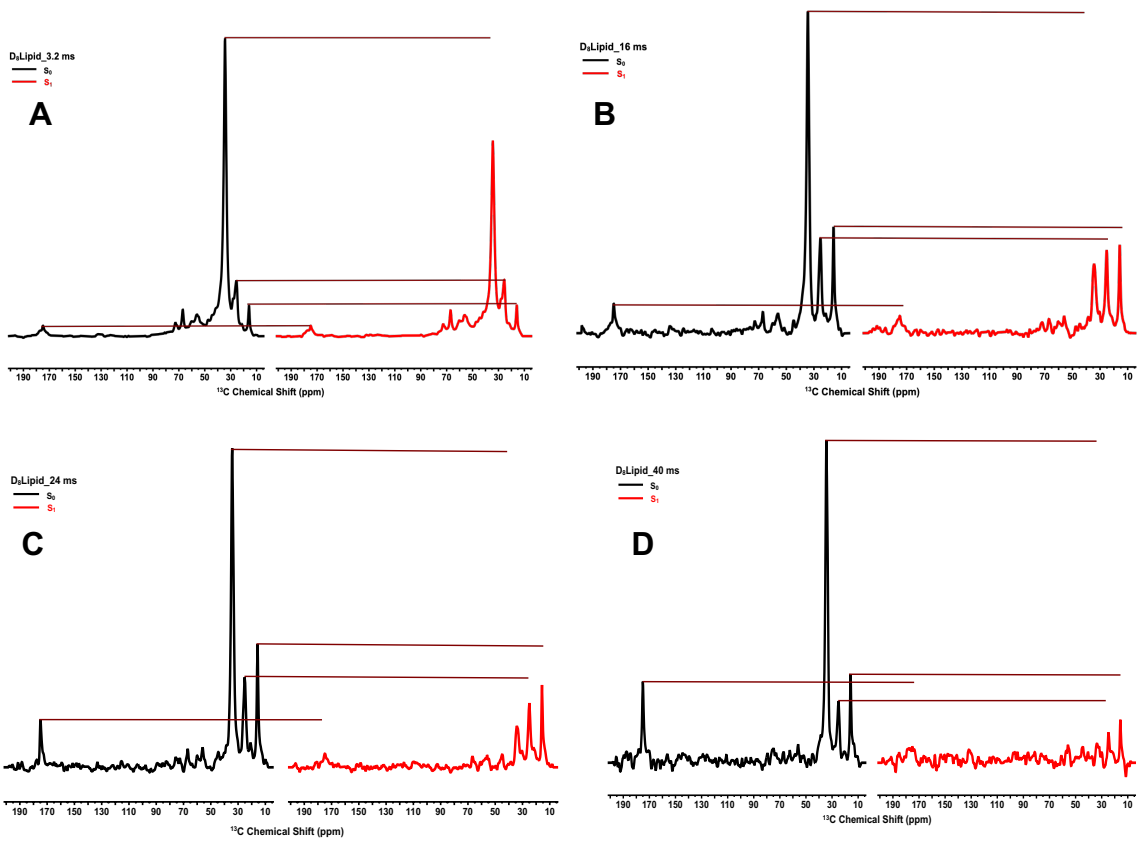


Figure 5.10: $^{13}C-^{2}H$ REDOR spectra of D_8 Lipid time at (A) 3.2 ms (B) 16 ms (C) 24 ms and (D) 40 ms. The black lines in Fig. (A-D) denoted S_0 spectrum and red lines are denoted S_1 spectrum. The dark brown line over S_0 to S_1 at 172-180 ppm, 30-38 ppm, 22-28 ppm and 12-18, respectively demonstrates the difference between S_0 and S_1 by which dephasing was calculated as $\Delta S/S_0 = S_0-S_1/S_0$.

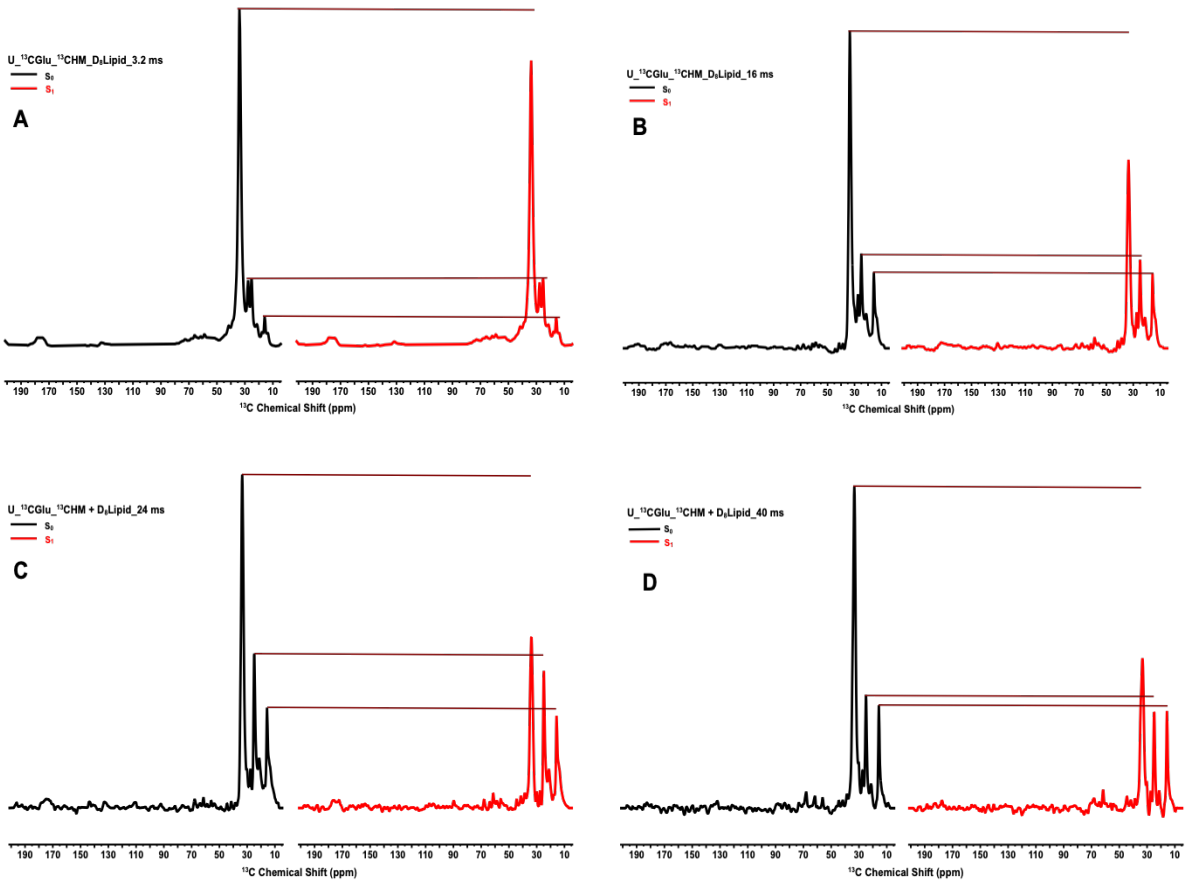


Figure 5.11: $^{13}C-^{2}H$ REDOR spectra of U_^{13}CGlu_^{13}CHM + D_8 Lipid (U_^{13}CGlu as carbon source for ^{13}CHM) with a dephasing time at (A) 3.2 ms (B) 16 ms (C) 24 ms and (D) 40 ms. The black lines in Fig. (A-D) denoted S_0 spectrum and red lines are denoted S_1 spectrum. The dark brown line over S_0 to S_1 at 30-38 ppm, 22-28 ppm and 12-18, respectively demonstrates the difference between S_0 and S_1 by which dephasing was calculated as $\Delta S/S_0 = S_0\text{-}S_1/S_0$.

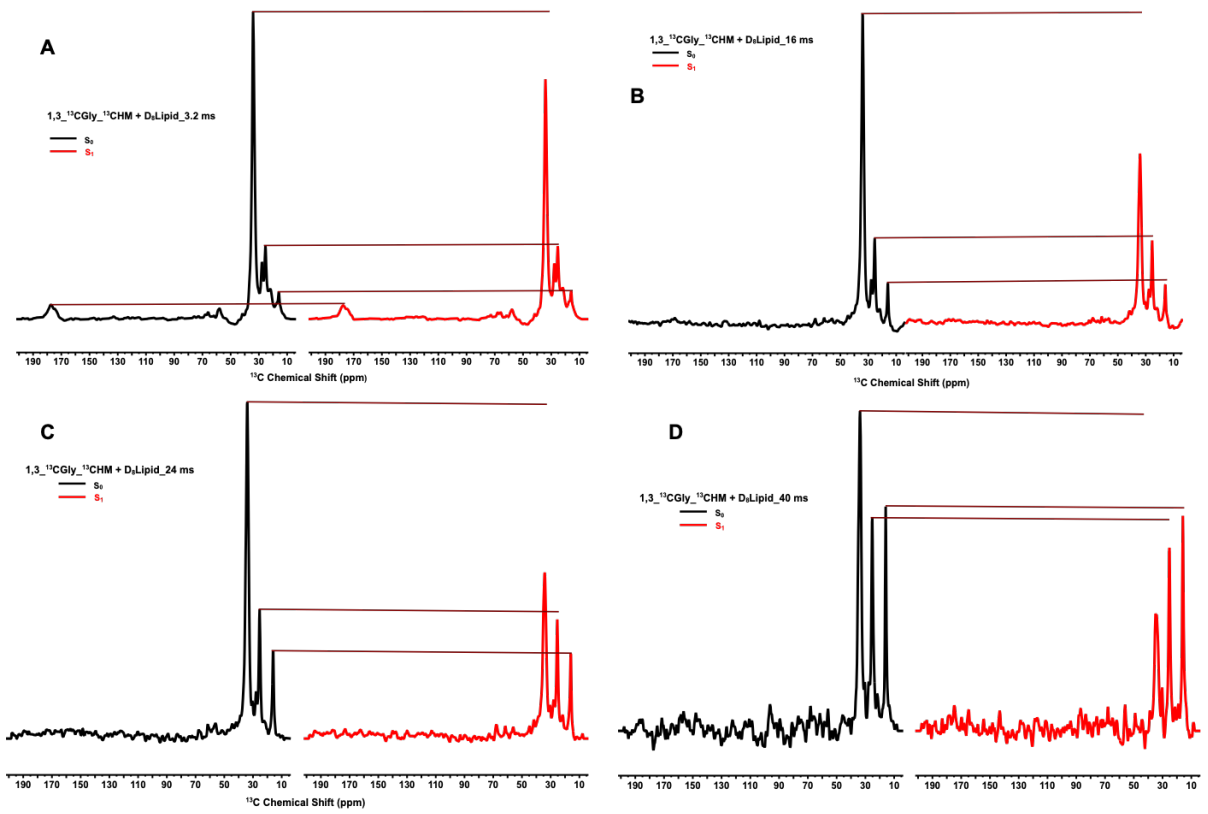


Figure 5.12: $^{13}C-^{2}H$ REDOR spectra of 1,3_^{13}CGly_^{13}CHM + D_8 Lipid (1,3_^{13}CGly as carbon source for ^{13}CHM) with a dephasing time at (A) 3.2 ms (B) 16 ms (C) 24 ms and (D) 40 ms. The black lines in Fig. (A-D) denoted S_0 spectrum and red lines are denoted S_1 spectrum. The dark brown line over S_0 to S_1 at 172-180 ppm, 30-38 ppm, 22-28 ppm and 12-18, respectively demonstrates the difference between S_0 and S_1 by which dephasing was calculated as $\Delta S/S_0 = S_0-S_1/S_0$.

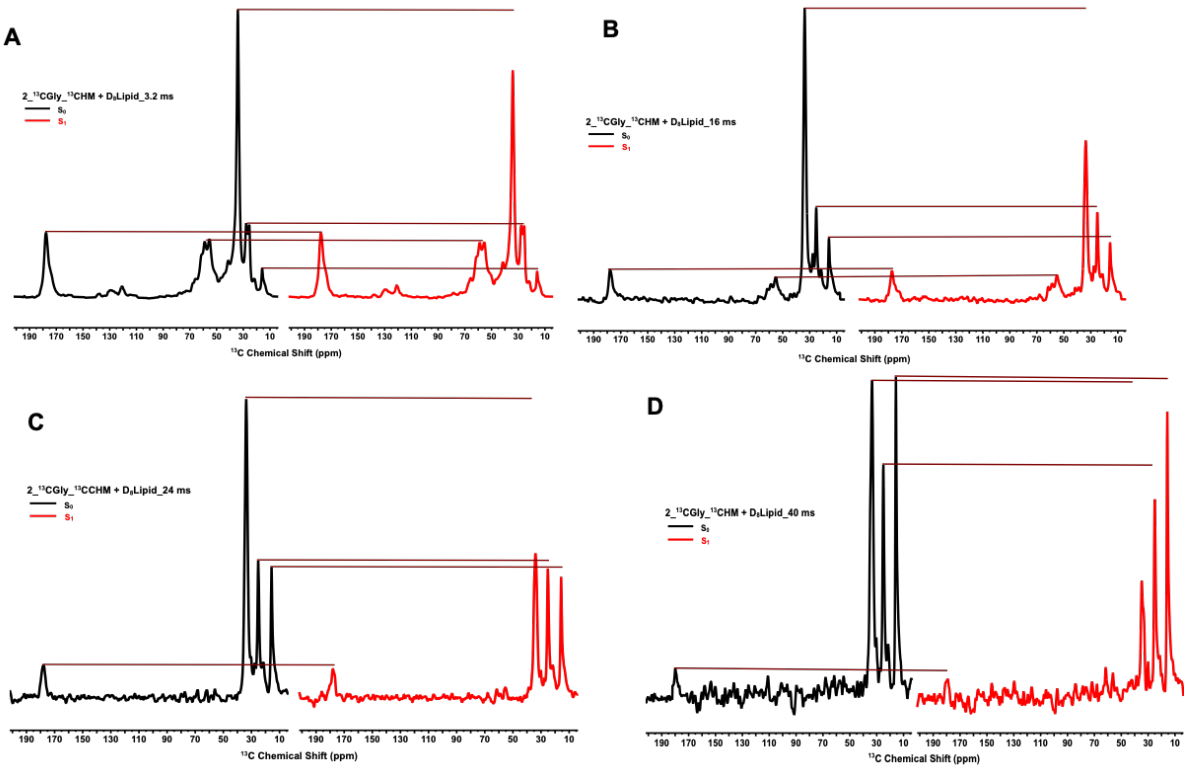


Figure 5.13: $^{13}C-^{2}H$ REDOR spectra of 2_^{13}CGly_^{13}CHM + D_8 Lipid (2_^{13}CGly as carbon source for ^{13}CHM) with a dephasing time at (A) 3.2 ms (B) 16 ms (C) 24 ms and (D) 40 ms. The black lines in Fig. (A-D) denoted S_0 spectrum and red lines are denoted S_1 spectrum. The dark brown line over S_0 to S_1 at 172-180 ppm, 52-60 ppm, 30-38 ppm, 22-28 ppm and 12-18, respectively demonstrates the difference between S_0 and S_1 by which dephasing was calculated as $\Delta S/S_0 = S_0-S_1/S_0$.

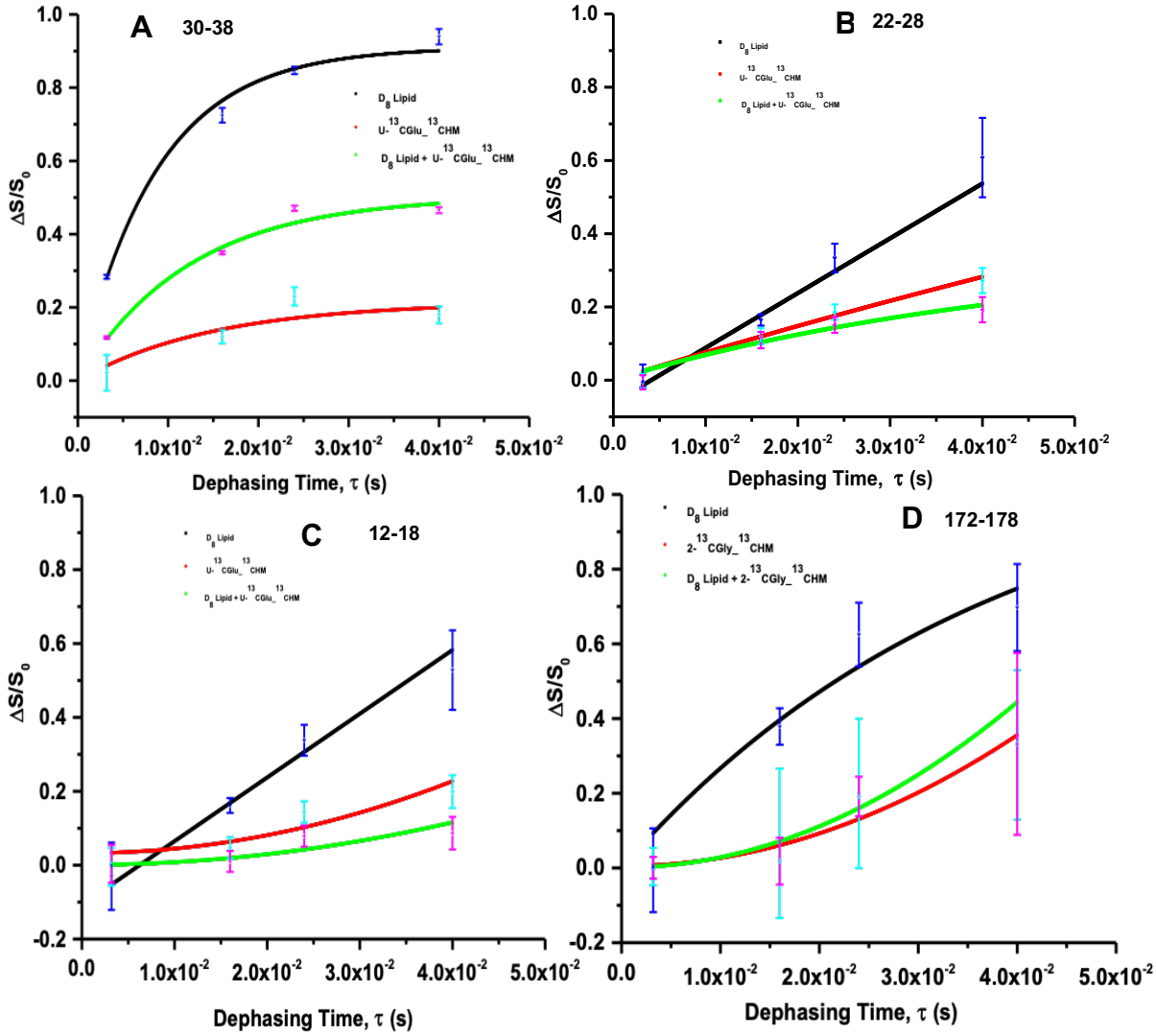


Figure 5.14: REDOR buildup curve of ($\Delta S/S_0$) vs. dephasing time (τ) at spectral region in (A) 30-38 ppm (B) 22-28 ppm (C) 12-18 ppm (D) 172-178 ppm, respectively. In (A-D) black fitted line is for D_{10}Lipid with blue error bars, green fitted line is for ^{13}CHM + D_{10}Lipid with pink error bars, and red fitted line is only for ^{13}CHM with turquoise blue error bars at 4 dephasing times (3.2, 16, 24 & 40 ms). Dephasing at several τ's for the buildup curve only for ^{13}CHM calculated using $\Delta S^{Protein}(\tau) / S_0^{Protein}(\tau) = 1 - S_1^{Protein}(\tau)/S_0^{Protein}(\tau) = 1 - \gamma^{Protein}(\tau)$ [E5.2]. The $(\Delta S/S_0)^{exp}$ of D_{10}Lipid in (A) at 30-38 ppm and in (D) at 172-178 ppm spectral region are fitted to a single exponential buildup $A\times(1-e^{-\gamma\tau})$ with A and γ are fitting parameters and (B) & C are fitted to a linear buildup $M + \gamma \times \tau$ with M and γ are fitting

Figure 5.14 (cont'd)

parameters. $(\Delta S/S_0)^{exp}$ of ^{13}CHM + D_{10}Lipid and $(\Delta S/S_0)^{Cal}$ of ^{13}CHM in (A) at 30-38 ppm, $(\Delta S/S_0)^{exp}$ of ^{13}CHM + D_{10}Lipid and $(\Delta S/S_0)^{Cal}$ of ^{13}CHM in (B) at 22-28 ppm spectral region are also fitted to a single exponential buildup $A\times(1-e^{-\gamma\tau})$. $(\Delta S/S_0)^{exp}$ of ^{13}CHM + D_{10}Lipid and $(\Delta S/S_0)^{Cal}$ of ^{3}CHM in (C) and $(\Delta S/S_0)^{exp}$ of ^{13}CHM + D_{10}Lipid and $(\Delta S/S_0)^{Cal}$ of ^{3}CHM in (D) are fitted to a parabolic buildup $M + \gamma^2 \times \tau^2$ with M and γ are fitting parameters.

Figure 5.14 shows the $(\Delta S/S_0)$ vs. dephasing time (τ) fitted curve of experimental D_8Lipid, experimental ^{13}CProtein + D_8Lipid and calculated ^{13}CProtein (lipid contribution has been omitted after calculation using E5.2) at (A) 30-38 ppm (B) 22-28 ppm (C) 12-18 ppm and (D) 172-178 ppm spectral region, respectively. At 30-38 ppm spectral region in Fig. 5.14A, $(\Delta S/S_0)^{exp}$ of D_8Lipid, $(\Delta S/S_0)^{exp}$ of U_^{13}CGlu_^{13}CHM + D_8Lipid and $(\Delta S/S_0)^{Cal}$ of U_^{13}CGlu_^{13}CHM are fitted to a single exponential buildup $A\times(1-e^{-\gamma\tau})$, respectively, where A and γ are fitting parameters with d= $3\gamma/2$=172 Hz and r (^{13}C-^{2}H)= $(4642/d)^{1/3}$=3.0± 0.07 Å for D_8Lipid, d= $3\gamma/2$=81 Hz and r (^{13}C-^{2}H)= $(4642/d)^{1/3}$=3.37± 0.26 Å for U_^{13}CGlu_^{13}CHM and d= $3\gamma/2$=66 Hz and r (^{13}C-^{2}H)= $(4642/d)^{1/3}$=3.60± 1.09 Å for U_^{13}CGlu_^{13}CHM. The r (^{13}C-^{2}H) value of U_^{13}CGlu_^{13}CHM within the range (two r's are not significantly different) of D_8Lipid's r (^{13}C-^{2}H) which indicates either ^{13}C in HM or D_{10}Lipid could contribute to magnetic dipolar interaction with $^{2}HD_8$Lipid to show dephasing (Table 5.10).

At 22-28 ppm spectral region in Fig. 5.14B, $(\Delta S/S_0)^{exp}$ of D_8Lipid is fitted to a linear buildup $M + \gamma \times \tau$ where M and γ are fitting parameters with d= $3\gamma/2$=22 Hz and r (^{13}C-^{2}H)= $(4642/d)^{1/3}$=5.91± 0.30 Å, $(\Delta S/S_0)^{exp}$ of U_^{13}CGlu_^{13}CHM + D_8Lipid and $(\Delta S/S_0)^{Cal}$ of U_^{13}CGlu_^{13}CHM are fitted to a single exponential buildup $A\times(1-e^{-\gamma\tau})$, where A and γ are fitting parameters with d= $3\gamma/2$=32 Hz and r (^{13}C-^{2}H)= $(4642/d)^{1/3}$=5.24± 2.35 Å for U_^{13}CGlu_^{13}CHM + D_8Lipid and d= $3\gamma/2$=7 Hz and r (^{13}C-^{2}H)= $(4642/d)^{1/3}$=8.62± 10 Å for U_^{13}CGlu_^{13}CHM. The r (^{13}C-^{2}H) value of U_^{13}CGlu_^{13}CHM has a large error value which within the range (two r's are not significantly different) of D_8Lipid's r (^{13}C-^{2}H) that indicates either ^{13}C in HM or D_{10}Lipid could contribute to magnetic dipolar interaction with $^{2}HD_8$Lipid to show dephasing (Table 5.10).

At 12-18 ppm spectral region in Fig. 5.14C, $(\Delta S/S_0)^{exp}$ of D_8Lipid is fitted to a linear buildup $M + \gamma \times \tau$ where M and γ are fitting parameters with d= $3\gamma/2$=26 Hz and r (^{13}C-^{2}H)=

$(4642/d)^{1/3}$=5.64± 0.25 Å, $(\Delta S/S_0)^{exp}$ of U_^{13}CGlu_^{13}CHM + D_8Lipid and $(\Delta S/S_0)^{Cal}$ of U_^{13}CGlu_^{13}CHM are fitted to a single parabolic buildup M + $\gamma^2 \times \tau^2$, where M and γ are fitting parameters with d= 3γ/2=13 Hz and r (^{13}C-^{2}H)= $(4642/d)^{1/3}$=7.14± .33 Å for U_^{13}CGlu_^{13}CHM + D_8Lipid and d= 3γ/2=17 Hz and r (^{13}C-^{2}H)= $(4642/d)^{1/3}$=6.55± .41 Å for U_^{13}CGlu_^{13}CHM. The error value of r (^{13}C-^{2}H) of U_^{13}CGlu_^{13}CHM is not within the close range (two r's are significantly different) of D_8Lipid's r (^{13}C-^{2}H) that indicates only protein contributed to magnetic dipolar interaction with $^{2}HD_8$Lipid to show dephasing (Table 5.10). Therefore, $^{\beta}$C-Ala, $^{\gamma}$C-Val, $^{\delta}$C-Ile, $^{\epsilon}$C-Met in ^{13}CHM are close enough to $^{2}HD_8$Lipid to show dephasing.

At 172-178 ppm spectral region in Fig. 5.14D, $(\Delta S/S_0)^{exp}$ of D_8Lipid is fitted to a single exponential buildup A×(1-$e^{-\gamma\tau}$) where A and γ are fitting parameters with d= 3γ/2=40 Hz and r (^{13}C-^{2}H)= $(4642/d)^{1/3}$=4.87± 0.16, $(\Delta S/S_0)^{exp}$ of 2_^{13}CGly_^{13}CHM + D_8Lipid and $(\Delta S/S_0)^{Cal}$ of 2_^{13}CGly_^{13}CHM are fitted to a single parabolic buildup M + $\gamma^2 \times \tau^2$, where M and γ are fitted parameters with d= 3γ/2=25 Hz and r (^{13}C-^{2}H)= $(4642/d)^{1/3}$=5.70 ± 0.15 Å for 2_^{13}CGly_^{13}CHM + D_8Lipid and d= 3γ/2=22 Hz and r (^{13}C-^{2}H)= $(4642/d)^{1/3}$=5.93 ± 0.12 Å for 2_^{13}CGly_^{13}CHM. The error value of r (^{13}C-^{2}H) of U_^{13}CGlu_^{13}CHM is not within the close range (two r's are significantly different) of D_8Lipid's r (^{13}C-^{2}H) that indicates only protein contributed to magnetic dipolar interaction with $^{2}HD_8$Lipid to show dephasing (Table 5.10). Therefore, >C=O's in ^{13}CHM are close enough to $^{2}HD_8$Lipid to show dephasing.

5.3.8: ^{13}CHM - to- $D_6{}^2$HChol REDOR NMR

REDOR spectra are shown for D_6Chol in Fig. 5.15, Uniform (U)_^{13}CGlucose_^{13}CHM + D_6Chol in Fig. 5.16 at four dephasing times (3.2, 16, 24 and 40 ms). Peak at 30-38 ppm spectral region for all spectra is assigned for lipid -CH_2 and also for protein $^{\beta}$C-Val/ $^{\beta}$C-Ile/ $^{\beta}$C-Phe/ $^{\gamma}$C-Met/ $^{\delta}$C-Lys. Peak at 22-28 ppm spectral region assigned for lipid -CH_2 (-CH_3) and also for protein $^{\gamma}$C-Ile/$^{\gamma}$C-Ile/$^{\gamma}$C-Lys/ $^{\gamma}$CThr/ $^{\gamma}$C-Arg. Peak at 12-18 ppm spectral region is the peak of lipid -CH_3 and also for protein $^{\beta}$C-Ala, $^{\gamma}$C-Val, $^{\delta}$C-Ile, $^{\epsilon}$C-Met. Spectral region at 172-180 ppm refers for both lipid and protein >C=O peak. U-^{13}CGlu_^{13}CHM + D_8Lipid and 1,3-^{13}CGly_^{13}CHM + D_8Lipid sample does not show >C=O peak at 172-180 ppm region at 16, 24 and 40 ms, respectively because there could be faster T2 relaxation of

the sample. There is a peak at 52-60 ppm region (Fig. 5.16) for U-^{13}CGlu_^{13}CHM + D_{10}Lipid which refers to $^{\alpha}$C-Ala/$^{\alpha}$C-Cys/$^{\alpha}$C-Thr/$^{\alpha}$C-Leu/$^{\alpha}$C-Val/$^{\alpha}$C-Lys/$^{\alpha}$C-Ile/$^{\alpha}$C-Met/$^{\alpha}$C-Phe/$^{\alpha}$C-Tyr/$^{\alpha}$C-Ser. Moreover, 180-188 ppm spectral region (Fig. 5.16) for U-^{13}CGlu_^{13}CHM + D_{10}Lipid refers to $^{\delta}$C-Glu, $^{\delta}$C-Asp peak. D_{10}Lipid does not show those peaks which indicates only protein contributes for this spectral region.

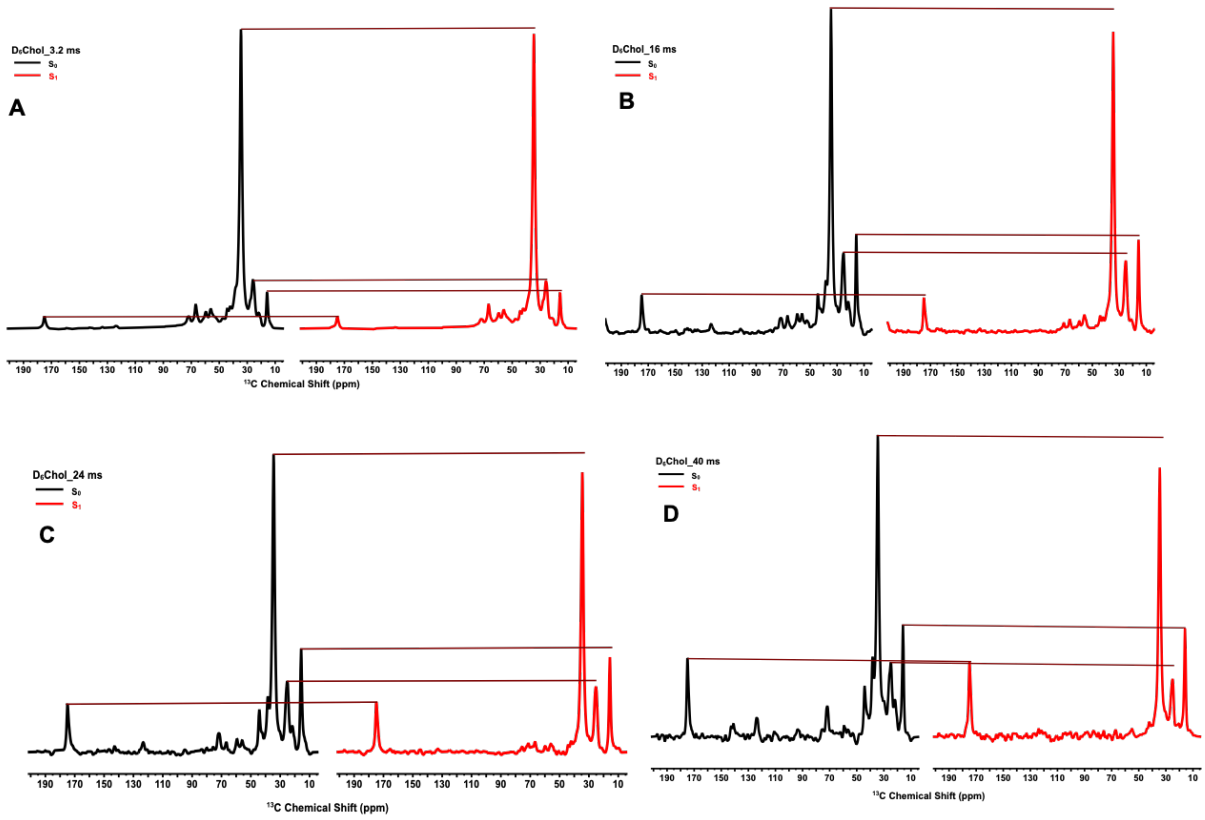


Figure 5.15: ^{13}C–^{2}H REDOR spectra of D_6 Chol Lipid time at (A) 3.2 ms (B) 16 ms (C) 24 ms and (D) 40 ms. The black lines in Fig. (A-D) denoted S_0 spectrum and red lines are denoted S_1 spectrum. The dark brown line over S_0 to S_1 at 172-180 ppm, 30-38 ppm, 22-28 ppm and 12-18, respectively demonstrates the difference between S_0 and S_1 by which dephasing was calculated as $\Delta S/S_0 = S_0\text{-}S_1/S_0$.

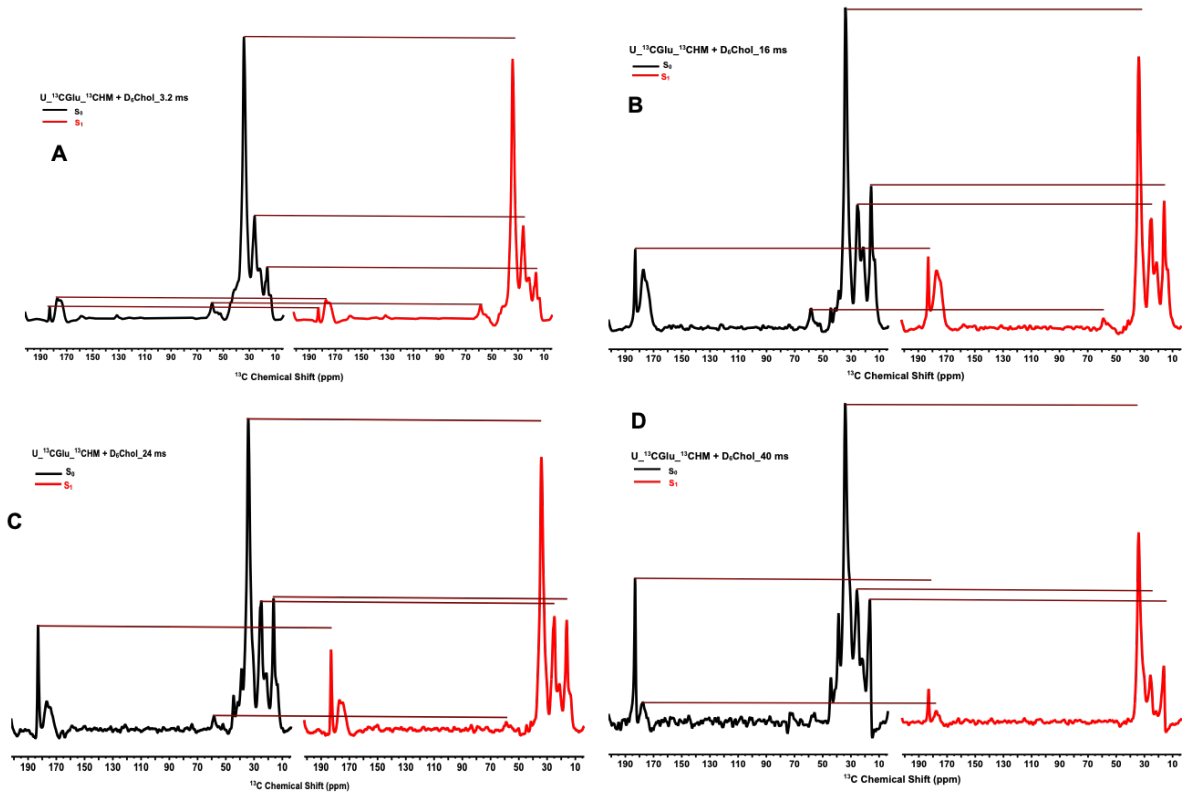


Figure 5.16: $^{13}C–^{2}H$ REDOR spectra of U_^{13}CGlu_^{13}CHM + D_6Chol (U_^{13}CGlu as carbon source for ^{13}CHM) with a dephasing time at (A) 3.2 ms (B) 16 ms (C) 24 ms and (D) 40 ms. The black lines in Fig. (A-D) denoted S_0 spectrum and red lines are denoted S_1 spectrum. The dark brown line over S_0 to S_1 at 180-188 ppm, 52-60 ppm, 30-38 ppm, 22-28 ppm and 12-18, respectively demonstrates the difference between S_0 and S_1 by which dephasing was calculated as $\Delta S/S_0 = S_0\text{-}S_1/S_0$.

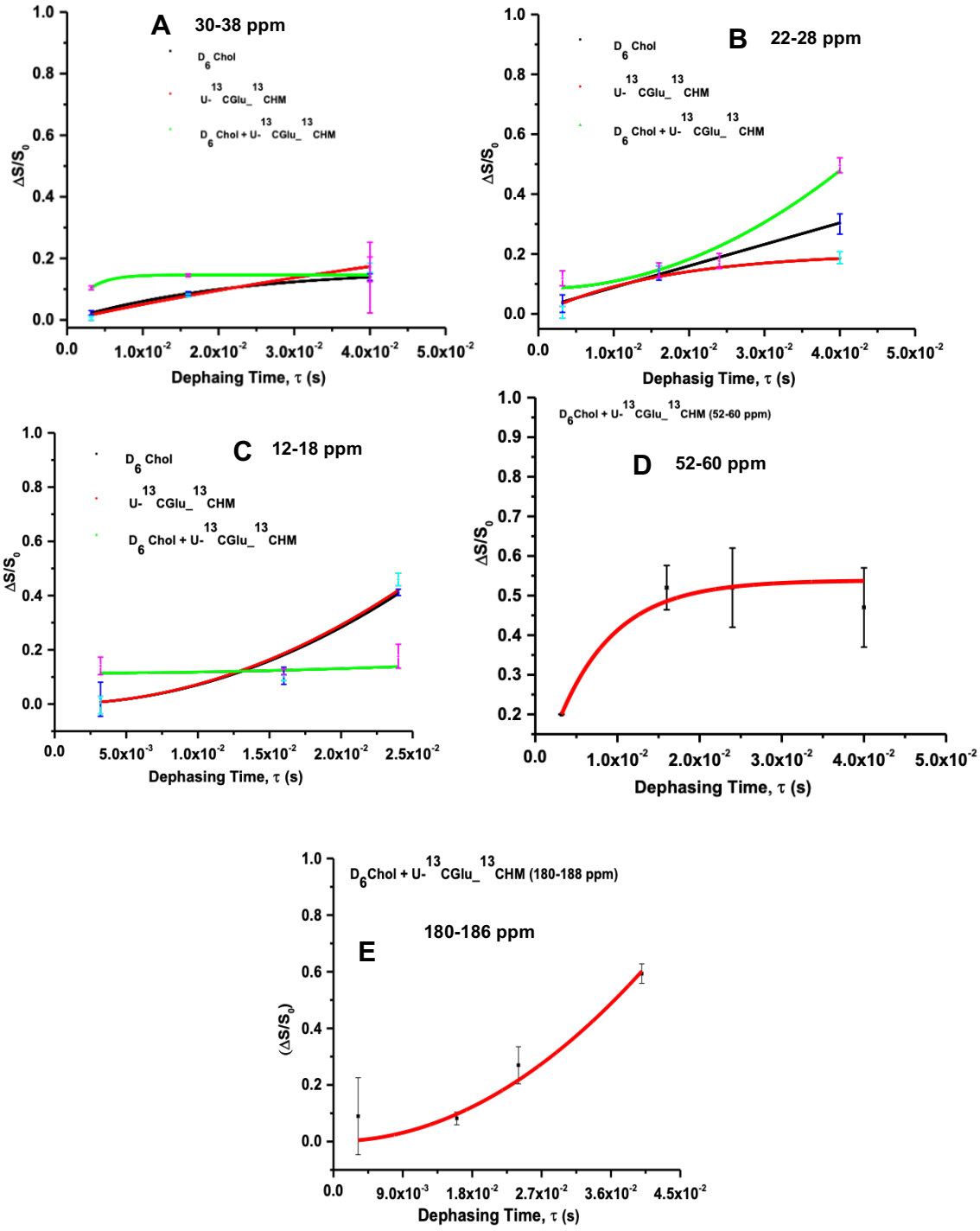


Figure 5.17: REDOR buildup curve of ($\Delta S/S_0$) vs. dephasing time (τ) at spectral region in (A) 30-38 ppm (B) 22-28 ppm (C) 12-18 ppm (D) 52-60 ppm (E) 180-188 ppm,

Figure 5.17 (cont'd)

respectively. In (A), $(\Delta S/S_0)^{exp}$ of D_6Chol (back line with cyan error bars) and U_^{13}CGlu_^{13}CHM + D_6Chol (blue line with pink error bars), respectively and $(\Delta S/S_0)^{Cal}$ of U_^{13}CGlu_^{13}CHM (red line with black error bars) are fitted to a single exponential buildup $A\times(1-e^{-\gamma\tau})$ with A and γ are fitting parameters. In (B), $(\Delta S/S_0)^{exp}$ of D_6Chol (back line with blue error bars) is fitted to a linear buildup $M + \gamma \times \tau$ with M and γ are fitting parameters, $(\Delta S/S_0)^{exp}$ of U_13CGlu_13CHM + D_6Chol (green line with pink error bars) is fitted to a parabolic buildup $M + \gamma^2 \times \tau^2$ where M and γ are fitting parameters and $(\Delta S/S_0)^{Cal}$ of U_13CGlu_13CHM (red line with cyan error bars) is fitted to a single exponential buildup $A\times(1-e^{-\gamma\tau})$. In (C), $(\Delta S/S_0)^{exp}$ of D_6Chol (back line with blue error bars) and $(\Delta S/S_0)^{Cal}$ of U_13CGlu_13CHM (red line with blue error bars) are fitted to a parabolic buildup $M + \gamma^2 \times \tau^2$ and $(\Delta S/S_0)^{exp}$ of U_13CGlu_13CHM + D_6Chol (green line with pink error bars) is fitted to a linear buildup $M + \gamma \times \tau$. In (D & E), $(\Delta S/S_0)^{Exp}$ of U_13CGlu_13CHM + D_6Chol at 52-60 ppm spectral region is fitted to a single exponential buildup $A\times(1-e^{-\gamma\tau})$ and at 180-188 ppm spectral region is fitted to a parabolic buildup $M + \gamma^2 \times \tau^2$; there is no peak in those spectral regions for D_6Chol (Fig. 5.15 & 5.16), therefore peak in this region only from ^{13}CHM.

Figure 5.17 (A) shows the $(\Delta S/S_0)$ vs. dephasing time (τ) fitted curve of experimental D_6Chol, U_13CGlu_13CHM + D_6Chol and calculated ^{13}CHM (lipid contribution has been omitted after calculation using E5.2) at 30-38 ppm spectral region, respectively in which all curves were fitted to a single exponential buildup $A\times(1-e^{-\gamma\tau})$, where A and γ are fitting parameters with $d= 3\gamma/2=137$ Hz and r (^{13}C-^{2}H)= $(4642/d)^{1/3}=3.23 \pm 0.02$ Å for D_6Chol, $d= 3\gamma/2=137$ Hz and r (^{13}C-^{2}H)= $(4642/d)^{1/3}=3.23 \pm 0.02$ Å for D_6Chol, $d= 3\gamma/2=593$ Hz and r (^{13}C-^{2}H)= $(4642/d)^{1/3}=1.98\pm0.03$ Å for U_^{13}CGlu_^{13}CHM + D_8Lipid and $d= 3\gamma/2=16$ Hz and r (^{13}C-^{2}H)= $(4642/d)^{1/3}=6.62\pm3.35$ Å for U_^{13}CGlu_^{13}CHM. This result indicates lipid contribution for ^{13}C to ^{2}H magnetic dipolar interaction is much significant compare to ^{13}CHM because r of calculated ^{13}CHM to lipid is within in the range in r of lipid ^{13}C to ^{2}H and smallest r of U_^{13}CGlu_^{13}CHM + D_8Lipid (less than lipid's r) could be contributed from lipid since of calculated r of ^{13}CHM is much larger than this sample.

In figure 5.17 (B) at 22-28 ppm spectral region, $(\Delta S/S_0)^{Exp}$ of D_6Chol is fitted to linear buildup, $(\Delta S/S_0)^{Exp}$ of U_13CGlu_13CHM + D_6Chol is fitted to a parabolic buildup and $(\Delta S/S_0)^{Cal}$ of U_13CGlu_13CHM is fitted to a single exponential buildup like mentioned in Fig. 5.17. Internuclear distance (r) between ^{13}C and ^{2}H of calculated U_^{13}CGlu_^{13}CHM is 3.75±1.02 which is significantly smaller than experimental r of D_6Chol (7.55±0.13) and U_13CGlu_13CHM + D_6Chol (5.81±0.1) (Table 5.1), respectively that indicates protein ^{13}C is much closer to lipid ^{2}H than lipid ^{13}C-^{2}H and protein ^{13}C has significant contribution to dephase lipid ^{2}H.

In figure 5.17 (C) at 12-18 ppm spectral region, $(\Delta S/S_0)^{Exp}$ of D_6Chol and $(\Delta S/S_0)^{Cal}$ of U_13CGlu_13CHM are fitted to parabolic buildup and $(\Delta S/S_0)^{Exp}$ of U_13CGlu_13CHM + D_6Chol is fitted to a linear buildup (there is no significant change of dephasing at different τ's). Internuclear distance (r) between ^{13}C and ^{2}H of calculated U_^{13}CGlu_^{13}CHM is 4.85±0.12 which is similar to experimental r of D_6Chol (4.88±0.04) (Table 5.1), that indicates protein and lipid ^{13}C are distinguishable to dephase lipid ^{2}H.

In figure 5.17 (D) at 52-60 ppm spectral region, $(\Delta S/S_0)^{Exp}$ of U_13CGlu_13CHM + D_6Chol is fitted to a single exponential buildup $A\times(1-e^{-\gamma\tau})$, where A and γ are fitting parameters with d= $3\gamma/2$=218 Hz and r (^{13}C-^{2}H)= $(4642/d)^{1/3}$=2.77 ± 0.08 Å. D_6Chol does not show any peak in this spectral region which indicates only protein ^{13}C (i.e. $^{\alpha}$C-Ala/$^{\alpha}$C-Cys/$^{\alpha}$C-Thr/$^{\alpha}$C-Leu/$^{\alpha}$C-Val/$^{\alpha}$C-Lys/$^{\alpha}$C-Ile/$^{\alpha}$C-Met/$^{\alpha}$C-Phe/$^{\alpha}$C-Tyr/$^{\alpha}$C-Ser) has contribution for this peak and ^{13}CHM has D_6Chol location (periphery in the lipid bilayer) in the membrane.

In figure 5.17 (E) at 180-186 ppm spectral region, $(\Delta S/S_0)^{Exp}$ of U_13CGlu_13CHM + D_6Chol is fitted to $M + \gamma^2 \times \tau^2$ where M and γ are fitting parameters with d= $3\gamma/2$=218 Hz and r (^{13}C-^{2}H)= $(4642/d)^{1/3}$=5.42 ± 0.06 Å. D_6Chol does not show any peak in this spectral region which indicates only protein ^{13}C ($^{\delta}$C-Glu, $^{\delta}$C-As) has contribution for this peak and ^{13}CHM has D_6Chol location (periphery in the lipid bilayer) in the membrane.

5.6: Discussion

This study was designed to find the location of soluble ectodomain (SE=HM=HP+MPER) of gp41 in HIV fusion. HM was ^{13}C labeled using carbon sources either by uniformed labeled glucose (U-^{13}CGlu) or selective labeled glycerol (1,3-^{13}CGly/2-^{13}CGly) and expressed in E. coli minimal medium through recombinant DNA technology. Expressed

protein (containing His-tag) from 1,3-^{13}CGly/2-^{13}CGly as carbon sources was purified by cobalt (Co^{2+}) resin protein affinity chromatography and protein (no His-tag) from U-^{13}CGlu as carbon sources was purified through successive washing by 8M urea/6M GuHCl. Monomer band was observed in gel for both purification protocol in Fig. 5.2. However, few other bands are observed for U-^{13}CGlu in Fig. 5.2 (a) that could be for impurities which also observe in MALDI spectrum (impurities band in between M^{+2} and M^{+1} peak) in figure 5.3 (a). Conformation of ^{13}C labeled in expressed protein was done by MALDI mass spectrometry in Fig. 5.3 where additional mass unit was observed for HM (M.W.12961) due to ^{13}C labeled in protein.

^{13}CHM was incorporated with either D10/D8 lipid or D6 Cholesterol to find the location of HM at different locations in membrane. REDOR dephasing was observed mainly in 172-178 ppm (>C=O peak), 52-60 ppm ($^{\alpha}$C-Ala/$^{\alpha}$C-Cys/$^{\alpha}$C-Thr/$^{\alpha}$C-Leu/$^{\alpha}$C-Val/$^{\alpha}$C-Lys/$^{\alpha}$C-Ile/$^{\alpha}$C-Met/$^{\alpha}$C-Phe/$^{\alpha}$C-Tyr/$^{\alpha}$C-Ser peak), 30-38 ppm ($^{\beta}$C-Val/ $^{\beta}$C-Ile/ $^{\beta}$C-Phe/ $^{\gamma}$C-Met/ $^{\delta}$C-Lys peak), 22-28 ppm ($^{\gamma}$C-Ile/$^{\gamma}$C-Ile/$^{\gamma}$C-Lys/ $^{\gamma}$CThr/ $^{\gamma}$C-Arg) and 12-18 ppm ($^{\beta}$C-Ala, $^{\gamma}$C-Val, $^{\delta}$C-Ile, $^{\epsilon}$C-Met). However, Lipid has significant contribution for dephasing at 172-178 ppm (>C=O peak), 30-38 ppm (-CH_2 peak), 22-28 ppm (-CH_2 (-CH_3) peak) and 12-18 ppm (-CH3 peak). This indicates internal lipid ^{2}H and natural abundant ^{13}C in lipid have close proximity for magnetic dipolar interaction to diphase in signal and internuclear distances [r (^{13}C-^{2}H] are show in table 5.1. Internuclear distances of lipid ^{13}C-^{2}H and protein ^{13}C-^{2}HLipid of few samples in various lipid regions are not significantly different (table 5.1) which indicates lipid contributes much to the dephasing. However, dephasing only by HM^{13}C to ^{2}H lipid in lipid+protein samples were calculated by E5.2 and all the fitting curves of protein contributions are shown in Fig. 5.18. Comparison of HM location in different lipid positions (D10=middle of the lipid bilayer, D8=middle of the leaflet of lipid bilayer, periphery of lipid bilayer, Fig. 2.1) also shown in Fig. 5.18.

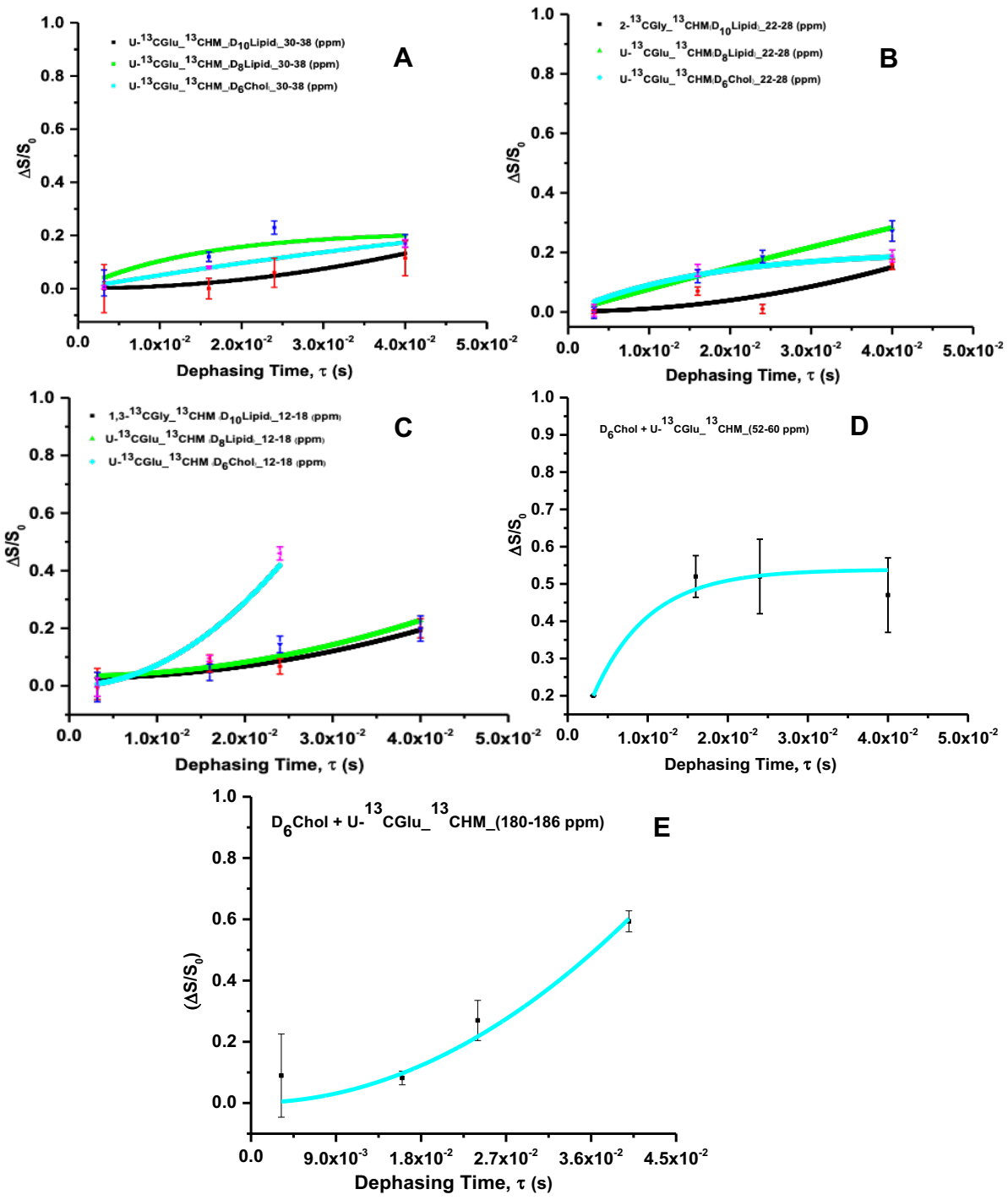


Figure 5.18: REDOR buildup fitted curve (collected from Fig. 5.8, 5.14 & 5.17) of ($\Delta S/S_0$) vs. dephasing time (τ) in comparing membrane location of HM from calculated ^{13}CGlu_^{13}CHM at spectral region in (A) 30-38 ppm (B) 22-28 ppm (C) 12-18 ppm for all

Figure 5.18 (cont'd)

lipid region in membrane (D_{10}, D_8 and D_6 Chol) and spectral region (D) 52-60 ppm (E) 180-188 ppm, respectively in D_6Chol region.

At 30-38 ppm region, $(\Delta S/S_0)^{Cal}$ values for U-^{13}CGlu_^{13}HM samples at different lipid region (either D_{10}/D_8/d_6Chol) are 0.14-0.19 at 40 ms and they are not statistically different (error bars are within same range). However, their fittings of $(\Delta S/S_0)^{Cal}$ vs. τ for different lipid regions are different by mathematical equation and internuclear distance's (r (^{13}C-^{2}H)) in D_{10} + ^{13}CHM samples are (6.99±.35 Å) with significant difference by standard deviations from D_{10}Lipid sample ((3.32±.07 Å, table 5.1). This phenomenon indicates ^{13}C in different $^{\beta}$C-Val/ $^{\beta}$C-Ile/ $^{\beta}$C-Phe/ $^{\gamma}$C-Met/ $^{\delta}$C-Lys of HM in D_{10} + ^{13}CHM samples might have closed to lipid ^{2}H. However, asymptotic values of dephasing (not significant buildup) by ^{13}CHM in D_{10} region is low and internuclear distances of ^{13}CProtein and ^{2}HLipid are much higher (~7 Å) than distance between lipid natural abundant ^{13}C and ^{2}H in lipid (~3 Å) which indicates those particular residues have shallow insertion with minor population in D_{10}lipid region[17,18,25]. However, internuclear distances of ^{13}C in HM to D_8Lipid/D_6 cholesterol region's ^{2}H has a large standard deviation at 30-38 ppm that are within the distance of lipid natural abundant ^{13}C to ^{2}H in lipid which indicates that in 30-38 ppm peak for D_8Lipid/D_6 cholesterol have more contributions by lipid itself for dephasing in the ^{13}CProtein+^{2}HLipid samples and ^{13}HM could not be found well located at those positions in lipid for those particular samples.

At 22-28 ppm region in Fig. 5.18, $(\Delta S/S_0)^{Cal}$ values for 2-^{13}CGlu_^{13}HM in D_{10} lipid region has lower buildup found after calculating by E5.2 and internuclear distance's (r (^{13}CHM-$^{2}HD_{10}$)) is 6.83±.43 Å (table 5.1). However, buildup of D_{10}Lipid (no protein) samples was not well fitted (Fig. 5.9) and internuclear distance was not calculated for $^{13}CD_{10}$-$^{2}HD_{10}$. On the other hand, r (^{13}CHM-$^{2}HD_8$) is 8.62±10 Å for U-^{13}CGlu_^{13}HM, $^{13}CD_8$-$^{2}HD_8$ is 5.91±.30 Å for D_8Lipid, r (^{13}CHM-$^{2}HD_6$Chol) is 3.75±1.02 Å for U-^{13}CGlu_^{13}HM and r ($^{13}CD_6$Chol-$^{2}HD_6$Chol) is 5.91±.30 Å for D_6Chol. It indicates protein contribution in D_8Lipid region since both internuclear distance values are in the same error range; therefore, it is hard to say whether $^{\gamma}$C-Ile/$^{\gamma}$C-Ile/$^{\gamma}$C-Lys/ $^{\gamma}$CThr/ $^{\gamma}$C-Arg residues in HM at 22-28 ppm in D8 lipid region have close contact with lipid. However, $^{\gamma}$C-Ile/$^{\gamma}$C-Ile/$^{\gamma}$C-Lys/ $^{\gamma}$CThr/ $^{\gamma}$C-Arg residues of HM

at 22-28 ppm region in D_6Chol have close contact since protein only samples internuclear distance's error is significantly different and protein internuclear distance, r (^{13}CHM-$^{2}HD_6$Chol) is much lower than r ($^{13}CD_6$Chol-$^{2}HD_6$Chol) which indicates those particular amino acid residues carbon mostly located in the lipid peripheral region (D_6 Chol) **[17,18,25].** Moreover, large internuclear distance at 22-28 ppm spectral region's (r (^{13}CHM-$^{2}HD_{10}$= 6.83±.43 Å) implies that amino acid residue of carbons for that particular peak region might have not been deeply inserted and HM amino acids residues do not have substantial position in D_{10}Lipid region**[17,18]**.

At 12-18 ppm region in Fig. 5.18, $(\Delta S/S_0)^{Cal}$ values for 1,3-^{13}CGly_^{13}HM in D_{10} lipid and U-^{13}CGlu_^{13}HM in D_8 lipid region has lower buildup (0.15-0.19 at 40 ms) found after calculating by E5.2 and internuclear distance's r (^{13}CHM-$^{2}HD_{10}$)) is 6.70±.27 Å and r (^{13}CHM-$^{2}HD_8$) is 6.55±.41 Å (table 5.1). Buildup of D_{10}Lipid (no protein) samples was very high (~0.76 at 40 ms) (Fig. 5.8D) and r ($^{13}CD_{10}$-$^{2}HD_{10}$) was 2.63±.06 Å. Moreover, buildup of D_8Lipid (no protein) samples was high (~0.6at 40 ms) (Fig. 5.14C) with linear fitting and r ($^{13}CD_8$-$^{2}HD_8$) was 5.64±.25 Å which is not significantly separated from protein only r (^{13}CHM-$^{2}HD_8$). Thus, the above results are conclusive as shallow insertion of $^{\beta}$C-Ala/$^{\gamma}$C-Val/$^{\delta}$C-Ile/$^{\epsilon}$C-Met residues in HM at 12-18 ppm in D_{10}/D_8Lipid[17,18]. $(\Delta S/S_0)^{Cal}$ values for U-^{13}CGlu_^{13}HM in D_6 Chol in D_6 Chol region has moderate buildup (~0.45 at 24 ms) found after calculating by E5.2 and internuclear distance's r (^{13}CHM-$^{2}HD_6$Chol)) is 4.85±.12 Å (table 5.1). r ($^{13}CD_6$Chol-$^{2}HD_6$Chol) is 4.88±0.04 Å for D_6Chol (no protein) at 12-18 ppm. Both internuclear distances from above values are not statistically different; therefore, it is hard to say either lipid ^{13}C or Protein ^{13}C contributes for dephasing at 12-18 ppm spectral region in D6Chol region.

At 52-60 ppm and 180-186 ppm in D_6Chol region in Fig. 5.18D&E show the significant buildup for U-^{13}CGlu_^{13}HM+D_6Chol. However, D_6Chol sample (no protein contribution) in these spectral regions does not show any peak (Fig. 5.15) which implies no lipid contributions had in those spectral peaks. The 52-60 ppm peak show the significant buildup (~0.58 at 40 ms) and r (^{13}CHM-$^{2}HD_{10}$))= 2.77±.08 Å which indicates α-^{13}C particular amino acids in HM ($^{\alpha}$C-Ala/$^{\alpha}$C-Cys/$^{\alpha}$C-Thr/$^{\alpha}$C-Leu/$^{\alpha}$C-Val/$^{\alpha}$C-Lys/$^{\alpha}$C-Ile/$^{\alpha}$C-Met/$^{\alpha}$C-Phe/$^{\alpha}$C-Tyr/$^{\alpha}$C-Ser) have close contact with the D_6Chol ^{2}Hlipid and it has a deep

insertion in this region (<3.0 Å) [17,18]. Thus HM has mostly peripheral contact in lipid bilayer. Moreover, The 180-186 ppm peak show the significant buildup (~0.55 at 40 ms) and r (^{13}CHM-^{2}HD$_{10}$))= 5.42±.06 Å which indicates ^{13}C ($^{\delta}$C-Glu, $^{\delta}$C-As) have close contact with the D$_6$Chol ^{2}Hlipid and it has a deep insertion in this region **[17,18].** Thus this data from the spectral region 180-186 ppm signifies HM has mostly peripheral contact in lipid bilayer.

L. Jia et al. (2015) showed antiparallel β-sheet registers are responsible for the distribution of HIV FP's (HFP) into various membrane location. In particular, an HFP registry's membrane insertion depth is probably determined by the lengths of its adjacent hydrophobic sections, which differ between registries. The distribution of HFP membrane sites may also differ between deep and shallow insertions. The activation energy of membrane fusion may be lowered and the membrane bilayer may be severely perturbed by the predominant deep insertion of HFP. L. Jia et al. (2015) showed PC:PG membranes with greater buildup with PC_d10 than with PC_d8 for HFP_G5C, HFP_L12C[17]. With a major population with a deep insertion in contact with d10 ^{2}Hs and a minor population with a shallower insertion in contact with d8 2Hs, this suggests that HFP has two distinct membrane sites. Our study also shows two distinct populations, as major and minor, in the membrane. The minor and major population is calculated from buildup of $(\Delta S/S_0)D_{10}$--to-$(\Delta S/S_0)D_{6,}$ $(\Delta S/S_0)D_8$-to-$(\Delta S/S_0)D_6$ and $(\Delta S/S_0)D_8$-to-$(\Delta S/S_0)D_{6,}$ respectively for U-^{13}CGlu_^{13}HM sample. The minor to major population is 2:5 for $(\Delta S/S_0)D_{10(12\text{-}18ppm)}$-to-$(\Delta S/S_0)D_{6(12\text{-}18ppm)}$, 2:5 for $(\Delta S/S_0)D_{8(12\text{-}18ppm)}$-to-$(\Delta S/S_0)D_{6(12\text{-}18ppm)}$, 1:1 for $(\Delta S/S_0)D_{10(12\text{-}18ppm)}$-to-$(\Delta S/S_0)D_{8(12\text{-}18ppm)}$ at τ=24 ms for 12-18 ppm spectral region. $(\Delta S/S_0)D_{10(30\text{-}38\ ppm)}$-to-$(\Delta S/S_0)D_{8(30\text{-}38\ ppm)}$-to-$(\Delta S/S_0)D_{6(30\text{-}38ppm)}$ is 1:1:1, $(\Delta S/S_0)D_{10(22\text{-}28\ ppm)}$-to-$(\Delta S/S_0)D_{8(22\text{-}28\ ppm)}$-to-$(\Delta S/S_0)D_{6(22\text{-}28ppm)}$ is 1:2:1, $(\Delta S/S_0)D_{10(22\text{-}28\ ppm)}$-to-$(\Delta S/S_0)D_{8(22\text{-}28\ ppm)}$-to-$(\Delta S/S_0)D_{6(52\text{-}60ppm)}$ is 1:2:5, $(\Delta S/S_0)D_{10(30\text{-}38\ ppm)}$-to-$(\Delta S/S_0)D_{8(30\text{-}38\ ppm)}$-to-$(\Delta S/S_0)D_{6(52\text{-}60ppm)}$ is 1:1:5, $(\Delta S/S_0)D_{10(22\text{-}28\ ppm)}$-to-$(\Delta S/S_0)D_{8(22\text{-}28\ ppm)}$-to-$(\Delta S/S_0)D_{6(180\text{-}186\ ppm)}$ is 1:2:5.5 and $(\Delta S/S_0)D_{10(30\text{-}38\ ppm)}$-to-$(\Delta S/S_0)D_{8(30\text{-}38\ ppm)}$-to-$(\Delta S/S_0)D_{6(180\text{-}186ppm)}$ is 1:1:5.5, respectively at τ=24 ms. Therefore, dephasing only by ^{13}CHM in D_{10}/D_8 lipid region has little buildup which indicates HM might have not been deeply inserted and HM amino acids residues do not have substantial position in D_{10}/D_8 lipid region. On the other hand, dephasing by ^{13}CHM in D$_6$CholLipid region has substantial buildup which indicates HM has major

population D_6Chollipid region. In other words, we can say HM has most locates in the periphery of the lipid bilayer.

REFERENCES

1. M. Pancera et al. Structure and immune recognition of trimeric pre-fusion HIV-1 Env. Nature 2014, 514, 455-461.
2. J.P. Julien et al. Crystal Structure of a Soluble Cleaved HIV-1 Envelope Trimer. Science 2013, 342, 6165, 1477-1483.
3. G. Ozorowski et al. Open and closed structures reveal allostery and pliability in the HIV-1 envelope spike. Nature 2017, 547, 560-363.
4. S. Liang et al. Efficient Fusion at Neutral pH by Human Immunodeficiency Virus gp41 Trimers Containing the Fusion Peptide and Transmembrane Domains. Biochemistry 2018, 57, 1219−1235.
5. D. C. Chan et al. Core Structure of gp41 from the HIV Envelope Glycoprotein. Cell 1997, 89, 263-273.
6. V. Buzon et al. Crystal Structure of HIV-1 gp41 Including Both Fusion Peptide and Membrane Proximal External Regions. PLoS Pathogens 2010, 6 (5), e1000880.
7. M. Caffrey et al. Three-dimensional solution structure of the 44 kDa ectodomain of SIV gp4. EMBO Journal 1998,17, 16, 4572–4584.
8. Yang, Z. N et al., The crystal structure of the SIV gp41 ectodomain at 1.47 A resolution. J. Struct. Biol., 1999, 126, 131−144.
9. M. Hong et al. Selective and extensive C-13 labeling of a membrane protein for solid-state NMR investigations, J. Biomol. 1999, NMR 14, 71–74.
10. D.M. LeMaster et al. Dynamical mapping of E. coli thioredoxin via C-13 NMR relaxation analysis, J. Am. Chem. Soc. 1996, 118 9255–9264.
11. F. Castellani et al. Structure of a protein determined by solid-state magic-angle-spinning NMR spectroscopy, Nature 2002, 420, 98–102.
12. A. Loquet et al. Supramolecular interactions probed by ^{13}C-^{13}C solid-state NMR spectroscopy, J. Am. Chem. Soc. 2010, 132,15164–15166.
13. A. Loquet et al. ^{13}C spin dilution for simplified and complete solid-state NMR resonance assignment of insoluble biological assemblies, J. Am. Chem. Soc. 2011, 133, 4722–4725.
14. V.A. Higman et al., Assigning large proteins in the solid state: a MAS NMR resonance assignment strategy using selectively and extensively ^{13}C labelled proteins, J. Biomol. 2009, NMR 44, 245–260.
15. M.J. Bayro et al. Dipolar truncation in magic-angle spinning NMR recoupling experiments, J. Chem. Phys. 2009, 130,114506.
16. G. A. Morris et al. Enhancement of nuclear magnetic resonance signals by polarization transfer, J. Am. Chem. Soc. 1979, 101, 3, 760–762.
17. L. Jia et al. REDOR solid-state NMR as a probe of the membrane locations of membrane-associated peptides and proteins. Journal of Magnetic Resonance 2015, 253, 154–165.

18. L. Xie, L. Jia, S. Liang, D.P. Weliky, Multiple locations of peptides in the hydrocarbon core of gel-phase membranes revealed by peptide ^{13}C to lipid ^{2}H rotational-echo double-resonance solid-state nuclear magnetic resonance, Biochemistry 54 (2015) 677–684.

19. K. Banerjee et al. Folded Monomers and Hexamers of the Ectodomain of the HIV gp41 Membrane Fusion Peptide: Potential Roles in Fusion and Synergy between the Fusion Peptide, Hairpin, and Membrane-Proximal External Region. Biochemistry 2014, 53, 7184-7198.

20. K. Sackett, M.J. Nethercott, R.F. Epand, R.M. Epand, D.R. Kindra, Y. Shai, D. P. Weliky, 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 (2010) 301–315.

21. A. Jacobs, C. Simon, M. Caffrey, Thermostability of the HIV gp41 wild-type and loop mutations, Protein Pept. Lett. 13 (2006) 477–480.

22. S. Boonstra et al., Hemagglutinin-Mediated Membrane Fusion: Biophysical Perspective, Annu. Rev. Biophys, 2018, 47, 153-173.

23. Gullion, T., Introduction to Rotational-Echo, Double-Resonance NMR. Concepts Magn. Resonance 1998, 10 (5), 277-289.

24. Zheng, Z., et al., Conformational flexibility and strand arrangements of the membrane-associated HIV fusion peptide trimer probed by solid-state NMR spectroscopy. Biochemistry, 2006. **45**(43): p. 12960-12975.

25. Schmick, S. D., and Weliky, D. P. (2010) Major antiparallel and minor parallel β sheet populations detected in the membraneassociated human immunodeficiency virus fusion peptide. Biochemistry 49, 10623−10635.

CHAPTER-06: Summary and Future Directions.

6.1: Summary

I have been working during the past few years on the large ectodomain part of HIVgp41. Gp41 is a single-pass integral HIV viral membrane protein involved in fusion catalysis. There is a 150-residue soluble ectodomain (SE) in HIV gp41 outside the virus. This soluble ectodomain (SE) adopts a soluble hairpin structure with an N-helix-turn-C-helix configuration. Two helices align antiparallelly and have van der Waals contact between them. Regarding the catalytic significance of the SE hairpin in HIV fusion, there are two competing hypotheses. While one hypothesis suggests that the hairpin is a "post-fusion" structure that plays no role in catalysis, the other hypothesis suggests that the hairpin plays a significant function in catalysis that is associated to binding to the membrane. This PhD research was focused on the catalytic role of SE and position of SE in membrane for HIV fusion.

In my first project in chapter three, we demonstrated that the final trimer of hairpins structure exhibits a dominant reduction of vesicle fusion, V513E (gp41 V2E in the absence of gp120), which is quantitatively similar to that of fusion and infection for gp160 (glycoprotein 160, a precursor protein that is cleaved into two smaller proteins: gp120 and gp41). This study also revealed correlation between low helicity and V2E fusion. Gp160-mediated fusion and infection were predicted to quantitatively similar with the study of reductions in FP_HM fusion and helicity vs. f_{V2E}. All V2E data have been global fitted to support the six WT gp41 (2 trimers) needed for fusion. These observations are explained by a scenario where the thermostable hairpin between the Fp in the target membrane and the Mper/transmembrane domain in the virus membrane compensates for the ~25 kcal/mol free energy for initial membrane apposition.

Chapter four showed the extension of first project in which role of hairpin and correlation between helicity vs. fusion further studied through characterization and fusion of G10V_FPHM and L9R_FPHM, respectively. This study demonstrated that there is less helicity in V2E/L9R than WT vs. less fusion in V2E/L9R than WT and similar helicity in WT/G10V vs. similar fusion in WT/G10V.

In chapter five for project two, the position of hairpin in membrane in HIV fusion was studied through measuring the proximity of protein with ^{13}C-labeled and lipid with ^{2}H labeled positioned either in middle /periphery/ in between of middle and periphery of the

membrane using the rotational-echo double-resonance (REDOR) solid-state NMR technique. The location of the hairpin in the membrane may provide additional insight into the membrane apposition (viral and host membrane) during fusion, since surface insertion forecasts far membrane apposition while deep SE insertion implies close membrane apposition. Furthermore, the locations of the surrounding lipid molecules in the membrane would probably differ greatly from their energetically stable position in an unperturbed bilayer, a hairpin contribution to the catalysis of membrane fusion is best understood if there is deep insertion. The production of ^{13}C proteins used a variety of carbon sources, and as a result, various labeled carbon peak assignments on the side chains of amino acids were observed in NMR. There was a significant REDOR dephasing for samples from different carbon sources, which shows that the soluble ectodomain has mostly peripheral location in the membrane. The production of ^{13}C proteins used a variety of carbon sources (25% Uniform labeled ^{13}CGlucose/1,3-^{13}CGlycerol/2-^{13}CGlycerol), and as a result, various labeled carbon peak assignments on the side chains of amino acids were observed in NMR. The soluble ectodomain is primarily located at the membrane's periphery, as evidenced by the considerable REDOR dephasing observed for samples from various carbon sources.

6.2: Future Works

Future studies could be done to express, purification, characterization and fusion study of other gp41 constructs i.e; V2E_G10V_FPHM (valine has been replaced by glutamic acid in 2 position and glycine has been replaced by valine in 10 position of FPHM), gp41 (512-716) and full length gp41 (512-856). Research of these constructs will further establish the role of gp41 hairpin in HIV fusion like shown in project 1 and extension of project 1. Dominance effect in fusion and helicity by double mutation (V2E and G10V) in V2E_G10V_FPHM could be done as project 1 in the future. This might get further ideas about the mystery of V2E mutation (great reduction of fusion by only a single point mutation in 2 positions of gp41) in HIV fusion and anomalies in behavior of G10V between gp160 and gp41 vesicle fusion. My colleague Tahmina Khatun in Dr. Weliky lab was trained by me for this research and started expression and purification of V2E_G10V_FPHM. Vesicle fusion study of gp41 (512-716) will give the idea about the role of amino acids residues in transmembrane (TM) domain in HIV fusion. Moreover,

vesicle fusion by full length gp41 (512-856) will mimic the role of gp41 in gp160 cell cell fusion. My colleague Forkan Saroar in Dr. Weliky lab was also trained by me for this research and started expression and purification of those constructs.

Future studies of project two could be done with site specific ^{13}C labeling of HM (only specific carbon atoms in certain residues are labeled with ^{13}C) and measuring the membrane location in lipid by REDOR solid state NMR. In project two we did HM labeling either uniformly by 25% uniform ^{13}C labeled Glucose or selectively either by 1,3-^{13}CGlycerol or 2-^{13}CGlycerol. However, HM labeling through these processes is not residue specific that would be enough to locate in the specific region in the lipid membrane ($D_{10}/D_8/D_6$Chol). Results of project two showed HM has membrane peripheral location (D_6Chol region). However, this study is semiquantitative in results and gives an approximate estimate (not fully accurate) that HM could be in the periphery of the lipid membrane based on measuring the dephasing of Protein+ D_6Chollipid sample. To get more quantitative estimation about HM location in D_6Chol region, site specific residue labeling will be a better approach because this will only be labeled ^{13}C in the amino acids that are located close to ^2H D_6Chol for dephasing. Therefore, future direction of project two will be site specific amino acid residues labeling of HM that are specifically close to ^2H D_6Chol for dephasing.