

SYNTHESIS OF HIV-1 GP41 INCLUDING FP AND MPER BY NATIVE CHEMICAL  
LIGATION WITH APPLICATIONS TO SSNMR. EXPRESSION, SOLUBILIZATION,  
AND PURIFICATION OF SARS-COV-2 SPIKE PROTEIN SUBUNIT 2

By

Robert John Wolfe

A DISSERTATION

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

Chemistry-Doctor of Philosophy

2022

## **ABSTRACT**

Human immunodeficiency virus type 1 (HIV-1) and coronavirus 2019 (Covid-19) have caused substantial risk to public health worldwide. Both HIV-1 and severe acute respiratory syndrome coronavirus type 2 (SARS-CoV-2) are membrane enveloped viruses which cause acquired immune deficiency syndrome (AIDS) and COVID-19 in humans, respectively. The process by which these viruses initiate the infection by membrane fusion of the viral and host cells is similar. However, they differ in which membrane protein is responsible for the fusion process. The HIV-1 glycoprotein 41 (Gp41) is a single-pass integral viral membrane protein containing a ~170-residue ectodomain that is important for membrane fusion between virus and host cells. The ectodomain includes the fusion peptide (FP), N-helical region (NHR), loop, C-helical region (CHR) and viral membrane-proximal external region (MPER). The ectodomain mediates joining (fusion) of the HIV-1 and host cell membranes, which is an initial step in infection. The ectodomain also adopts different structures, including a final hyperthermostable state. Some data supports a fusion role for this final state. Like HIV-1, SARS2-CoV-2 is enveloped by a membrane that is obtained during budding from an infected host cell. Infection of a new cell requires fusion of the virus membrane with a membrane of the target cell and subsequent deposition of the viral nucleocapsid in the cytoplasm. This process is catalyzed by the Spike (S) protein subunit 2 (S2). Amino acid sequences are conserved for a particular virus and its variants. Amino acid sequence is different for distinct viruses so that by understanding the similarities and differences between the domains of fusion proteins we can expand our knowledge of the mechanism of membrane fusion.

My research has focused on the production and characterization of several membrane protein constructs (with or without FP) and the characterization of S2\_816-1273, a large S2 protein construct containing all regions of the SARS-CoV-2 spike protein including the FP, transmembrane (TM) and the cytoplasmic tail (CT). Biophysical characterization comparisons between S2\_816-1273 and shorter constructs including S2\_903-998SGGRGG1153-1207 and S2\_903-998SGGRGG1163-1207 were performed using circular dichroism (CD) spectroscopy, size exclusion chromatography (SEC), mass spectrometry (MS), and vesicle fusion assays.

My work can be applied in future research to synthesize site-specific  $^{13}\text{C}$  and  $^2\text{H}$  labeled large protein constructs since this is the first time that large Gp41 constructs, including full MPER, have been synthesized in mg quantities using a combination of bacterial expression system and solid phase peptide synthesis. Furthermore, this methodology is applicable to many proteins inside of the lipid bilayer that cannot be easily characterized by other methods, such as crystallography.

Dedicated to Jesus, leg day, and my mamma.

## **ACKNOWLEDGEMENTS**

- I would like to thank my advisor Dr. David P. Weliky for his support and guidance throughout my research career as a graduate student. He has given me the freedom to try new techniques and experiments to explore and innovate as a scientist. I would like to thank him for teaching me to think critically and for his suggestions on how to interact with the scientific community efficiently. I am grateful to him for his patience and for his role in molding me into a responsible scientist.

- I would like to acknowledge Dr. Kevin Walker who guided me to join with Dr. Weliky's research group. I would also like to thank my other committee members including Dr. Babak Borhan, and Dr. John McCracken for their support.

- I am grateful for the past and present members of the Weliky group. I would like to thank Dr. Punsisi Ratnayake and Dr. Koyeli Banerjee for teaching me basic biochemical and molecular biology techniques in the lab. I would like to thank Dr. Li Xie, Dr. Lihui Jia, and Dr. Shuang Liang for teaching me how to perform solid state NMR and always helping me with research. I would like to thank present group members of the lab, Dr. Ujjayini Ghosh, Noel Chau, Yijin Zhang, and MD Rokonujjaman for being so friendly and helpful. They all made graduate life a journey filled with fun and laughter.

- I would like to thank Dr. Jim Geiger and his students Courtney Bingham and Nona Ehyaei for teaching me FPLC and how to make competent cells, and how to repack FPLC columns. I would also like to thank them for useful discussions about protein purification and expression. I would like to thank Dr. Tony Schillmiller for teaching me to use MALDI mass spectrometry to identify proteins and peptides. The instrument proved to be especially important for my research. I would like to extend thanks to Dr. Babak

Borhan and his lab for usage of the CD instrument and for helping with advice relating to repair of laboratory equipment. Finally, I would like to thank Dr. Lee Kroos and his lab for allowing me use of the western blot imager and for helpful advice in western blotting technique.

- I would like to thank my family for their dedication, good wishes, blessings, and encouragement. I would like to thank my father for always teaching me. I would like to think my sister Heather Wolfe for guidance and help whenever I needed it. My little sister Jade and my mother have always been there for me. My journey through graduate school would not have been possible without the support and encouragement of my family.

## TABLE OF CONTENTS

KEY TO ABBREVIATIONS .....	ix
CHAPTER 1: HUMAN IMMUNODEFICIENCY VIRUS MEMBRANE FUSION	
PROTEIN GP41 .....	1
1.1. Introduction .....	2
1.2. SARS-CoV-2 .....	16
1.3. Possible Mechanism of Membrane Fusion .....	27
1.4. Mechanism of Fusion of HIV-1 and Host Cell Membranes Caused by Gp41 .....	29
1.5. Native Chemical Ligation .....	35
1.6. Introduction of Solid-State NMR Techniques for Membrane Proteins .....	38
REFERENCES .....	43
CHAPTER 2: MATERIALS AND METHODS.....	
2.1. Introduction .....	56
2.2. Solid Phase Peptide Synthesis .....	56
2.3. Molecular Subcloning .....	58
2.4. Protein Expression .....	66
2.5. Solubilization and Purification of Expressed Protein .....	74
2.6. SDS PAGE (Sodium Dodecyl Sulphate Poly Acrylamide Gel Electrophoresis) .....	79
2.7. Gel Filtration Chromatography (SEC- Size Exclusion Chromatography) .....	80
2.8. CD Spectroscopy .....	81
2.9. Western Blots .....	82
2.10. Lipid Mixing Assays .....	82
2.11. Native Chemical Ligation .....	84
2.12. Solid State NMR Sample Preparation .....	86
2.13. Solid State NMR .....	87
REFERENCES .....	89
CHAPTER 3: PRODUCTION AND ISOTOPIC LABELING OF A LARGE GP41 ECTODOMAIN CONSTRUCT BY NATIVE CHEMICAL LIGATION BETWEEN THE FUSION PEPTIDE AND SOLUBLE ECTODOMAIN .....	
3.1. Introduction .....	94
3.2. Materials and Methods .....	98
3.3. Results .....	107
3.4. Discussion .....	132
3.5. Summary .....	136
REFERENCES .....	137
CHAPTER 4: APPLICATIONS OF NATIVE CHEMICAL LIGATION OF GP41 ECTODOMAIN TO STRUCTURAL ANALYSIS BY REDOR NMR .....	
4.1. Introduction .....	144
4.2. Materials and Methods .....	145

4.3. Results and Discussion .....	154
4.4. Discussion .....	162
4.5. Summary .....	163
REFERENCES .....	164
CHAPTER 5: EXPRESSION, PURIFICATION, SOLUBILIZATION, AND CHARACTERIZATION OF SARS-COV-2 PROTEIN CONSTRUCTS PRODUCED FROM <i>E.COLI</i> .....	
5.1. Introduction .....	170
5.2. Materials and Methods .....	171
5.3. Results and Discussion .....	172
5.4. Summary .....	186
REFERENCES .....	199
CHAPTER 6: SUMMARY AND FUTURE DIRECTIONS .....	
REFERENCES .....	209
APPENDIX A: Tables of NMR Values .....	213
APPENDIX A: Tables of NMR Values .....	
	214



## KEY TO ABBREVIATIONS

AA Amino acid

AB Aqueous binding

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

CoV coronavirus

CP cross-polarization

Cryo-EM Cryo-electron microscopy

CT Cytoplasmic tail

CXCR4 C-X-C Chemokine receptor type 4

DCM Dichloromethane

DEPBT 3-(Diethylphosphoryloxy)-1,2,3-benzotriazin-4(3H)-one

DIEA N, N-Diisopropylethylamine

DM n-Decyl- $\beta$ -D-Maltoside

DMF Dimethylformamide

DNA Deoxyribonucleic acid

dNTPs deoxynucleotide triphosphates

DOTAP 1,2-dioleoyl-3-trimethylammonium-propane (chloride salt)

DPPC n-Dodecylphosphocholine

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

DTPC 1,2-di-O-tetradecyl-sn-glycero-3-phosphocholine

E. coli Escherichia coli

EDTA Ethylenediamine tetraacetic acid

Endo Endodomain

EPR Electron paramagnetic resonance

ESR Electron spin resonance

Fmoc Fluoroenylmethyloxycarbonyl

FP Fusion peptide

FPHM FP+HM

GPCR G protein coupled receptor

gp160 Glycoprotein 160

gp140 Glycoprotein 140

gp120 Glycoprotein 120

gp41 Glycoprotein 41

GuHCl Guanidine hydrochloride

GUVs giant unilamellar vesicles

IB Inclusion body

HA hemagglutinin

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

IPTG Isopropyl  $\beta$ -D-1-thiogalactopyranoside

LB Luria-Bertani broth

LUVs Large unilamellar vesicles

MALDI-MS matrix assisted laser desorption ionization mass spectrometry

MAS Magic angle spinning

MERS-CoV Middle East respiratory syndrome

MES 2-(N-morpholino) ethanesulfonic acid

MPAA S-Trityl- $\beta$ -mercaptopropionic acid

MPER membrane-proximal external region

mRNA messenger RNA

NCL Native chemical ligation

nCoV Novel coronavirus

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)

NTA Nitrilotriacetic acid

dipalmitoylphosphatidylethanolamine

OC Organic co-solubilization

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

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

RNA Ribonucleic acid

RP Recombinant protein

RP-HPLC Reversed-phase HPLC

S spike protein

Sarkosyl Sodium lauryl sarcosinate

SARS1 Severe acute respiratory syndrome

SARS-CoV 2 severe acute respiratory syndrome coronavirus type 2

SDS Sodium dodecyl sulfate

SDS-PAGE Sodium dodecyl sulfate polyacrylamide gel electrophoresis

SEC Size-exclusion chromatography

SHB Six helix bundle

SSNMR Solid-state nuclear magnetic resonance

SUVs small unilamellar vesicles

S1 Spike protein subunit 1

S2 Spike protein subunit 2

t-Boc tert-Butyloxycarbonyl

TCEP Tris(2-carboxyethyl) phosphine hydrochloride

TFA Trifluoroacetic acid

THF Tetrahydrofuran

TM Transmembrane domain

UNAIDS The Joint United Nations Programme on HIV/AIDS

V2E Mutation of the Val-2 to Glu-2

WHO World health organization

**CHAPTER 1:**  
**HUMAN IMMUNODEFICIENCY VIRUS MEMBRANE FUSION PROTEIN GP41**

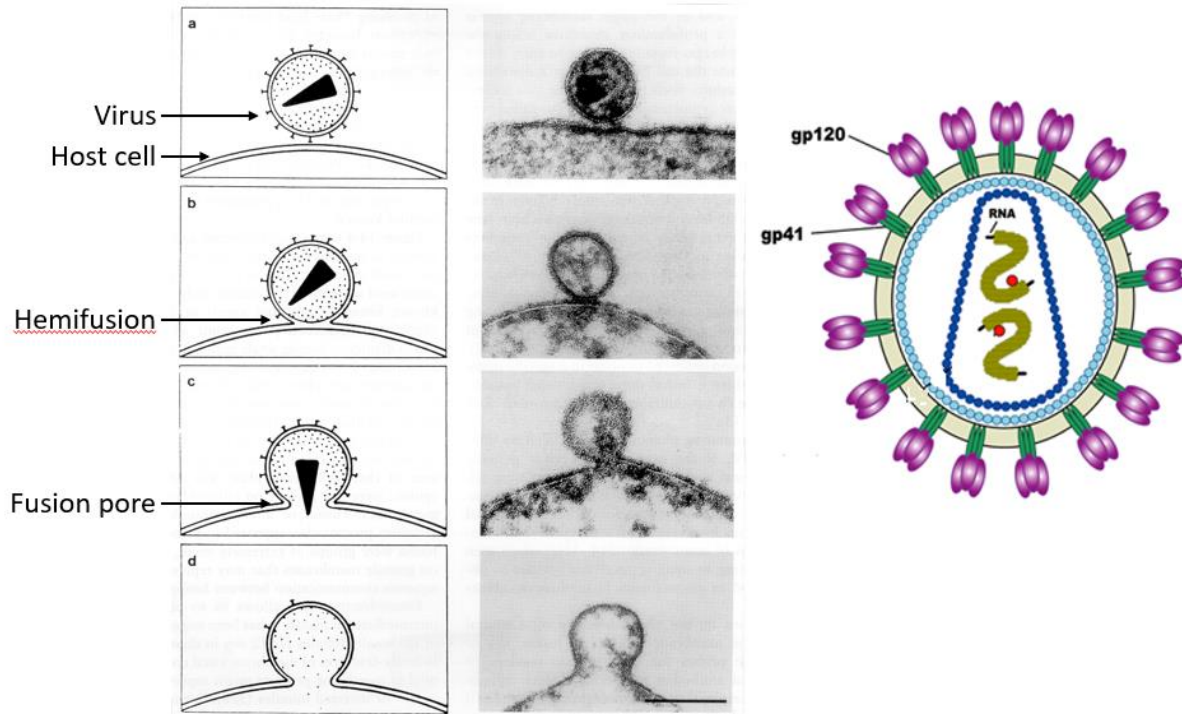
## 1.1. Introduction

Acquired Immune Deficiency Syndrome (AIDS) is a disease of the human immune system caused by the Human Immunodeficiency Virus (HIV). According to The Joint United Nations Programme on HIV/AIDS known as UNAIDS (126), over 35 million people have died from AIDS worldwide since the first cases were reported over 35 years ago (126). By the end of 2019, there were over 38 million people living with AIDS worldwide with more than 1.7 million new infections, and it is estimated that almost 700,000 people died of the illness in 2019. (126) HIV-1 is a membrane-enveloped virus, or a virus that has an outer wrapping called the envelope. The envelope comes from the infected cell, or host cell, in a process called “budding off”. (1,2) During the budding process, newly formed virus particles become enveloped by becoming wrapped in an outer coat that is made from the host cells plasma membrane.

The gp41 plays a significant role in fusion between viral and host cell membranes (6) (**Figure 1.1**). Following the binding of the HIV-1 virion to the CD4 receptors and coreceptors, it has been proposed that viral entry may occur in one of two ways:

(1) Fusion of the viral and the host cell plasma membrane catalyzed by gp41 and then release of the viral genetic material into the host cell which eventually integrates with the host genome (1,2).

(2) Infection by endocytic pathway has also been proposed. The virus enters the cell cytoplasm through the endosomes and then fusion with the endosome membrane is catalyzed by Gp41 occurs followed by release of viral genetic material into the cell cytoplasm (3,4).

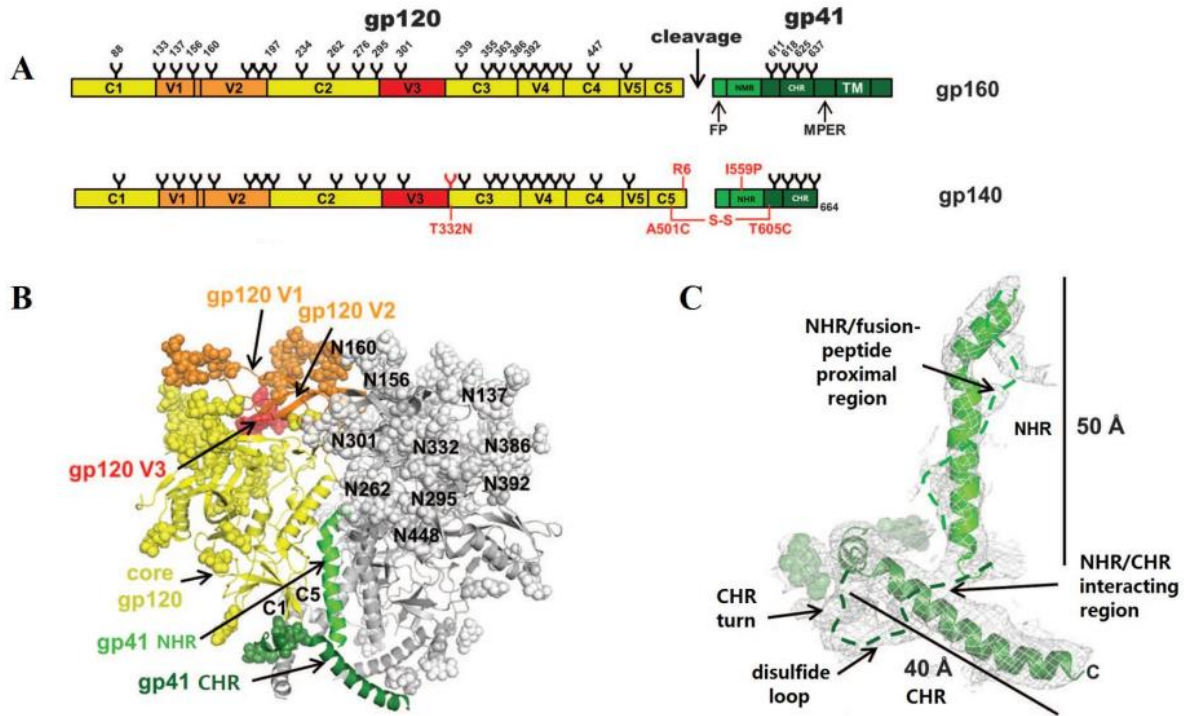


**Figure 1.1** (Left) Infection model of HIV, (Right) Electron microscopy (6) pictures of HIV infection process, (a) Binding of HIV and host cell, (b) Hemifusion of viral and host cell membrane, (c) Pore formation, and (d) complete fusion and entry of viral genetic material into the host cell (6). (Right)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

The process of membrane fusion between a virion and host cell is mediated by HIV-1 membrane glycoprotein 160 (gp160). A single virion can have as many as 25 gp160 trimers, but typically have around 15. Gp160 (**Figure 1.2**) comprises a non-covalently associated outer receptor binding subunit, glycoprotein 120 (gp120), and an inner fusion protein, glycoprotein 41 (gp41) (1). The gp160 in the HIV-1 membrane is activated for fusion by interactions with Cluster of Differentiation 4 (CD4) receptor, and C-X-C Chemokine Receptor Type 4 (CXCR4) or C-C Chemokine Receptor Type 5 (CCR5) coreceptor of the T-cell or macrophage cell membrane. Gp120 binds to the CD4 receptor



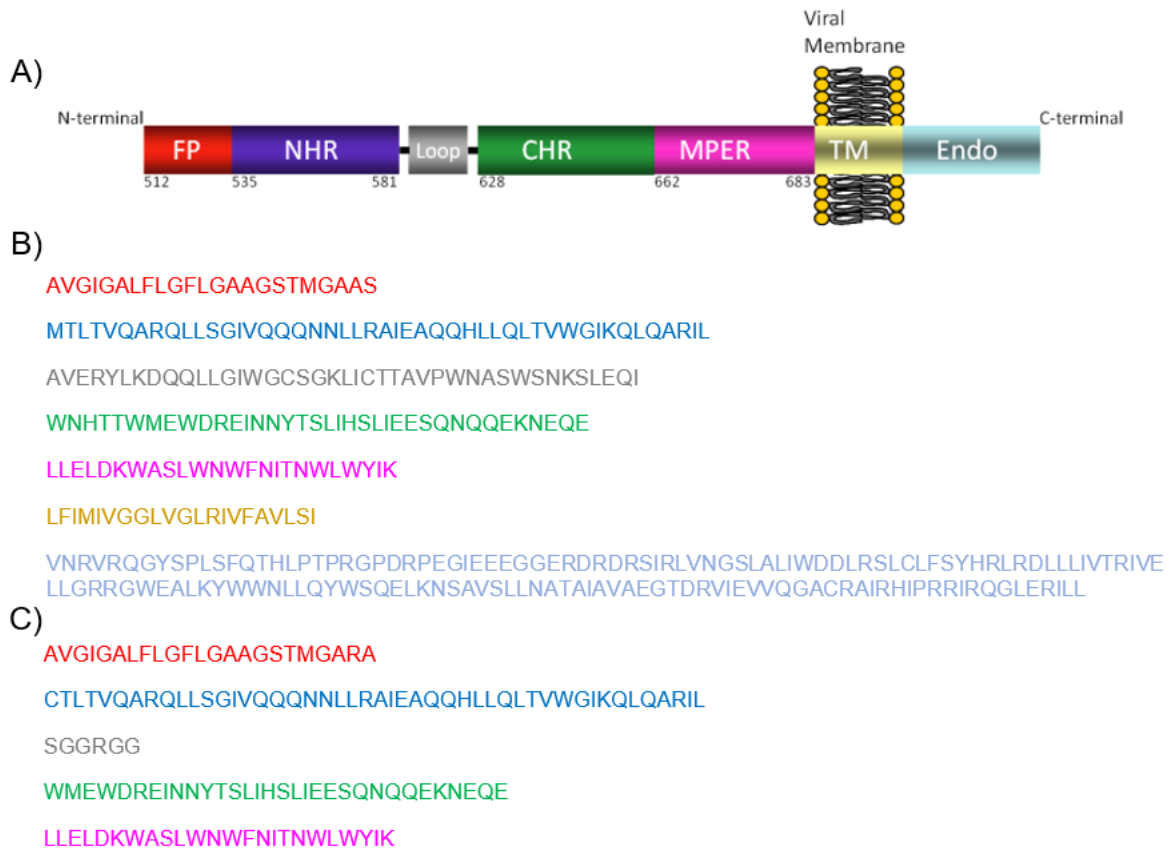
and coreceptors to produce a conformational change in gp120 that causes it to dissociate, leaving gp41 exposed (1).



**Figure 1.2** (A) Schematic of the HIV gp140 construct (68) studied in comparison to full-length gp160. N-linked glycans are shown and numbered on their respective Asn residues. The FP, NHR, CHR, MPER, transmembrane (TM), and endodomain elements in gp41 are indicated. The mutations are shown in red, as well as the added N332 glycan site. The color coding is preserved in (B) and (C). (B) Side view of the gp140 trimer. The main domains are labeled correspond to panel A and glycans are shown as spheres. (C) gp41 in gp140 complex. Secondary structure determination was ambiguous at the dashed line areas.

Gp41 consists of different regions such as ectodomain with amino acid residues 512-683, comprising of N-terminal fusion peptide (FP), an N-terminal helix region (NHR), an immunodominant loop region, a C-terminal helix region (CHR), membrane proximal

ectodomain region (MPER), the transmembrane region (TM) amino acid residues 684-705, and the endodomain cytoplasmic tail (CT) amino acid residues 705-856. Each region of gp41 plays a unique role (76) (**Figure 1.3**).



**Figure 1.3** (A) Schematic diagrams (76) of HIV gp41 and corresponding colors: 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, yellow; and endo ≡ endodomain, cyan. (B) Full amino acid sequence for protein in panel A.

### Figure 1.3 (cont'd)

(C) Amino acid sequences for HM with colors matching segments based on panel A. The sequence is from the HXB2 laboratory strain of HIV and has the gp160 precursor residue numbering, 1-511 for gp120 and 512-856 for gp41.

#### 1.1.1. Fusion Peptide

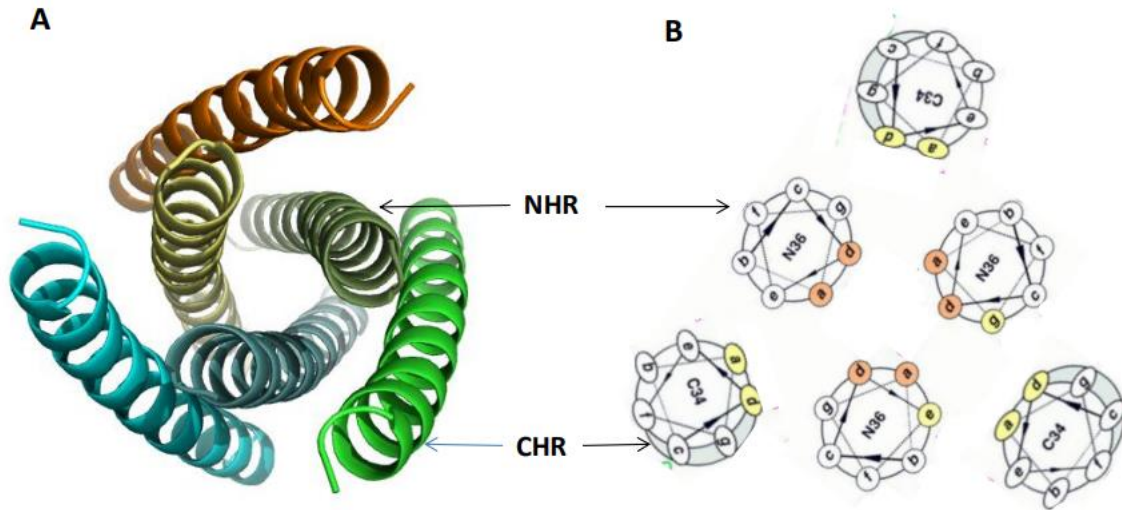
The FP region is a Gly rich hydrophobic N-terminal domain of gp41. There is no atomic resolution crystal structure of the FP region; however, some nuclear magnetic resonance (NMR) structural studies have been done with FP associated with detergent micelles. (127) While the FP is unstructured in solution, it forms  $\alpha$ -helical structure in the presence of detergents sodium dodecyl sulfate (SDS) or n-dodecyl phosphocholine (DPC) (7-9). There have been contradictory structural studies suggesting a  $\beta$ -sheet structure for the FP in a physiologically relevant membrane environment (10,11), whereas another study concludes that the FP region takes on either  $\alpha$ -helical monomeric structure or a  $\beta$ -sheet structure at low and high peptide: lipid concentration, respectively (12). Difference in FP conformation has been observed with the presence or absence of cholesterol in the lipid membrane. The FP adopts  $\alpha$ -helical structure in the absence of cholesterol, and  $\beta$ -sheet structure in the presence of cholesterol (11). Studies of the FP domain indicate that oligomerization may be important for fusion activity. Lipid mixing assays and analytical ultracentrifugation studies of FPmonomer, FPdimer, and FPtrimer shows maximum fusion activity for the trimer FP construct (13). Mutation of the Val-2 to Glu-2 (V2E) in the N-terminal end shows inhibition of fusion activity and syncytia formation. The effect of V2E mutation is such that even in the presence of excess wild type gp160, the membrane fusion and HIV infection is inhibited revealing the functional

oligomeric property of FP domain (14). These studies suggest that oligomerization of FP is essential for fusion of viral and host cell membrane. The study of the interaction between synthetic FP and phosphatidyl glycerol (PG) large unilamellar vesicles (LUVs) has been done by lipid mixing assays, as well as by vesicle binding and leakage experiments. These studies suggest that wild type FP can penetrate through the vesicle monolayer and cause permeabilization (12). A similar study performed with mutant V2E FP peptides showed no destabilization of PG vesicles (15). Additionally, solid state nuclear magnetic resonance (SSNMR) studies of FP in the lipid membranes show maximum membrane insertion depth and fusogenicity of trimer FP compared to that of the monomeric form or mutated (V2E) FP (16). Based on these studies, it could be proposed that FP inserts into the host membrane during the fusion process.

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

The high-resolution crystal structure of the NHR and CHR without the loop region shows a highly helical trimer structure forming a hairpin conformation (17,18). The regions studied were residues 546-581 and 628-661, where numbering begins with the NHR. This trimer NHR-CHR structure, also known as six-helix bundle (SHB), is proposed to be the structure of the gp41 core region after fusion is complete (**Figure 1.4**). At the center of the SHB are parallel NHR trimers held together by hydrophobic interactions primarily between the Ile or Leu residues. The amino acid sequence of the NHR and the CHR comprise of 4-3 heptad repeats of hydrophobic residues. Heptad repeats are formed by repeats of hydrophobic residues in the 1st and 4th residue of a seven-residue helical turn. There may be multiple helical turns per heptad repeat, depending on the size of the protein. The CHR helices pack on the outside grooves of the NHR trimer core in

antiparallel orientation. (22) The interaction between the NHR and CHR is predominantly hydrophobic. Some of the hydrophobic interactions include: (a) W628, W631 of CHR interacts with W571 of NHR, and (b) I635 of CHR interacts with L565 and L568 of NHR. In addition, Q653 of CHR has intramolecular hydrogen bond with Q551 of NHR and Q653 forms intermolecular hydrogen bond with the backbone CO oxygen of V549 of NHR. There is an inter-molecular salt bridge between the NHR and CHR between the K574 and D632. The NHR and CHR SHB is thermally stable with melting temperatures of  $\sim 80^{\circ}\text{C}$  for shorter ectodomain constructs with no loop (equimolar mixture of CHR and NHR peptides), whereas melting temperatures up to  $110^{\circ}\text{C}$  have been observed for longer constructs (535-581SGGRGG628-662) (19,20). This thermostability has supported the SHB as the final gp41 structure during fusion. Recent solution NMR structural study of NHR-CHR (546-655) in DPC micelles suggests monomeric structure with much less interaction between the NHR and CHR. The  $^{13}\text{C}$   $\alpha$  chemical shifts of all the residues of NHR-CHR construct were compared with those of the NHR and CHR residues of the individual peptides (127). It was observed that in the  $^{13}\text{C}$   $\alpha$  chemical shift of the NHR-CHR construct exactly overlapped with the individual NHR and CHR peptides. The same group conducted the study to understand the interaction of the NHR-CHR protein construct with detergent molecules by paramagnetic relaxation enhancement (128). It was observed that both the NHR and CHR helices are amphipathic with the hydrophobic residues lying on one side of the helix and the hydrophilic residues on the other side of each helix. Based on this observation, it is proposed that both NHR and CHR can be embedded at the lipid-water interface (21).



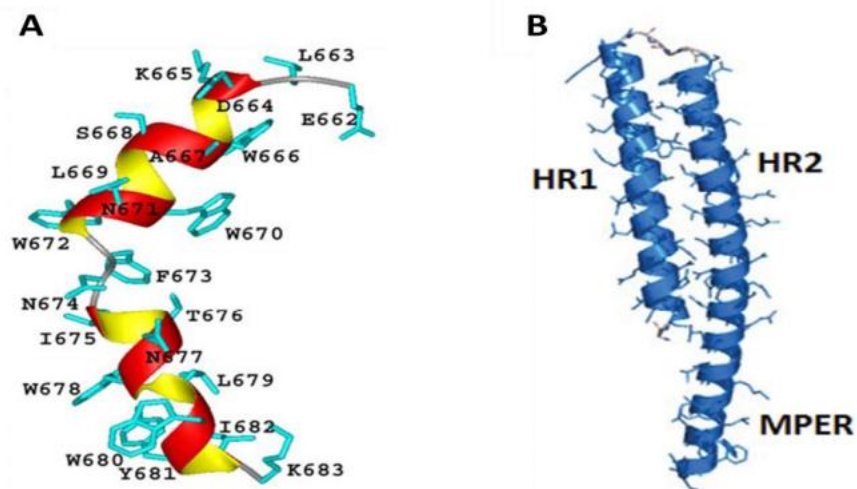
**Figure 1.4** (A) Six-helix bundle of gp41 comprising NHR and CHR and (B) NHR and CHR showing heptad repeat amino acid positions as orange and yellow, respectively (22). N36 is the 36 residues in NHR (546-581) and C34 is the 34 residues in the CHR (628-661).

### 1.1.3. Membrane Proximal External Region

The MPER is the last ectodomain region following the CHR towards the C-terminal end, comprising of amino acid residues 662 – 683. The two primary characteristics of the MPER make it a region of great interest: the amino acid residues are highly conserved (23-25), and the region comprises of epitopes for three broadly neutralizing antibodies (bNAbs) (23,25). Broadly neutralizing antibodies are antibodies which can stop infection caused by a broad range of HIV-1 strains. HIV-1 is the most common type of HIV and accounts for 95% of infections. (126) The immune response of the host is largely non-functional due to non-exposure of neutralizing sequence-conserved epitopes. Very few broadly neutralizing antibodies are elicited by the epitopes contained in the highly conserved MPER domain (23,25). Even though the broadly neutralizing antibodies are

efficient enough to inhibit the fusion process by binding to the MPER epitopes, still there is no success by injecting the antigen into the host body to elicit antibody (26). Efforts to produce broadly neutralizing antibodies to the conserved epitopes of MPER proved to be ineffective. The conserved neutralizing epitopes are reported to be poor immunogens, as they mimic host antigens leading to the depletion of immune tolerance (27).

Structural characterization of the MPER peptide by electron paramagnetic resonance (EPR) and NMR in both lipid and detergent environments, respectively, shows a kinked helical structure (**Figure 1.5**). A recently published high resolution crystal structure of MPER with NHR-CHR (residues 547-575GGGGS630-675) shows helical structure (29). EPR and NMR structural studies show a conformational change of the MPER (662-683) after binding to the antibodies. By EPR studies, the membrane insertion depth has been determined for spin labeled MPER bound to 4E10 antibody with respect to POPC (1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine and POPG 1-palmitoyl-2-oleoyl-sn-glycero-3-[phospho-rac-(1-glycerol)] (sodium salt) membranes. The difference in  $^{15}\text{N}$  chemical shifts of amides of the residues of MPER in bound and unbound state suggests conformational change after binding to antibodies. (28,129) Additionally, NMR cross-saturation transfer experiments, in which magnetization is transferred from  $^1\text{H}$  of methyl groups of the bound antibodies to the amides of residues of perdeuterated MPER, further confirm the conformational change by  $^{13}\text{C}$  chemical shift of the residues W672, F673, N674, I675, T676, N671, N677, L679, Y681, I682, and K683 of MPER in contact with the antibodies (28,30). The kink between W672-K683 changes to a continuous helical structure upon binding to antibody 4E10 as observed in the crystal structure of NHR-CHR-MPER (29) (residues 547-575GGGGS630-675). There appears to be a change from kinked to non-kinked MPER with antibody-binding. This change of confirmation in the MPER may be relevant to the inhibition of fusion discussed earlier.



**Figure 1.5** Models of MPER and HR. (A) MPER662-683 in DPC micelles at pH 6.6 showing a L-shape kink between two  $\alpha$ -helices (28); (B) Crystallography result of NHR547–575-CHR630-662-MPER663-675 monomer (29). The CHR and MPER form a continuous helix and the MPER in this structure ends at I675.

The viral life cycle is comprised of two stages – (1) early stage comprise of viral attachment, transfer of viral genetic material ribonucleic acid (RNA) to the host cell, reverse transcription of viral RNA to viral deoxyribonucleic acid (DNA) in the host cell cytoplasm, insertion of viral DNA to host nucleus, and integration of viral DNA to host cell DNA in the host nucleus and (2) late stage is comprised of transcription, translation, assembly of viral proteins and genetic material, budding, maturation, and new virus formation or syncytia formation with neighboring host cells (32). Enfuvirtide (fusion inhibitor drug) is a peptide mimic of the CHR+MPER domain that copies much of the amino acid sequence. It is proposed that enfuvirtide binds to NHR in the early steps of gp41 conformational change during membrane fusion (33).

It is thought that HIV enters the host cell through several separate but cooperative steps including attachment, co-receptor binding, and fusion. (1-3) HIV predominantly



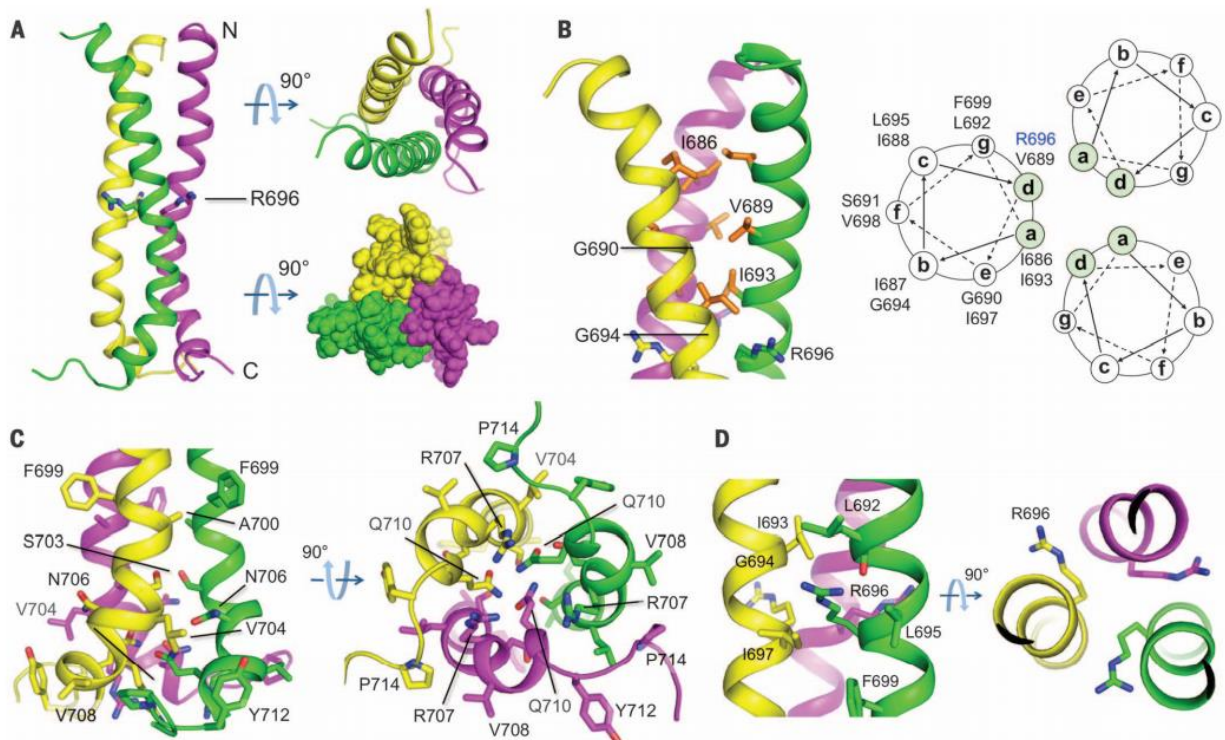
infects T cells carrying the CD4 antigen through an initial association of the viral envelope glycoprotein gp120 with the CD4 receptor on the host cell. After this initial attachment, a conformational change is believed to occur in gp120 that allows its further association with host-cell chemokine co-receptors CCR5 and CXCR4. Subsequently, a conformational change in the second viral envelope glycoprotein gp41 allows it to insert the hydrophobic N terminus into the host-cell membrane. The CHR domain of gp41 then folds back on itself and associates with the NHR domain. (1) This process (known as gp41 zipping) leads to fusion of the viral and host-cell membranes and infection of the cell. However, in the presence of a fusion inhibitor, such as enfuvirtide, an association between the fusion inhibitor and gp41 prevents the successful completion of gp41 zipping, thereby blocking infection (33). The membrane fusion model is explained in greater detail in **section 1.3**.

The peptide amino acid sequence of the enfuvirtide fusion inhibitor drug is YTSLIHSLIEESQNQQEKNEQELLELDKWASLWNWF. The binding of the broadly neutralizing antibodies 2F5, 4E10, and Z13 to the MPER epitopes reduces the gp160 fusion activity (25,34,35). Therefore, binding of the antibodies or peptide enfuvirtide inhibits the fusion of viral and host cell membrane, thus blocking the very first step of viral life cycle. Studies of MPER and membrane interaction revealed membrane permeability and fusion (36-38). FP and MPER peptides interact and act synergistically to enhance membrane fusion (35,36,39,40,76).

#### 1.1.4. Transmembrane Domain

The TM domain anchors the gp41 into the viral membrane and comprises of 22 amino acids (684-706). According to computer simulation data, it acquires an  $\alpha$ -helical structure (41) and tends to oligomerize into trimers (42). Substitution of the TM domain with C-terminal glycosyl- phosphatidylinositol lipid anchor reduces the fusion activity considerably (44,45). To the contrary, another study of fusion activity and infectivity by complete substitution of gp41 TM by cytoplasmic 180 residues at the C—terminal from cellular protein cluster of differentiation-22 (CD22) showed no change in function (46).

Recent NMR structural data (77) of a TM construct (677-716) shows a tightly assembled trimer  $\sim 54$  Å long, with the conserved arginine (R696) near its midpoint. The structure shows a packing arrangement not seen in any other known TM helix dimers or trimers. The 677-716 N- and C-terminal halves have different modes of assembly, with an intervening kink (**Figure 1.6**) (77). The N-terminal region is a conventional three-chain coiled-coil formed by residues 686 to 696. The C-terminal is held together by a network of polar contacts, mainly involving R707 and Q710, at the trimer interface of the kinked helical segments (residues 704 to 712). R696, near the middle of each TM helix, produces three unbalanced charges at the center of the membrane. It points toward the threefold axis of the trimer, while the rest of the side chain bends away from the axis. The backbone carbonyl of L692 may form a hydrogen bond with one of the guanidinium NH<sub>2</sub> group of R696. Other groups have suggested that rather than have a threefold axis of symmetry, the TM of gp41 is monomeric (78).



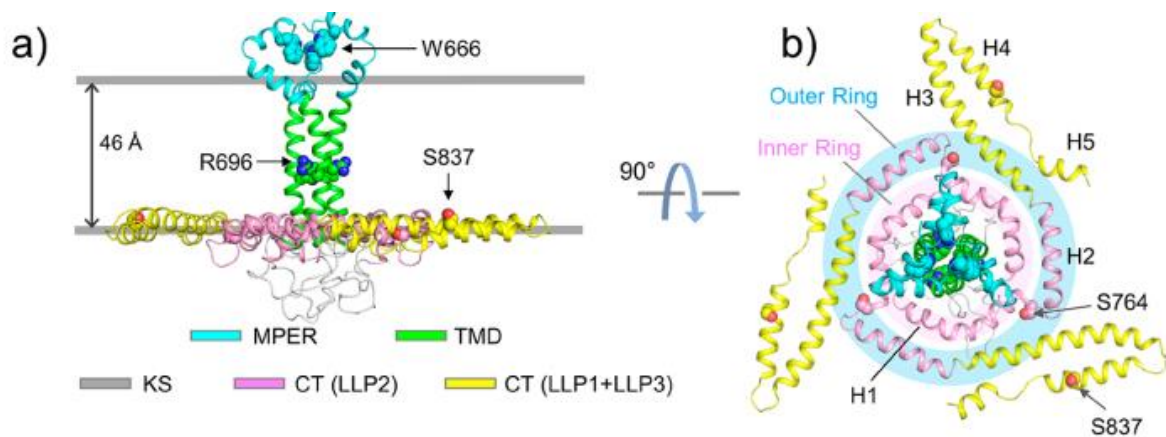
**Figure 1.6** NMR structure of the gp41 HIV-1 (677-716) trimer in DMPC: DHPC 1: 2

bicelles (77). (A) Ribbon representation of the lowest-energy structure from the calculated trimer. N-terminal residues begin at the top. (B) The N-terminal half of the structure with hydrophobic residues (orange) arranged in the coiled-coil pattern (right panel). Hydrophobic interaction causes the threefold axis of the trimer. (C) The C-terminal half of the structure showing an array of polar residues that form the C-terminal hydrophilic core. The network of polar contacts is hypothesized to stabilize the trimer. (D) Enlarged middle region of the structure showing the intra-membrane R696 and its surrounding hydrophobic residues, as well as the backbone oxygen of L692.

### 1.1.5. Cytoplasmic Tail

The gp41 CT is the C-terminal region after the TM domain and comprise of ~150 amino acids. The CT forms a large trimeric baseplate around the TM region, this baseplate (**Figure 1.7**) supports the MPER-TM structure as it spans the membrane (79).

According to the computational modeled structure, gp41 Endo is a single-pass membrane protein (41). However, another monoclonal antibody binding study shows that CT has at least one membrane spanning domain possibly exposing the epitope on the surface of the virion (47). The study of HIV-1 particle entry during virus-cell fusion by fluorescence assay showed immature virus particles were less active than mature virus particles. The inactivity of the immature virus particles is attributed to interaction of CT with matrix protein, and only dissociation of CT from matrix protein and nucleocapsid protein leads to gp41-mediated fusion (48). Functional studies of the CT by mutation reveal that it plays a significant role in incorporation of gp160 into the viral membrane (49,50). Interaction of cytoplasmic endodomain with viral matrix protein has been suggested by mutagenesis of the CT (51,52).



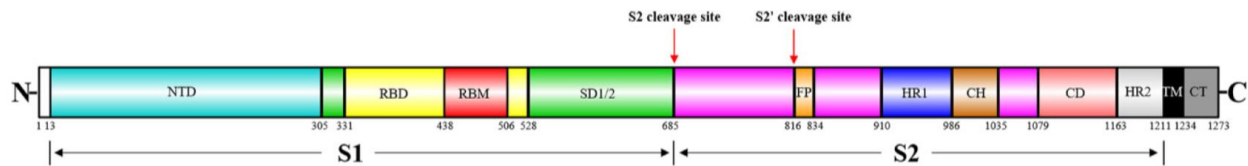
**Figure 1.7** Structure of the entire membrane region of the HIV-1 Env, including CT (79).

(a) Structural model of the MPER-TMD-CT trimer bicelles derived from integrated NMR. (b) Top view (from the MPER) of the trimeric complex showing the inner and outer rings of the baseplate, shaded in pink and blue, respectively.

## 1.2. SARS-CoV-2

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is a zoonotic virus that is the pathogen of the COVID-19 pandemic (81,82). SARS-CoV-2 is enveloped by a membrane that is obtained during budding from an infected host cell. Infection of a new cell requires fusion of the virus membrane with a membrane of the target cell and subsequent deposition of the viral nucleocapsid in the cytoplasm (82,83). This process is catalyzed by the Spike (S) protein subunit 2 (S2). There is not homology between fusion proteins of different virus families. However, we hypothesize that methods used in producing and characterizing gp41 will apply to S2 (76). There is a single gene that codes for the S protein subunit 1 (S1) and S2 with respective residues 1-685 and 686-1273. The S1 subunit and the large S2 ectodomain (residues 687-1207) are outside the virus, followed by the viral transmembrane domain (TM, 1208-1234) and then endodomain (CT, 1235-1273) inside the virus (**Figure 1.8**). The ectodomains of three S2s forms a trimer with three bound S1s (84). Target cells are identified by binding between S1 and the extracellular domain of angiotensin-converting enzyme 2 (ACE2). If there has also been proteolytic cleavage between S1 and S2, the S1 domains separate from the S2 ectodomains, followed by a large structural change of the ectodomains to form a final hairpin structure (**Figure 1.9**). The cleavage and ectodomain structural change are required for fusion, but there is not clear data about the relative timings of the ectodomain vs. membrane topological changes during fusion. (125) For initial infection of respiratory epithelial cells by SARS-CoV-2, cellular proteases may do S1/S2 cleavage so that fusion occurs with the plasma membrane (85). For systemic infection of cells in other tissues, there may be endocytosis after S1/ACE2 binding followed by endosome maturation with

reduction of pH < 6, activation of cathepsin L proteases at the low pH, S1/S2 cleavage (93), and then fusion with the endosome membrane. Both fusion pathways result in deposition of the nucleocapsid in the cytoplasm. The initial trimeric S state and final hairpin structure of the S2 protein are the basis for identifying S2 as a “class I” fusion protein (1).

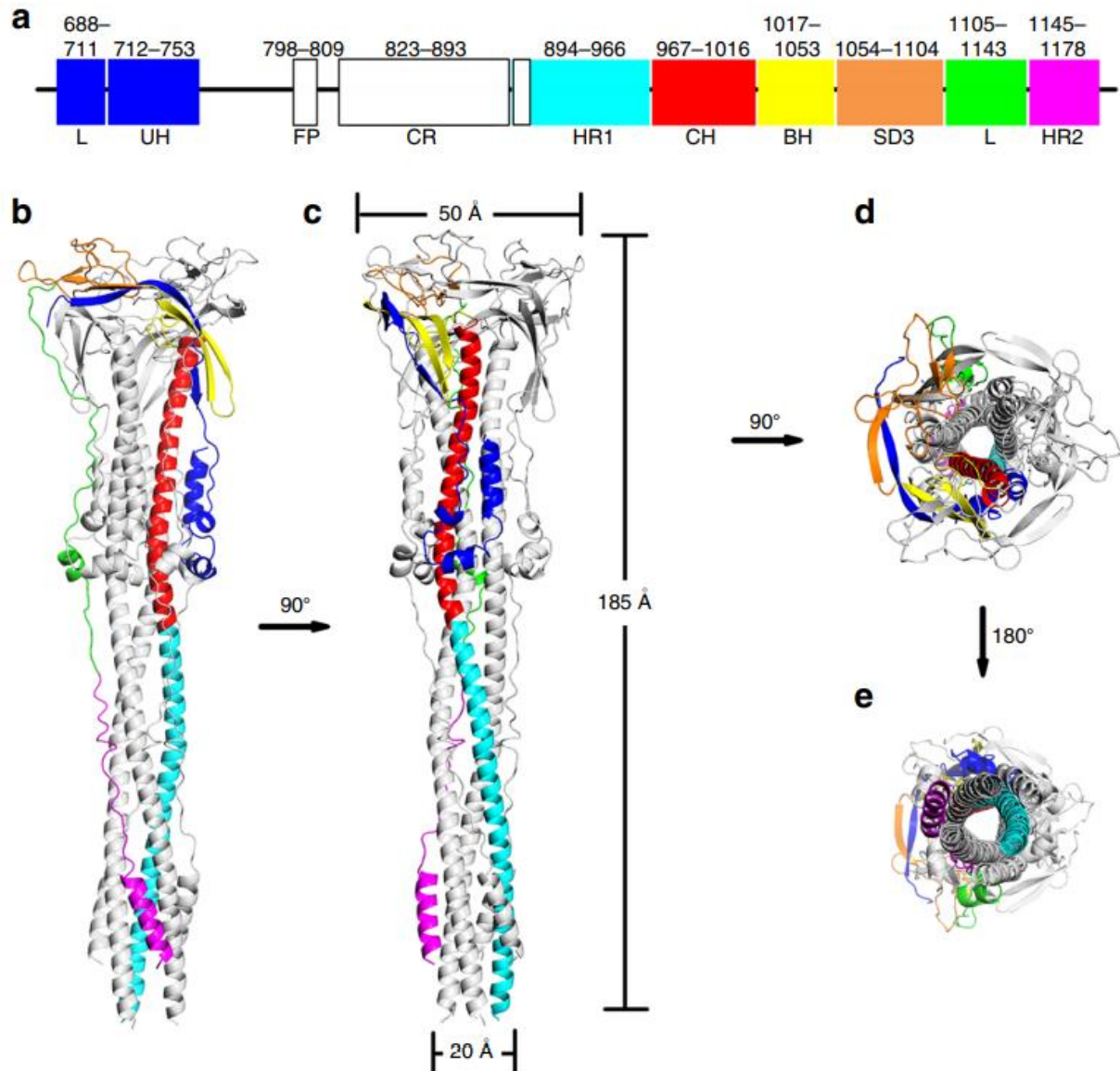


**Figure 1.8** Schematic representation of the domain arrangement of SARS CoV-2 S

protein. The S1 dissociates from the S2 and is involved in receptor binding. The S1 contains the NTD (N-terminal domain), RBD (receptor-binding domain), RBM (receptor-binding motif), and SD1/2 (subdomain 1 and 2). The S2 is involved in membrane fusion and contains the FP (fusion peptide), HR1 (heptad repeat 1), CH (central helix), CD (connector domain), HR2 (heptad repeat 2), transmembrane domain (TM), and the cytoplasmic tail (CT).

HIV gp41 and influenza hemagglutinin (Ha2) are other class I fusion proteins, although there is not sequence homology among the proteins (1). For Gp41 and Ha2, the first ~20 of the ~30 residues N-terminal of the hairpin structure are considered a “fusion peptide” (FP) domain that is thought to bind to the target membrane early in the fusion process. Subsequent formation of ectodomain hairpin structure brings the virus and target membranes in close apposition. The FP binding and hairpin formation could catalyze fusion, as there is a calculated energy barrier of ~25 kcal/mole for apposition (90). For SARS-CoV-2 S2, there are ~225 residues N-terminal of the hairpin structure and at least five distinct ~20-residue segments have been proposed as a FP (92,118). The epitopes

of many neutralizing antibodies may be in the S2 region that is N-terminal of the hairpin, as this was observed for antibodies from convalescent patients of the 2002-2004 SARS epidemic. The S2 sequence of SARS-CoV-2 has ~90% sequence identity with S2 of the severe acute respiratory syndrome 1 (SARS1) viral pathogen of the earlier epidemic. Antibody binding to an epitope of the constructs of the present study could support fusion inhibition as the basis for neutralization whereas binding to the epitope in the initial S1/S2 complex could support other reasons for neutralization like prevention of S1/S2 separation (86,105-109).



**Figure 1.9** A Schematic diagram (a) of the SARS-CoV S2 subunit ectodomain (125).

The uncolored region represents the undetermined region in the structure. There is a linker region (L), Upstream helices (UH), FP, connecting region (CR), heptad repeats 1 and 2 (HR1, HR2), central helix (CH),  $\beta$ -hairpin (BH), subdomain 3 (SD3). Structures (b) and (c) are the post-fusion six helix bundle.



### **Figure 1.9 (cont'd)**

Structures (d) and (e) are the top view of the protein. Numbering here is the SARS CoV-1 numbering which is highly conserved with SARS CoV-2. In SARS-CoV S2 the FP begins at 798 and in SARS CoV-2 S2 the FP begins at 816.

#### **1.2.1. Fusion Peptide**

The FP is a short segment of 15–20 conserved amino acids of the viral family, composed mainly of hydrophobic residues, which anchor to the target membrane when the S protein adopts the pre-hairpin conformation. Pre-hairpin describes the extended ectodomain state before fusion occurs (1). Previous research has shown that the FP of the SARS-CoV-2 S2 plays an essential role in mediating membrane fusion by disrupting and connecting lipid bilayers of the host cell membrane (84,85,90).

From the perspective of membrane fusion, viral fusion peptides are perhaps the most critical region of virus envelope glycoproteins as they directly bind and disrupt target host cell membranes. Thus, the corresponding regions have been extensively studied in viral fusion proteins. While there are no strict definitions for what constitutes a viral fusion peptide, several criteria must be met to designate a segment of viral fusion protein a fusion peptide (1). Often, the region corresponding to an FP is composed of hydrophobic amino acids, particularly enriched in glycine (G) and alanine (A) residues. It is important to note that fusion peptides can contain a few charged residues as well as bulky hydrophobic residues such as tryptophan's (W) (1). Another feature of fusion peptides is that they correspond to regions which are extremely sensitive to point mutations in the context of the full-length fusion protein. A single residue substitution within an FP often results in loss of fusion activity (1). Although viral fusion protein

sequences tend to vary greatly amongst different viral families, they are extremely well conserved, with typically >90% sequence identity within a given family (1,120). The capacity of a synthetic peptide to induce lipid-mixing and fusion of liposomes is another criterion that can help in identifying regions hypothesized to be FPs (20). Furthermore, the study of the effects of such peptides on lipid bilayer ordering using powerful spectroscopy techniques such as electron spin resonance (ESR) has also been successfully used to characterize other cellular and viral FPs such as those of influenza virus and HIV (29,127-129).

Within each major class of viral fusion proteins (I, II, and III), the associated fusion peptides share basic characteristics (1). Class I fusion peptides are usually, but not always, enriched in alanine and/or glycine residues. They can either be N-terminal, located immediately downstream of the activating cleavage site) or internal depending on their positions relative to the cleavage site. For class II and III fusion proteins, the fusion peptides are not released by proteolytic cleavage, but instead form so-called internal “fusion loops” which can be bipartite as in the case of class III fusion proteins (1).

Identifying and locating the coronavirus fusion peptide has been most extensively performed using SARS-CoV spike as a model (92). Coronaviruses (CoV) are a large family of viruses that cause illness ranging from the common cold to more severe diseases such as Middle East Respiratory Syndrome (MERS-CoV) and severe acute respiratory syndrome (SARS-CoV). A novel coronavirus (nCoV) is a new strain that has not been previously identified in humans (82). Initial work, using a peptide library derived from the SARS-CoV S protein, showed that three regions (R1, R2, and R3)

within the S2 domain displayed membrane-interacting properties in experiments measuring vesicle membrane leakage. R1 (876–904) was determined to be upstream of HR1, R2 was located between heptad repeat 1 (HR1) and heptad repeat 2 (HR2) (1095–1110) and R3 (1208–1220) was found to be situated proximal to the transmembrane region (125). Mutations within the region spanning residues 870–901, which approximately corresponds to the above-mentioned R1 region, were found to decrease cell-cell fusion in S-expression mediated syncytia formation assays (102).

These initial studies have identified regions of S2 with the capacity to interact with membranes (103,104). They also suggest that several regions within S2 could act in a concerted fashion to mediate the membrane fusion process. Further studies have been conducted that take a more functional approach and consider other criteria that define fusion peptides, such as proximity to a cleavage site and sequence conservation (113-115). The starting point for both studies was the identification of a second cleavage site in the SARS-CoV S S2 domain called S2' (84). The SARS-CoV S protein is proteolytically activated sequentially at the S1/S2 and S2' sites. The segment located immediately downstream of the SARS-CoV S2' cleavage site, <sup>816</sup>SFIEDLLFNKVT<sub>LADAGF</sub><sup>833</sup>, displayed features of a fusion peptide. In the SARS-CoV S2 the FP begins at 798 and in SARS CoV-2 S2 the FP is speculated to begin at 816 (84). We will refer to the S2 protein using the SARS CoV-2 numbering unless otherwise noted. The highly conserved <sup>821</sup>LLF<sub>823</sub> correspond to the beginning of a major antigenic determinant of SARS-CoV S protein (821–846), capable of eliciting the generation of neutralizing antibodies. Mutagenesis analysis of the S2'-proximal 816–833 segment in the context of full-length protein demonstrated its importance in mediating membrane fusion.

Subsequent structural analysis of the isolated peptide by circular dichroism (CD) spectral analysis and vesicle lipid mixing assays confirmed the role of the segment as a fusion peptide (102).

### **1.2.2. Heptad Repeat Domains HR1 and HR2**

Insertion of the fusion peptide in the target membrane has been shown for the influenza virus fusion peptide (121). It is hypothesized that other fusion peptides from various viral families function similarly. After insertion of the viral fusion peptide into the target cellular membrane, the more N terminal heptad repeat region (HR1) of S2 folds into a trimeric helical coiled-coil structure. It is thought that the C-terminal heptad repeat region (HR2) dissociates into monomers and packs against the grooves of the HR1 trimer in an antiparallel manner. Prior to dissociation, the HR1 is part of the extended spike protein trimer. This packing is driven by hydrophobic interactions between HR1 and HR2 and forms the viral fusion core, also known as the six-helix bundle (94). The formation of the fusion core brings the viral and target membranes together so membrane fusion can commence. Thus, disrupting fusion core formation would seem to be a valid anti-fusion strategy. This process is discussed further in **section 1.3**.

Peptides derived from the HIV-1 gp41 HR2 (107) region could bind to the HR1 region in the pre-intermediate state, preventing the gp41 HR2 from binding and forming the fusion core. By competitively binding to the HR1 region, these HR2 peptides can potentially inhibit HIV-1 infection at nanomolar concentrations. One of these peptides, DP-178, was tested in proof of principle clinical studies and it was demonstrated that patients receiving peptide treatment displayed viral load reduction (110,111). Further clinical studies established the long-term safety and efficiency of DP-178 and eventually it was

approved by the US Food and Drug Administration for HIV/AIDS treatment and named Fuzeon (enfuvirtide), becoming the first fusion inhibitor drug (32,33). In addition, a previously developed pan-coronavirus fusion inhibitor (EK1) against HR1 in SARS spike to inhibit membrane fusion was also found to inhibit membrane fusion during infection by SARS CoV-2 and MERS-CoV (88). CoV specific HR peptides are required for inhibiting CoV fusion (110,111). Circular-dichroism (CD), a technique that provides information on the fraction of structured and disordered content of a peptide, showed that HR2 peptides bound to viral HR1 regions have more alpha-helical content than only HR2, suggesting that HR2 peptides form a stable structure when bound to HR1 (**Figure 1.10**) (125). These results suggest that HR2 peptides can compete with the virus fusion protein's own HR2 region to bind HR1 and prevent fusion. A similar tale can be told for MERS-CoV HR2-based peptides. MERS-CoV fusion core formation has been investigated and found that an analogous HR2 peptide inhibited infection. CD experiments indicated that the HR2 peptide binds to HR1 (109,110).



### **1.2.3. Transmembrane Domain and Cytoplasmic Tail**

The transmembrane (TM) domain of the S protein is known to be highly conserved in SARS-CoV-2 and its close relatives (118). This conservation is also reflected in the hydrophobicity profile among SARS-CoV-2 and its close relatives (1). The TM domain consists of the following three parts: a juxtamembrane aromatic part, a central hydrophobic part, and a cysteine-rich part. It is followed by a highly hydrophilic cytoplasmic tail (CT) which anchors the spike inside the viral membrane (83,84).

The tryptophan residues in the aromatic part are strongly conserved among SARS CoV-2 and related coronaviruses, suggesting their functional importance. Replacing them even by another aromatic residue such as phenylalanine will severely impact the efficiency of viral infection (1,118). However, this finding was not supported in another study in which replacing tryptophan by phenylalanine was tolerated (118). The central hydrophobic part of the TM forms a helix. Because S proteins form homotrimers, there are three transmembrane helices interacting with each other. The TM and the C-terminus contribute to the stabilization of the trimeric structure which is important for membrane fusion (118). Destabilization of the trimeric structure is associated with reduced fusogenicity and infectivity. Replacing hydrophobic residues in the central part by hydrophilic ones such as lysine decreases the efficiency of an infection (118). Cysteine residues immediately proximal to the membrane (1235-1243) are palmitoylated (122); replacing them by other amino acids inhibits membrane fusion. In contrast, replacing cysteine residues in the last half of the cysteine-rich part does not inhibit membrane fusion (1247-1254). During the cell-to-cell infection stage, the membrane-proximal cysteine-rich part, and the cytoplasmic tail anchor the C-terminus of S inside the infected cell, while the

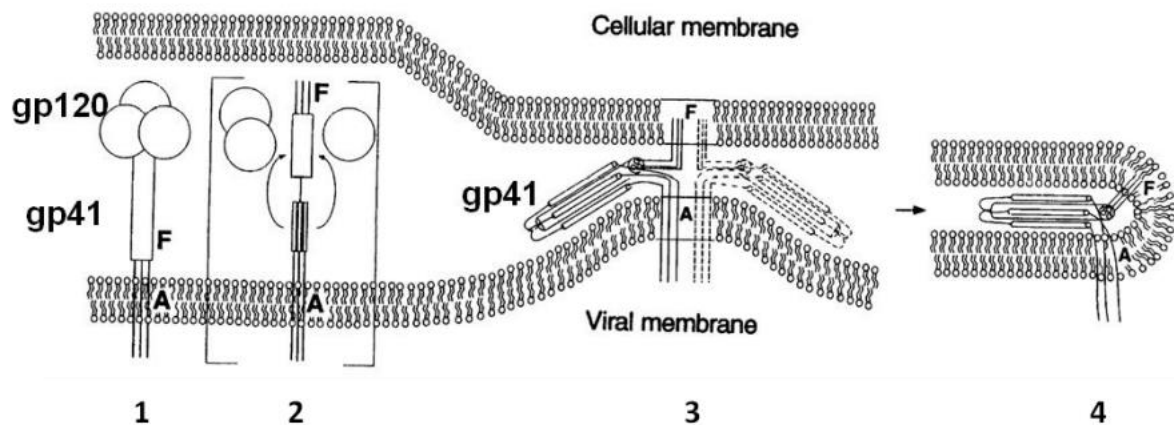
N-terminal of the S2 penetrates the target cell membrane to anchor the N-terminus inside, which is typical of viral fusion proteins. The conformational changes of S2 helps to bring the membranes of infected and target cells close together to facilitate cell-cell membrane fusion and viral entry. The anchor provided by the cysteine-rich part and CT is enhanced by the membrane-actin linker which, upon phosphorylation, links specific transmembrane proteins such as the S homotrimer to actin to reinforce the anchor inside the cell (123). This occurs through an unknown mechanism (131).

### **1.3. Possible Mechanism of Membrane Fusion**

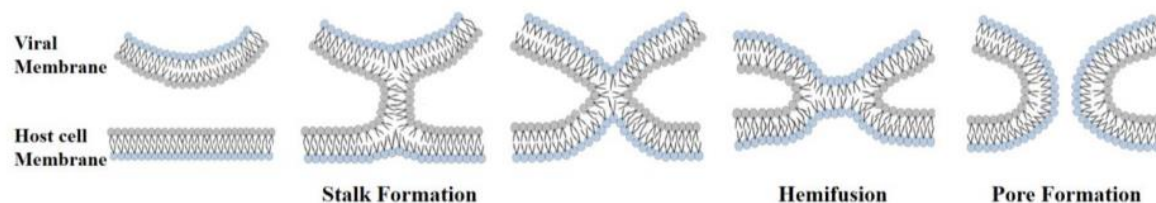
Fusion of two membranes is a crucial step for viral entry into host cells. Membrane fusion proceeds through the following steps: (1) contact between the outer leaflet membrane surface, (2) stalk formation, (3) hemifusion – where the outer leaflets of the two membranes combine to form a single membrane first, and the distal membrane leaflets remain separate until the opening of a fusion pore (**Figure 1.12**), and (4) pore formation (**Figure 1.11, Figure 1.12**). The whole process of membrane fusion occurs isothermally, and X-ray measurements have estimated the free energy minimum distance between biologically relevant membranes to be between 2 and 3 nm (53). Membranes with no net charge on the lipid molecules are at equilibrium due to hydrophobic interactions and repulsive interactions (short range). When the membranes are closer than 3 nm, energy is required to overcome the repulsive force between the outer leaflets of the two membranes (54). In this membrane environment, water molecules tend to bind to the polar head groups of the lipid molecules by hydrogen bonding. For membrane fusion between viral and host cells to occur, energy is required to displace the bound water molecules. This energy barrier is referred to as the desolvation energy. The



desolvation energy is incorporated into the activation energy required to form the stalk, which is primarily composed of lipid bilayer. The distal membrane leaflet does not undergo lipid mixing during stalk formation, though in the next stage the lipid molecules of the proximal membrane leaflet rearrange (1). The stalk formation is followed by a hemifusion stage (54,55). There is not significant exposure to water during the structural transition from stalk to hemifusion (1). The activation energy of the stalk formation is higher than that of hemifusion formation because more energy is required for the desolvation penalty to be paid during stalk formation relative to the curvature energy (Gaussian curvature elastic energy) of hemifusion formation (124). Hemifusion formation is followed by pore formation (**Figure 1.12**).



**Figure 1.11** Schematic diagram of Model 1. (1) Trimer gp120 and gp41 at pre-fusion state; (2) displacement of gp120 and PHS formation; (3) SHB formation; (4) gp41 at the SHB post-fusion state following pore formation. 'A' represents the transmembrane domain and 'F' represent the FP domain (1).



**Figure 1.12** Stages of membrane fusion.

Stalk and hemifusion are two intermediates in one model of membrane fusion, and as such they have defined structures and occur at local free energy minima. It could be that the energy required for all the stages of membrane fusion are being supplied by the energy released during the conformational changes in the viral membrane protein. To date there is no experimental evidence of this energy exchange between membranes and gp41 protein folding, though it has been proposed that the folding of gp41 NHR-CHR ectodomain is an exothermic process releasing about 65 kcal/mol of enthalpy and that enthalpy is utilized to overcome the activation energy for stalk formation during membrane fusion (57). However, this hypothesis of membrane fusion caused by protein folding is not consistent with our observation of fusogenicity of already folded hairpin conformation of gp41 ectodomain. The gp41 ectodomain core sequence comprising of NHR-short loop-CHR (HP) forms stable folded hairpin conformation at pH 3.2 and other pH values, and the gp41 with hairpin structure causes vesicle fusion (76).

#### **1.4. Mechanism of Fusion of HIV-1 and Host Cell Membranes Caused by Gp41**

HIV-1 and host cell membrane fuse to enable the transfer of viral genetic material into the host cell and further use the host cell machinery to complete the life cycle and viral replication (59). The Env glycoprotein gp160 plays a significant role in mediating the fusion between the two membranes. Gp160 comprises of non-covalently associated gp120 and gp41. Gp120 has a strong affinity for CD4 receptors and co-receptors

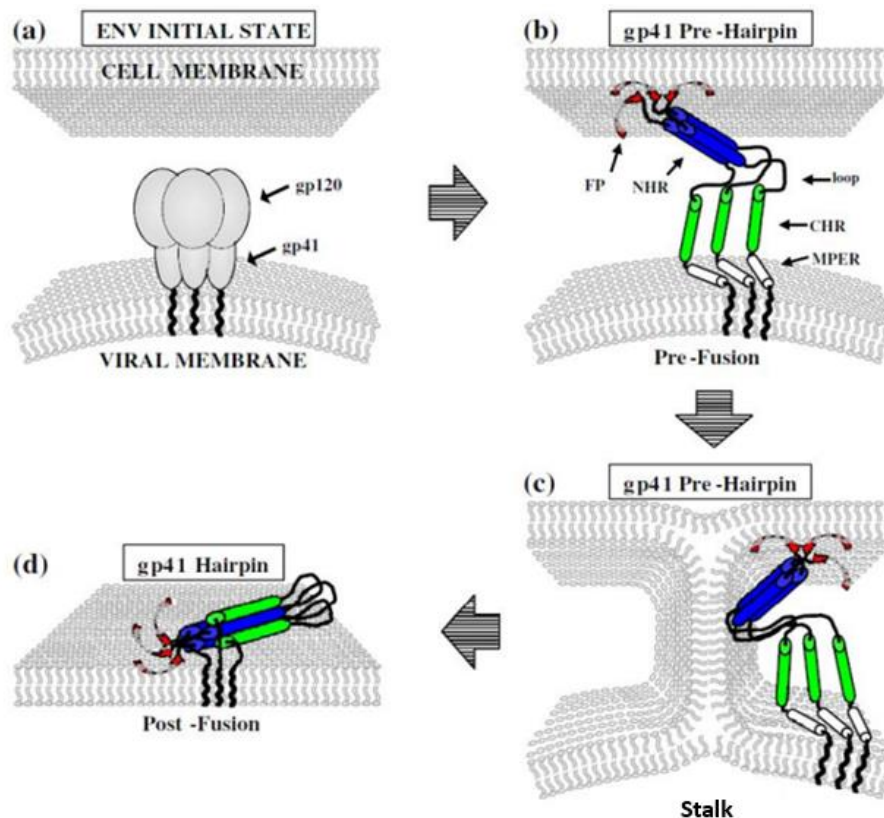
CXCR4/CCR5 (60,1). After binding, gp120 dissociates to fully expose the fusion efficient gp41 (61). A recent cryoelectron tomography study of virion particles with Env gp160 bound to HIV neutralizing proteins showed a conformational change of gp120 still attached to gp41 (62). Another cryoelectron tomography study of conformational change of gp120 upon binding with CD4 receptor and coreceptor ligand showed gp140 (gp120+gp41 ectodomain truncated) intact structure (63). However, this gp140 structure is stabilized by non-native disulfide bonds between gp120 and gp41 residues. Therefore, to date there is no clear structural evidence whether gp120 completely dissociates from gp41. There are two proposed models for the fusion of membrane caused by gp41 (1).

#### **1.4.1. Model I**

After the dissociation of gp120 **Figure 1.11-1**, gp41 undergoes a conformational change to expose the FP towards the host membrane. In this elongated intermediate stage, the gp41s are a trimer bundle **Figure 1.11-2** (1). This is followed by the formation of the NHR and CHR SHB with close apposition of the N-terminal FP and C-terminal TM regions, which consequently brings the two membranes close to each other **Figure 1.11-3** (1). Viral and host cell membrane begin stalk formation, followed by hemifusion, and then complete fusion. The gp41 FP and TM then lie in the same membrane following fusion **Figure 1.11-4** (18). The folding of the gp41 from an intermediate state to the final SHB folded state satisfies the energy requirements of about 25 kcal/mol activation energy for fusion to take place.

#### 1.4.2. Model II

According to this model, after the gp120 moves away from gp41 **Figure 1.13-a**, the gp41 forms an extended pre-hairpin intermediate (PHI) structure and the FP inserts into the host membrane **Figure 1.13-b**. The hemifusion state is followed by fusion pore formation by folding of PHI to SHB **Figure 1.13-d** (65). The experimental evidence supporting the pore formation before SHB formation is based on cell/cell fusion assays where the closing of pores after the addition of NHR or CHR peptides was observed (66). Experimental studies of vesicle fusion assay of 512-582 and (512-581SGGRGG628-667) at physiological pH show rapid and negligible fusion, respectively (67). These data support a model that the most effective fusion is induced by small, exposed FP oligomers.



**Figure 1.13** Schematic diagram of Model 2 (65).

### Figure 1.13 (cont'd)

(a) Trimer gp120 and gp41 at pre-fusion state; (b) PHI formation and FP inserted in host cell membrane; (c) Viral and host cell membrane stalk formation; (d) gp41 at the SHB post-fusion state. FP is shown in red, NHR in blue, CHR in green and MPER in white.

These models are for vesicle fusion assays with and without the FP present

**Figure 1.3.** These data are consistent with a model that initiates membrane fusion following FP insertion into the host cell (65).

#### 1.4.3. Caveats of Both Models

In model I, the sequence of protein folding stages is inconsistent with the experimental observations of fusion inhibition and syncytia formation by binding of NHR or CHR peptides, (66) or CHR+MPER peptide (32). Model I show stalk formation occurring before formation of the SHB. Fusion inhibited gp41 is prevented from forming SHB and model II shows stalk formation prior to SHB formation. In both the models I and II, it has been presumed that the PHI is a trimer structure with NHR forming the core. There is no structural evidence of the PHI state of gp41. It is informative however, to examine the PHI structure of a similar protein.

In the prefusion state of recombinant parainfluenza virus 5 (PIV5), HR1 lies along the protein surface, while HR2 forms a trimeric coiled coil stalk adjacent to the viral membrane (116). In the pre-hairpin intermediate state, HR1 is believed to detach from the protein surface and refold into a long, extended three-stranded coiled coil projecting towards the target bilayer, while HR2 remains essentially intact and anchored near the virus surface (1). During formation of the hairpin, the HR1 trimer remains constant, while the original HR2 bundle dissociates and its individual chains bind along the exterior of the

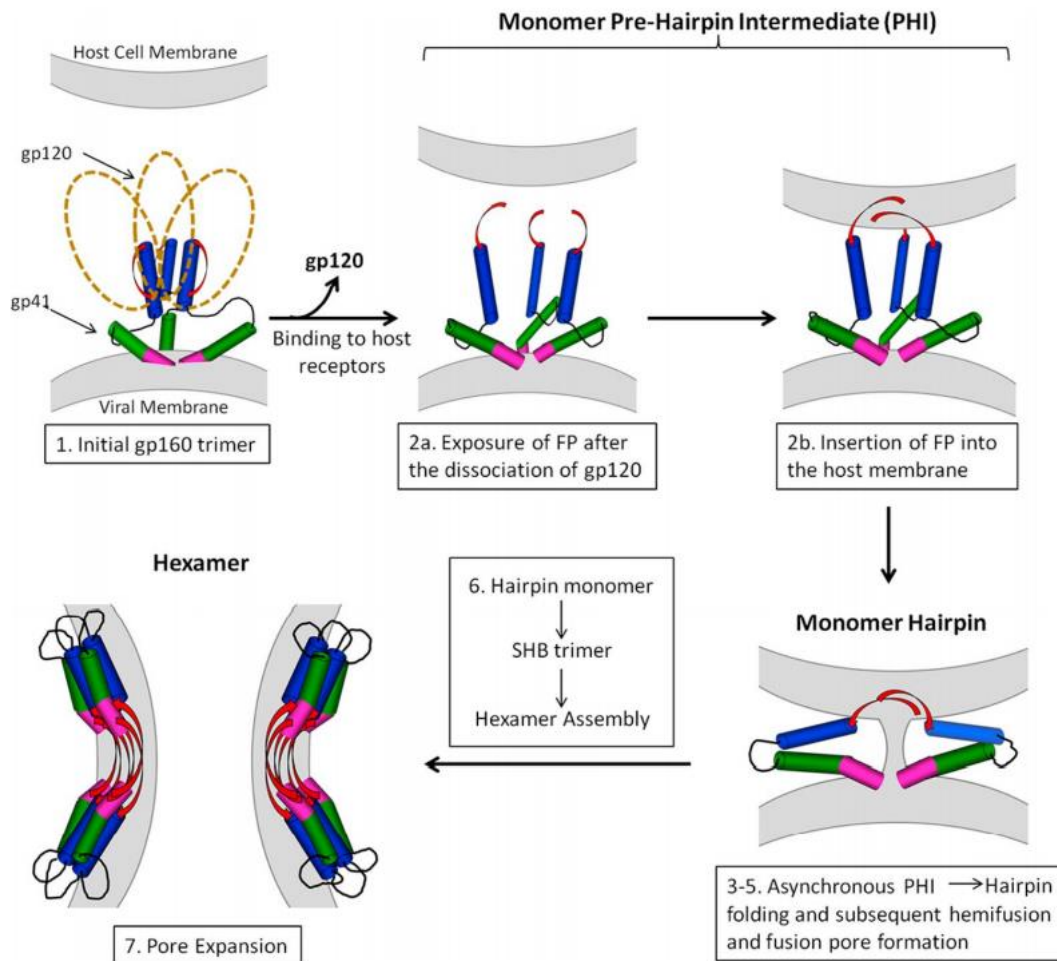
HR1 trimer, pulling the viral membrane toward the target bilayer until the postfusion trimer of helical hairpins is fully formed.

The loop domain (581-628) of HIV-1 gp41 is typically invariant in the pre- and post-fusion structures (1). For our purposes, the loop was truncated and replaced with an SGGRGG linker. The domain appears to form a hinged hub about which HR1 and HR2 pivot during these conformational rearrangements.

It is possible that the PHI could be monomeric ectodomain gp41. In that case the inhibitors would prevent the formation of the SHB by inhibiting formation of trimers. Recent x-ray crystal and cryo-EM structure of gp140 (gp120+gp41 without the FP and MPER) show loosely packed NHR domains **Figure 1.2**. This is presumably the state prior to membrane fusion (68).

#### **1.4.4. Model III**

This model begins with gp120 dissociation and gp41 exposure **Figure 1.14-1**. Following exposure of gp41 there is dissociation of the gp41 ectodomain into monomers and formation of extended PHI, followed by FP insertion into the host cell membrane **Figure 1.14-2** (76). Monomers then fold and bring the two membranes close together **Figure 1.14- (3-5)**. This folding leads to membrane hemifusion, and initial pore formation **Figure 1.14-6**. The hairpin monomers then form SHB trimer which leads to hexamer ectodomain assembly and finally fusion pore expansion **Figure 1.14-7** (76).



**Figure 1.14** Membrane fusion Model 3 of gp41 ectodomain monomer and hexamer (76). The different domains of gp41 are color coded the same as Figure 1.3 and the TM and endodomain are not shown. One of the monomers is not displayed in steps 3–5. The initial gp41 structure of step 1 and the final SHB structure of step 7 are based on high-resolution structures.

In Model 3, the PHI initializes the hemifusion and fusion pore formation through the protein folding transition. The PHI to hairpin folding was also supported by hyperthermostable ectodomain monomer. The stable hexamer is also consistent with the requirement of multiple gp160 trimers for membrane fusion and HIV infection (76).

#### 1.4.5. Stoichiometry of Fusion Proteins at the Site of Membrane Fusion

Structural and Functional studies of gp120 and gp41 have led us to understand that model 3 is the most probable fusion mechanism of viral and host cell membrane fusion. However, each membrane fusion event may require more than one gp160 trimer. Mathematical estimation of number of trimers per virion concludes that there are 5-8 functional trimers per virion that effectively participate in viral entry (72,73). Cryoelectron microscopy tomography study on wild-type HIV-1 virions show about 14 gp160 protein spikes on the viral membranes surface. According to cryo-electron tomography study with virions, the Env trimers associate as a claw like structure when the HIV-1 particles interact with CD-4 positive cells. There are about 5-7 gp160 trimers (“entry claw”) which meet target host cells (75). Following insertion into lipid bilayers, the FP of different gp41 proteins tend to align themselves antiparallel to each other. Recent solid state NMR studies of membrane associated FP in gp41 ectodomain FP-HP show antiparallel sheet arrangement. It is proposed that this antiparallel arrangement is possible by interleaved FP strands of two gp41 trimers (65).

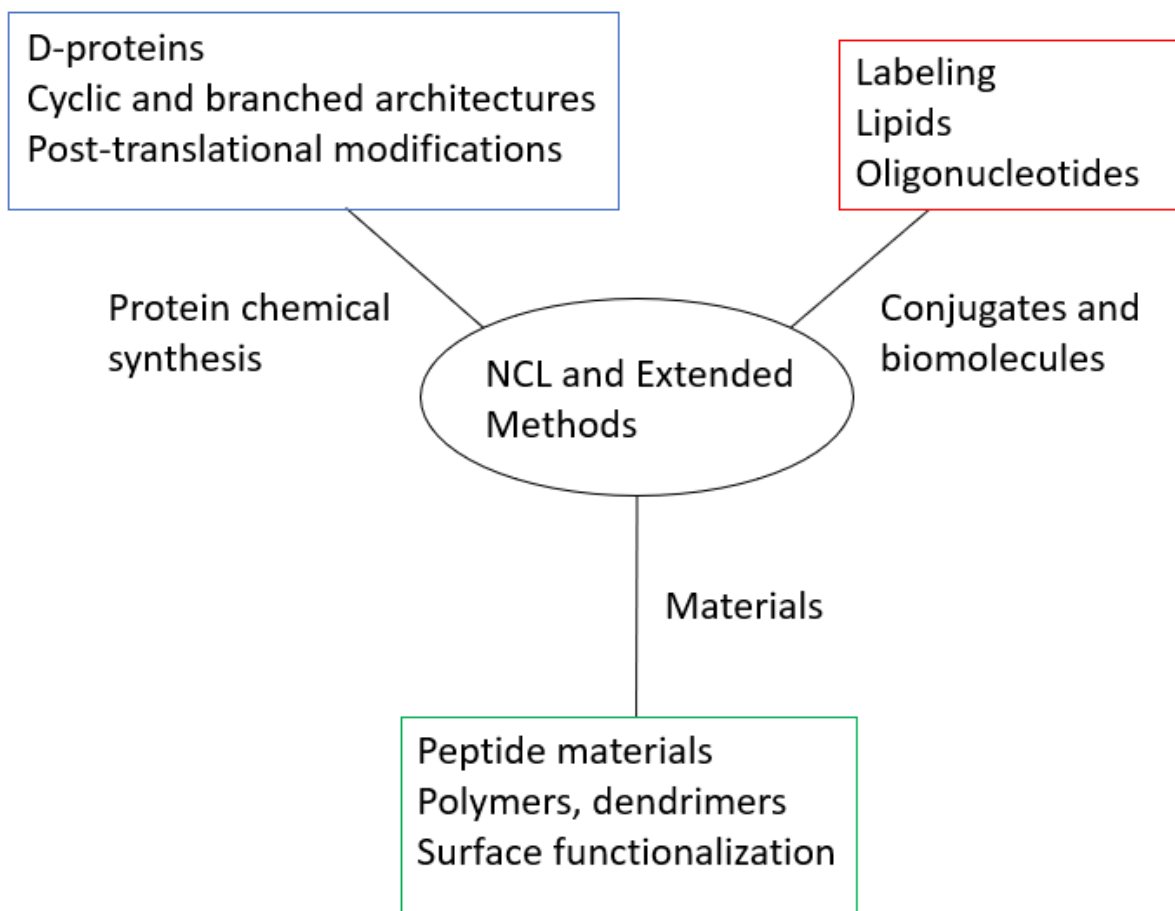
#### 1.5. Native Chemical Ligation

Since 1994, native chemical ligation (NCL) has found wide synthetic applications in chemical biology, medicinal chemistry, and material science. NCL is a chemical ligation reaction in which two small peptides are ligated together using a cysteine amino acid on the N-terminus of one protein and the C-terminus of another protein **Section 2.11** (117). According to the PCS database, over 700 proteins of biological relevance were synthesized using NCL and extended methods (117). Continuous and significant advances in the field have recently culminated with the report of synthetic objects of



exceptional size such as fully functional synthetic analogues of bacterial polymerases (~350 amino acids) (117). Beyond their size, the diversity of synthetic targets also reflects the versatility of chemical approaches based on NCL for accessing proteins that can be produced by living systems only with great difficulty (**Figure 1.15**). Among the latter, cyclic peptides have been the subject of numerous synthetic studies (117). The synthesis of branched architectures can be facilitated by ligation approaches that provide full control over the number of units connected and their attachment site. Other post-translational modifications are also accessible. More recent work has resulted in the emergence of applications involving the preparation of D-proteins, stereoisomer proteins (117). Because D-proteins cannot be produced by living systems, this is a purely synthetic application which best illustrates the usefulness of ligation-based approaches. Beyond the total chemical synthesis of peptides and proteins, the NCL reaction has largely transcended its initial scope of application by enabling straightforward access to conjugates and biomolecules. The attractiveness of NCL as a synthetic method in fact not only lies in its ability to proceed in mild conditions compatible with a wide range of substrates but also stems from a high chemoselectivity allowing for specific (bio)conjugation (117). These properties were used early on to produce membrane-anchored proteins by reacting recombinant protein thioesters with cysteine-containing phospholipids. In the same way, peptide–oligonucleotide conjugates have been obtained. Thioester-based ligation reactions have also proved to be a valuable tool for the incorporation of fluorescent probes in proteins (117). In another application, NCL enabled the selective modification of complex samples for the identification of N-terminally homocysteinyllated peptides by proteomic approaches (117). It is also worth mentioning

that NCL has been used to produce complex lipids and phospholipid bilayers (117). A recent application of this concept to the reconstitution of G protein-coupled receptors (GPCR) by in situ production of the proteoliposomes was described (117). Our work describes a method for producing site-specific isotopically labeled large protein constructs for NMR structural studies.



**Figure 1.15** Synthetic applications of NCL and extended methods.

## **1.6. Introduction of Solid-State NMR Techniques for Membrane Proteins**

Membrane proteins are the proteins that interact with biological membrane, and they have crucial functions such as membrane receptors and enzymes. However, due to the difficulty of growing membrane protein crystals, only a small fraction of membrane proteins has high resolution crystal structures (130). The difficulty for crystallography is due in part to producing enough folded protein for crystallography trials. Cryo-electron microscopy (Cryo-EM) for membrane protein structure generally works better for larger proteins (>100kDa). Solid-state nuclear magnetic resonance (NMR) is a powerful technique to determine high resolution structure and function of biomolecules. Compared to widely used X-ray crystallography and liquid-state NMR, solid-state NMR is specifically suitable for membrane proteins, large proteins, protein aggregates and nucleic acids that cannot be crystallized or that are too large for solution NMR spectroscopy (119). Typically, the limit for protein molecular weight in solution NMR is approximately 35 kDa.

### **1.6.1. Magic-Angle Spinning (MAS)**

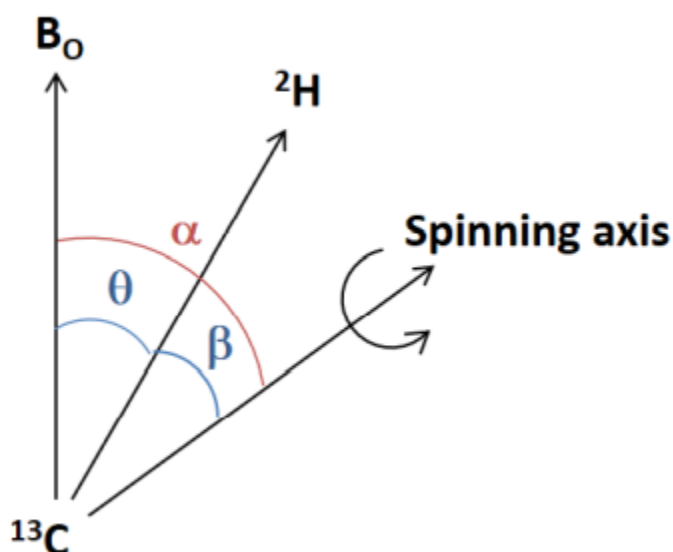
In solution state NMR spectra, anisotropic effects are rarely observed because of the rapid tumbling of the molecules in solution (119). Thus, the orientational dependency of the NMR frequency with respect to the external magnetic field  $B_0$  is rapidly averaged out and results in sharp peaks (119). For solid samples or biomolecules, the tumbling is much slower which results in broad lines in NMR spectra since each of the orientations in the sample contribute to different spectral frequencies. Magic-angle spinning (MAS) (119) is a routinely used technique to achieve high resolution spectra. It can remove the effects of chemical shift anisotropy to assist in the removal of heteronuclear dipolar-

coupling effects. It is also used to narrow lines from quadrupolar nuclei and removing the effects of homonuclear dipolar coupling.

In solid-state NMR MAS experiments, samples are packed in a cylindrical rotor and spun at high speed by an angle  $\alpha$  with respect to  $B_0$ , and  $\beta$  is the angle between the  $^{13}\text{C} - ^2\text{H}$  internuclear vector and spinning axis (**Figure 1.16**). When the sample is spun at  $\alpha = 54.7^\circ$ , then  $\theta$  varies with time as the molecule rotates with the sample.  $\theta$  is the angle between the spin vector and  $B_0$  and is relevant to the dipolar coupling (119). The average of  $(3\cos^2\theta - 1)$  over each rotor period is:

$$[3\cos^2\theta - 1] = \frac{1}{2} (3\cos^2\alpha - 1) (3\cos^2\beta - 1) \text{ (equation 1.1)}$$

MAS set  $\alpha = 54.7^\circ$  so that  $[3\cos^2\theta - 1] = 0$ . Spinning sidebands begin to appear in the spectra when the spinning of the sample is above 3 kHz (119).



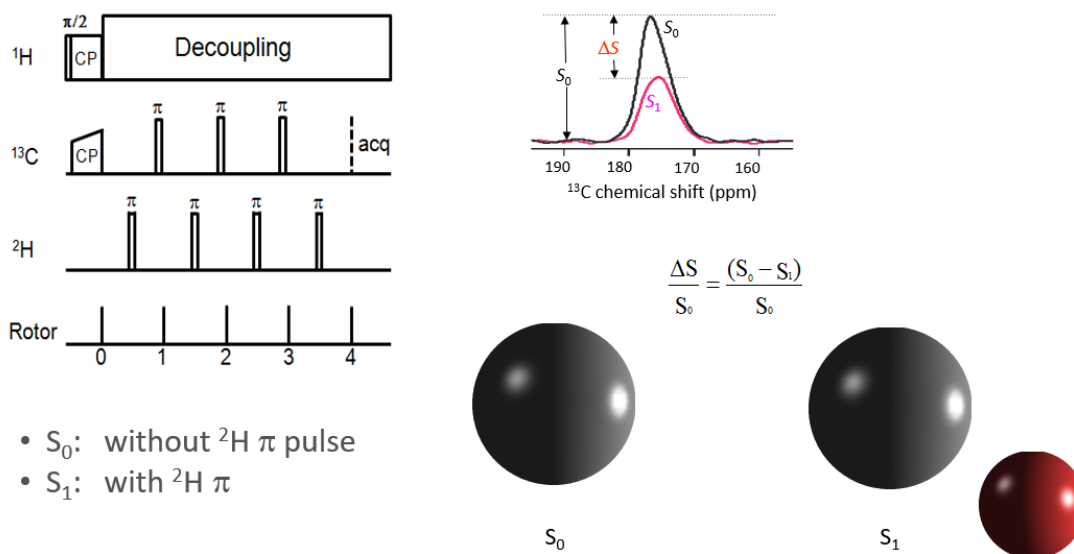
**Figure 1.16** Geometry of the  $^{13}\text{C} - ^2\text{H}$  vector in solid state NMR sample under MAS.

**Figure 1.16 (cont'd)**

The sample is spun rapidly in a cylindrical rotor about a spinning axis oriented at the magic angle ( $\alpha = 54.7^\circ$ ) with respect to external magnetic field  $B_0$ .

### 1.6.2. $^{13}\text{C}$ - $^2\text{H}$ Rotational-Echo Double-Resonance NMR (REDOR)

Rotational echo double resonance (REDOR) was developed by Gullion, and Schafer and it is widely used MAS NMR techniques for studying molecular structure in solid-state (**Figure 1.17**) (119). As discussed previously, MAS can average out the heteronuclear dipolar interactions. In REDOR experiments, the heteronuclear dipolar interactions can be recovered by using simple rotor-synchronized  $\pi$  pulses (119). Since the dipolar interaction is inversely proportional to the cube of the internuclear distance, the distances can be easily obtained. Another advantage of REDOR is the simplicity of its pulse sequence and data analysis.



**Figure 1.17**  $^{13}\text{C}$ - $^2\text{H}$  REDOR NMR pulse sequence.

### Figure 1.17 (cont'd)

The columns represent the  $\pi/2$  or  $\pi$  pulses. CP = cross polarization that transfers  $^1\text{H}$  transverse magnetization to  $^{13}\text{C}$  and can enhance the  $^{13}\text{C}$  signal. The CP is followed by a  $^{13}\text{C}$ - $^2\text{H}$  dipolar evolution for a period which is called dephasing time ( $\tau$ ). Adjacent  $^{13}\text{C}$   $\pi$  pulses are separated by one rotor period as are adjacent  $^2\text{H}$   $\pi$  pulses.  $^{13}\text{C}$  is the detecting channel. A typical REDOR spectra is displayed in the upper right.  $S_0$  is best thought of as just the  $^{13}\text{C}$  spectra whereas  $S_1$  is the  $^{13}\text{C}$  spectra under influence of proximal deuterium atoms.

$^{13}\text{C}$ - $^2\text{H}$  REDOR is a three-channel experiment. At the beginning of the sequence, a  $\pi/2$  pulse is applied to rotate the  $^1\text{H}$  magnetization from the  $B_0$  direction to the transverse plane. Pulses are applied on the  $^1\text{H}$  and  $^{13}\text{C}$  channels. Then a  $^1\text{H}$ - $^{13}\text{C}$  cross polarization (CP) pulse sequence is applied to transfers  $^1\text{H}$  transverse magnetization to  $^{13}\text{C}$  nucleus and to enhance the  $^{13}\text{C}$  signal (119). Due to the various orientation of molecules, chemical bonds, internuclear vectors with respect to  $B_0$ , and the nuclei in a powder sample have a distribution of Larmor frequencies and the resonance offset fields  $(\omega_L - \omega)/\gamma$  where  $\omega_L$  is the Larmor frequency,  $\omega$  is the angular frequency with respect to  $B_0$ , and  $\gamma$  is the gyromagnetic ratio. Thus, a ramp CP on the  $^{13}\text{C}$  channel is used to increase the efficiency of the magnetization transfer (119).

After the  $^1\text{H}$ - $^{13}\text{C}$  CP, there is a dephasing period ( $\tau$ ) during which a series of  $^{13}\text{C}$  and  $^2\text{H}$   $\pi$  pulses and a  $^1\text{H}$  decoupling field are applied. For each  $\tau$ , two separate spectra are collected, denoted as  $S_0$  and  $S_1$ . In the  $S_0$  experiment, only  $^{13}\text{C}$   $\pi$  pulses are applied at the end of each rotor period except at the end of  $\tau$ . In the  $S_1$  experiment,  $^{13}\text{C}$   $\pi$  pulses

are applied at the end of each rotor period while  $^2\text{H}$   $\pi$  pulses are applied in the middle of each rotor period. This is followed by the  $^{13}\text{C}$  signal acquisition (119).

The  $^{13}\text{C}$ - $^2\text{H}$  dipolar coupling and chemical shift anisotropy are averaged out by MAS. In REDOR experiments, the function of the  $^{13}\text{C}$  and  $^2\text{H}$   $\pi$  pulses are to flip the spin by  $180^\circ$ . The effect of the  $^{13}\text{C}$  pulses in  $S_0$  are to provide a  $^{13}\text{C}$  NMR spectrum while in  $S_1$  the combined  $^{13}\text{C}$  and  $^2\text{H}$  pulses produce the  $^{13}\text{C}$  NMR spectrum under the influence dipolar coupling of nearby  $^2\text{H}$ . This means that the  $S_1$   $^{13}\text{C}$  signal will be reduced depending on the proximity of nearby  $^2\text{H}$  (119).

## REFERENCES

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



13. Yang, R., Prorok, M., Castellino, F. J., and Weliky, D. P. (2004) A Trimeric HIV-1 Fusion Peptide Construct Which Does Not Self-Associate in Aqueous Solution and Which Has 15-Fold Higher Membrane Fusion Rate. *JACS Communications* 126, 14722-14723
14. FREED, E., DELWART, E. L., GARY L. BUCHSCHACHER, J., and PANGANIBAN, A. T. (1992) A mutation in the human immunodeficiency virus type 1 transmembrane glycoprotein gp41 dominantly interferes with fusion and infectivity. *Proc Natl Acad Sci USA* 89, 70-74
15. Pereira, F. B., Goni, F. M., and Nieva, J. L. (1995) Liposome Destabilization Induced by the Hiv-1 Fusion Peptide Effect of a Single Amino-Acid Substitution. *Febs Lett* 362, 243- 246
16. Qiang, W., Sun, Y., and Weliky, D. P. (2009) A strong correlation between fusogenicity and membrane insertion depth of the HIV fusion peptide. *P Natl Acad Sci USA* 106, 15314-15319
17. Chan, D. C., and Kim, P. S. (1998) HIV Entry and Its Inhibition. *Cell* 93, 681–684
18. W, W., A, D., SC, H., JJ, S., and DC, W. (1997) Atomic structure of the ectodomain from HIV-1 gp41. *Nature* 387, 426-430
19. Lu, M., Ji, H., and Shen, S. (1999) Subdomain folding and biological activity of the core structure from human immunodeficiency virus type 1 gp41: implications for viral membrane fusion. *J. Virol.* 73, 4433-4438
20. Sackett, K., Nethercott, M. J., Epand, R. F., Epand, R. M., Kindra, D. R., Shai, Y., and Weliky, D. P. (2010) Comparative analysis of membrane-associated fusion peptide secondary structure and lipid mixing function of HIV gp41 constructs that model the early pre-hairpin intermediate and final hairpin conformations. *J. Mol. Biol.* 397, 301-315
21. Roche, J., Louis, J. M., Grishaev, A., Ying, J. F., and Bax, A. (2014) Dissociation of the trimeric gp41 ectodomain at the lipid-water interface suggests an active role in HIV-1 Env-mediated membrane fusion. *Proc Natl Acad Sci USA* 111, 3425-3430
22. Chan, D. C., Fass, D., Berger, J. M., and Kim, P. S. (1997) Core structure of gp41 from the HIV envelope glycoprotein. *Cell* 89, 263-273
23. Muster, T., Steindl, F., Purtscher, M., Trkola, A., Klima, A., Himmler, G., Rucker, F., and Katinger, H. (1993) A Conserved Neutralizing Epitope on Gp41 of Human Immunodeficiency-Virus Type-1. *Journal of Virology* 67, 6642-6647
24. Salzwedel, K., West, J. T., and Hunter, E. (1999) A conserved tryptophan-rich motif in the membrane-proximal region of the human immunodeficiency virus type 1 gp41 ectodomain is important for Env-mediated fusion and virus infectivity. *Journal of Virology* 73, 2469-248030

25. Zwick, M. B., Labrijn, A. F., Wang, M., Spenlehauer, C., Saphire, E. O., Binley, J. M., Moore, J. P., Stiegler, G., Katinger, H., Burton, D. R., and Parren, P. W. H. I. (2001) Broadly neutralizing antibodies targeted to the membrane-proximal external region of human immunodeficiency virus type 1 glycoprotein gp41. *Journal of Virology* 75, 10892-10905
26. Montero, M., van Houten, N. E., Wang, X., and Scott, J. K. (2008) The membrane proximal external region of the human immunodeficiency virus type 1 envelope: Dominant site of antibody neutralization and target for vaccine design (vol 72, pg 54, 2008). *Microbiol Mol Biol R* 72, 378-378
27. Kelsoe, G., Verkoczy, L., and Haynes, B. F. (2014) Immune System Regulation in the Induction of Broadly Neutralizing HIV-1 Antibodies. *Vaccines* 2, 1-14
28. Sun, Z. Y. J., Oh, K. J., Kim, M. Y., Yu, J., Brusic, V., Song, L. K., Qiao, Z. S., Wang, J. H., Wagner, G., and Reinherz, E. L. (2008) HIV-1 broadly neutralizing antibody extracts its epitope from a kinked gp41 ectodomain region on the viral membrane. *Immunity* 28, 52-63
29. Shi, W., Bohon, J., Han, D. P., Habte, H., Qin, Y., Cho, M. W., and Chance, M. R. (2010) Structural Characterization of HIV gp41 with the Membrane-proximal External Region. *Journal of Biological Chemistry* 285, 24290–24298
30. Song, L., Sun, Z.-Y. J., Coleman, K. E., Zwick, M. B., Gach, J. S., Wang, J.-h., Reinherz, E. L., Wagner, G., and Kim, M. (2009) Broadly neutralizing anti-HIV-1 antibodies disrupt a hinge-related function of gp41 at the membrane interface. *PNAS* 106, 9057–9062
31. Julien, J. P., Bryson, S., Nieva, J. L., and Pai, E. F. (2008) Structural Details of HIV-1 Recognition by the Broadly Neutralizing Monoclonal Antibody 2F5: Epitope Conformation, Antigen-Recognition Loop Mobility, and Anion-Binding Site. *Journal of Molecular Biology* 384, 377-392
32. Wild, C. T., Shugars, D. C., Greenwell, T. K., Mcdanal, C. B., and Matthews, T. J. (1994) Peptides Corresponding to a Predictive Alpha-Helical Domain of Human Immunodeficiency-Virus Type-1 Gp41 Are Potent Inhibitors of Virus-Infection. *P Natl Acad Sci USA* 91, 9770-9774
33. Matthews, T., Salgo, M., Greenberg, M., Chung, J., DeMasi, R., and Bolognesi, D. (2004) Enfuvirtide: The first therapy to inhibit the entry of HIV-1 into host CD4 lymphocytes. *Nat Rev Drug Discov* 3, 215-225
34. Zwick, M. B., Jensen, R., Church, S., Wang, M., Stiegler, G., Kunert, R., Katinger, H., and Burton, D. R. (2005) Anti-Human Immunodeficiency Virus Type 1 (HIV-1) Antibodies 2F5 and 4E10 Require Surprisingly Few Crucial Residues in the Membrane-Proximal

External Region of Glycoprotein gp41 To Neutralize HIV-1. *Journal of Virology* 79, 1252–126131

35. Lorizate, M., Gomara, M. J., de la Torre, B. G., Andreu, D., and Nieva, J. L. (2006) Membrane-transferring sequences of the HIV-1 gp41 ectodomain assemble into an immunogenic complex. *Journal of Molecular Biology* 360, 45-55

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

37. Apellaniz, B., Nir, S., and Nieva, J. L. (2009) Distinct Mechanisms of Lipid Bilayer Perturbation Induced by Peptides Derived from the Membrane-Proximal External Region of HIV-1 gp41. *Biochemistry* 48, 5320-5331

38. Apellaniz, B., Garcia-Saez, A. J., Nir, S., and Nieva, J. L. (2011) Destabilization exerted by peptides derived from the membrane-proximal external region of HIV-1 gp41 in lipid vesicles supporting fluid phase coexistence. *Bba-Biomembranes* 1808, 1797-1805

39. Bellamy-McIntyre, A. K., Lay, C. S., Bar, S., Maerz, A. L., Talbo, G. H., Drummer, H. E., and Pombourios, P. (2007) Functional links between the fusion peptide-proximal polar segment and membrane-proximal region of human immunodeficiency virus gp41 in distinct phases of membrane fusion. *Journal of Biological Chemistry* 282, 23104-23116

40. Lorizate, M., de la Arada, I., Huarte, N., Sanchez-Martinez, S., de la Torre, B. G., Andreu, D., Arrondo, J. L. R., and Nieva, J. L. (2006) Structural analysis and assembly of the HIV-1 gp41 amino-terminal fusion peptide and the pretransmembrane amphipathic-interface sequence. *Biochemistry* 45, 14337-14346

41. Gallaher, W. R., Ball, J. M., Garry, R. F., Griffin, M. C., and Montelaro, R. C. (1989) A General-Model for the Transmembrane Proteins of Hiv and Other Retroviruses. *Aids Res Hum Retrov* 5, 431-440

42. Kim, J. H., Hartley, T. L., Curran, A. R., and Engelman, D. M. (2009) Molecular dynamics studies of the transmembrane domain of gp41 from HIV-1. *Bba-Biomembranes* 1788, 1804-1812

43. Owens, R. J., Burke, C., and Rose, J. K. (1994) Mutations in the Membrane-Spanning Domain of the Human-Immunodeficiency-Virus Envelope Glycoprotein That Affect Fusion Activity. *Journal of Virology* 68, 570-574

44. Salzwedel, K., Johnston, P. B., Roberts, S. J., Dubay, J. W., and Hunter, E. (1993) Expression and Characterization of Glycophospholipid-Anchored Human Immunodeficiency-Virus Type-1 Envelope Glycoproteins. *Journal of Virology* 67, 5279-5288

45. Weiss, C. D., and White, J. M. (1993) Characterization of Stable Chinese-Hamster Ovary Cells Expressing Wild-Type, Secreted, and Glycosylphosphatidylinositol-Anchored 32 Human-Immunodeficiency-Virus Type-1 Envelope Glycoprotein. *Journal of Virology* 67, 7060-7066
46. Wilk, T., Pfeiffer, T., Bukovsky, A., Moldenhauer, G., and Bosch, V. (1996) Glycoprotein incorporation and HIV-1 infectivity despite exchange of the gp160 membrane-spanning domain. *Virology* 218, 269-274
47. Cleveland, S. M., McLain, L., Cheung, L., Jones, T. D., Hollier, M., and Dimmock, N. J. (2003) A region of the C-terminal tail of the gp41 envelope glycoprotein of human immunodeficiency virus type 1 contains a neutralizing epitope: evidence for its exposure on the surface of the virion. *J Gen Virol* 84, 591-602
48. Wyma, D. J., Jiang, J. Y., Shi, J., Zhou, J., Lineberger, J. E., Miller, M. D., and Aiken, C. (2004) Coupling of human immunodeficiency virus type 1 fusion to virion maturation: a novel role of the gp41 cytoplasmic tail. *Journal of Virology* 78, 3429-3435
49. Piller, S. C., Dubay, J. W., Derdeyn, C. A., and Hunter, E. (2000) Mutational analysis of conserved domains within the cytoplasmic tail of gp41 from human immunodeficiency virus type 1: Effects on glycoprotein incorporation and infectivity. *Journal of Virology* 74, 11717-11723
50. Yu, X. F., Yuan, X., Mclane, M. F., Lee, T. H., and Essex, M. (1993) Mutations in the Cytoplasmic Domain of Human-Immunodeficiency-Virus Type-1 Transmembrane Protein Impair the Incorporation of Env Proteins into Mature Virions. *Journal of Virology* 67, 213-221
51. Freed, E. O., and Martin, M. A. (1996) Domains of the human immunodeficiency virus type 1 matrix and gp41 cytoplasmic tail required for envelope incorporation into virions. *Journal of Virology* 70, 341-351
52. Dorfman, T., Mammano, F., Haseltine, W. A., and Gottlinger, H. G. (1994) Role of the Matrix Protein in the Virion Association of the Human-Immunodeficiency-Virus Type-1 Envelope Glycoprotein. *Journal of Virology* 68, 1689-1696
53. Rand, R. P., and Parsegian, V. A. (1989) Hydration Forces between Phospholipid-Bilayers. *Biochim Biophys Acta* 988, 351-376
54. Cohen, F. S., and Melikyan, G. B. (2004) The energetics of membrane fusion from binding, through hemifusion, pore formation, and pore enlargement. *J Membrane Biol* 199, 1-14
55. Chernomordik, L. V., Zimmerberg, J., and Kozlov, M. M. (2006) Membranes of the world unite! *J Cell Biol* 175, 201-20733

56. Kuzmin, P. I., Zimmerberg, J., Chizmadzhev, Y. A., and Cohen, F. S. (2001) A quantitative model for membrane fusion based on low-energy intermediates. *Proc Natl Acad Sci USA* 98, 7235-7240
57. Sackett, K., Nethercott, M. J., Epand, R. F., Epand, R. M., Kindra, D. R., Shai, Y., and Weliky, D. P. (2010) Comparative Analysis of Membrane-Associated Fusion Peptide Secondary Structure and Lipid Mixing Function of HIV gp41 Constructs that Model the Early Pre-Hairpin Intermediate and Final Hairpin Conformations. *Journal of Molecular Biology* 397, 301–315
58. Zhou, F., and Schulten, K. (1994) Molecular-Dynamics Studies of a Dipe Membrane Bilayer and Phospholipase-A2 Dipe Membrane Complex. *Biophys J* 66, A399-A399
59. Grove, J., and Marsh, M. (2011) The cell biology of receptor-mediated virus entry. *J Cell Biol* 195, 1071-1082
60. Sattentau, Q. J., and Weiss, R. A. (1988) The Cd4 Antigen - Physiological Ligand and Hiv Receptor. *Cell* 52, 631-633
61. Moore, J. P., Mckeating, J. A., Weiss, R. A., and Sattentau, Q. J. (1990) Dissociation of Gp120 from Hiv-1 Virions Induced by Soluble Cd4. *Science* 250, 1139-1142
62. Meyerson, J. R., Tran, E. E., Kuybeda, O., Chen, W., Dimitrov, D. S., Gorlani, A., Verrips, T., Lifson, J. D., and Subramaniam, S. (2013) Molecular structures of trimeric HIV-1 Env in complex with small antibody derivatives. *Proc Natl Acad Sci U S A* 110, 513-518
63. Harris, A., Borgnia, M. J., Shi, D., Bartesaghi, A., He, H., Pejchal, R., Kang, Y. K., Depetris, R., Marozsan, A. J., Sanders, R. W., Klasse, P. J., Milne, J. L. S., Wilson, I. A., Olson, W. C., Moore, J. P., and Subramaniam, S. (2011) Trimeric HIV-1 glycoprotein gp140 immunogens and native HIV-1 envelope glycoproteins display the same closed and open quaternary molecular architectures. *Proc Natl Acad Sci USA* 108, 11440–11445
64. Weissenhorn, W., Dessen, A., Harrison, S. C., Skehel, J. J., and Wiley, D. C. (1997) Atomic structure of the ectodomain from HIV-1 gp41. *Nature* 387, 426-430
65. Sackett, K., Nethercott, M. J., Zheng, Z., and Weliky, D. P. (2014) Solid-State NMR Spectroscopy of the HIV gp41 Membrane Fusion Protein Supports Intermolecular Antiparallel  $\beta$  Sheet Fusion Peptide Structure in the Final Six-Helix Bundle State. *Journal of Molecular Biology* 426, 1077-1094
66. Markosyan, R. M., Cohen, F. S., and Melikyan, G. B. (2003) HIV-1 Envelope Proteins Complete Their Folding into Six-helix Bundles Immediately after Fusion Pore Formation. *Molecular Biology of the Cell* 14, 926–93834

67. Sackett, K., Nethercott, M. J., Shai, Y., and Weliky, D. P. (2009) Hairpin Folding of HIV gp41 Abrogates Lipid Mixing Function at Physiologic pH and Inhibits Lipid Mixing by Exposed gp41 Constructs. *Biochemistry* 48, 2714–2722
68. Julien, J. P., Cupo, A., Sok, D., Stanfield, R. L., Lyumkis, D., Deller, M. C., Klasse, P. J., Burton, D. R., Sanders, R. W., Moore, J. P., Ward, A. B., and Wilson, I. A. (2013) Crystal Structure of a Soluble Cleaved HIV-1 Envelope Trimer. *Science* 342, 1477-1483
69. Harris, A. K., Bartesaghi, A., Milne, J. L. S., and Subramaniam, S. (2013) HIV-1 Envelope Glycoprotein Trimers Display Open Quaternary Conformation When Bound to the gp41 Membrane-Proximal External-Region-Directed Broadly Neutralizing Antibody Z13e1. *Journal of Virology* 87, 7191–7196
70. Lyumkis, D., Julien, J. P., de Val, N., Cupo, A., Potter, C. S., Klasse, P. J., Burton, D. R., Sanders, R. W., Moore, J. P., Carragher, B., Wilson, I. A., and Ward, A. B. (2013) Cryo-EM Structure of a Fully Glycosylated Soluble Cleaved HIV-1 Envelope Trimer. *Science* 342, 1484-1490
71. Sanders, R. W., Vesanen, M., Schuelke, N., Master, A., Schiffner, L., Kalyanaraman, R., Paluch, M., Berkhout, B., Maddon, P. J., Olson, W. C., Lu, M., and Moore, J. P. (2002) Stabilization of the soluble, cleaved, trimeric form of the envelope glycoprotein complex of human immunodeficiency virus type 1. *Journal of Virology* 76, 8875-8889
72. Klasse, P. J. (2007) Modeling how many envelope glycoprotein trimers per virion participate in human immunodeficiency virus infectivity and its neutralization by antibody. *Virology* 369, 245-262
73. Magnus, C., and Regoes, R. R. (2012) Analysis of the Subunit Stoichiometries in Viral Entry. *Plos One* 7, e33441
74. Zhu, P., Liu, J., Jr, J. B., Chertova, E., Lifson, J. D., Grise', H., Ofek, G. A., Taylor, K. A., and Roux, K. H. (2006) Distribution and three-dimensional structure of AIDS virus envelope spikes. *Nature* 441, 847-852
75. Sougrat, R., Bartesaghi, A., Lifson, J. D., Bennett, A. E., Bess, J. W., Zabransky, D. J., and Subramaniam, S. (2007) Electron Tomography of the Contact between T Cells and SIV/HIV-1: Implications for Viral Entry. *PLoS Pathogens* 3, 0570-0581
76. 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.
77. Jyoti Dev, James J. Chou, Et.al. (2016) Structural basis for membrane anchoring of HIV-1 envelope spike, *Science* 355, 6295.

78. Sai Chaitanya Chiliveri, and Ad Bax. (2018) Tilted, Uninterrupted, Monomeric HIV-1 gp41 Transmembrane Helix from Residual Dipolar Couplings, *J. Am. Chem. Soc.* 2018, 140, 34–37.
79. 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.
80. Gisanddata. ArcGIS, John Hopkins University & Medicine <https://gisanddata.maps.arcgis.com>. (2021)
81. U.S. Department of Health and Human Services. Centers for Disease Control and Prevention. [https://www.cdc.gov/nchs/nvss/vsrr/covid19/excess\\_deaths.htm](https://www.cdc.gov/nchs/nvss/vsrr/covid19/excess_deaths.htm). (2021)
82. Wu, F., Zhao, S., Yu, B., Chen, Y. M., Wang, W., Song, Z. G., . . . Zhang, Y. Z. (2020) A new coronavirus associated with human respiratory disease in China, *Nature* 579, 265-269.
83. Lu, R. J., Zhao, X., Li, J., Niu, P. H., Yang, B., Wu, H. L., . . . Tan, W. J. (2020) Genomic characterization and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding, *Lancet* 395, 565-574.
84. Wrapp, D., Wang, N., Corbett, K. S., Goldsmith, J. A., Hsieh, C.-L., Abiona, O., . . . McLellan, J. S. (2020) Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation, *Science* 367, 1260-1263.
85. Ng, M. L., Tan, S. H., See, E. E., Ooi, E. E., and Ling, A. E. (2003) Early events of SARS coronavirus infection in Vero cells, *J. Med. Virol.* 71, 323-331.
86. Zhang, H., Wang, G. W., Li, H., Nie, Y. C., Shi, X. L., Lian, G. W., . . . Deng, H. K. (2004) Identification of an antigenic determinant on the S2 domain of the severe acute respiratory syndrome coronavirus spike glycoprotein capable of inducing neutralizing antibodies, *J. Virol.* 78, 6938-6945.
87. Zhong, X. F., Yang, H. H., Guo, Z. F., Sin, W. Y. F., Chen, W., Xu, J. J., . . . Guo, Z. H. (2005) B-cell responses in patients who have recovered from severe acute respiratory syndrome target a dominant site in the S2 domain of the surface spike glycoprotein, *J. Virol.* 79, 3401-3408.
88. Xia, S., Liu, M. Q., Wang, C., Xu, W., Lan, Q. S., Feng, S. L., . . . Lu, L. (2020) Inhibition of SARS-CoV-2 (previously 2019-nCoV) infection by a highly potent pan-coronavirus fusion inhibitor targeting its spike protein that harbors a high capacity to mediate membrane fusion, *Cell Research* 30, 343-355.
89. Heald-Sargent, T., and Gallagher, T. (2012) Ready, Set, Fuse! The Coronavirus Spike Protein and Acquisition of Fusion Competence, *Viruses-Basel* 4, 557-580.

90. Tang, T., Bidon, M., Jaimes, J. A., Whittaker, G. R., and Daniel, S. (2020) Coronavirus membrane fusion mechanism offers a potential target for antiviral development, *Antiviral Research* 178, 104792.
91. Aydin, H., Al-Khooly, D., and Lee, J. E. (2014) Influence of hydrophobic and electrostatic residues on SARS coronavirus S2 protein stability: Insights into mechanisms of general viral fusion and inhibitor design, *Prot. Sci.* 23, 603-617.
92. Lai, A. L., Millet, J. K., Daniel, S., Freed, J. H., and Whittaker, G. R. (2017) The SARS-CoV Fusion Peptide Forms an Extended Bipartite Fusion Platform that Perturbs Membrane Order in a Calcium-Dependent Manner, *J. Mol. Biol.* 429, 3875-3892.
93. Belouzard, S., Chu, V. C., and Whittaker, G. R. (2009) Activation of the SARS coronavirus spike protein via sequential proteolytic cleavage at two distinct sites, *Proc. Natl. Acad. Sci. U.S.A.* 106, 5871-5876.
94. Duquerroy, S., Vigouroux, A. N., Rottier, P. J. M., Rey, F. A., and Bosch, B. J. (2005) Central ions and lateral asparagine/glutamine zippers stabilize the post-fusion hairpin conformation of the SARS coronavirus spike glycoprotein, *Virology* 335, 276-285.
95. Guillen, J., Kinnunen, P. K. J., and Villalain, J. (2008) Membrane insertion of the three main membranotropic sequences from SARS-CoV S2 glycoprotein, *Biochim. Biophys. Acta* 1778, 2765-2774.
96. Guillen, J., Perez-Berna, A. J., Moreno, M. R., and Villalain, J. (2008) A second SARS-CoV S2 glycoprotein internal membrane-active peptide. Biophysical characterization and membrane interaction, *Biochemistry* 47, 8214-8224.
97. Guillen, J., de Almeida, R. F. M., Prieto, M., and Villalain, J. (2008) Structural and dynamic characterization of the interaction of the putative fusion peptide of the S2SARS-CoV virus protein with lipid membranes, *J. Phys. Chem. B* 112, 6997-7007.
98. Madu, I. G., Roth, S. L., Belouzard, S., and Whittaker, G. R. (2009) Characterization of a highly conserved domain within the severe acute respiratory syndrome coronavirus spike protein S2 domain with characteristics of a viral fusion peptide, *J. Virol.* 83, 7411-7421.
99. Madu, I. G., Belouzard, S., and Whittaker, G. R. (2009) SARS-coronavirus spike S2 domain flanked by cysteine residues C822 and C833 is important for activation of membrane fusion, *Virology* 393, 265-271.
100. Mahajan, M., and Bhattacharjya, S. (2015) NMR structures and localization of the potential fusion peptides and the pre-transmembrane region of SARS-CoV: Implications in membrane fusion, *Biochim. Biophys. Acta* 1848, 721- 730.



101. Basso, L. G. M., Vicente, E. F., Crusca, E., Jr., Cilli, E. M., and Costa-Filho, A. J. (2016) SARS-CoV fusion peptides induce membrane surface ordering and curvature, *Sci. Rep.* 6.
102. Millet, J. K., and Whittaker, G. R. (2018) Physiological and molecular triggers for SARS-CoV membrane fusion and entry into host cells, *Virology* 517, 3-8.
103. Mahajan, M., Chatterjee, D., Bhuvaneswari, K., Pillay, S., and Bhattacharjya, S. (2018) NMR structure and localization of a large fragment of the SARS-CoV fusion protein: Implications in viral cell fusion, *Biochim. Biophys. Acta* 1860, 407-415.
104. Meher, G., Bhattacharjya, S., and Chakraborty, H. (2019) Membrane cholesterol modulates oligomeric status and peptide-membrane interaction of severe acute respiratory syndrome coronavirus fusion peptide, *J. Phys. Chem. B* 123, 10654-10662.
105. Duan, J. Z., Yan, X. Y., Guo, X. M., Cao, W. C., Han, W., Qi, C., . . . Jin, G. (2005) A human SARS-CoV neutralizing antibody against epitope on S2 protein, *Biochem. Biophys. Res. Comm.* 333, 186-193.
106. Keng, C. T., Zhang, A., Shen, S., Lip, K. M., Fielding, B. C., Tan, T. H. P., . . . Tan, Y. J. (2005) Amino acids 1055 to 1192 in the S2 region of severe acute respiratory syndrome coronavirus S protein induce neutralizing antibodies: Implications for the development of vaccines and antiviral agents, *J. Virol.* 79, 3289-3296.
107. Lip, K. M., Shen, S., Yang, X. M., Keng, C. T., Zhang, A. H., Oh, H. L. J., . . . Tan, Y. J. (2006) Monoclonal antibodies targeting the HR2 domain and the region immediately upstream of the HR2 of the S protein neutralize in vitro infection of severe acute respiratory syndrome coronavirus, *J. Virol.* 80, 941-950.
108. Tripet, B., Kao, D. J., Jeffers, S. A., Holmes, K. V., and Hodges, R. S. (2006) Template-based coiled-coil antigens elicit neutralizing antibodies to the SARS-coronavirus, *J. Struct. Biol.* 155, 176-194.
109. Elshabrawy, H. A., Coughlin, M. M., Baker, S. C., and Prabhakar, B. S. (2012) Human monoclonal antibodies against highly conserved HR1 and HR2 domains of the SARS-CoV spike protein are more broadly neutralizing, *Plos One* 7.
110. Ni, L., Zhu, J. Q., Zhang, J. J., Yan, M., Gao, G. F., and Tien, P. (2005) Design of recombinant protein-based SARS-CoV entry inhibitors targeting the heptad-repeat regions of the spike protein S2 domain, *Biochem. Biophys. Res. Comm.* 330, 39-45.
111. Xia, S., Yan, L., Xu, W., Agrawal, A. S., Algaissi, A., Tseng, C.-T. K., . . . Lu, L. (2019) A pan-coronavirus fusion inhibitor targeting the HR1 domain of human coronavirus spike, *Science. Advances* 5.

112. Walls, A. C., Tortorici, M. A., Snijder, J., Xiong, X., Bosch, B.-J., Rey, F. A., and Veesler, D. (2017) Tectonic conformational changes of a coronavirus spike glycoprotein promote membrane fusion, *Proc. Natl. Acad. Sci. U.S.A.* 114, 11157-11162.
113. Petit, C. M., Melancon, J. M., Chouljenko, V. N., Colgrove, R., Farzan, M., Knipe, D. M., and Kousoulas, K. G. (2005) Genetic analysis of the SARS-coronavirus spike glycoprotein functional domains involved in cell-surface expression and cell-to-cell fusion, *Virology* 341, 215-230.
114. Petit, C. M., Chouljenko, V. N., Iyer, A., Colgrove, R., Farzan, M., Knipe, D. M., and Kousoulas, K. G. (2007) Palmitoylation of the cysteine-rich endodomain of the SARS-coronavirus spike glycoprotein is important for spike-mediated cell fusion, *Virology* 360, 264-274.
115. Shulla, A., and Gallagher, T. (2009) Role of spike protein endodomains in regulating coronavirus entry, *J. Biol. Chem.* 284, 32725-32734.
116. DeGrado, William F., Capture and imaging of a prehairpin fusion intermediate of the paramyxovirus PIV5, 20992–20997, *PNAS*, December 27, 2011, 108, 52
117. Melnyk, Oleg, Native Chemical Ligation and Extended Methods: Mechanisms, Catalysis, Scope, and Limitations, *Chem. Rev.* 2019, 119, 7328–7443.
118. Xia, Xuhua. Domains and Functions of Spike Protein in SARS-Cov-2 in the Context of Vaccine Design. *Viruses* 2021, 13, 109. <https://doi.org/10.3390/v13010109>
119. Harper, James K., An Explanation of Magic Angle Spinning NMR Experiments in the Time Domain. *Magnetic Resonance*. 2009, 34:5 249-263
120. Whittaker, Gary. Phylogenic Analysis and Structural Modeling of SARS-CoV-2 Spike Protein Reveals an Evolutionary Distinct and Proteolytically Sensitive Activation Loop. *J Mol Biol.* 2020 May 1; 432(10): 3309–3325.
121. Expand, Richard M., Fusion Peptides, and the Mechanism of Viral Fusion. *Biochemica et Biophysica Acta (BBA)* 1614,1, July 2003., 116-121.
122. Petit, C.M.; Chouljenko, V.N.; Iyer, A.; Colgrove, R.; Farzan, M.; Knipe, D.M.; Kousoulas, K.G. Palmitoylation of the cysteine-rich endodomain of the SARS-coronavirus spike glycoprotein is important for spike-mediated cell fusion. *Virology* 2007, 360, 264–274.
123. Millet, J.K.; Kien, F.; Cheung, C.Y.; Siu, Y.L.; Chan, W.L.; Li, H.; Leung, H.L.; Jaume, M.; Bruzzone, R.; Peiris, J.S.; et al. Ezrin interacts with the SARS coronavirus Spike protein and restrains infection at the entry stage. *PLoS ONE* 2012, 7, e49566

124. Siegel, David. The Gaussian Curvature Elastic Energy of Intermediates in Membrane Fusion. *Biophys J.* 2008 Dec 1; 95(11): 5200–5215.
125. Zhang, X. Cryo-EM analysis of the post-fusion structure of the SARS-CoV spike glycoprotein. *Nature Communications.* 11, Article number: 3618 (2020)
126. United Nations Joint Programme on AIDS. UNAIDS. <https://www.unaids.org/en> (2022)
127. 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
128. 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.
129. Sun Z. Y., et al., Disruption of helix-capping residues 671 and 674 reveals a role in HIV-1 entry for a specialized hinge segment of the membrane proximal external region of gp41. *J. Mol. Biol.* **426**, 1095–1108 (2014)
130. RCSB, RCSB Protein Data Bank. [RCSB PDB: Homepage](https://www.rcsb.org/) (2022)
131. Tiziana Crepaldi, Alexis Gautreau, Paolo M. Comoglio, Daniel Louvard, and Monique Arpin. Ezrin is an Effector of Hepatocyte Growth Factor-Mediated Migration and Morphogenesis in Epithelial Cells. *J Cell Biol.* 1997 Jul 28; **138**(2): 423–434.

**CHAPTER 2:**  
**MATERIALS AND METHODS**

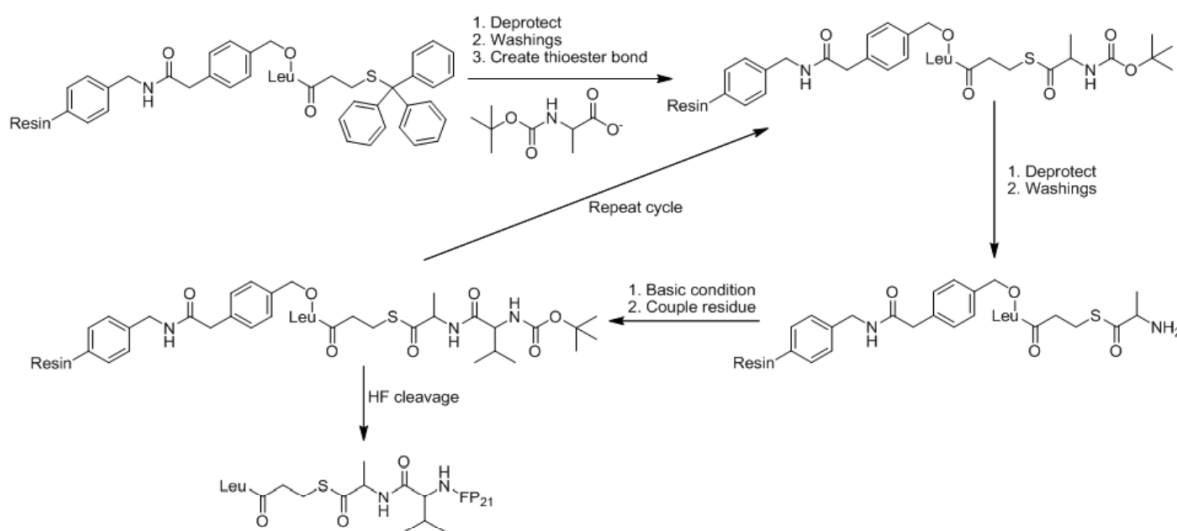
## 2.1. Introduction

Membrane proteins represent less than 3% of the total protein structures known to date based on structures reported in Protein Data Bank (40). This is due to the challenges in investigating the structures of membrane proteins. Obtaining enough folded protein, which is in a non-aggregated oligomeric state, is necessary to perform any of the structural or functional studies. For instance, some functional assays such as lipid mixing assays performed to assess activity of proteins require at least 10  $\mu$ L of a protein sample that is 40  $\mu$ M in concentration per assay trial (1,2). For a structural study such as a crystallography experiment, 10-20 mg protein is required (3). Exploring methods that can express, solubilize, and purify protein in large quantities is an important endeavor in the characterization of proteins. Preparing protein for characterization experiments require the successful completion of multiple steps.

## 2.2. Solid Phase Peptide Synthesis

Our previous experience has shown that there is less protein expressed when the FP region is included in the protein construct. The FP region is a hydrophobic domain and could easily promote aggregation in the solution. The FP has affinity to bind to cell membranes, and so binding of FP to *E. coli* membranes during the purification process leads to more difficulties in the isolation process. In addition, for this research it is important to have site-specific isotopic labeling in the FP but not in the HM. Solid phase peptide synthesis known as tert-Butoxycarbonyl (t-Boc) method can be done based on the principles of the work of Merrifield (5,6). According to this method there is sequential building of peptide chains, with the peptides being anchored to a solid insoluble resin support (**Figure 2.1**). To the resin (solid phase support) the first amino acid is covalently bonded with a C-terminal linker group and the N-terminus is protected. This is followed

by series of washes to remove the contaminants and deprotect the N-terminus of the attached amino acid. Then a new amino acid is added with an unprotected C-terminal to get attached to the N-terminal of the previous amino acid and build up the sequence. This whole cycle is repeated until the required sequence is created (5). The C-terminal was modified consisting of a thioester. The prepared peptide can be cleaved from the resin by breaking the covalent bond between the resin and the first amino acid using hydrofluoric acid.



**Figure 2.1** Schematic representation of t-Boc solid phase peptide synthesis.

In our work, solid-phase peptide synthesis was done manually using t-butoxycarbonyl (t-boc) chemistry and S-trityl- $\beta$ -mercaptopropionyl-p-methylbenzhydrylamine resin (230 mg, 0.88 meq/g). Sidechain protecting groups for amino acids in the SPPS reaction include: His, and Arg, tosyl; Asp benzyl ester; Lys, carboxybenzyl; Ser, and Thr benzyl. A liquid or liquid solution with reagents was added to the resin in a 40 mL Teflon vessel with cap, filter, stopcock, and nozzle, followed by shaking the vessel, and then drainage of the liquid from the vessel. Synthesis began with

resin-swelling in CH<sub>2</sub>Cl<sub>2</sub> (3 mL, 1 h) followed by trityl-group cleavage in 95:2.5:2.5 (v:v) trifluoroacetic acid (TFA):H<sub>2</sub>O:triisopropylsilane (10 mL, 4 min, 2×).

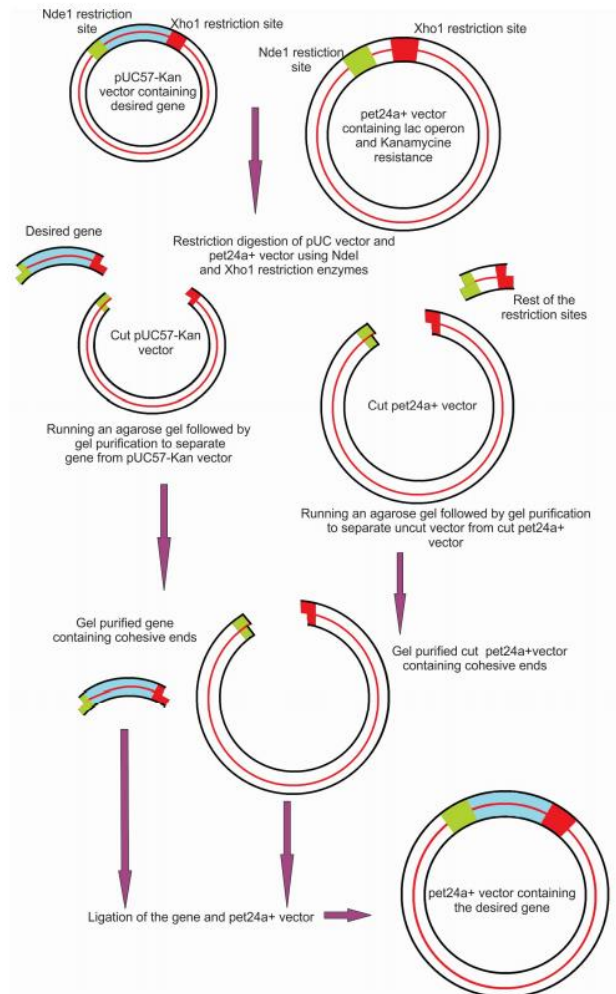
The first cycle of coupling amino acid (Ala-534) to the resin began with resin washing with CH<sub>2</sub>Cl<sub>2</sub> (3 mL, 1 min, 5×) and then 5% N,N-diisopropylethylamine (DIEA) in CH<sub>2</sub>Cl<sub>2</sub> (3 mL, 1 min, 3×), with concurrent reaction in a flask between t-boc-Ala (6.8 mmole) and activator 3-(diethoxyphosphoryloxy)-1,2,3-benzotriazin-4(3H)-one (DEPBT) (6.4 mmole) in tetrahydrofuran (THF, 5.5 mL). DIEA (1.1 mL) was added, and the total solution used to couple Ala-534 to the resin (4 h), followed by washes with CH<sub>2</sub>Cl<sub>2</sub> (3 mL, 1 min, 5×).

Deprotection (t-boc cleavage) was done with 50:48:2 (v:v) TFA:CH<sub>2</sub>Cl<sub>2</sub>:anisole (3 mL) followed by coupling of Arg-533 to the resin using a similar procedure as above, but with 3.4 mmole amino acid and 3.2 mmole DEPBT in 2.8 mL THF. There were then sequential deprotection/coupling cycles to complete the synthesis. There was a final deprotection followed by washing with CH<sub>2</sub>Cl<sub>2</sub> (3 mL, 1 min, 5×) and 5% DIEA in CH<sub>2</sub>Cl<sub>2</sub> (3 mL, 1 min, 3×), and drying overnight in a vacuum desiccator. Peptide was cleaved from the resin using hydrofluoric acid (HF) at Midwest Biotech Corporation (Fishers, IN). Following cleavage, the peptide can be purified by reverse phase high performance liquid chromatography (RP-HPLC) and verified by matrix assisted laser desorption ionization mass spectrometry (MALDI-MS).

### **2.3. Molecular Subcloning**

Protein expression in *E. Coli* requires the insertion of the targeted DNA that encodes for the protein of interest into the expression vector (**Figure 2.2**). Genes were ordered from Genscript Inc. (NJ). Genscript provides the gene in their vector, named

pUC57-Kan. The objective of subcloning is to cut the gene (insert) which is originally in pUC57-Kan vector and ligate it into cut pET24a+ expression vector. The term ligation means joining of the two DNA fragments in this case gene (insert) with the cut pET24a+ vector through the action of an enzyme called T4 DNA ligase.



**Figure 2.2** Schematic diagram for process of subcloning.

Before ordering the gene from the company, the restriction sites to be used for the subcloning need to be decided. NdeI is typically used at the 5' end of the gene sequence and XhoI restriction site at the 3' end of the gene sequence. Use of XhoI and NdeI



restriction sites is preferred rather than using other restriction sites to prevent the occurrence of non-native sequences in the expressed proteins.

(a) Restriction site correlating to the N-terminus:

Use of BamH1 instead of Nde1 will result unnecessary T7 tag (MetAlaSerMetThrGlyGlyGlnMetGlyArg) to be present before the N-terminus of protein of interest. This will result in the protein expressed to possess a T7 tag followed by the protein sequence. T7 tag is a tag used for immunoaffinity purification. In immunoaffinity purification, antibody which is covalently attached to a resin is added to a cell lysate. These antibodies can bind with proteins containing T7 tag. If the immunoaffinity purification is to be implemented for the purification of the expressed protein, the BamH1 site can be used instead of the Nde1 restriction site. For most of our protein constructs, an N-terminal tag was not used to aid in protein purification because we wanted to avoid reducing the function of the FP at the N-terminus. In some cases, 5N-terminal tags were used then later cleaved.

(b) Restriction site correlating to the C-terminus:

Selection of the second restriction site Xho1 at 3' end depends on the position of His-tag. Since it was planned to purify the expressed protein using immobilized metal affinity chromatography (IMAC) which requires the presence of His tag in the protein being expressed, this restriction site was used. Usage of Xho1 results only L (Leucine) E (Glutamate) amino acids to appear before the tag after the native protein is expressed. Use of Sal 1 instead of Xho1 will cause unnecessary non-native amino acid sequence to appear on C-terminal of the protein. Use of Sal1 will cause ArgGlnAlaCysGlyArgThrArgAla to appear in the C-terminus of the protein before the tag.

Therefore, usage of Xho1 at 3' end is preferred over the usage of any other restriction site since it only adds up amino acid L and E before the tag.

Usage of Xho1 and Nde1 is preferred over the usage of any other restriction sites because these two restriction sites eliminate the addition of long non-native sequences to either N-terminal or C-terminal of the protein being expressed. Nucleotide sequence corresponding to the Nde1 is CATATG and that for Xho1 is CTCGAG, which code for (H) Histidine (M) Methionine and (LE) respectively.

### **2.3.1. Basic Steps Involved in Subcloning**

1) Transformation of gene (insert), which is in pUC57-Kan vector, and the empty (pet24a+) expression vector was performed separately.

2) Extraction of the plasmid (pUC57-Kan) containing the gene of interest (insert) and extraction of the empty expression vector (pet24a+) from the overnight grown cell cultures was performed. This process is commonly referred to as miniprep and commercially available plasmid extraction kits were used in the process. (Promega plasmid extraction kit).

3) Restriction digestion of the pUC57-Kan plasmid which contains the gene of interest (insert) and the vector (pet24a+) using restriction enzyme Nde1 and Xho1. Restriction enzymes can identify specific DNA sequences and cut the DNA, causing the formation of cohesive ends. Selected restriction sites should only be present in the 5' and 3' ends of the gene and vector, and those should be absent anywhere else in the gene and vector sequence. If the selected restriction sites (in this case Xho1 and Nde1) are present in multiple different places of the gene sequence and vector sequence, this will result in the digestion of the gene and vector into several different pieces.

Commonly used restriction sites are localized into one area known as “multiple cloning site”. In pet24a+ vector, multiple cloning is present between the Nde1 and Xho1 sites. Restriction sites present in the multiple cloning site do not appear anywhere else in the vector sequence. Therefore, usage of the restriction sites in the multiple cloning site guarantees that the vector will not cut into several pieces during the restriction digest.

4) Two preparative DNA gels (agarose gels) were run to isolate cut pet24a+ vector from uncut pet24A+ vector, and to isolate gene (insert) from pUC57-Kan vector. Depending on the size of the vector and insert, percentage of the agarose needed for the DNA gel preparation varies. A 0.5% agarose gel was used for the separation of cut pet24a+ vector from uncut pet24a+ vector (about 5.3 kbp), and a 1% agarose gel was used for the isolation of gene of interest (insert) (about 500 bp) from pUC57-Kan vector (about 2.5 kbp). Small amounts of agarose result in large pore size in the gel, which are useful in the isolation of constructs having large numbers of base pairs. Similarly, increased agarose content in the gel results in small pore size, enabling the separation of constructs with small number of base pairs.

(5) Setting up a ligation reaction. Ligation reactions must be set up by considering the molar ratio of the empty vector and the insert. Typically, the amount of empty vector used for the ligation reaction is 80-200 ng. Generally, three ligation reactions are set up using the insert: vector molar ratios 1:1, 3:1, and 1:3. First, the insert and vector were mixed at these ratios. Next, this mixture is combined with T4 DNA ligase enzyme and its buffer and stored at 16 °C overnight for ligation to take place. T4 DNA ligase and its specific buffer (50 mM Tris-HCl, 10 mM MgCl<sub>2</sub>, 1 mM ATP, 10 mM DTT, pH of buffer is 7.4) were purchased from New England Biolabs Inc.

(6) Transformation of ligation product. Transformed cells were plated onto agar plates containing the antibiotic (kanamycin) to which the pet24a+ vector displays resistance. BL21DE3 competent cells (50  $\mu$ L) were taken out of a -80°C freezer and thawed on ice, and a volume of 1 to 5  $\mu$ L of ligation product was added. Ligation product and competent cells were gently mixed by flicking the bottom of the microcentrifuge tube. Then the competent cell/ligation product mixture was placed on ice for 20-30 min. The sample was then placed in a 42°C water bath for a heat shock of 45 seconds. By exposing cells to a sudden increase in temperature, a pressure difference between the outside and the inside of the cell is created to induce the formation of pores through which supercoiled plasmid DNA can enter (8). The sample was then put on ice for an additional 2 minutes. The competent cell/ ligation mixture was diluted to 500  $\mu$ L with LB (Luria Broth), and the sample was placed in an incubator at 37°C for 1 hour. The incubator was shaken at 225 rpm speed. After 1 hour, the competent cell/ ligation cell mixture was plated on agar plate which contains the kanamycin antibiotic. Agar plates were incubated overnight at 37 °C.

(7) Once the initial colonies appeared on the agar plate, individual colonies were extracted from the plate and added to separate Eppendorf tubes with 1 mL of LB per sample. Colonies were grown in 1 mL of LB medium with antibiotic, with 1 $\mu$ L of 50mg/mL antibiotic for every 1 mL of LB. Cells were then transferred to 50 mL of LB and grown overnight at 37 °C.

(8) Finally, the plasmid was extracted using a Promega Wizard Plus DNA miniprep kit and submitted to a DNA sequencing facility to confirm the presence of the desired gene in the pet24a+ vector.

### **2.3.2. Site Directed Mutagenesis**

Insertion site directed mutagenesis was initiated with a short primer which contains the required mutation. Primers were designed complementary to the template DNA sequence near the insertion site to enable hybridization of the primer to the gene of interest. One can choose the insertion site to be anywhere in the DNA sequence. Two primers i.e., forward and reverse were designed complementary to the gene of interest DNA double strand. The primers with the added nucleotides for insertion mutation elongate using DNA polymerase. The primer to plasmid ratio should be 3:1 but may be as high as 7:1. The ratio should be varied to optimize the specific polymerase chain reaction (PCR). The amplification of the required DNA was performed by PCR. The various steps of PCR comprised of:

1. Initialization – The initialization is done by a hot start at ~95 °C for 5 min. This is done to activate the thermally stable Pfu DNA polymerase (34).

2. Denaturation – This short step of heating (~95 °C) is done for DNA melting of the template DNA by breaking the H-bonds between the nucleotides of the double strand DNA (34). In this step single strand template DNA is obtained.

3. Annealing – In this step the primers anneal with the single strand DNA. To encourage binding of the primers to the template, the temperature is lowered, typically to ~5 °C below the primer melting temperature. Primers were designed to be GC rich to enable better binding to the template DNA. After primer-template DNA hybrid formation, synthesis of the whole DNA starts (34).

4. Elongation – After binding of the primer to the template DNA, the DNA polymerase elongates the DNA complementary to the template DNA by adding deoxynucleotide triphosphates (dNTPs). The optimum temperature is typically 72 °C (34).

5. Final Elongation – This step is also done at 72 °C to ensure that all single stranded DNA is completely synthesized. While performing an insertion mutation, various annealing temperatures were applied during PCR based on the melting temperature of the DNA. Substitutions of up to 3 nucleotides could be done at one time. To verify whether the desired DNA has been produced or not, the final PCR product was transformed, and the DNA was submitted for sequencing.

### **2.3.3. Producing Chemically Competent Cells**

On day 1, a frozen glycerol stock of bacterial cells (BL21DE3 PLysS) was streaked onto an LB plate (chloramphenicol was used as an antibiotic for BL21DE3 PLysS). Work was performed in a sterile environment. Plate was grown overnight at 37°C.

On day 2 several solutions were autoclaved including 1 L LB, 1 L of 100 mM CaCl<sub>2</sub>, 1 L of 100 mM MgCl<sub>2</sub>, 100 mL of 85 mM CaCl<sub>2</sub> 15% glycerol v/v, centrifuge bottles and caps with microfuge tubes. The 100 mM CaCl<sub>2</sub>, 100 mM MgCl<sub>2</sub>, 85 mM CaCl<sub>2</sub> 15% glycerol v/v were chilled overnight. A starter culture of cells was prepared from a single colony of E. coli from the LB plate and inoculated a 50 mL starter culture of LB (chloramphenicol). Culture was grown at 37°C in a shaker overnight.

On day 3, 300 mL of LB media was inoculated with 1 mL of the 50 mL starter culture and grown at 37°C while shaking. The OD<sub>600</sub> was measured every hour, then every 15-20 minutes when the OD reached 0.2. When the OD<sub>600</sub> reached 0.35-0.4, the cells were immediately put on ice. The culture was chilled for 20-30 minutes, with

occasional swirling to ensure even cooling. Centrifuge bottles were also placed on ice at this time. It is important not to let the OD get any higher than 0.4. The OD should be carefully monitored and checked often, especially when it gets above 0.2, as the cells grow exponentially. It usually takes about 3 hours to reach an OD of 0.35 when using a 10 mL starter culture. It is also particularly important to keep the cells at 4°C for the remainder of the procedure to prevent premature cell growth. The cells, and any bottles or solutions that they encounter, must be pre-chilled to 4°C. The culture was split into six parts by pouring about 50 mL into ice cold centrifuge bottles. Cells were harvested by centrifugation at 3000g for 15 minutes at 4°C. The supernatant was decanted and gently resuspended in about 20 mL of ice cold MgCl<sub>2</sub>. Suspensions were combined into one centrifuge bottle. Cells were harvested by centrifugation at 2000g for 15 minutes at 4°C. 6. The supernatant was decanted, and the pellet resuspended in about 20 mL of ice cold CaCl<sub>2</sub>. This suspension was kept on ice for at least 20 minutes. 1.5 mL microfuge tubes were placed on ice and chilled. Cells were harvested by centrifugation at 2000g for 15 minutes at 4°C. The supernatant was decanted, and the pellet resuspended in ~50 mL of ice cold 85 mM CaCl<sub>2</sub>, 15% glycerol. The suspension was transferred to a 50 mL conical tube. Cells were harvested by centrifugation at 1000g for 15 minutes at 4°C. The supernatant was decanted, and the pellet resuspended in 2 mL of ice cold 85 mM CaCl<sub>2</sub>, 15% glycerol. 50 µL of suspension was pipetted into sterile 1.5 mL microfuge tubes and snap frozen with liquid nitrogen. Cells were then stored at -80 °C.

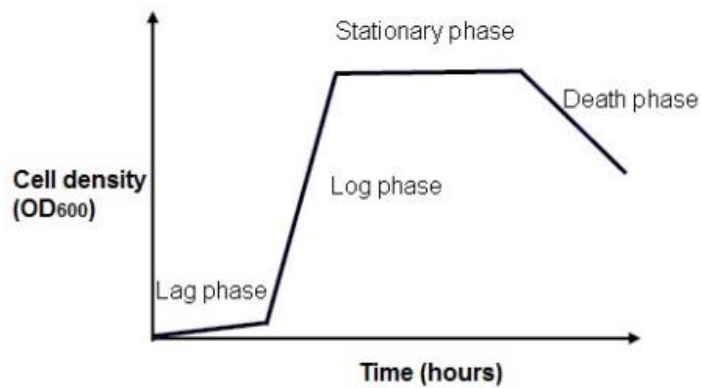
## **2.4. Protein Expression**

Bacterial expression systems represent the most used and cheapest methods of protein synthesis for structural and functional studies. By performing molecular cloning as

discussed above, a DNA sequence that encodes for the target protein is inserted to the expression vector. This vector is then transformed into a host cell (*E. coli*) and protein synthesis is induced by the addition of inducer such as Isopropyl  $\beta$ -D-1-thiogalactopyranoside (IPTG).

After inoculation of the *E. coli* cells into a culture medium, the cell density can be graphically represented as graphs known as growth curves. OD<sub>600</sub> (optical density) is used to estimate *E. coli* cell density. When OD<sub>600</sub> reaches 1, it is established that about  $1 \times 10^9$  cells per 1 mL of culture medium is present. The growth curve (**Figure 2.3**) of *E. coli* consists of 4 phases namely lag phase, log phase, stationary phase, and death phase. At the lag phase there is only a limited number of cells. During log phase, *E. coli* cells grow exponentially with time due to the presence of plenty of nutrients in the culture medium. Therefore, they are in the optimum state to produce protein (34). In the stationary phase, nutrients in the medium become limited (due to being consumed in the previous two phases) and metabolic byproducts accumulate and start resulting in cell death. The accumulation of metabolic byproducts such as ethanol, lactate, and acetate results in lowering of the pH of the medium (*E. coli* cell growth is maximum in neutral pH). Investigation of *E. coli* cell growth with time is done by measuring the OD<sub>600</sub> with time. Measured OD<sub>600</sub> vs time is monitored and the time at which IPTG is to be added to induce protein expression is chosen. In general, IPTG addition is done when OD<sub>600</sub> reaches 0.5-0.8 where *E. coli* growth is in the log phase.

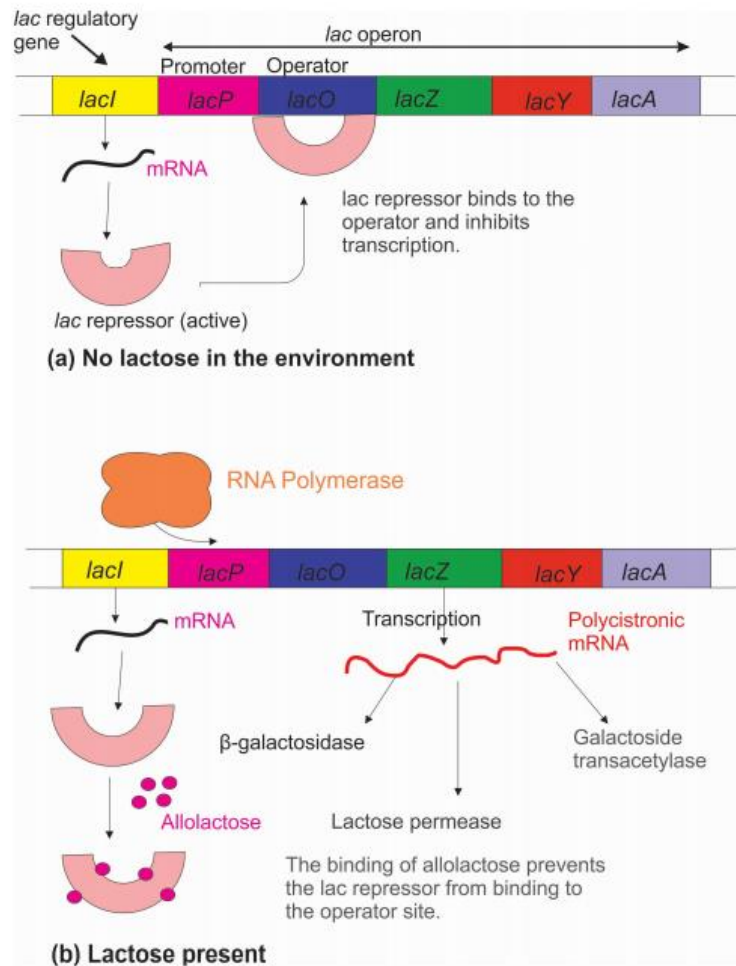




**Figure 2.3** Diagram of the four phases of E. Coli growth.

### 2.4.1 Lac Operon

An operon is a cluster of bacterial genes which is controlled under a single promoter. A promoter is a region of DNA that initiates the transcription of a particular gene. The lac operon is natively found in E. coli and other enteric bacteria and used to transport and metabolize lactose in an efficient manner (7). The lac operon (**Figure 2.4**) consists of promoter, operator, terminator, and three other genes known as lacZ, lacY, and lacA. Lac Z encodes for  $\beta$ -galactosidase enzyme. This enzyme cleaves the disaccharide lactose into glucose and galactose, and it converts lactose into allolactose (7). LacY encodes for lactose permease enzyme, which facilitates transport of lactose into the cells. LacA encodes galactoside O-acetyltransferase, an enzyme that transfers an acetyl group from acetyl-CoA to  $\beta$ -galactosidase (7,34).



**Figure 2.4** Diagram showing the functionality of Lac operon (a) In the absence of lactose (b) In the presence of lactose.

In the absence of lactose, the regulatory gene *lacI* produces *lac* repressor protein and it can bind with operator. Once repressor protein binds to the operator it will prevent the transcription of messenger RNA (mRNA) for *lacZ*, *lacY*, and *lacA*. In the presence of lactose, conversion of lactose to allolactose by  $\beta$ -galactosidase occurs. The allolactose formed binds with *lac* repressor protein causing conformational change which prevents the binding of repressor to operator. This allows RNA polymerase to transcribe *lac* genes (7,34).

In recombinant protein expression, the gene sequence of the protein that needs to be expressed is placed following the operon to take advantage of the lac operon. The protein expression is induced by addition of IPTG (a non-hydrostable compound like allolactose). IPTG binds with repressor protein and inhibits the binding of repressor protein with operator that leads to production of recombinant protein (34).

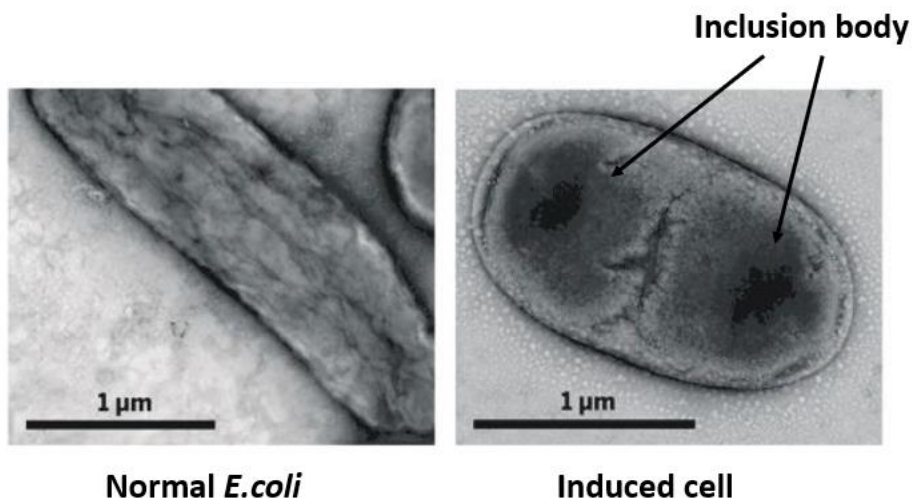
#### **2.4.2. pET Vector System**

The pet24a+ vector (**Figure 2.5**) was used in this project to take the advantage of the lac operator. This vector basically consists of the lacI gene, T7 promoter, lac operator, and T7 terminator (**Figure 2.6**) (35). Bacteriophage T7 promoter is non-native to the bacterial genome, and it can only recognize T7 RNA polymerase, although not bacterial RNA polymerase. Efficiency of T7 RNA polymerase is greater compared to bacterial RNA polymerase since it can transcribe DNA eight times faster than bacterial RNA polymerase (8). T7 RNA polymerase is not produced by regular bacterial cells and BL21 (DE3) competent cells that we have used in our studies already have the gene that codes for T7 RNA polymerase. T7 RNA polymerase specifically binds with T7 promoter and transcribes the DNA that is downstream to the T7 promoter. In the absence of the transducer molecule IPTG, T7 RNA polymerase is not produced and therefore gene of interest is not transcribed. Upon addition of IPTG, IPTG induces the expression of RNA polymerase and thereby the target protein. Pet vectors expression of a targeted protein is tightly controlled by T7 promoter and T7 RNA polymerase (35).



### 2.4.3. Inclusion Bodies

Overexpression of many recombinant proteins using *E. coli* results in the accumulation of proteins in inclusion bodies (**Figure 2.7**). In the inclusion bodies, in addition to the overexpressed proteins, phospholipids from the *E. coli* membrane, other *E. coli* and proteins and RNA are present. Inclusion bodies produced in bacteria have different shapes such as spherical, ellipsoidal, cylindrical, and even tear-shaped with the size ranging from 50-700 nm (9). Proteins present in inclusion body have either amorphous (devoid of any structural regularity) or paracrystalline (short to medium term ordering in the structure) nature (10,11). Amorphous proteins are unfolded, while paracrystalline are folded. The presence of inclusion bodies in bacterial cells can be easily identified using transmission electron microscopy.



**Figure 2.7** (Left) Transmission electron microscopy pictures of uninduced bacterial cells (Right) Transmission electron microscopy pictures of bacterial cells induced to produce protein (38).

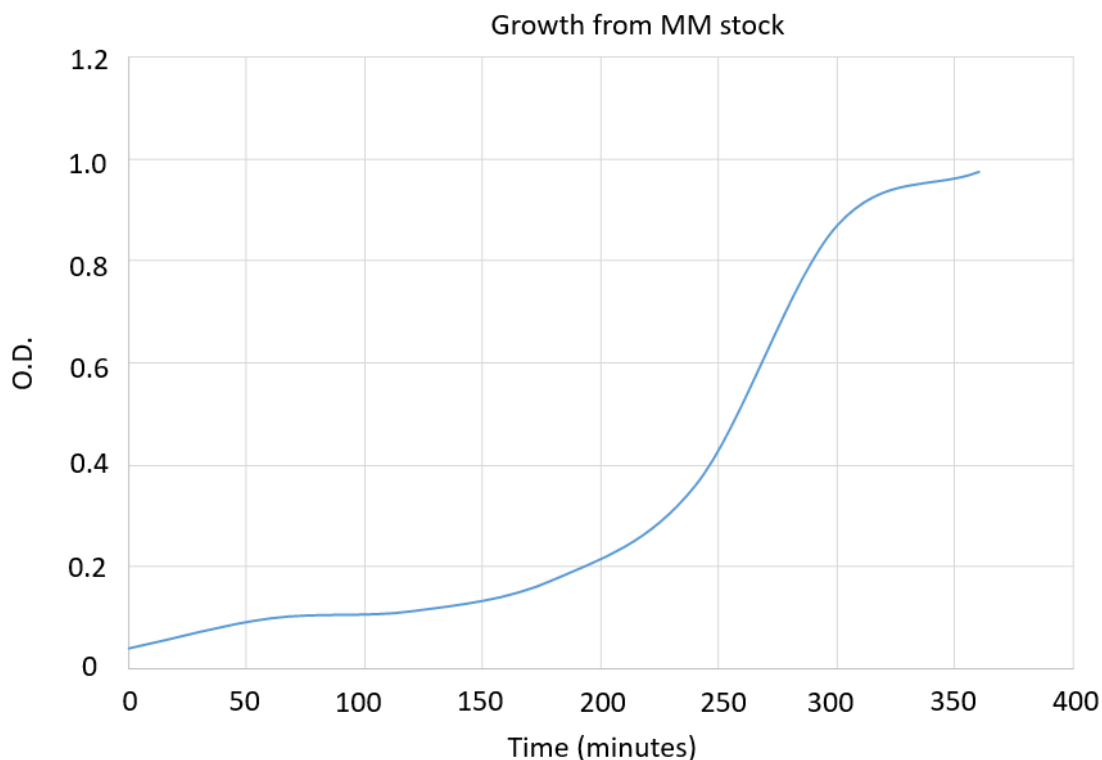
Inclusion bodies can be solubilized using denaturants such as 8M urea, 6M guanidinium hydrochloride or ionic detergents (0.5% sarkosyl, 0.5% SDS) (12,13).

Multiple other detergents have also been used. There are some advantages of using protein purified from inclusion bodies for structural and functional studies. (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. Using high speed centrifugation (15000 rpm, 30 minutes) after cell lysis it is possible to separate inclusion bodies from the less dense cell debris. However, the pellet formed at this stage will contain other subcellular components of high density such as ribosomes. (3) With effective isolation and purification, protein that is more than 95% pure can be obtained (14).

#### **2.4.4. Expression Proteins with Stable Isotope Labeling**

Isotopically labeled protein was produced by growing *E. coli* containing protein construct in isotopically labeled media followed by induced expression of the protein in labeled media. Protein expression from cells using the pUC19 vector was also attempted, but experienced expression in the absence of IPTG (leaky expression) that made the vector ineffective for isotopic labeling. Bacteria was initially grown in rich media. A single media switch followed by induction with IPTG yielded no HM protein. A protocol was adapted for our protein expression that involves a double media switch (41). *E. coli* was first grown in rich media, then a small fraction (about 20%) of the cell pellet from the rich media growth was transferred to a flask with minimal media and unlabeled glucose for a second growth period (**Figure 2.8**). This second growth period in minimal media may be necessary to produce *E. coli* with metabolic systems that allow for better use of the limited nutrients present in the minimal media. The *E. coli* from the second growth phase in minimal media was then transferred to a flask containing minimal media and labeled

glucose. If the protein needed non-exchangeable labeling with deuterium, D<sub>2</sub>O was added to the minimal media.



**Figure 2.8** Graph of *E. coli* growth in minimal media.

## 2.5. Solubilization and Purification of Expressed Protein

When optimizing a protein solubilization and purification protocol, it is necessary to check all possible places where protein can exist. Most of the time overexpressed proteins are aggregated as inclusion bodies (14). First, the cells were sonicated in the presence of phosphate buffered saline (PBS) to break the cells and release the soluble proteins into the medium. The lysate was centrifuged at 15000 rpm (35000g) for 30 min at 4°C. The pellet resulting from this centrifugation contains the inclusion body fraction and the other heavy cell debris. The soluble proteins and cell membrane are in the supernatant. Soluble proteins can be isolated from the purification of this supernatant. Ultracentrifugation (speed of 45000 rpm or approximately 300000g for 1 hour at 4°C) of

this supernatant results in the membrane fraction of the E. Coli being separated as a pellet. Membrane proteins are more likely to be present in this fraction (14).

For membrane protein isolation, detergents and chaotropic agents are used as solubilizing reagents (12,13,15). Several characteristics of membrane proteins make their solubilization and purification challenging compared to soluble proteins: (1) Due to the presence of the hydrophobic domains, membrane proteins tend to accumulate on cell membranes and therefore cells possess only a limited membrane surface area to store these membrane proteins in the correctly folded state. (2) Expressed protein is often stored as an insoluble aggregate known as inclusion bodies and denaturing detergents (Ex: 0.5% sarkosyl, 0.5% SDS) or denaturing reagents (e.g: 8M urea, 6M guanidium hydrochloride) must be used for solubilization. Furthermore, proteins extracted from inclusion bodies must be refolded before they are used for any structural or functional experiments. (3) Downstream protein purification procedures including immobilized metal affinity chromatography (IMAC), ion exchange chromatography, and gel filtration chromatography techniques can fail due to several reasons.

(a) To succeed with IMAC chromatography the histidine tag must be exposed to the metal ion. Due to the hydrophobic nature of the membrane protein often limited or no exposure of histidine tag occurs resulting in no targeted protein binding with the metal ion causing the purification to fail. When exposed to aqueous solvent, hydrophobic domains of membrane proteins tend to fold in such a way that they are least exposed to solvent while the hydrophilic domains are more exposed. In such circumstances, if the tag is attached to a hydrophobic domain, it can often move with the hydrophobic domain rendering it unexposed to the solvent. This lack of exposure causes the above



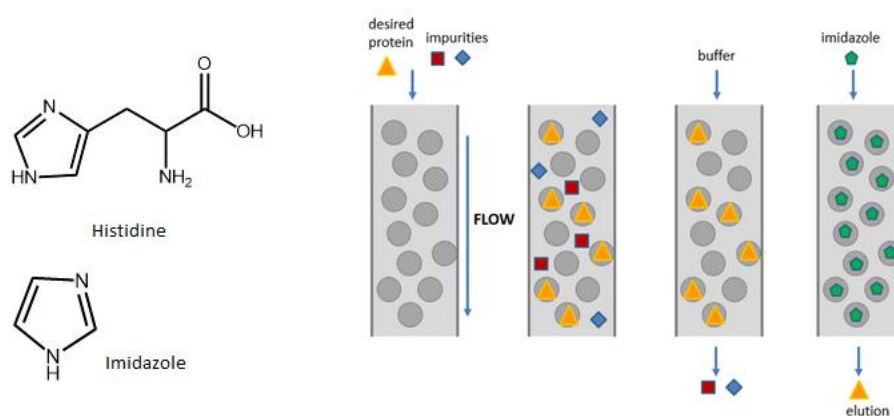
phenomena and makes His tag-based purification of membrane proteins challenging. Using a longer linker region (addition of a glycine repeat) between the protein sequence and His tag sequence can mitigate this issue to some extent (15,16). Our constructs feature a 6-Glycine repeat in either the N-terminus or C-terminus to increase the amount of protein that can be purified by IMAC.

(b) Ion exchange chromatography can fail in the presence of charged detergents since detergent molecules compete with protein molecules for the binding with the ion exchange resin when both detergent molecules and the protein molecules are similarly charged.

(c) Some membrane proteins can form aggregates in the loading solution for the gel filtration columns if the concentration of detergent being used is less than the critical micelle concentration (CMC) level of the detergent. Different detergents must be screened to identify the ones in which no aggregation occurs. Protein aggregation can occur even when the detergent concentration is above the CMC. The screening of detergents is typically above the CMC. Clogging of the gel filtration column due to protein aggregation will cause increased backpressure. In addition, protein is either precipitated at the front end of the column or ends up in the void volume, perhaps with other proteins. Denaturing solutions such as 8M urea or 6M guanidine hydrochloride can be used to solubilize precipitates and therefore clean the column followed by 0.5M NaOH to remove salts and other contaminants. Proteins used in this work were typically solubilized in 6M guanidine hydrochloride before size exclusion chromatography.

### 2.5.1. Immobilized Metal Affinity Chromatography (IMAC)

In immobilized metal affinity chromatography (IMAC) (**Figure 2.9**), histidine residues present in the surface of the protein molecule and immobilized metal ions form weak coordinate bonds. Immobilized metal ions such as  $\text{Ni}^{2+}$ ,  $\text{Cu}^{2+}$ , and  $\text{Co}^{2+}$  form coordinate bonds with chelating compounds of nitrilotriacetic acid (NTA) or iminodiacetic acid (IDA). Some of the metal coordination sites that remain contain water or buffer and can undergo reversible exchange with sidechain electron donor groups of histidine present in His-tagged protein (16). Elution of the protein from the affinity matrix can be achieved by (1) using a competitive displacement agent such as imidazole, (2) using a low pH buffer or (3) using a strong chelating agent such as ethylenediamine tetraacetic acid (EDTA).



**Figure 2.9** Affinity Chromatography separation of desired protein from cell lysis

impurities. Target protein and cell impurities are mixed with affinity resin. Impurities are washed away from target protein which binds to affinity resin. Purified protein is eluted from affinity column by imidazole containing buffer.

It is easy to elute His-tagged protein (poly-histidine tag with six histidine residues) attached to metal ions with a small molecule such as imidazole since imidazole also can

display the same interaction shown by the His tag. Therefore, the His-tagged protein is detached and elutes from the metal while imidazole binds with the metal. Using 250-300 mM imidazole helps to elute the protein from the affinity resin. The pKa of the histidine imidazole ring is about 6.0. At low pH, histidine become protonated at the amine side chain and cannot form coordinate bonds with metal ions since its lone pairs are donated to the H<sup>+</sup>. Therefore, a His-tagged protein will elute from the resin when the eluent is at low pH conditions. In addition, strong chelating agents like EDTA form a complex with metal ion and disrupt the interaction with the metal ion and His-tag proteins, causing the protein to elute.

### **2.5.2. Ion Exchange Chromatography**

The basis for ion exchange chromatography is the ionic interaction between protein molecules and the ion exchange resin. The separation is driven by the binding of protein molecules to the oppositely charged groups in the ion exchange resin. The isoelectric point (pI) of a protein is where its net charge is zero and it depends on the number of different ionizable sidechains of amino acid residues typically at the protein surface. Depending on the pH and pI values, proteins have either a net positive or negative charge that enable their separation by ion exchange. Anion exchange is performed using positively charged resin with the loading buffer pH values above the pI of the protein, so the protein is negatively charged. Cation exchange is performed with the loading buffer pH values below the pI value, so the protein is positively charged relative to a negatively charged resin. To elute the proteins from the ion exchange resin, a salt gradient or manipulation of the pH can be used. Molecules that are weakly bound to the resin elute at low concentration of salt, while a higher salt concentration is

necessary to elute strongly bound proteins. Raising the pH of the mobile phase above the isoelectric point of a given protein causes the protein to become negatively charged. Therefore, this method can be used to elute the protein off a cation exchange resin when performing cation exchange chromatography. The increased pH causes both cation ion exchange resin and protein to become similarly charged, causing protein to elute off the column. Similarly, in anion exchange chromatography lowering the pH of the mobile phase causes protein molecules to be protonated and positively charged causing it to elute out from the positively charged resin.

## **2.6. SDS PAGE (Sodium Dodecyl Sulphate Poly Acrylamide Gel Electrophoresis)**

SDS PAGE allows mass selective separation of protein molecules. SDS (Sodium dodecyl sulphate) feature a dodecyl aliphatic chain and anionic sulfate head. When SDS detergent that coats over the proteins it results in protein denaturation. Proteins bind to SDS in the same ratio of 1.4 g of SDS per 1g of protein (about one SDS molecule per every 2 amino acids). Positively charged and hydrophobic residues bind SDS. This causes proteins to attain an intrinsic negative charge. Therefore, SDS treated proteins have similar charge-mass ratio.

SDS-bound protein molecules migrate in the SDS PAGE gel depending primarily on their masses, since all protein molecules have the same negative charge-to-mass ratio due to the presence of SDS. This allows us to determine molecular mass of the protein. By changing the ratio of acrylamide to bisacrylamide one can change the crosslinking pore size of the SDS PAGE matrix. Pore size in polyacrylamide gels is determined with the values of %T [total polyacrylamide percentage (w/v)] and %C bis [the ratio of bis to monomer (w/w)]. Higher percentage polyacrylamide gels are used to increase the

separation resolution of the small proteins since the pore size is small. Molecules with higher mass will experience more resistance to migration when the pore size of the gel is smaller (34).

## **2.7. Gel Filtration Chromatography (SEC-Size Exclusion Chromatography)**

Gel filtration chromatography is used to separate proteins and oligonucleotides based on their size. Gel filtration bed matrix contains a variety of pore sizes prepared by the crosslinking of polysaccharides dextran and agarose. Molecules present in the mobile phase pass through the stationary phase (column matrix) allowing molecules to diffuse into the pores. Larger molecules do not enter the pores of the column matrix since pore size is too small for them to penetrate and therefore move through the column faster and elute earlier. Small molecules can enter the column pores and move through the column slowly while larger molecules are excluded from the pores and move quickly. In gel filtration chromatography the exclusion limit refers to the molecular mass (or size) beyond which the molecules start to elute at void volume of the column. The permeation limit refers to the lower limit of molecular mass (or size) of molecules that causes all molecules below this mass to elute as a single band. The Superdex 200 column utilized in our lab has an exclusion limit of 1.3 MDa and permeation limit of 6.5 kDa. Using gel filtration chromatography, the molecular weight of the proteins in their native states can be determined. In membrane protein studies the usage of detergent is common. Detergents will form micelles with these proteins, and when the protein elutes from the column its observed mass is often greater than its actual molecular mass. Therefore, the observed mass is the sum of that from the detergent molecules and from the protein molecule (17). For the determination of the molecular mass of an unknown protein, first, a set of protein

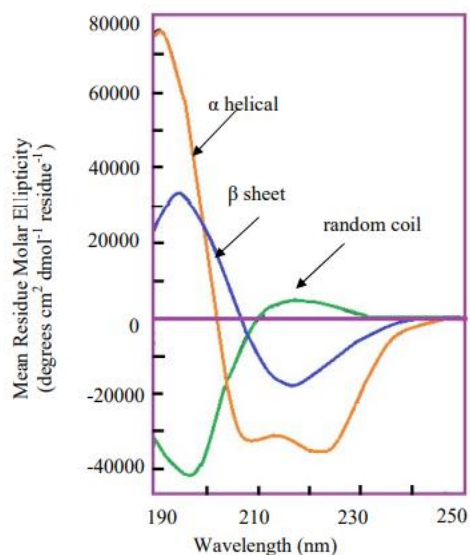
calibration standards of known molecular weight are run. Then the unknown protein sample can be run on the column and its partition coefficient can be calculated.

## 2.8. CD Spectroscopy

CD spectroscopy (**Figure 2.10**) is used to find global secondary structure of protein molecules. In CD spectroscopy, differential absorption of the left- and right- circularly polarized light absorbed by a chiral molecule is measured (21). A CD plot shows the mean residue molar ellipticity vs. wavelength. The mathematical relationship between molar ellipticity (in  $\text{deg}\cdot\text{cm}^2/\text{dmol}$ )  $[\Theta]$  is given by,

$$[\Theta] = m^\circ \cdot M / (10 \cdot L \cdot C) \text{ (equation 2.1)}$$

where  $m^\circ$  is the millidegrees obtained experimentally,  $M$  is the molecular weight of the protein in g/mol,  $L$  is the path length of the cell in cm, and  $C$  is the concentration of the sample in g/L. Alpha-helix, beta-sheet, and random coil structures give characteristic shapes of CD spectrum in the region of 190-250 nm (21).



**Figure 2.10** Characteristic CD curves for different secondary structures of protein molecules.

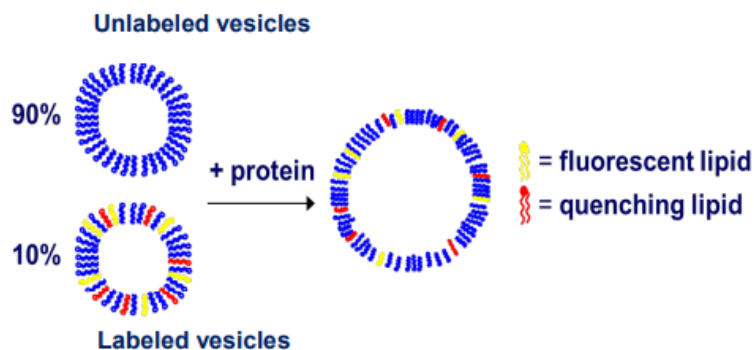
## 2.9. Western Blots

Western blotting is a useful technique to identify a specific protein from a complex mixture of proteins. In this technique, first gel electrophoresis (SDS PAGE) is used to separate proteins by mass. Then the proteins are then transferred to a nitrocellulose membrane. The next step is known as blocking, where a milk solution (dry milk solubilized in SDS solution) is added to the membrane to saturate the nitrocellulose paper with protein. Then an antibody which can specifically bind with the protein of interest is added and detection of the signal is carried out. In the normal procedure, first, a primary antibody that specifically recognizes the protein of interest is added, then a secondary antibody linked to a reporter enzyme is added. The secondary antibody binds to the primary antibody and produces a specific signal color such as, in the presence of specific substrate.

Most expressed proteins in our lab have a C-terminal histidine tag which aids in affinity purification. For the detection of these proteins, an anti-His antibody was used. This anti-His antibody is conjugated to horseradish peroxidase (HRP) enzyme, which reacts with a chemiluminescence substrate. This reaction results in a brown color, allowing the detection of the His-tagged proteins.

## 2.10. Lipid Mixing Assays

A fluorescence quenching assay (**Figure 2.11**) was used to compare the fusogenicity, the extent of lipid vesicle fusion, among different protein constructs. This assay depends on the interaction between a donor molecule and an acceptor molecule when the emission band of the donor molecule overlaps with the excitation band of the acceptor when the two molecules are in proximity.

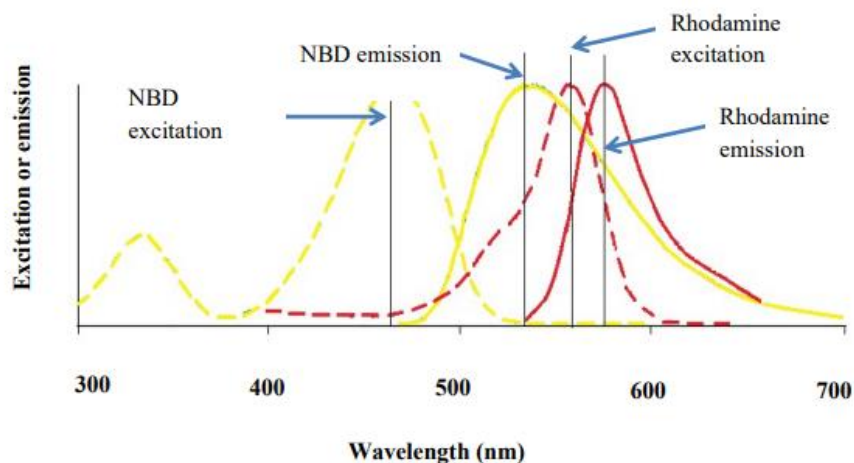


**Figure 2.11** Principle of lipid mixing. A small fraction of labeled vesicles containing both quenching lipid (acceptor) and fluorescent lipid (donor) is mixed with unlabeled vesicles.

Fusion associated dilution results in increased fluorescence signal.

The efficiency of this energy transfer is inversely proportional to the sixth power of the distance between donor and acceptor. For the fluorescence quenching assay used in this work, [N-(7-nitro-2,1,3-benzoxadiazol-4-yl) (ammonium salt) dipalmitoylphosphatidylethanolamine (N-NBD-PE) is the energy donor and [N-(lissamine rhodamine B sulfonyl) (ammonium salt) dipalmitoylphosphatidyl-ethanolamine] (N-Rh-PE) is the energy acceptor (37). The head group of the phosphatidylethanolamine lipid is modified to add either NBD fluorophore or rhodamine quencher (**Figure 2.12**).





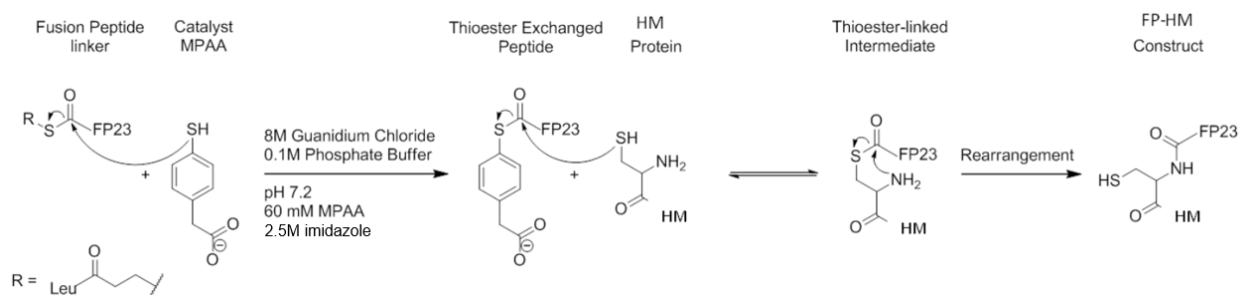
**Figure 2.12** Principle of fluorescence quenching assay. NBD excitation happens around 467 nm and emits around 530nm. When donor (NBD) and acceptor (Rhodamine) are in proximity emission signal of donor is effectively absorbed by acceptor molecule. In the presence of protein, vesicular fusion happens, resulting in increased donor acceptor distance. This increase of distance results in a decrease of quenching by the acceptor.

Increased fluorescence signal being observed is the result.

## 2.11. Native Chemical Ligation

Native chemical ligation (NCL) is an efficient technique to couple protein fragments (1). This method can be used for the synthesis of larger polypeptide by ligating synthetic peptides to a heterologously expressed protein (18,19). The principle (18) on which native chemical ligation works is the synthetic peptide is made with the thioester at the  $\alpha$ -carboxyl group that is attacked nucleophilically by the side chain of the N-terminal cysteine residue at the N-terminal of the expressed gp41 NHR+CHR+MPER (HM) protein (**Figure 2.13**). The thio-exchange product initially undergoes a rapid intramolecular reaction due to the favorable five-membered ring arrangement of the  $\alpha$ -amino group of the second peptide (expressed HM). This finally yields a product with a native peptide bond at the site of ligation. Both the ligated peptides (synthetic peptide +

expressed protein) are in unprotected form and needs no further manipulation. Some factors on which the rate of ligation depends include: (1) the nature of the thiol leaving group and (2) side chain steric conditions (3,5). The alkyl thioesters which are used in solid phase peptide synthesis are comparatively less reactive. This encourages the use of a catalyst in this reaction, and the rate of the reaction depends on the choice of the catalyst. It has been shown that 4-mercaptophenyl acetic acid (MPAA) catalyzes the ligation reaction to promote the in-situ formation of the more reactive thioester moiety (6,19).



**Figure 2.13** Mechanism of native chemical ligation. The FP23 that is synthesized by t-Boc SPPS is shown having a thiol on the C-terminus. To make the reaction more rapid, a catalyst MPAA is used. MPAA undergoes thioester exchange, and the HM with the N-terminal Cys can easily form thioester-linked intermediate. The more stable peptide bond is then formed following S>N acyl rearrangement.

To obtain the FP+HM (FP-HM) construct, the expressed HM protein was ligated to the FP. The FP was dissolved in the ligation buffer (8M Guanidine chloride, 2.5M imidazole, 0.1 M phosphate buffer, and pH 7). The peptide was reacted with MPAA catalyst to modify the thioester linked to the C-terminus of the fusion peptide (19,20), and the mixture was incubated for 30 min. Similarly, HM was dissolved in the ligation buffer including Tris(2-carboxyethyl)- phosphine (TCEP) (2mM) to maintain cysteine residue in

their reduced state, and the two solutions were mixed. The reaction was allowed to proceed for 2 days in an inert (Ar) atmosphere, stirring, and ambient temperature.

## **2.12. Solid State NMR Sample Preparation**

Some protein samples were prepared for NMR without lipid. In the case of non-lipid samples, the proteins were dialyzed against distilled, deionized water to remove salts, and the proteins were precipitated and centrifuged under vacuum to decant most of the water. The samples were then freeze dried and packed into an NMR rotor and rehydrated overnight.

For lipid containing samples, 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 a 8:2:5 mole ratio. The cholesterol mole fraction in both compositions is close to that of the plasma membrane of host cells of HIV (2,4,12). This composition was chosen because (1) the phosphocholine lipids are a major fraction in HIV-1 virus host cell plasma membrane, and (2) the host cell plasma membrane has a charge of approximately -1, which is mimicked by the 4:1 DPPC:DPPG charge (1,2,12). Lipid (50  $\mu$ mol) was dissolved in 2 mL of a 9:1 chloroform and methanol solution, and the solvent was then removed by dry nitrogen gas flow and vacuum pumping overnight. Hydration of the lipid film was done using 3 mL of 10 mM 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES) and 5mM 2-(N-morpholino) ethane sulfonic acid (MES) buffer at pH 5.0 (39) and was followed by 10 freeze-thaw cycles to make a homogenous suspension of unilamellar vesicles. The lipid-buffer suspension was extruded 10 times through a polycarbonate membrane with 100 nm pore size to get large unilamellar vesicles in the retentate. Small unilamellar

vesicles (SUV) are typically below 100 nm diameter, large unilamellar vesicles (LUV) are between 100 nm and 1  $\mu$ m diameter, and giant unilamellar vesicles (GUV) are above 1  $\mu$ m diameter. Target protein (1  $\mu$ mol) was dissolved in 10 mL of the HEPES/MES buffer and was added to lipid vesicles drop by drop, then agitate overnight. The peptide-lipid complex was pelleted down by ultra-centrifugation at 160000g for 4 h. The proteo-liposome complex pellet was lyophilized overnight. Lyophilization helps to reduce sample lost when the sample is packed into NMR rotor. The sample was packed into the NMR rotor and rehydrated with 10  $\mu$ L of the HEPES/MES buffer overnight at room temperature.

### **2.13. Solid State NMR**

Rotational Echo Double Resonance (REDOR) solid state NMR was used in this study. Spectra were obtained from a 9.4 T Agilent Infinity Plus spectrometer and triple-resonance MAS probe tuned to  $^1\text{H}$ ,  $^{13}\text{C}$ , and  $^2\text{H}$  frequencies.

For all REDOR experiments, the rotor was spun at 10 kHz with bearing gas and -50°C, cooling by nitrogen gas flow and the expected sample temperature is about -30°C. The REDOR experiment collects two sets of data:  $S_0$  and  $S_1$ . In both experiments (1)  $^1\text{H}$   $\pi/2$  pulse; (2)  $^1\text{H}$ -to- $^{13}\text{C}$  cross-polarization (CP); (3)  $^1\text{H}$  decoupling; (4)  $^{13}\text{C}$   $\pi$  pulses at the end of each rotor period; and (5)  $^{13}\text{C}$  detection.  $^2\text{H}$   $\pi$  pulses are applied in the middle of each rotor period in  $S_1$  experiment but absent in  $S_0$  experiment. The spectra were processed with 100 Hz Gaussian line broadening and referenced to the adamantane methylene chemical shift, which is 40.5 ppm. The  $S_0$  and  $S_1$  intensities were calculated with a 3-ppm width of the  $^{13}\text{CO}$  peak. The uncertainties were the RMSD's of 6 spectral noise regions with 3-ppm widths.

I4 peptide AEAAAKEAAAKEAAKAW which has a regular helical structure was used as a standard sample. The reference I4 peptide was  $^{13}\text{CO}$  labeled at A9 and  $^2\text{H}_\alpha$  labeled at A8. The isolated  $^{13}\text{CO}$ - $^2\text{H}$  spin pairs all have the same separation  $r$  of 5.0 Å and with a corresponding dipolar coupling  $D$  of 37 Hz.

## REFERENCES

1. 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 gp41 with predominant  $\beta$  sheet fusion peptide confirmation. *Biochemica et Biophysica Acta (BBA) – Biomembranes*, 2015, 1848(1, Part B): p. 289-298.
2. 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.
3. Benvenuti, M. and S. Mangani, Crystallization of soluble proteins in vapor diffusion for x-ray crystallography. *Nature Protocols*, 2007. 2(7): p. 1633-1651.
4. Sackett, K., Wexler-Cohen, Y., and Shai, Y. (2006) Characterization of the HIVN-terminal fusion peptide-containing region in context of key gp41 fusion conformations. *Journal of Biological Chemistry* 281, 21755-21762
5. Merrifield, R. B. (1963) Solid Phase Peptide Synthesis .1. Synthesis of a Tetrapeptide. *J Am Chem Soc* 85, 2149
6. 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
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. Iost, I., J. Guillerez, and M. Dreyfus, Bacteriophage-T7 RNA-Polymerase travels far ahead of ribosomes invivo. *Journal of Bacteriology*, 1992. 174(2): p. 619-622.
9. Paternal, S. and R. Komel, Active Protein Aggregates Produced in Escherichia coli. *International Journal of Molecular Sciences*, 2011. 12(11): p. 8275-8287
10. Taylor, G., et al., Size and density of protein inclusion-bodies. *Bio-Technology*, 1986. 4(6): p. 553-557.
11. Singh, A., et al., Protein recovery from inclusion bodies of Escherichia coli using mild solubilization process. *Microbial Cell Factories*, 2015. 14.
12. 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.

13. 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 inclusion bodies in whole bacterial cells and structural characterization of purified and membrane associated Fgp41. *Biochemistry* 2011. 50 p. 10013-10026

14. 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.

15. 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.

16. Bornhorst, J.A. and J.J. Falke, [16] Purification of Proteins Using Polyhistidine Affinity Tags. *Methods in Enzymology*, 2000. 326: p. 245-254.

17. 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.

18. Dawson, P. E., Muir, T. W., Clarklewis, I., and Kent, S. B. H. (1994) Synthesis of proteins by native chemical ligation. *Science* 266, 776-779

19. Johnson, E. C. B., and Kent, S. B. H. (2006) Insights into the mechanism and catalysis of the native chemical ligation reaction. *J Am Chem Soc* 128, 6640-6646

20. 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

21. Greenfield, N. and Fasman, G. D. (1969) Computed Circular Dichroism Spectra for Evaluation of Protein Conformation. *Biochemistry*, 1969. 8, 4108

22. Liu, J., Deng, Y. Q., Dey, A. K., Moore, J. P., and Lu, M. (2009) Structure of the HIV-1 gp41 Membrane-Proximal Ectodomain Region in a Putative Prefusion Conformation. *Biochemistry* 48, 2915-2923

23. 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

24. 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

25. 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
26. Buzon, V., Natrajan, G., Schibli, D., Campelo, F., Kozlov, M. M., and Weissenhorn, W. (2010) Crystal Structure of HIV-1 gp41 Including Both Fusion Peptide and Membrane Proximal External Regions. *Plos Pathogens* 664
27. Song, L., Sun, Z.-Y. J., Coleman, K. E., Zwick, M. B., Gach, J. S., Wang, J.-h., Reinherz, E. L., Wagner, G., and Kim, M. (2009) Broadly neutralizing anti-HIV-1 antibodies disrupt a hinge-related function of gp41 at the membrane interface. *PNAS* 106, 9057–9062
28. Sackett, K., Nethercott, M. J., Epand, R. F., Epand, R. M., Kindra, D. R., Shai, Y., and Weliky, D. P. (2010) Comparative Analysis of Membrane-Associated Fusion Peptide Secondary Structure and Lipid Mixing Function of HIV gp41 Constructs that Model the Early Pre-Hairpin Intermediate and Final Hairpin Conformations. *Journal of Molecular Biology* 397, 301–315
29. Bodner, M. L., Gabrys, C. M., Parkanzky, P. D., Yang, J., Duskin, C. A., and Weliky, D. P. (2004) Temperature dependence and resonance assignment of C-13 NMR spectra of selectively and uniformly labeled fusion peptides associated with membranes. *Magn Reson Chem* 42, 187-194
30. Yang, J., and Weliky, D. P. (2003) Solid-state nuclear magnetic resonance evidence for parallel and antiparallel strand arrangements in the membrane-associated HIV-1 fusion peptide. *Biochemistry* 42, 11879-11890
31. Yang, J., Parkanzky, P. D., Bodner, M. L., Duskin, C. A., and Weliky, D. P. (2002) Application of REDOR subtraction for filtered MAS observation of labeled backbone carbons of membrane-bound fusion peptides. *J Magn Reson* 159, 101-110
32. Zhang, H. Y., Neal, S., and Wishart, D. S. (2003) RefDB: A database of uniformly referenced protein chemical shifts. *J Biomol Nmr* 25, 173-195
33. Gullion, T. (1998) Introduction to rotational-echo, double-resonance NMR. *Concept Magnetic Res* 10, 277-289
34. Alberts, B., et al., Molecular Biology of the Cell. *Third Edition ed.* 1994.
35. Novagen. Inc, pET-24a-d(+) Vectors, in pET-24a-d(+) Technical Bulletin
36. Wang, L., Towards revealing the structure of bacterial inclusion bodies. *Prion*, 2009. 3(3): p. 139-145



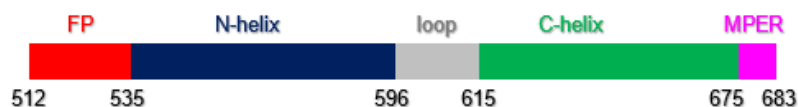
37. Struck, D.K., D. Hoekstra, and R.E. Pagano, Use of Resonance energy transfer to monitor membrane fusion. *Biochemistry*, 1981. 20: p. 4093-4099.
38. Weliky, D. *Chemical and Engineering News, Science/Technology*, **2008** 86:41
39. Yamaguchi, T. Koga, M. Fujita, Y. Kimoto, E. Effect of pH on Membrane Fluidity of Human erythrocytes. *J Biochem.* 1982 Apr;91(4): 1299-304.
40. RCSB, RCSB Protein Data Bank. [RCSB PDB: Homepage](#) (2022)
41. Raffaello Verardi, Nathaniel J Traaseth, Larry R Masterson, Vitaly V Vostrikov, Gianluigi Veglia. Isotope labeling for solution and solid-state NMR spectroscopy of membrane proteins. *Adv. Exp. Med Biol.* **2012**; 992: 35-62.

**CHAPTER 3:**  
**PRODUCTION AND ISOTOPIC LABELING OF A LARGE GP41 ECTODOMAIN**  
**CONSTRUCT BY NATIVE CHEMICAL LIGATION BETWEEN THE FUSION PEPTIDE**  
**AND SOLUBLE ECTODOMAIN**

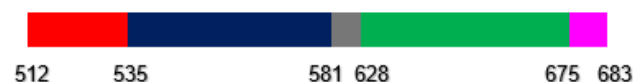
### 3.1. Introduction

Human immunodeficiency virus (HIV) is a membrane-enveloped virus whose initial infection of host cells begins with membrane fusion through a process initiated by gp160 (3,6,7). The gp160 glycoprotein complex is comprised of two noncovalently associated subunits, gp120 and gp41. The membrane of the virion has clusters of three gp160 comprised of three gp120 and three gp41 molecules. HIV gp41 has three major domains. The ectodomain, the transmembrane domain, and the cytoplasmic domain work together to mediate the membrane fusion functions of the gp41 protein. The gp41 is originally part of a larger noncovalent trimer of heterodimer complex with gp120. The gp120 recognizes target host cells by binding to CD4 and co-receptor proteins which causes dissociation from the gp41 ectodomain and subsequent structural rearrangement of the ectodomain. The ectodomain itself is subdivided into different domains with defined structures and functions. **Figure 3.1** displays the various protein ectodomain constructs we are investigating for this study. The domains include the fusion peptide (FP), the N-helical (NHR), and C-helical regions (CHR), a truncated loop, and the membrane-proximal external region (MPER). The FP includes 16 apolar residues at the N-terminus. Virus/cell fusion is impaired when there are deletions and mutations present in the FP. The “N-helical” and “C-helical” regions are each ~60-residue continuous helices in the final state structure. The MPER is proposed to bind to the viral membrane and is adjacent to the transmembrane domain. Increased rates and extents of vesicle fusion have been experimentally observed in ectodomain constructs containing both the FP and the MPER however, there is little evidence of close contact between the FP and MPER. The synergistic effects of the FP and MPER together in increasing the final fusion extent in lipid vesicles indicates that there may be some interactions between these domains (48).

### gp41 ectodomain



### FP\_HM protein

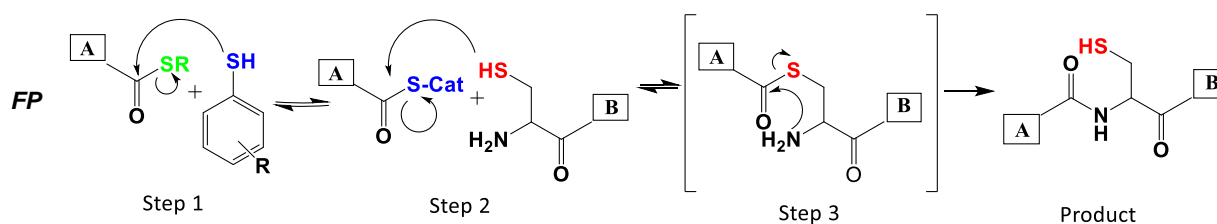


### FP\_HM sequence

AVGIGALFLGFLGAAGSTMGARACTLTVQARQLLSGIVQQQNNLLRAIEAQQHLLQLTVWGIKQLQARIL -  
SGGRGGMEWDREINNYSLSLIHSLIEESQNQKEKNEQELLELDKWASLWNWFNTNWLWYIK

**Figure 3.1** Schematic diagram of the HIV gp41 ectodomain and the related FP\_HM construct with domains color-coded as: FP  $\equiv$  fusion peptide, red; N-helix, blue; loop, gray; C-helix, green; and MPER  $\equiv$  membrane-proximal external region, pink. The residue numbering is for gp160 (with gp120 and gp41 subunits) from the HXB2 laboratory strain of HIV, with 535 and 596 as the approximate termini of the N-helix and 615 and 675 as the approximate termini of the C-helix of the helical hairpin structure of the soluble ectodomain. In some other studies, the MPER starts around residue 662. For the FP\_HM and HM constructs, residues 582-627 from the native N-helix, loop, and C-helix structural regions are replaced by non-native SGGRGG. Protein solubility is improved with this replacement, and helical hairpin structure and hyperthermostability of the soluble ectodomain are retained, with  $T_m > 100$  °C. The FP\_HM sequence also has the S534A and M535C mutations which are needed for native chemical ligation between the FP and HM protein segments. The non-native residues are underlined in the FP\_HM sequence.

The native chemical ligation reaction was introduced in 1994 and is the selective formation of a native amide bond between two protein fragments in aqueous solution (4). The NCL reaction is typically between a fragment A with a C-terminal thioester and a fragment B with a N-terminal cysteine. **Figure 3.2** displays a proposed literature mechanism for the NCL reaction. Step 1 is a nucleophilic attack by a thiol catalyst on the thioester carbonyl of peptide A to produce an activated thioester. Step 2 is transthioesterification between the activated peptide A and the N-terminal cysteine sidechain of peptide B. Step 3 is amine nucleophilic attack on the thioester carbonyl, followed by five-membered ring formation that occurs in the transition state at the thioester carbonyl. The resulting native amide bond that form is irreversible. Computation supports that both the thiol-thioester exchange and the transthioesterification proceed through a concerted anionic 2<sup>nd</sup> order nucleophilic substitution (SN2) mechanism, and that rearrangement goes by addition-elimination mechanism. The deprotonated thiol acts as the anion (5). To our knowledge, no ligation intermediates have been isolated experimentally at this time.



**Figure 3.2** Possible mechanism of native chemical ligation between protein segment A with C-terminal thioester and protein segment B with N-terminal Cys, with catalysis by an aryl thiol. For the present study, A is the HIV gp41 FP with (CO)S(CH<sub>2</sub>CH<sub>2</sub>COOH) thioester, B is the HM construct of the gp41 soluble ectodomain, and the catalyst is 4-mercaptophenylacetic acid (MPAA).

### **Figure 3.2 (cont'd)**

Step 1 is nucleophilic attack by the thiol group of the catalyst MPAA and exchange with the FP thioester. Step 2 is nucleophilic attack by the Cys thiol and exchange with the MPAA thiol. Step 3 is nucleophilic attack by the Cys amine and intramolecular rearrangement to form an amide bond.

The main benefit of the solid phase peptide synthesis (SPPS) reaction relative to biosynthesis is the addition of amino acids one at a time with the possibility of specific and selective labeling of individual residues. Any of these single amino acids may be specifically labeled at one or more atoms in the amino acid. The NCL reaction is then used in combination with the SPPS to synthesize larger protein constructs that retain the labeling selectivity of the SPPS fragment. In contrast, protein synthesis solely by bacteria is not as selective for residue-specific labeling as chemical synthesis. This reduced selectivity of isotopic labeling can create some ambiguity in the assignment of the signals in the solid state nuclear magnetic resonance (SSNMR) spectra.

Multiple proteins have been successfully synthesized by NCL ranging across the different classes of proteins including redox proteins, intracellular proteins, membrane proteins, and enzymes (10,11,14,19). Therefore, the methods described in this paper can have broad application across several proteins where SSNMR measurements are desired (36-38).

This study describes the native chemical ligation synthesis of the gp41 ectodomain FPHM including separation of FPHM protein from the HM segment. The data presented here explores some aspects of the reactivity of the HM protein towards NCL. These data also provide a synthetic blueprint for future experimentation with segmentally, or single

atom isotopically labeled gp41 ectodomain. Our group has previously explored NCL with smaller gp41 constructs, we present here the largest constructs of gp41 produced from NCL to date.

### **3.2. Materials and Methods**

Materials were purchased from the following companies: DNA – GenScript (Piscataway, NJ); Escherichia coli BL21(DE3) strain – Novagen (Gibbstown, NJ); Luria-Bertani (LB) medium- Dot Scientific (Burton, MI); isopropyl  $\beta$ -D-thiogalactopyranoside (IPTG) and tris-(carboxyethyl) phosphine (TCEP) – Goldio (St. Louis, MO); Co<sup>2+</sup>-resin – Thermo Scientific (Waltham, MA); D-Glucose-1,2,3,4,5,6-<sup>13</sup>C<sub>6</sub> (98%) – Synthese (Ontario, Canada). Most other materials were obtained from Sigma-Aldrich (St. Louis, MO).

#### **3.2.1. Fusion Peptides**

FP peptides used in this study include FP  $\equiv$  gp41<sub>512-534(S534A)</sub> and non-native C-terminal (CO)S(CH<sub>2</sub>CH<sub>2</sub>COOH) for ligation with HM. The S534A mutation reduces sidechain volume and may increase ligation rate (1,2). In many cases, a non-native N-terminal H<sub>6</sub>G<sub>6</sub>D<sub>4</sub>K tag was included, which consists of H<sub>6</sub> for FPHM binding to Co<sup>2+</sup>-resin and D<sub>4</sub>K for enterokinase cleavage of FPHM from the resin. G<sub>6</sub> is an unstructured spacer that increases exposure of the H<sub>6</sub> and D<sub>4</sub>K segments. Increased exposure of the H<sub>6</sub> tag prevent the possibility of the tag being hidden from the resin by other parts of the protein.

Solid-phase peptide synthesis was done manually using t-butoxycarbonyl (t-boc) chemistry and S-trityl- $\beta$ -mercaptopropinoyl-p-methyl-benzhydrylamine resin (230 mg, 0.88 meq/g). Sidechain protecting groups for amino acids including: His, and Arg, tosyl; Asp benzyl ester; Lys, carboxybenzyl; Ser, and Thr benzyl. A liquid or liquid solution with

reagents was added to the resin in a 40 mL Teflon vessel with cap, filter, stopcock, and nozzle, followed by shaking the vessel, and then drainage of the liquid from the vessel. Synthesis began with resin-swelling in  $\text{CH}_2\text{Cl}_2$  (3 mL, 1 h) followed by trityl-group cleavage in 95:2.5:2.5 (v:v) TFA:H<sub>2</sub>O:triisopropylsilane (10 mL, 4 minutes, 2×).

The first cycle of coupling amino acid (Ala-534) to the resin began with resin washing with  $\text{CH}_2\text{Cl}_2$  (3 mL, 1 min, 5×) and then 5% N,N-diisopropylethylamine (DIEA) in  $\text{CH}_2\text{Cl}_2$  (3 mL, 1 min, 3×), with concurrent reaction in a flask between t-boc-Ala (6.8 mmole) and activator 3-(diethoxyphosphoryloxy)-1,2,3-benzotriazin-4(3H)-one (DEPBT) (6.4 mmole) in tetrahydrofuran (THF, 5.5 mL). DIEA (1.1 mL) was added, and the total solution used to couple Ala-534 to the resin (4 h), followed by washes with  $\text{CH}_2\text{Cl}_2$  (3 mL, 1 min, 5×).

Deprotection (t-boc cleavage) was done with 50:48:2 (v:v:v) TFA: $\text{CH}_2\text{Cl}_2$ :anisole (3 mL) followed by coupling of Arg-533 to the resin using a similar procedure as above, but with 3.4 mmol amino acid and 3.2 mmol DEPBT in 2.8 mL THF. There were then sequential deprotection/coupling cycles to complete the synthesis. There was a final deprotection followed by washing with  $\text{CH}_2\text{Cl}_2$  (3 mL, 1 min, 5×) and 5% DIEA in  $\text{CH}_2\text{Cl}_2$  (3 mL, 1 min, 3×), and drying overnight in a vacuum desiccator. Peptide was cleaved from the resin using HF at Midwest Biotech Corporation (Fishers, IN).

### **3.2.2. Molecular Biology and Protein Expression and Separation**

The FP<sub>HM</sub> amino acid sequence is shown in **Figure 3.1** and is based on the HXB2 laboratory strain of HIV. The FP sequence is gp41<sub>512-534(S534A)</sub> and the HM sequence is gp41<sub>535-581(M535C)</sub>-SGGRGG-gp41<sub>628-683</sub>. Like the gp41 soluble ectodomain (SE), HM adopts helical hairpin structure that is hyperthermostable with melting



temperature,  $T_m > 100\text{ }^{\circ}\text{C}$ . Replacement of residues 582-627 with SGGRGG improves solubility (35,46). The N-terminal Cys of HM is used for ligation with the C-terminal thioester of FP.

Several related proteins were produced by expression in *E. coli* bacteria, BL21(DE3) strain, using vectors with inserts. Proteins included HM\_G4H4 in the pGEM-t vector and HM\_G6LEH6 in the pET-24a(+) vector, which have been previously described. HM was produced in the pET-24a(+) vector after replacement of the first Gly codon with a stop codon. Mutants HM<sup>3</sup>\_G4H4 (W628A, W631R, D632A) and HM<sup>4</sup>\_G4H4 (W628A, W631R, D632A, Q652A) were produced in the pGEM-t vector by sequential site-directed mutagenesis. The mutations in the C-helix region were intended to destabilize helical hairpin structure by reducing binding interactions between C- and N-helices.

H<sub>10</sub>S<sub>2</sub>GHID<sub>4</sub>KHM\_FPHM was produced in the pET-19b vector using sub-cloning with addition of 5 ng insert DNA coding for FPHM to a 50  $\mu\text{L}$  suspension of *E. coli* competent-cells, incubation on ice for 30 min, heat shock in a 42  $^{\circ}\text{C}$  bath without shaking for 50 s, and then ice for 2 min. LB medium (450  $\mu\text{L}$ ) was added followed by incubation at 37  $^{\circ}\text{C}$  for 1 h. The cell suspension (20-200  $\mu\text{L}$ ) was added to kanamycin-resistant selective plates at 37  $^{\circ}\text{C}$ , followed by incubation overnight. A single colony from a plate was added to a flask that contained 25 mL LB medium and 50 mg/mL kanamycin, followed by growth to  $\text{OD}_{600} \approx 0.8$ . The plasmid DNA was isolated and purified from an aliquot of the suspension with a Wizard Plus Minipreps kit (Promega – Madison, WI), with subsequent sequencing that confirmed the insert. The remaining suspension was divided into 1 mL aliquots with subsequent addition of 0.6 mL 50% glycerol, and storage at -80  $^{\circ}\text{C}$ .

Production of unlabeled protein began with addition of 1.5 mL *E. Coli* glycerol stock to 50 mL LB medium followed by growth for 3 h. This and all other growths were done at 37 °C in shake flasks at 180 rpm, and with 50µL of 50 mg/mL kanamycin in the medium. The culture was added to 1 L fresh LB medium, followed by: (1) growth for 2 h to OD<sub>600</sub> ≈ 0.8; (2) addition of 2 mM IPTG and induction of protein expression overnight at 37 °C; and (3) harvesting the cell pellet after centrifugation at 9000g for 10 min.

Production of labeled protein began with growth in LB for 3 h followed by harvesting the cell pellet. Minimal medium was prepared by mixing autoclaved aqueous solutions that included: (1) 50 mL with M9 salts (0.34 g Na<sub>2</sub>HPO<sub>4</sub>, 0.15 g KH<sub>2</sub>PO<sub>4</sub>, 0.03 g NaCl, and 0.05 g NH<sub>4</sub>Cl); (2) 1 mL with 0.1 M CaCl<sub>2</sub>; and (3) 1 mL with 1 M MgSO<sub>4</sub>, as well as 0.5 mL minimal essential medium (MEM) vitamin, and 200 mg glucose. The cell pellet was suspended in 10 mL minimal medium, and a 2 mL aliquot of this suspension was then added to the remaining minimal medium, followed by 4 h growth. Individual 5 mL aliquots of the suspension were then added to four separate flasks that each contained 50 mL fresh minimal medium, followed by 2 h growth, induction of expression, and harvesting of the cell pellet. Cells in this step of the growth process could be made into glycerol stock in an analogous manner to the LB glycerol stocks. The minimal media glycerol stocks could then be used to grow cells in minimal media to O.D. 0.8 followed by induction with IPTG and expression. Protein with fractional <sup>13</sup>C-labeling was produced using a minimal medium containing a mixture of unlabeled- and 1,2,3,4,5,6-<sup>13</sup>C D-glucose. Protein with fractional <sup>2</sup>H-labeling was produced using minimal medium that contained 1,2,3,4,5,6-<sup>2</sup>H D-glucose and a mixture of H<sub>2</sub>O and D<sub>2</sub>O. The cells were grown in labeled media, and

therefore all protein produced by the *E. coli* under these conditions were isotopically labeled.

Separation of inclusion body-rich material began with: (1) tip-sonication in ~30 mL PBS per 10g cells at pH 7.4 in a 50 mL beaker in an ice bath; (2) centrifugation at 27000g for 30 min and harvesting the new pellet; and (3) 2× repetition of the sonication/centrifugation/harvesting steps with a resultant pellet 1. The next step was tip sonication of the pellet 1 in ~30 mL PBS at pH 7.4 with 6 M GuCl, followed by centrifugation. Much of pellet 1 was solubilized, as evidenced by a new pellet: pellet 1 volume ratio  $< \frac{1}{2}$ . The supernatant was dialyzed against deionized water overnight at 10 °C, using a 10 kDa cutoff membrane, with accompanying precipitation of HM-enriched material. This was followed by centrifugation at 11700g for 40 min, and the harvesting of pellet 2. There were two variants of the remaining procedure that are denoted A and B, where protocol A includes an additional purification step.

For protocol A, much of pellet 2 was dissolved by vortexing in ~10 mL PBS at pH 8 with 8 M urea, followed by centrifugation at 11700g for 40 minutes followed by dialysis at 10 °C of the supernatant against water (one day total with three changes to fresh water) with accompanying precipitation by adding 0.2g NaCl to initiate precipitation. Precipitation is usually performed after transferring protein in solution to a conical vial. The suspension was centrifuged and pellet 3A was harvested. For protocol B, pellet 2 was desalted by vortexing in ~50 mL fresh deionized water, followed by centrifugation and harvesting pellet 3B. Pellets 3A and 3B were sometimes lyophilized and stored at low temperature prior to using them in ligations or NMR experiments. Cells that expressed HM with a histidine tag were sometimes subjected to a previously published protocol which included

sonication in PBS, centrifugation, and harvesting the pellet (3×), solubilization of the final pellet in PBS with 6 M GuHCl, and Co<sup>2+</sup>-affinity chromatography.

### **3.2.3. Characterization of Proteins by Mass Spectrometry and Solid-State NMR (SSNMR) Spectroscopy**

Matrix assisted laser desorption ionization time of flight mass spectrometry (MALDI-TOF-MS) was done using a Kratos Analytical Axima-CFR Plus instrument. Protein (~0.1 mg) was vortexed in 1 mL of 98% formic acid, and a 2 µL aliquot was then mixed by pipette with 4 µL of a solution containing α-cyano-4-hydrocinnamic acid (10 mg/mL) in 3:1 acetonitrile:0.1% TFA. A 2 µL aliquot of the mixed solution was transferred to the MALDI plate, dried, and then subjected to MALDI-TOF in linear positive mode. There were often M<sup>+</sup> and M<sup>2+</sup> signals, with assignments to a single chemical species done using  $(m/z)_{M^+} \approx 2 \times (m/z)_{M^{2+}}$ .

SSNMR experiments were done using a 9.4 T Agilent Infinity Plus spectrometer and a magic angle spinning (MAS) probe equipped for a 4 mm diameter rotor and tuned simultaneously to <sup>1</sup>H, <sup>13</sup>C, and <sup>2</sup>H NMR frequencies. Rotational-echo double-resonance (REDOR) data were acquired using a pulse sequence: (1) <sup>1</sup>H π/2 pulse; (2) <sup>1</sup>H-<sup>13</sup>C cross polarization (CP); (3) dephasing period of duration τ; and (4) <sup>13</sup>C detection. S<sub>0</sub> and S<sub>1</sub> REDOR data were acquired alternately and differed in the pulses applied during the dephasing period. For both S<sub>0</sub> and S<sub>1</sub> there was a <sup>13</sup>C π pulse at the end of each rotor cycle except for the last one and for S<sub>1</sub>, there was also a <sup>2</sup>H π pulse at the midpoint of each cycle. Typical parameters included: (1) 10 kHz MAS frequency and 1.5 ms CP contact time; (2) 50 kHz rf fields for <sup>1</sup>H π/2 pulse and CP; (3) 55-66 kHz <sup>13</sup>C CP ramp; (4) 60 kHz <sup>13</sup>C π pulses, 100 kHz <sup>2</sup>H π pulses with XY-8 phase cycling applied to all π pulses;

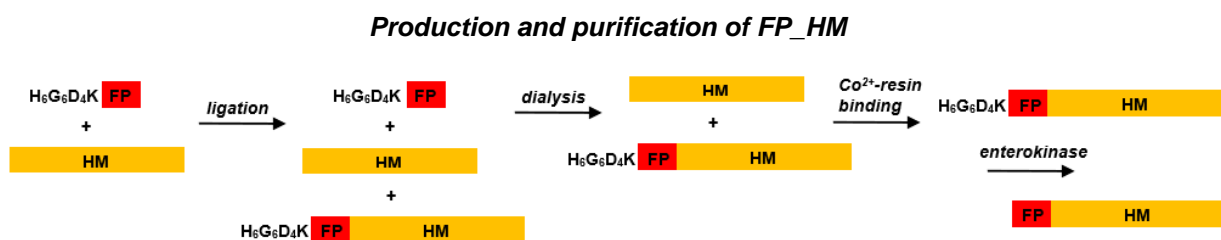
and ~70 kHz two-pulse phase-modulated  $^1\text{H}$  decoupling during dephasing and acquisition. Typical recycle delays were 1 s ( $\tau = 2, 8, 16$  ms), 1.5 s ( $\tau = 24, 32$  ms), and 2 s ( $\tau = 40$  and 48 ms). Typical numbers of summed  $S_0$  or  $S_1$  scans were ~4000, 7000, 12000, 25000, 32000, 40000, and 50000 for  $\tau = 2, 8, 16, 24, 32, 40$  and 48 ms, respectively.  $^{13}\text{C}$  chemical shift referencing was done externally using the methylene peak of adamantane at 40.5 ppm.

#### **3.2.4. Native Chemical Ligation and Purification**

Ligation buffer was prepared with GuHCl (19.1 g), imidazole (4.3 g), 2.11 mL of 1 M  $\text{Na}_2\text{HPO}_4$ , 0.43 mL of 1 M  $\text{NaH}_2\text{PO}_4$ , and  $\text{H}_2\text{O}$  in ~25 mL total volume. The solution was heated to 100 °C for 1 minute and vortexed, cooled to room temperature, and ~5 mL water then added to achieve clarity. Typical solute concentrations were: GuHCl, 6.4 M, imidazole, 2.0 M, and phosphate, 1.4 mM. The FP solution was typically prepared with a FP variant like  $\text{H}_6\text{G}_6\text{D}_4\text{K\_FP}$  (~1.2 mg, ~0.3  $\mu\text{mol}$ ) and 4-mercaptophenylacetic acid (MPAA) catalyst (7.5 mg, 45  $\mu\text{mol}$ ) dissolved in 0.9 mL ligation buffer followed by constant stirring for 30 minutes. A HM construct in ligation buffer (~5 mg/mL) was prepared and ~0.2 mL (~0.07  $\mu\text{mol}$ ) transferred to the FP/MPAA solution. The reducing agent tris-(carboxyethyl) phosphine (TCEP, 0.50 M) solution was prepared in ligation buffer and 0.2 mL then transferred to the FP + HM solution, followed by addition of ligation buffer to achieve a total volume of 1.5 mL with pH adjusted to 6.8. The approximate solute concentrations were  $\text{H}_6\text{G}_6\text{D}_4\text{K\_FP}$ , 200  $\mu\text{M}$ ; HM, 50  $\mu\text{M}$ ; MPAA, 30 mM; and TCEP, 67 mM. The ligation reaction was done by stirring the solution overnight, and pH typically increase to ~7.0 during this time.

**Figure 3.3** shows a separation scheme for HM from FPHM ligation product. Separation of ligation product began with overnight dialysis of the reaction against 8 M urea using a membrane with 10 kDa cutoff, which removed H<sub>6</sub>G<sub>6</sub>D<sub>4</sub>K\_FP. The H<sub>6</sub>G<sub>6</sub>D<sub>4</sub>K\_FP\_HM and HM were then precipitated by dialysis against water for one day, with one water change. The dialysis suspension was centrifuged (48000g, 20 min), and the solid pellet harvested and then solubilized by vortexing in 3 mL buffer that contained PBS, pH 7.4, and 6 M GuHCl. A 1 mL suspension of Co<sup>2+</sup>-resin in ethanol (HisPur, Thermo-Fisher) was added to a small plastic column, and excess ethanol removed by filtration. The resin was then washed with PBS, pH 7.4, with 8 M urea and 10 mM imidazole (1 mL, 3×) and then PBS, pH 7.4, with 6 M GuHCl (1 mL, 3×), followed by suspension in 3 mL of the latter buffer. The resin suspension was transferred to the protein solution, and the mixture then agitated overnight at 10 °C to achieve H<sub>6</sub>G<sub>6</sub>D<sub>4</sub>K\_FP\_HM binding to the resin. Excess liquid with unbound protein was removed by filtration, and the resin then washed with 6.7 M GuHCl, 2.1 M urea, and 0.5 M TCEP at pH 3.5 (1 mL, 1×), where the lower pH aided unbinding of HM from the resin. The resin was then washed with 50 mM tris(hydroxymethyl)aminomethane (Tris) buffer, pH 8.0, with 10 mM CaCl<sub>2</sub> and 0.1% tween-20 (1 mL, 4×). The resin was suspended in 3 mL of the latter buffer, and transferred by Pasteur pipette to a larger vial, and buffer added to achieve 5 mL total volume. Enterokinase (50 units/50 µL; EKMax, Thermo-Fisher) was added to the resin suspension followed by overnight incubation at 37 °C without agitation. The cleavage between D<sub>4</sub>K and FPHM released FPHM into the solution. The resin suspension was filtered, followed by washes with PBS buffer, pH 8.0, with 8 M urea (1 mL, 3×). The filtrate and washes were combined and then dialyzed against water which

precipitated the protein. The protein suspension was centrifuged (11600g, 40 min) and the protein pellet harvested.



**Figure 3.3** Schematic diagram of the production and purification of FP\_HM.

H<sub>6</sub>G<sub>6</sub>D<sub>4</sub>K\_FP and HM are produced and then reacted to produce H<sub>6</sub>G<sub>6</sub>D<sub>4</sub>K\_FP\_HM.

Unreacted H<sub>6</sub>G<sub>6</sub>D<sub>4</sub>K\_FP is removed by dialysis and H<sub>6</sub>G<sub>6</sub>D<sub>4</sub>K\_FP\_HM then preferentially bound to Co<sup>2+</sup>-resin via the H<sub>6</sub> tag. Enterokinase cleaves the peptide bond between K and N-terminal A of FP, and releases FP\_HM.

### 3.2.5. Anti-H<sub>5</sub> Western Blot

A cassette was prepared with a gel from SDS-PAGE, nitrocellulose membrane, foam inserts, and casing. Buffer (100 mL) was prepared with 250 mM Tris at pH 8.3, 1.94 M glycine, and 1% SDS (w/v) and then diluted 10× with 700 mL H<sub>2</sub>O and 200 mL MeOH. The cassette was placed in a tub with this solution, with cooling achieved by placing a plastic bottle filled with ice in the solution and placing the tub in a larger ice bath. Protein was transferred to the nitrocellulose membrane by applying 100 V for 1 h. The nitrocellulose membrane was shaken for 1 h in a solution made from mixing: (1) 200 mM Tris buffer at pH 8.0 with 1.37 M NaCl, 2.0 mL; (2) H<sub>2</sub>O, 18 mL; (3) 50% Tween-20, 0.04 mL; and (4) nonfat dried milk, 1.0 g. The membrane was then washed with the (1) + (2) + (3) solution without milk. The membrane was shaken for 1 h in 10 mL of the latter solution with 10 μL anti-H<sub>5</sub>-HRP-conjugate antibody, and then washed for 5 min with the solution

without antibody (4×). The membrane was then immersed in 5 mL Clarity ECL chemiluminescence substrate and analyzed by digital imaging software.

### **3.3. Results**

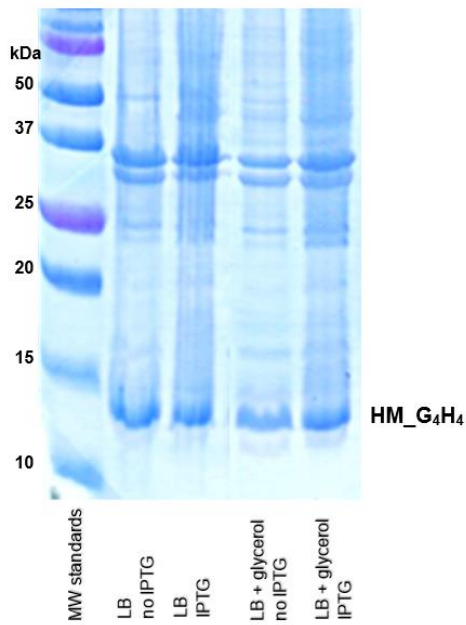
Our strategy for production of FPHM which includes the feature of the non-native H<sub>6</sub> tag on H<sub>6</sub>G<sub>6</sub>D<sub>4</sub>K\_FP rather than HM so that the H<sub>6</sub>G<sub>6</sub>D<sub>4</sub>K\_FPHM product but not HM would bind tightly to Co<sup>2+</sup>-resin during purification. FP peptides were synthesized with high purity using t-boc chemistry, as evidenced by mass spectrometry.

#### **3.3.1. IPTG-Controlled Expression and Bacterial Growth in Minimal Medium for Labeled Protein**

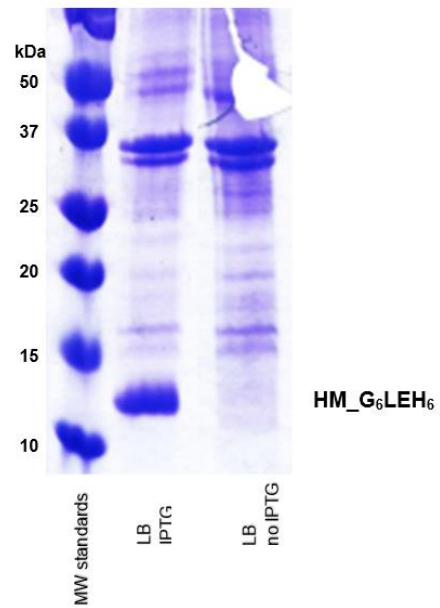
Controlled isotopic labeling of the HM protein is one goal of the project and requires strong regulation of expression by IPTG and high-yield expression in minimal medium. Initial expression was done in LB medium using *E. coli* cells with a previously created pGEM-t vector that produced HM\_G<sub>4</sub>H<sub>4</sub> **Figure 3.4**. After sonication in PBS with 6 M GuHCl, SDS-PAGE of cell lysates showed a ~13 kDa band whose intensity is approximately IPTG-independent, and inconsistent with IPTG-regulated expression. Regulation was demonstrated using *E. coli* cells with a different previously made pET-24a(+) vector that produced HM\_G<sub>6</sub>LEH<sub>6</sub>.



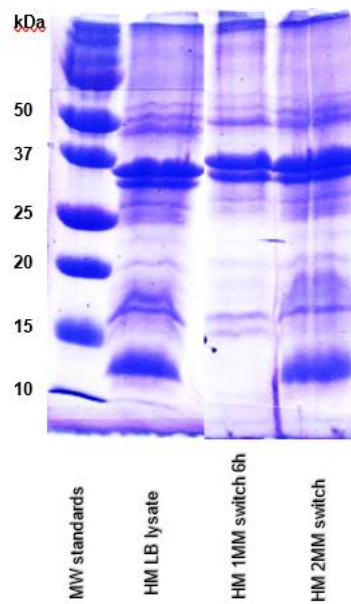
**A** SDS-PAGE of lysates from cells with pGEM-t vector expressing HM\_G4H4



**B** SDS-PAGE of lysates from cells with pET-24a(+) vector expressing HM\_G6LEH6



**C** Gel showing expression with 2 vs. 1 minimal medium switches.



**Figure 3.4** SDS-PAGE of lysates of *E. coli* cells that contained a vector with an insert for the HM protein.

### Figure 3.4 (cont'd)

The BL21(DE3) cells had either **(A)** a pGEM-t vector that expressed HM\_G<sub>4</sub>H<sub>4</sub> or **(B)** a pET-24a(+) vector that expressed HM\_G<sub>6</sub>LEH<sub>6</sub>. Cells were grown in 200 mL culture to OD<sub>600</sub> ≈ 0.8 followed by expression overnight, sonication of the cell pellet in PBS at pH 8.0 + 8 M urea, and then SDS-PAGE. The growth and expression medium always included LB and sometimes 10 g/L glycerol, and there was either no IPTG or 2 mM IPTG added during the expression period. After the MW standards, the left-to-right lanes of gel **A** have growth and expression conditions that include: LB, no IPTG; LB, IPTG; LB + glycerol, no IPTG; LB + glycerol, 2 mM IPTG. All lanes have a ~13 kDa band that is assigned to HM\_G<sub>4</sub>H<sub>4</sub> (MW = 13.7 kDa). There isn't much dependence of band intensity on the presence vs. absence of IPTG. After the MW standards, the left-to-right lanes of gel **B** have growth and expression conditions that include: LB, IPTG; LB, no IPTG. The lane for expression with IPTG has a ~13 kDa band that is assigned to HM\_G<sub>6</sub>LEH<sub>6</sub> (MW = 14.4 kDa). This band isn't apparent in the lane for expression without IPTG. **C** has growth and expression conditions that include: LB, IPTG; Minimal media (MM), IPTG, one switch; MM, IPTG, 2 switches. No expression is observed when using only one minimal media switch. A strong HM monomer band is observed when using 2 minimal media switches.

We have observed that bacterial growth is much slower in minimal vs. LB medium (**Figure 2.8**) and is also slower in fresh minimal medium after initial growth in minimal vs. LB medium. In the present study, bacterial growth was slower in minimal medium after (1) growth in LB medium and then switch to minimal medium vs. (2) growth in LB medium, switch to minimal medium, growth, and then switch to fresh minimal medium, i.e. a

doubling time of ~60 min vs. ~30 min for (1) vs. (2). The improved growth with protocol 2 could be due to better adaptation of the bacteria to synthesis of amino acids after two vs. one minimal medium switch. Protocol 2 was also important for high-yield expression, as evidenced by a ~12 kDa band in SDS-PAGE of the cell lysate vs. no band in the lysate from protocol 1 (**Figure 3.4C**). Minimal media cells were also made into glycerol stocks by following protocol 2 and growing cells in the second minimal media to an O.D. of 0.8 before making the stocks. A 1.6 mL glycerol stock added to 50 mL fresh minimal media will typically reach an O.D of 0.8 in ~5 h.

### **3.3.2. Separation of HM in Inclusion Bodies**

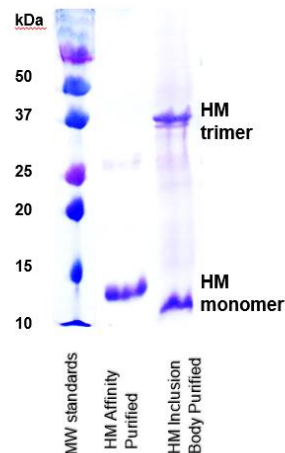
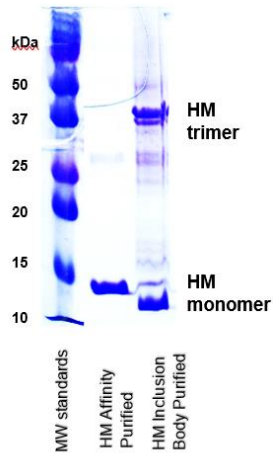
The pET-24a(+) plasmid was mutated with a stop codon to produce HM rather than HM\_G<sub>6</sub>LEH<sub>6</sub>, and the mutation was confirmed by DNA sequencing. Bacterial cultures were grown followed by HM expression. A purification protocol was developed to separate HM in inclusion bodies from the rest of the cellular material. The protocol included 3× cycles of sonication of the cell pellet in PBS, centrifugation, and harvesting the cell pellet. The final yellowish pellet was sonicated in PBS with 6 M GuHCl followed by centrifugation which resulted in a smaller yellow pellet and a clear supernatant. Dialysis of the supernatant against water resulted in the precipitation of white solid, followed by centrifugation, and harvesting of the pellet (protocol B). Further purification was sometimes done by solubilization of the pellet in PBS with 8 M urea, followed by centrifugation which resulted in a semi-solid pellet and clear supernatant (protocol A). The supernatant was dialyzed against water, with accompanying precipitation of HM-enriched material with addition of NaCl (approximately 10 mg/mL), followed by centrifugation and harvesting the pellet. SDS-PAGE of the protocol A or protocol B pellets solubilized in 8 M

urea show bands that are assigned to HM monomer and trimer (**Figure 3.5**). HM assignment was supported by MALDI mass spectrometry (**Figure 3.6**) and by subjecting gp41 protein bands from similar gels to trypsin digestion, followed by chromatographic separation and mass spectrometry of the peptides (**Figure 3.7**). A strong trimer band is unusual in SDS-PAGE but is consistent with our previous observation using size-exclusion chromatography of a significant trimer fraction for gp41 proteins in 0.2% SDS at pH 7.4 (59). The above protocols typically resulted in ~5 g wet cell mass/L culture and ~10 mg purified HM/L culture, with similar yields from HM expression in LB medium or in minimal medium with 4 g/L glucose.

### A Expression in LB medium

Expression in LB medium  
Protocol **A** separation

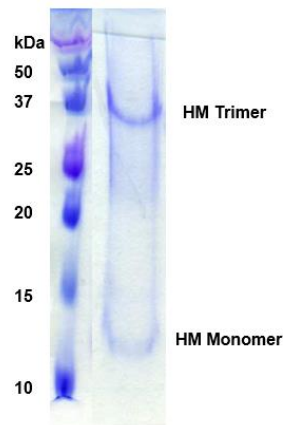
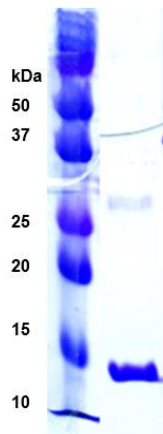
Expression in LB medium  
Protocol **B** separation



### B Expression in minimal medium

HM expressed in minimal media  
and purified by affinity  
chromatography

HM expressed in minimal media  
and purified by protocol **B**



**Figure 3.5** SDS-PAGE of separated material of *E. coli* cells that had expressed HM. For panel A, expression was in LB medium, and the protocol A separation was used with a final extraction into 8 M urea. For panel B, expression was in minimal medium with 4 g/L

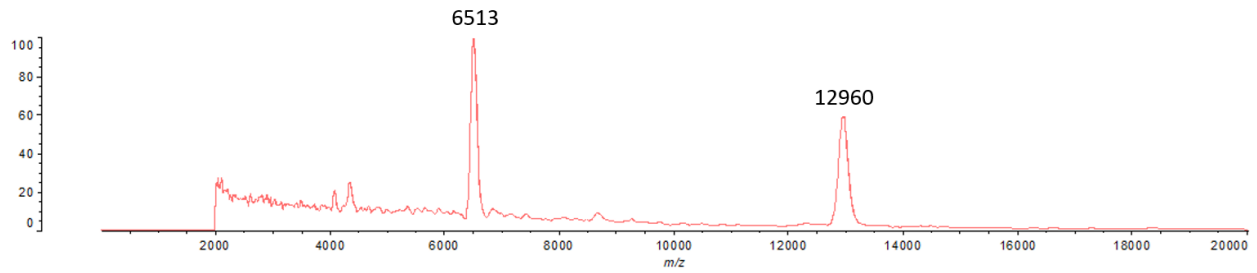
### Figure 3.5 (cont'd)

glucose and the protocol B separation was used without the final extraction. There is a prominent band in both gels with  $MW \approx 35$  kDa and a band in both gels with  $MW \approx 12$  kDa. Both bands are observed in gels of replicate separations and in gels of HM mutants. HM proteins containing the G<sub>6</sub>LEH<sub>6</sub> c-terminal affinity tag also show a faint band at  $\sim 26$  kDa. These bands are assigned to HM monomer ( $MW = 13.0$  kDa) HM dimer ( $MW = 26$  kDa) and HM trimer ( $MW = 38.9$  kDa). Bands in similar gels were subjected to trypsin digestion followed by chromatographic separation and mass spectrometry of the peptides. Many of the peptide masses matched the calculated masses of peptide segments of HM.

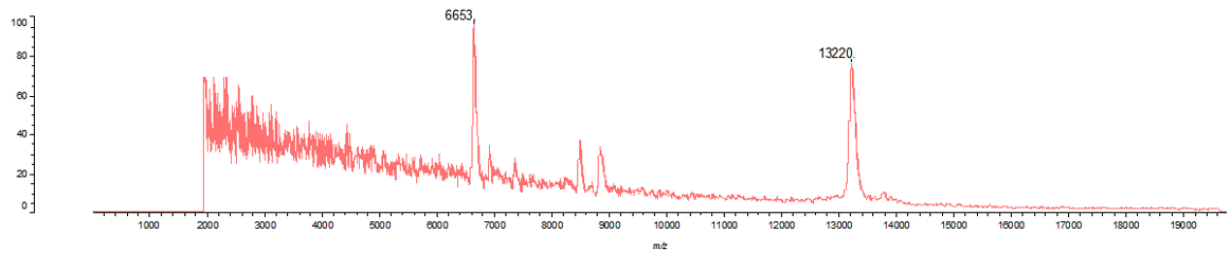
Assignment of the 35 kDa band to trimer is consistent with the trimer as a major species in earlier size-exclusion chromatography of HM constructs in SDS at neutral pH. Monomer bands are at different positions in the gel due to presence or absence of His-tag respectively.

## MALDI Mass Spectrometry

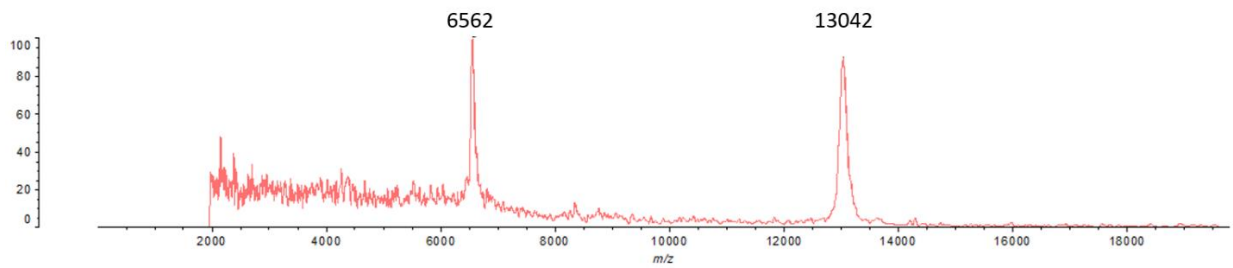
### A Unlabeled HM



### B $^2\text{H}$ labeled (scaling)



### C $^{13}\text{C}$ labeled



**Figure 3.6** MALDI mass spectra of material from separation of inclusion bodies using protocol A with a final extraction in 8 M urea. Expression of HM was in: (A) LB medium.

### Figure 3.6 (cont'd)

(B) minimal medium with 4 g/L 1,2,3,4,5,6,6-<sup>2</sup>H D-glucose and 50% D<sub>2</sub>O. (C) minimal medium with 3 g/L unlabeled and 1 g/L 1,2,3,4,5,6,6-<sup>13</sup>C D-glucose. The two dominant peaks are assigned to HM<sup>+</sup> and HM<sup>2+</sup> ions which for unlabeled HM have calculated m/z = 12962 and 6481, respectively. Peaks in the panel A spectra are assigned to HM<sup>+</sup> and HM<sup>2+</sup> ions based on their significant intensities and on similarities with the calculated mass values of 12962 Da and 6481 Da. The most intense peaks in the panel B and C spectra are also assigned to HM<sup>+</sup> and HM<sup>2+</sup>. The percent labeling is estimated as  $\Delta M_{\text{exp}}/\Delta M_{\text{lab}} \times 100$ , where  $\Delta M_{\text{exp}}$  is the difference between experimental m/z in labeled and unlabeled media, and  $\Delta M_{\text{lab}}$  is the difference in calculated m/z between fully labeled and unlabeled HM. For HM<sup>+</sup>,  $\Delta M_{\text{lab}}$  is 910 Da for <sup>2</sup>H labeling and 580 Da for <sup>13</sup>C-labeling and. The estimated <sup>2</sup>H labeling from panel B is 29% using  $\Delta M_{\text{exp, HM}^+}$  and 31% using  $\Delta M_{\text{exp, HM}^{2+}}$ . The estimated <sup>13</sup>C labeling from panel C is 14% using  $\Delta M_{\text{exp, HM}^+}$  and 17% using  $\Delta M_{\text{exp, HM}^{2+}}$ .



## Proteomics Data for FPHM

### A) Data for FPHM pet24a+ monomer

ENV\_HV1BR (100%), 97,489.7 Da

Envelope glycoprotein gp160 OS=Human immunodeficiency virus type 1 group M subtype B (isolate BRU/LAI) GN=env PE=1 SV=1  
4 exclusive unique peptides, 8 exclusive unique spectra, 95/861 amino acids (11% coverage)

MRVKEKYQHL	WRWGKWKWTM	LLGILMICS	TEKLWVTVY	GVPVWKEATT	TLFCASDAKA	YDTEVHNWVA	THACVPTDPN
PQEVVLNVVT	ENFNMWKNDM	VEQMHEDIIS	LWDQSLKPCV	KLTPLCVSLK	CTDLGNATNT	NSSNTNSSSG	EMMMEKGEIK
NCSFNISTSI	RGKVQKEYAF	FYKLDIIPID	NDTTSYTLTS	CNTSVITQAC	PKVSFEPIPI	HYCAPAGFAI	LKCNKTFNG
TGPCNTNVSTV	QCTHGIRPVV	STQLLLNGSL	AEEEVVIRSA	NFTDNAKTII	VQLNQSVIEIN	CTRPNNNTRK	SIRIQRGPGR
AFVTIGKIGN	MRQAHNCISR	AKWNATLKQI	ASKLREQFGN	NKTIIFKQSS	GGDPEIVTHS	FNCGGEFFYC	NSTQLFNSTW
FNSTWSTEGS	NNTGSDTIT	LPCRIKQFIN	MWQEVGKAMY	APPISGQIRC	SSNITGLLLT	RDGGNNNNGS	EIFRPGGGDM
RDNRSELYK	YKVVKIEPLG	VAPTKAKRRV	VQREKRAVGI	GALFLGFLGA	AGSTMGARSM	TLTVQARQLL	SGIVQQQNNL
LRAIEAQOHL	LQLTVWGIIKQ	LQARILAYER	YKDDQQLLGI	WGCSGKLICT	TAVPWNASWS	NKSLEQIWNH	MTWMEWDREI
NNYTSLIHSL	IEESQNQQEK	NEQELLELDK	WASLWNWFFNI	TNWLWYIKIF	IMIVGGLVGL	RIVFAVLSIV	NRVRQGYSP
SFQTHLPTPR	GPDRPEGIEE	EGGERDRDRS	IRLVNGSLAL	IWDDLRLSLC	FSYHRLRDL	LIVTRIVELL	GRRGWEALKY
WWNLLQYWSQ	ELKNSAVSLL	NATAIAVAEG	TDRVIEVVQG	ACRAIRH1PR	RIRQGLERIL	L	

### B) Data for FPHM pet24a+ trimer

ENV\_HV1BR (100%), 97,489.7 Da

Envelope glycoprotein gp160 OS=Human immunodeficiency virus type 1 group M subtype B (isolate BRU/LAI) GN=env PE=1 SV=1  
2 exclusive unique peptides, 2 exclusive unique spectra, 6 total spectra, 63/861 amino acids (7% coverage)

MRVKEKYQHL	WRWGKWKWTM	LLGILMICS	TEKLWVTVY	GVPVWKEATT	TLFCASDAKA	YDTEVHNWVA	THACVPTDPN
PQEVVLNVVT	ENFNMWKNDM	VEQMHEDIIS	LWDQSLKPCV	KLTPLCVSLK	CTDLGNATNT	NSSNTNSSSG	EMMMEKGEIK
NCSFNISTSI	RGKVQKEYAF	FYKLDIIPID	NDTTSYTLTS	CNTSVITQAC	PKVSFEPIPI	HYCAPAGFAI	LKCNKTFNG
TGPCNTNVSTV	QCTHGIRPVV	STQLLLNGSL	AEEEVVIRSA	NFTDNAKTII	VQLNQSVIEIN	CTRPNNNTRK	SIRIQRGPGR
AFVTIGKIGN	MRQAHNCISR	AKWNATLKQI	ASKLREQFGN	NKTIIFKQSS	GGDPEIVTHS	FNCGGEFFYC	NSTQLFNSTW
FNSTWSTEGS	NNTGSDTIT	LPCRIKQFIN	MWQEVGKAMY	APPISGQIRC	SSNITGLLLT	RDGGNNNNGS	EIFRPGGGDM
RDNRSELYK	YKVVKIEPLG	VAPTKAKRRV	VQREKRAVGI	GALFLGFLGA	AGSTMGARSM	TLTVQARQLL	SGIVQQQNNL
LRAIEAQOHL	LQLTVWGIIKQ	LQARILAYER	YKDDQQLLGI	WGCSGKLICT	TAVPWNASWS	NKSLEQIWNH	MTWMEWDREI
NNYTSLIHSL	IEESQNQQEK	NEQELLELDK	WASLWNWFFNI	TNWLWYIKIF	IMIVGGLVGL	RIVFAVLSIV	NRVRQGYSP
SFQTHLPTPR	GPDRPEGIEE	EGGERDRDRS	IRLVNGSLAL	IWDDLRLSLC	FSYHRLRDL	LIVTRIVELL	GRRGWEALKY
WWNLLQYWSQ	ELKNSAVSLL	NATAIAVAEG	TDRVIEVVQG	ACRAIRH1PR	RIRQGLERIL	L	

### C) Data for FPHM pet19b Ek cleave monomer

ENV\_HV1BR (100%), 97,489.7 Da

Envelope glycoprotein gp160 OS=Human immunodeficiency virus type 1 group M subtype B (isolate BRU/LAI) GN=env PE=1 SV=1  
2 exclusive unique peptides, 2 exclusive unique spectra, 10 total spectra, 63/861 amino acids (7% coverage)

MRVKEKYQHL	WRWGKWKWTM	LLGILMICS	TEKLWVTVY	GVPVWKEATT	TLFCASDAKA	YDTEVHNWVA	THACVPTDPN
PQEVVLNVVT	ENFNMWKNDM	VEQMHEDIIS	LWDQSLKPCV	KLTPLCVSLK	CTDLGNATNT	NSSNTNSSSG	EMMMEKGEIK
NCSFNISTSI	RGKVQKEYAF	FYKLDIIPID	NDTTSYTLTS	CNTSVITQAC	PKVSFEPIPI	HYCAPAGFAI	LKCNKTFNG
TGPCNTNVSTV	QCTHGIRPVV	STQLLLNGSL	AEEEVVIRSA	NFTDNAKTII	VQLNQSVIEIN	CTRPNNNTRK	SIRIQRGPGR
AFVTIGKIGN	MRQAHNCISR	AKWNATLKQI	ASKLREQFGN	NKTIIFKQSS	GGDPEIVTHS	FNCGGEFFYC	NSTQLFNSTW
FNSTWSTEGS	NNTGSDTIT	LPCRIKQFIN	MWQEVGKAMY	APPISGQIRC	SSNITGLLLT	RDGGNNNNGS	EIFRPGGGDM
RDNRSELYK	YKVVKIEPLG	VAPTKAKRRV	VQREKRAVGI	GALFLGFLGA	AGSTMGARSM	TLTVQARQLL	SGIVQQQNNL
LRAIEAQOHL	LQLTVWGIIKQ	LQARILAYER	YKDDQQLLGI	WGCSGKLICT	TAVPWNASWS	NKSLEQIWNH	MTWMEWDREI
NNYTSLIHSL	IEESQNQQEK	NEQELLELDK	WASLWNWFFNI	TNWLWYIKIF	IMIVGGLVGL	RIVFAVLSIV	NRVRQGYSP
SFQTHLPTPR	GPDRPEGIEE	EGGERDRDRS	IRLVNGSLAL	IWDDLRLSLC	FSYHRLRDL	LIVTRIVELL	GRRGWEALKY
WWNLLQYWSQ	ELKNSAVSLL	NATAIAVAEG	TDRVIEVVQG	ACRAIRH1PR	RIRQGLERIL	L	

### D) Data FPHM pet19b Ek cleave trimer

ENV\_HV1BR (100%), 97,489.7 Da

Envelope glycoprotein gp160 OS=Human immunodeficiency virus type 1 group M subtype B (isolate BRU/LAI) GN=env PE=1 SV=1  
1 exclusive unique peptides, 1 exclusive unique spectra, 5 total spectra, 63/861 amino acids (6% coverage)

MRVKEKYQHL	WRWGKWKWTM	LLGILMICS	TEKLWVTVY	GVPVWKEATT	TLFCASDAKA	YDTEVHNWVA	THACVPTDPN
PQEVVLNVVT	ENFNMWKNDM	VEQMHEDIIS	LWDQSLKPCV	KLTPLCVSLK	CTDLGNATNT	NSSNTNSSSG	EMMMEKGEIK
NCSFNISTSI	RGKVQKEYAF	FYKLDIIPID	NDTTSYTLTS	CNTSVITQAC	PKVSFEPIPI	HYCAPAGFAI	LKCNKTFNG
TGPCNTNVSTV	QCTHGIRPVV	STQLLLNGSL	AEEEVVIRSA	NFTDNAKTII	VQLNQSVIEIN	CTRPNNNTRK	SIRIQRGPGR
AFVTIGKIGN	MRQAHNCISR	AKWNATLKQI	ASKLREQFGN	NKTIIFKQSS	GGDPEIVTHS	FNCGGEFFYC	NSTQLFNSTW
FNSTWSTEGS	NNTGSDTIT	LPCRIKQFIN	MWQEVGKAMY	APPISGQIRC	SSNITGLLLT	RDGGNNNNGS	EIFRPGGGDM
RDNRSELYK	YKVVKIEPLG	VAPTKAKRRV	VQREKRAVGI	GALFLGFLGA	AGSTMGARSM	TLTVQARQLL	SGIVQQQNNL
LRAIEAQOHL	LQLTVWGIIKQ	LQARILAYER	YKDDQQLLGI	WGCSGKLICT	TAVPWNASWS	NKSLEQIWNH	MTWMEWDREI
NNYTSLIHSL	IEESQNQQEK	NEQELLELDK	WASLWNWFFNI	TNWLWYIKIF	IMIVGGLVGL	RIVFAVLSIV	NRVRQGYSP
SFQTHLPTPR	GPDRPEGIEE	EGGERDRDRS	IRLVNGSLAL	IWDDLRLSLC	FSYHRLRDL	LIVTRIVELL	GRRGWEALKY
WWNLLQYWSQ	ELKNSAVSLL	NATAIAVAEG	TDRVIEVVQG	ACRAIRH1PR	RIRQGLERIL	L	

### E) Data FPHM pet19b after separation and cleavage

ENV\_HV1BR (100%), 97,489.7 Da

Envelope glycoprotein gp160 OS=Human immunodeficiency virus type 1 group M subtype B (isolate BRU/LAI) GN=env PE=1 SV=1  
2 exclusive unique peptides, 2 exclusive unique spectra, 63/861 amino acids (6% coverage)

MRVKEKYQHL	WRWGKWKWTM	LLGILMICS	TEKLWVTVY	GVPVWKEATT	TLFCASDAKA	YDTEVHNWVA	THACVPTDPN
PQEVVLNVVT	ENFNMWKNDM	VEQMHEDIIS	LWDQSLKPCV	KLTPLCVSLK	CTDLGNATNT	NSSNTNSSSG	EMMMEKGEIK
NCSFNISTSI	RGKVQKEYAF	FYKLDIIPID	NDTTSYTLTS	CNTSVITQAC	PKVSFEPIPI	HYCAPAGFAI	LKCNKTFNG
TGPCNTNVSTV	QCTHGIRPVV	STQLLLNGSL	AEEEVVIRSA	NFTDNAKTII	VQLNQSVIEIN	CTRPNNNTRK	SIRIQRGPGR
AFVTIGKIGN	MRQAHNCISR	AKWNATLKQI	ASKLREQFGN	NKTIIFKQSS	GGDPEIVTHS	FNCGGEFFYC	NSTQLFNSTW
FNSTWSTEGS	NNTGSDTIT	LPCRIKQFIN	MWQEVGKAMY	APPISGQIRC	SSNITGLLLT	RDGGNNNNGS	EIFRPGGGDM
RDNRSELYK	YKVVKIEPLG	VAPTKAKRRV	VQREKRAVGI	GALFLGFLGA	AGSTMGARSM	TLTVQARQLL	SGIVQQQNNL
LRAIEAQOHL	LQLTVWGIIKQ	LQARILAYER	YKDDQQLLGI	WGCSGKLICT	TAVPWNASWS	NKSLEQIWNH	MTWMEWDREI
NNYTSLIHSL	IEESQNQQEK	NEQELLELDK	WASLWNWFFNI	TNWLWYIKIF	IMIVGGLVGL	RIVFAVLSIV	NRVRQGYSP
SFQTHLPTPR	GPDRPEGIEE	EGGERDRDRS	IRLVNGSLAL	IWDDLRLSLC	FSYHRLRDL	LIVTRIVELL	GRRGWEALKY
WWNLLQYWSQ	ELKNSAVSLL	NATAIAVAEG	TDRVIEVVQG	ACRAIRH1PR	RIRQGLERIL	L	

Figure 3.7 Proteomics data for gp41 constructs.

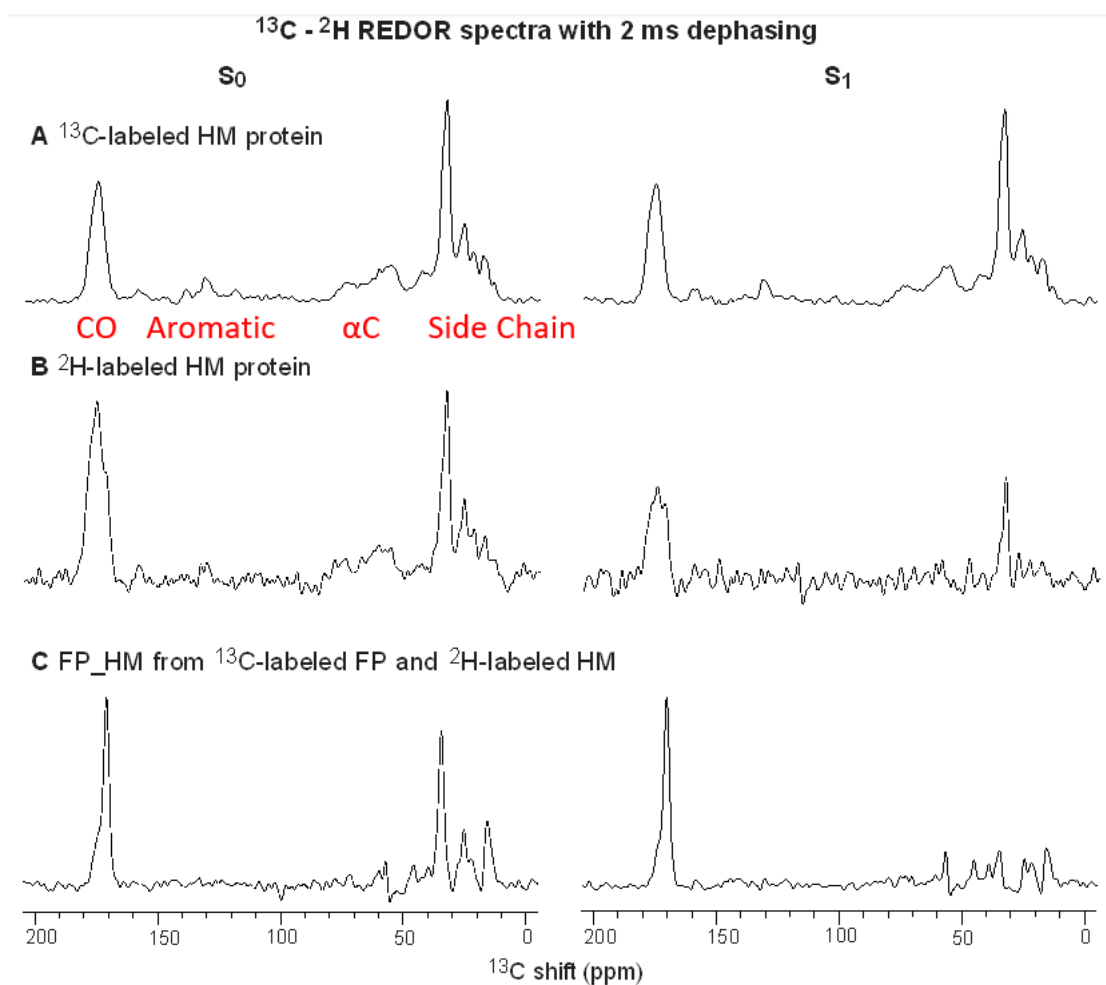
HM protein with fractional  $^{13}\text{C}$  labeling was produced using expression in minimal medium with a mixture of unlabeled and 1,2,3,4,5,6- $^{13}\text{C}$  D-glucose. Purified HM yield was approximately independent of the fraction of  $^{13}\text{C}$  D-glucose in the medium. The  $^{13}\text{C}$  labeling of HM was estimated using MALDI mass spectrometry (**Figure 3.6**). Peaks were assigned to  $\text{HM}^+$  or  $\text{HM}^{2+}$  ions, and percent  $^{13}\text{C}$  labeling was estimated as  $\Delta M_{\text{exp}}/\Delta M_{\text{lab}} \times 100$ , where  $\Delta M_{\text{exp}}$  is the difference between experimental m/z in labeled and unlabeled media, and  $\Delta M_{\text{lab}}$  is the difference in calculated m/z between fully labeled and unlabeled HM. Comparative values of  $^{13}\text{C}$  D-glucose in the medium vs.  $^{13}\text{C}$  labeling of HM include 25% vs. 14%, and 100% vs. 82%.

HM protein with fractional  $^2\text{H}$  labeling was produced using 1,2,3,4,5,6- $^2\text{H}$  D-glucose and a mixture of  $\text{D}_2\text{O}$  and  $\text{H}_2\text{O}$ . Subsequent purification steps were done with  $\text{H}_2\text{O}$ , so that exchangeable hydrogens were  $^1\text{H}$ s. For expression media with  $\text{D}_2\text{O}$  in the 0-50% range, purified HM yields were like those from medium with unlabeled glucose and  $\text{H}_2\text{O}$ , while yield was reduced from medium with >50%  $\text{D}_2\text{O}$ . The percent  $^2\text{H}$  labeling of HM was estimated using an approach like that for  $^{13}\text{C}$ -HM. Comparative values of  $\text{D}_2\text{O}$  in the medium vs.  $^2\text{H}$  labeling of HM include 0% vs. 18%, 50% vs. 40%, and 90% vs. 54%. These values are for non-exchangeable hydrogens, and additional purification steps cause all exchangeable  $^2\text{H}$  to be  $^1\text{H}$ .

It appears to be a common trend that  $\{2 \times (m/z)_{2+} - (m/z)_{1+} > 0\}$ . This calculated difference for each mass spectra in **Figure 3.6** are 66 m/z for **Figure 3.6A**, 86 for **Figure 3.6B**, and 82 for **Figure 3.6C**. The difference between the  $(m/z)_{2+}$  peak and the  $(m/z)_{1+}$  peak may correspond with some sodium or potassium adduct. 66 m/z corresponds to

approximately 3 Na ions, 86 m/z and 82 m/z are more closely approximated by the mass of 2 K ions.

Isotopically labeled and lyophilized HM was also probed by  $^{13}\text{C}$ - $^2\text{H}$  REDOR NMR (**Figure 3.8**). REDOR probes proximal ( $<8 \text{ \AA}$ )  $^{13}\text{C}$  and  $^2\text{H}$  nuclei, using attenuation of REDOR  $^{13}\text{C}$   $S_1$  vs.  $S_0$  signals. This attenuation is quantified as  $\Delta S/S_0$ , where  $\Delta S$  is the difference between  $S_0$  and  $S_1$  signal intensities. For  $^{13}\text{C}$ 's proximal to  $^2\text{H}$ 's,  $\Delta S/S_0$  increases as a function of the experimental dephasing time  $\tau$ . **Figure 3.8B** displays representative REDOR  $^{13}\text{C}$  spectra of HM with 54%  $^2\text{H}$  labeling based on mass spectroscopy. The 1% natural abundance  $^{13}\text{C}$  nuclei are randomly distributed throughout HM. The high attenuation of  $S_1$  intensity in all spectral regions evidence that most  $^{13}\text{C}$ 's have proximal  $^2\text{H}$ 's, i.e., the  $^2\text{H}$ 's are also randomly distributed throughout the non-exchangeable sites of HM. **Figure 3.8A** displays representative REDOR spectra of HM with 14%  $^{13}\text{C}$  labeling. There is negligible attenuation of  $S_1$  signals, which correlates with only 0.01%  $^2\text{H}$  natural abundance. REDOR spectra were also acquired for HM with 82%  $^{13}\text{C}$ -labeling, and there was also negligible attenuation of  $S_1$  vs.  $S_0$  signals. These samples exhibited reduced  $S_0$  signals at longer  $\tau$ , likely because of shorter  $^{13}\text{C}$ - $^{13}\text{C}$  distances and consequent larger  $^{13}\text{C}$ - $^{13}\text{C}$  dipolar couplings that are recoupled by the rotor-synchronized  $^{13}\text{C}$   $\pi$  pulses.



**Figure 3.8**  $^{13}\text{C}$ - $^2\text{H}$  rotational-echo double-resonance (REDOR) solid-state NMR spectra of lyophilized HM without lipid with either (A) 13.5%  $^{13}\text{C}$ -labeling or (B) 54%  $^2\text{H}$ -labeling, with percent labeling determined from mass spectrometry. Panel **A** and **B** proteins were prepared from HM purified by solubilization, as they have no His-tag. Panels **A** and **B** have no lipid present in the sample.

The samples are lyophilized overnight following purification and packed into rotors before rehydration. The reduction in  $^{13}\text{C}$  S<sub>1</sub> vs. S<sub>0</sub> signals probes the presence of  $^2\text{H}$

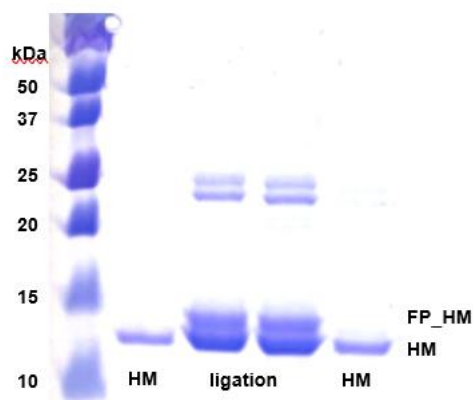
### Figure 3.8 (cont'd)

nuclei less than about 8 Å from the  $^{13}\text{C}$  nuclei. The panel **A** signals are from labeled  $^{13}\text{C}$  nuclei and do not show significant reductions in  $S_1$  vs.  $S_0$  signal intensities, which is consistent with 0.01%  $^2\text{H}$  natural abundance. The panel **B** signals are from natural abundance  $^{13}\text{C}$  nuclei and show significant reductions in  $S_1$  vs.  $S_0$  signal intensities, which is consistent with a large amount  $^2\text{H}$  labeling. A higher scan count is used due to the lower quantity of  $^{13}\text{C}$ . The panel **C** signals are from natural abundance  $^{13}\text{C}$  nuclei in HM and 100%  $^{13}\text{C}$  labeled Gly-5,10,16. The REDOR shows significant reductions in  $S_1$  vs.  $S_0$  signal intensities for aliphatic carbons (0-50ppm), which is consistent with a large amount  $^2\text{H}$  labeling. The CO signal (170ppm) does not show significant reduction in  $S_1$  vs.  $S_0$  signal intensities as the FP section of the Protein has only natural abundance  $^2\text{H}$ . This protein sample is in 8:2:5 DMPC: DMPG: Chol that has been lyophilized and rehydrated. Dephasing times are 2ms each and other dephasing times are shown (**Figure 3.11**). General Parameters include ~5 mg lyophilized protein, 10 kHz magic angle spinning frequency, 65 scans/spectrum in panel A, 35355 scans/spectrum in panel B, and 10000 scans/spectrum in panel C, 150 Hz Gaussian line broadening, and ambient temperature.

#### 3.3.3. Importance of Imidazole for Ligation Yield

Ligation between FP and HM constructs was done under different conditions followed by dialysis against 8 M urea and then SDS-PAGE to assess relative quantities of FPHM vs. HM. **Figure 3.9** displays a gel for ligation in 6.4 M GuHCl, 2.0 M imidazole, and 30 mM MPAA, which are conditions that reproducibly showed the highest FPHM:HM ratio  $\approx 0.4$ . The GuHCl aids protein solubilization, but there is probably at least partial

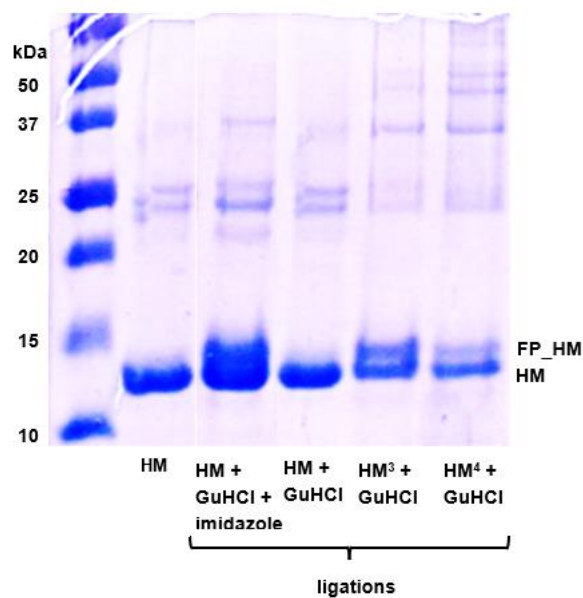
retention of the native helical HM structure, as evidenced by earlier CD spectra in GuHCl (48). The highest FP\_HM yield was achieved with [imidazole]/[MPAA] > 10, and there was no FP\_HM gel band after reaction in buffer with GuHCl and either imidazole or MPAA (**Figure 3.10**). There was ligation reaction in buffer with GuHCl and MPAA for HM<sup>3</sup> or HM<sup>4</sup> which have mutations in the C-helix region intended to destabilize interactions between the C- and N-helices (**Figure 3.10**). The positive correlations of yield of FPHM with imidazole for wild type (WT)-HM, and with reduced hairpin stability for mutant HM<sup>3</sup> and HM<sup>4</sup> suggest that imidazole may improve ligation yield by enhancing HM unfolding and consequently increasing exposure of reactive termini. It would be interesting to test this hypothesis in future work. The denaturant function of imidazole is consistent with highest WT FP\_HM yield when [imidazole] > 0.4 M. However, to our knowledge, there is not literature that supports imidazole as denaturant. In addition, yields varied significantly (30-45%) between replicate reactions in buffer with 6.4 M GuHCl and 2.0 M urea, which are both denaturants. Imidazole might also directly participate in the ligation reaction by nucleophilic substitution of the MPAA thioester (proposed intermediate in step 2 of **Figure 3.2**). The imidazole-(CO) is then attacked by the N-terminal Cys of HM, with imidazole as the leaving group. Enhanced ligation rate by this mechanism might be aided by the similarities between the pK<sub>a</sub> values of MPAA and imidazole, respectively 6.6 and 6.9, and the neutral pH of the reaction. Similar pK<sub>a</sub> would facilitate exchange between FP-MPAA and FP-imidazole catalytic species.



**Figure 3.9** SDS-PAGE showing the extent of native chemical ligation between FP with C-terminal thioester -COS-CH<sub>2</sub>-CH<sub>2</sub>-COOH (MW = 2.1 kDa) and HM\_G<sub>4</sub>H<sub>4</sub> (MW = 13.7

kDa) that had been purified by Co<sup>2+</sup>-resin affinity chromatography. After overnight ligation reaction at ambient temperature and neutral pH with ~400  $\mu$ M FP, 50  $\mu$ M HM, 30 mM MPAA, 67 mM TCEP, 6.7 M GuHCl and 2.0 M imidazole, the reaction solution was dialyzed against 8 M urea and analyzed by SDS-PAGE. The gel lanes from left-to-right are MW standards, HM\_G<sub>4</sub>H<sub>4</sub>, ligation mixture (2 replicate lanes), and HM\_G<sub>4</sub>H<sub>4</sub>.

The ligation mixture lanes show bands that are assigned to HM\_G<sub>4</sub>H<sub>4</sub> and FP\_HM\_G<sub>4</sub>H<sub>4</sub>. The ratio of HM\_G<sub>4</sub>H<sub>4</sub>:FP\_HM\_G<sub>4</sub>H<sub>4</sub> band intensities is ~0.4. Weaker bands at ~24 kDa may be dimer protein.



**Figure 3.10** SDS-PAGE showing the effects of imidazole and HM mutations on the extent of native chemical ligation. All reactions were done overnight at ambient temperature and neutral pH with FP with C-terminal (CO)S(CH<sub>2</sub>CH<sub>2</sub>COOH) and HM<sub>G4H4</sub> that had been purified by Co<sup>2+</sup>-resin affinity chromatography. The reaction solutions were then dialyzed against 8 M urea followed by SDS-PAGE. The gel lanes from left-to-right are: (1) MW standards; (2) HM<sub>G4H4</sub>; and ligations with: (3) HM<sub>G4H4</sub>, 6.7 M GuHCl, and 2.1 M imidazole; (4) HM<sub>G4H4</sub> and 6.7 M GuHCl; (5) HM<sup>3</sup><sub>G4H4</sub> and 6.7 M GuHCl; and (6) HM<sup>4</sup><sub>G4H4</sub> and 6.7 M GuHCl. Bands are identified for HM<sub>G4H4</sub> and for FP\_HM<sub>G4H4</sub>. Ligation (3) was done under the same conditions as the ligation in **Figure 3.8** and exhibits relative HM to FP\_HM band intensities that are like those in the **Figure 3.8** gel. The M<sup>3</sup> (W628A, W631R, D632A), and M<sup>4</sup> (W628A, W631R, D632A, Q652A) HM proteins have mutations in the C-helix region that are intended to destabilize binding of C-helices with N-helices. Ligations (4-6) are done without imidazole, and there is significant FP\_HM product for ligations (5) and (6) with mutant HM but not ligation (4) with WT HM.



### Figure 3.10 (cont'd)

These results suggest that the ligation rate is increased by protein unfolding, likely because of increased exposure of the N-terminal Cys. The results are also consistent with imidazole-induced increase in unfolded protein.

#### 3.3.4. Importance of High [FP]:[HM] and Low [HM] for Ligation Yield

FP\_HM yield was highest with FP:HM molar ratio  $\geq 4$  and with [HM]  $\approx 50 \mu\text{M}$  (**Figure 3.9**). Yield was lower with FP:HM  $\approx 1$  or with [HM]  $\approx 1 \text{ mM}$ . These observations may be related to earlier size-exclusion chromatography (SEC) data showing that  $70 \mu\text{M}$  HM in  $6 \text{ M}$  GuHCl is predominantly folded hexamers, with smaller populations of dodecamers and larger aggregates (48). The terminal cysteines of these hexamers, dodecamers, and larger aggregates may have low exposures and consequent slow ligation rates. We hypothesize that FPHM is primarily a reaction product of FP with monomer HM with exposed cysteine. Monomer HM was not detected by SEC with [HM]  $\approx 70 \mu\text{M}$  but could exist at low concentration for smaller [HM]. Highest FP\_HM yield with  $\geq 4$ -fold molar excess of FP vs. HM may be due to the even larger excess of FP vs. monomer HM and the consequent higher concentration of the proposed intermediate in step 3 of the **Figure 3.2** reaction scheme. This pushes the reaction more towards the FPHM.

#### 3.3.5. Effects of Other Parameters on Ligation Yield

Other FP + HM ligation conditions were examined, and some of these conditions were also examined for FP + HP ligation, where HP lacks the 17 C-terminal residues of HM but still adopts hyperthermostable helical hairpin structure (35). The catalyst [MPAA]  $\approx 30 \text{ mM}$  and [HM]  $\approx 50 \mu\text{M}$  correspond to MPAA:HM ratio  $\approx 600$ , and higher ratio with

[MPAA] = 60 mM did not reproducibly increase the product yield. No reaction was observed with other catalysts including 2-mercaptoethanesulfonate, sodium salt, an alkyl thiol, and thiophenol, an aryl thiol, and with 2-nitrophenol and 4-nitrophenol which would form oxy- rather thio- ester intermediates (**Figure 3.2**). Yield was not improved by reaction with higher [TCEP]  $\approx$  130 mM vs. 67 mM. FP\_HP yields were similar for reaction times in the 3 h – 3 day range and FP\_HM yields were similar for reaction overnight and reaction for three days. Reactions at 40 °C and 50 °C did not show higher yields relative to reaction at ambient temperature. The reaction extent  $\approx$  40% of the present study could be limited by unreactive oligomers/aggregates of HM that do not contribute to the reactive monomer population.

### 3.3.6. Low pH Wash for Separation of FP\_HM

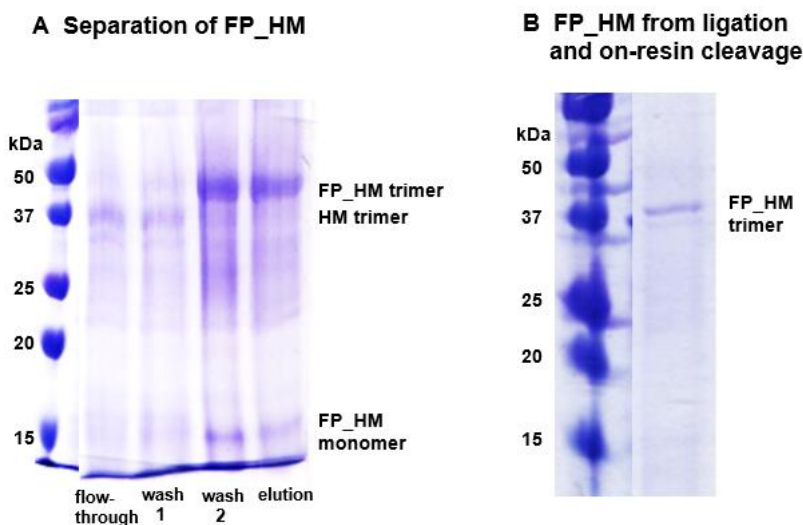
Earlier studies reported that the ligation product FP\_HP is separable from FP and HP by reverse-phase HPLC (48). RP-HPLC of the FP + HM reaction mixture showed co-elution of HM and FP\_HM, as assessed by SDS-PAGE and mass spectrometry of the chromatographic peaks. We therefore developed the approach shown in **Figure 3.3** for separation of FP\_HM and HM using affinity chromatography. We had to consider several competing factors in developing the protocol: (1) affinity chromatography relies on tightly bound and monomeric FP\_HM and unbound HM (2) gp41 SE constructs are often monomeric at pH < 4 but are typically associated as trimers or larger oligomers or aggregates at neutral pH (38), and (3) His-tags bind tightly to the Co<sup>2+</sup> resin near neutral pH and binding decreases as pH is reduced. We first observed that H<sub>6</sub>G<sub>6</sub>D<sub>4</sub>K\_FP\_HM was tightly bound to the resin only near neutral pH. After dialysis to remove unreacted H<sub>6</sub>G<sub>6</sub>D<sub>4</sub>K\_FP, the H<sub>6</sub>G<sub>6</sub>D<sub>4</sub>K\_FP\_HM + HM mixture was well-solubilized in PBS buffer +

GuHCl in which the proteins were mostly associated as hexamers (48). The solution was mixed with  $\text{Co}^{2+}$ - resin, followed by standard washes and then elutions. SDS-PAGE of the eluents showed approximately equal-intensity bands for H<sub>6</sub>G<sub>6</sub>D<sub>4</sub>K\_FP\_HM and HM, i.e., failure of separation, which is consistent with hexamers that are mostly mixtures of the two proteins.

Recombinant H<sub>10</sub>S<sub>2</sub>GHID<sub>4</sub>K\_FP\_HM was then created as a test system to develop the separation protocol. The FP\_HM DNA was sub-cloned in the pET-19b vector, and subsequent expression in *E. coli*, separation of the solid inclusion bodies from cellular material, and then solubilization, and  $\text{Co}^{2+}$ -affinity chromatography. Mixed hexamers of H<sub>10</sub>S<sub>2</sub>GHID<sub>4</sub>K\_FP\_HM and HM were prepared by: (1) combining the two proteins in 1:2 molar ratio in ligation buffer at 50  $\mu\text{M}$  total protein concentration; (2) protein precipitation by dialysis against water; and (3) solubilization of the precipitate in PBS + 6 M GuHCl. The protein solution was combined with  $\text{Co}^{2+}$ - resin and mixed overnight at 10 °C, followed by filtration. Conditions were then tested for combined preferences for both release of HM and binding of H<sub>10</sub>S<sub>2</sub>GHID<sub>4</sub>K\_FP\_HM to the resin. This testing was executed experimentally by resin washes with different solutions followed by filtration, elution of bound protein with imidazole, and then SDS-PAGE analysis of washes and elutions. One wash of a pH 3.5 solution preferentially releases HM, whereas two washes release significant H<sub>10</sub>S<sub>2</sub>GHID<sub>4</sub>K\_FP\_HM. This was subsequently observed for ligation reaction mixtures with HM and H<sub>6</sub>G<sub>6</sub>D<sub>4</sub>K\_HM (**Figure 3.11A**). Release of HM with one low pH wash could be due to: (1) dissociation of hexamers or (2) reduced His-tag binding to the resin so that hexamers with more HM and fewer FP\_HM molecules are released. Release of FP\_HM with the second wash reflects reduced His-tag binding at low pH.

Wash solutions with other conditions or additives did not preferentially remove HM. This included pH 5 solution, and pH 7 solutions with either N-lauroylsarcosine, sodium dodecyl sulfate, or dodecylphosphocholine detergent. The latter detergent increases monomer concentration for some gp41 constructs.

protein.



**Figure 3.11** SDS-PAGE of (A) separation of H<sub>6</sub>G<sub>6</sub>D<sub>4</sub>K\_FP\_HM (MW = 17.5 kDa) and HM (MW = 13.0 kDa) and (B) FP\_HM (MW = 15.1 kDa) from enterokinase treatment of H<sub>6</sub>G<sub>6</sub>D<sub>4</sub>K\_FP\_HM bound to Co<sup>2+</sup>-resin. (A) A 1.5 mL ligation reaction between 1.2 mg H<sub>6</sub>G<sub>6</sub>D<sub>4</sub>K\_FP and 1 mg HM was followed by dialysis to remove unreacted H<sub>6</sub>G<sub>6</sub>D<sub>4</sub>K\_FP, mixing with Co<sup>2+</sup>-resin, 2× resin washes with low pH buffer to promote formation of monomer protein and release of HM from the resin, and then elution with neutral pH buffer with 250 mM imidazole. The gel lanes from left-to-right are MW standards, resin flow-through, first wash, second wash, and elution.

The flow-through and first wash lanes have a band at ~36 kDa that are assigned to HM trimer, and the second wash and elution lanes have significant bands at ~15 and 45 kDa that are respectively assigned to H<sub>6</sub>G<sub>6</sub>D<sub>4</sub>K\_FP\_HM monomer and trimer. (B)

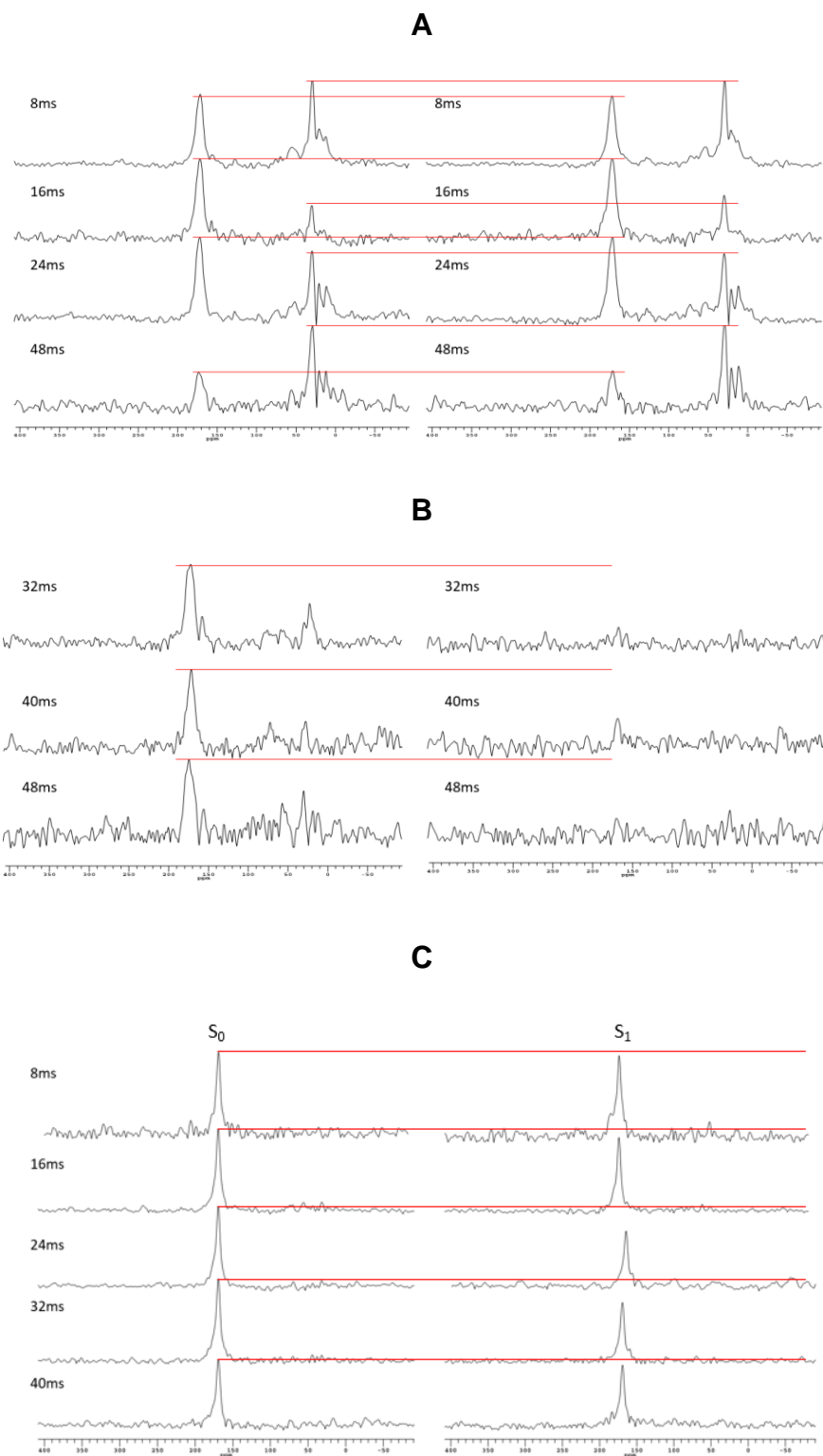
Ligation reaction and purification were like panel A except: (1) there was only one wash of the resin with low pH buffer; and (2) elution was replaced with resin treatment with enterokinase following replacement of separation buffer with enterokinase buffer. The gel lanes are MW standards and the resin wash solution after enterokinase treatment. The band at ~ 39 kDa is assigned to FP\_HM trimer. Some of the panel A and B gel bands were subjected to proteolytic digestion and mass-spectrometric analysis of the digested peptides. The assigned peptides are consistent with the band assignments of the HM trimer, H<sub>6</sub>G<sub>6</sub>D<sub>4</sub>K\_FP\_HM monomer and trimer, and FP\_HM trimer.

Enterokinase cleavage C-terminal of D<sub>4</sub>K was tested using purified H<sub>10</sub>S<sub>2</sub>GHID<sub>4</sub>K\_FP\_HM that was bound to Co<sup>2+</sup>-resin. The resin was washed and then incubated overnight at 37 °C with enterokinase, followed by washes that contained FP\_HM released by cleavage. FP\_HM was precipitated by dialysis against water and then resolubilized in PBS + 8 M urea, followed by SDS-PAGE.

The full FP\_HM synthesis was done using the ligation and purification protocols presented in Materials and Methods. The FP\_HM was pure, and yield was ~10% relative to the HM limiting reactant (**Figure 3.11B**). The G<sub>6</sub> flexible linker in the FP N-terminal H<sub>6</sub>G<sub>6</sub>D<sub>4</sub>K tag may be important for yield, both for H<sub>6</sub> exposure for resin binding, and for D<sub>4</sub>K exposure for enterokinase cleavage. This idea is supported by comparison with recombinant H<sub>10</sub>S<sub>2</sub>GHID<sub>4</sub>K\_FP\_HM with S<sub>2</sub>GHI rather than G<sub>6</sub> linker. The (moles FP\_HM after cleavage)/(moles H<sub>10</sub>S<sub>2</sub>GHID<sub>4</sub>K\_FP\_HM initially added to the resin) is only ~10%.

Larger quantities of FP\_HM were produced using parallel reactions and purifications at the smaller scale. Reaction and purification (12×) were done using FP with <sup>13</sup>CO labels at G<sub>5</sub>, G<sub>10</sub>, and G<sub>16</sub>, followed by pooling of the cleavage product solutions,

and precipitation of FP\_HM product by dialysis against water. This yielded ~1.5 mg FP\_HM which was then reconstituted in membrane using a published protocol (37). **Figure 3.8C** shows an example  $^{13}\text{C}$  REDOR spectrum of this sample. Additional dephasing times are shown in **Figure 3.12**. There does not appear to be significant dephasing in the REDOR spectra indicating that FP does not come into close contact with MPER, in accordance with circular dichroism data estimating numbers of helical residues in gp41 constructs (59).



**Figure 3.12** Supplementary REDOR data for **Figure 3.7A, B, and C**. Figure 3.7A is  $^{13}\text{C}$  labeled and shows strong signals in  $S_0$  with no dephasing in  $S_1$ . In **B**, and **C** aliphatic

### Figure 3.12 (cont'd)

signals are not strongly observed. In **B** aliphatic signal is present but weak. No signal is observed in **B** for higher dephasing times. Aliphatic signals are not observed in **C** due to the large abundance of  $^{13}\text{C}$  CO, however there is a strong carbonyl signal observed around 170ppm. There is small dephasing observed at 24, and 32ms dephasing time.

Similar results are observed when separating chemically synthesized H<sub>6</sub>G<sub>6</sub>D<sub>4</sub>K\_FP\_HM from HM. The product of the native chemical ligation was dialyzed against pure water to precipitate the solid protein. The solid protein was solubilized in 6M GuHCl buffer and mixed with Co<sup>2+</sup> resin overnight at 10<sup>0</sup>C with tumbling. (**Figure 3.11A**) shows the SDS-PAGE gel of the separation products of H<sub>6</sub>G<sub>6</sub>D<sub>4</sub>K\_FP\_HM from HM. The lanes representing the non-binding protein and the first wash with separation buffer contain mostly HM protein. There are also trimer bands corresponding to FP\_HM which did not bind to the resin. The second wash with separation buffer contains mostly FP\_HM protein as does the final elution of protein from the column. Cleavage of the H<sub>6</sub>G<sub>6</sub>D<sub>4</sub>K\_FP\_HM on the column by enterokinase results in a single trimer band corresponding to FP\_HM. The identity of trimeric FP\_HM is confirmed by proteolysis followed by mass spectrometric detection of peptides expected from the FP\_HM sequence, including the S535A mutation, and the addition of N-terminal cysteine on the HM necessary to facilitate the native chemical ligation (**Figure 3.13**).



Enzyme: Trypsin/P: cuts C-term side of KR.  
 Fixed modifications: **Carbamidomethyl (C)**  
 Variable modifications: **Oxidation (M)**

Protein sequence coverage: 34%

Matched peptides shown in **bold red**.

1 HHHHHHGGGG GGDDDDKAVG IGALFLGFLG AAGSTMGARA **CTLTVQARQL**  
 51 **LSGIVQQQNN** LLRAIEAQQH **LQLTVWGIK** QLQARILSGG RGGWMEWDRE  
 101 INNYTSLIHS LIEESQNQQE **KNEQELLELD** KWASLWNWFN ITNWLWYIK

Unformatted sequence string: **149 residues** (for pasting into other applications).

Sort by ☒ residue number ☐ increasing mass ☐ decreasing mass  
 Show ☒ matched peptides only ☐ predicted peptides also  
 Show ☒ uncorrected delta ☐ delta corrected for 13C

Query	Start - End	Observed	Mr(expt)	Mr(calc)	ppm	M Score	Expect	Rank	U	Peptide
<a href="#">2204</a>	40 - 48	510.2718	1018.5291	1018.5230	6.03	0	53	0.0026	1	<b>R.ACTLTVQAR.Q</b>
<a href="#">5734</a>	49 - 63	862.4965	1722.9785	1722.9740	2.58	0	74	7.5e-06	1	<b>R.QLLSGIVQQQNNLLR.A</b>
<a href="#">5735</a>	49 - 63	862.4982	1722.9818	1722.9740	4.51	0	77	3.3e-06	1	<b>R.QLLSGIVQQQNNLLR.A</b>
<a href="#">5736</a>	49 - 63	862.4985	1722.9825	1722.9740	4.89	0	90	1.5e-07	1	<b>R.QLLSGIVQQQNNLLR.A</b>
<a href="#">5737</a>	49 - 63	575.3351	1722.9834	1722.9740	5.45	0	54	0.00059	1	<b>R.QLLSGIVQQQNNLLR.A</b>
<a href="#">5738</a>	49 - 63	575.3354	1722.9843	1722.9740	5.92	0	65	4.3e-05	1	<b>R.QLLSGIVQQQNNLLR.A</b>
<a href="#">5745</a>	49 - 63	862.9900	1723.9654	1722.9740	575	0	65	6.9e-05	1	<b>R.QLLSGIVQQQNNLLR.A</b>
<a href="#">6641</a>	64 - 80	974.5596	1947.1046	1947.0942	5.34	0	99	1.7e-08	1	<b>R.AIEAQQHLLQLTVWGIK.Q</b>
<a href="#">6642</a>	64 - 80	650.0425	1947.1057	1947.0942	5.91	0	55	0.00044	1	<b>R.AIEAQQHLLQLTVWGIK.Q</b>
<a href="#">3628</a>	122 - 131	615.8176	1229.6206	1229.6139	5.47	0	49	0.0047	1	<b>K.NEQELLELDK.W</b>

Figure 3.13 Proteomics data for NCL product.

### 3.4. Discussion

This paper describes a synthetic approach to produce and purify the FPHM construct from bacterially expressed HM using native chemical ligation.

#### 3.4.1. Isotopically Labeled Constructs

Isotopically labeled HM was produced by growth of *E. coli* containing pET-24a(+) vector with HM protein insert in labeled media followed by induced expression of the protein in labeled media. Other vectors that were tried including pUC19 experienced leaky expression that made the vector ineffective for isotopic labeling. Initial bacterial growth was done in rich media. A single media switch from LB to minimal media followed by expression induction with IPTG yielded no HM protein. A protocol was adapted to our protein expression that involves a double media switch (61). For this protocol, *E. coli* was first grown in rich media, then a small fraction (about 20%) of the cell pellet from the rich media growth was transferred to a flask with minimal media and unlabeled glucose for a second growth period. This second growth period in minimal media may be necessary to

produce *E. coli* with metabolic systems better adapted for the use of limited nutrients present in the minimal media (61). The *E. coli* from the second growth phase in minimal media was then transferred to a flask containing minimal media and labeled glucose. This transfer allowed for the *E. Coli* to express isotopically labeled protein membrane protein.

### **3.4.2. Native Chemical Ligation Reactivity**

The fusion peptide was produced from solid phase peptide synthesis. We produced FP constructs both with and without an N-terminal affinity tag. FP without the affinity tag was used in studies aimed at increasing the yield of the native chemical ligation reaction. Several parameters of the ligation reaction were examined including (1) reaction pH, (2) concentration of protein, and (3) different native chemical ligation catalysts. Reaction temperature was also examined but had no effect on the overall conversion of HM to FPHM (45% at best), and FPHM was made at temperatures above 50° C. HM protein is hyperthermostable, so increasing the temperature is unlikely to cause any increase in denaturation of the protein that would improve the NCL reaction. The decrease in reactivity may instead be due to thermal degradation of other reagents including MPAA and ligation buffer.

The concentration of the HM protein in ligation buffer was determined to have a large effect on the overall conversion of HM to FPHM. HM protein concentrations of less than 50  $\mu$ M have a higher reactivity, and greater overall conversion to the product FPHM than sample that were ligated with concentrations above 50  $\mu$ M HM protein present in the sample. This may be in part due to the existence of an equilibrium between monomeric and trimeric forms of HM. It is likely that only the monomer is reactive towards native chemical ligation, due to steric factors.

Several ligation catalysts were examined to determine which catalyst produced the highest yields of FPHM. These catalysts included aryl thiols, as well as phenols to examine if an oxo-ester mediated native chemical ligation reaction was possible with this system. 4-mercaptophenylacetic acid (MPAA) proved to be the best catalyst overall as it catalyzes the native chemical ligation faster and more completely than thiophenol. No reaction was observed with 2-nitrophenol, or 4-nitrophenol indicating that an oxo-mediated native chemical ligation is not efficient with this chemical system. We used ~250 molar equivalents of catalyst relative to the starting amount of HM. The large quantity of catalyst is used to drive the initial equilibrium in the first step of the ligation as far to the product side as possible to achieve the highest possible yield of product FPHM (**Figure 3.2**). Using more than 250 molar equivalents does not have a large effect on ligation but yield drops off sharply at less than 250 molar equivalents.

#### **3.4.3. Native Chemical Ligation with Imidazole and M4\_HM**

Imidazole was found to play a key role in the success of the native chemical ligation reaction. We first found the importance of imidazole when attempting to perform the NCL reaction in 8M GuHCl and PBS without imidazole. The sample produced no FPHM under these conditions. Imidazole was screened as a catalyst based on previous findings that acyl-imidazole functions as an effective acyl donor similarly to MPAA-thioester (60). Imidazole does not catalyze the native chemical ligation reaction when MPAA is absent from the reaction mixture. In addition, when imidazole is combined with MPAA the native chemical ligation reaction proceeds to approximately a 40% yield.

We investigated two possible reasons for the necessity of imidazole in activating this system towards NCL. The first is that imidazole is a co-catalyst and that an equilibrium

exists between the acyl-imidazole species and the MPAA thioester activating one of the species more toward attack by the cysteine. This is very likely due to acyl-imidazole being known electrophiles (60). Another possibility is that the nitrogen on the imidazole disrupts intra-molecular hydrogen bonding in the HM and stabilizes the pre-hairpin intermediate, which may be more reactive towards native chemical ligation due to less potential steric interaction between CHR or MPER and the FP. However, there is no known literature to support this hypothesis.

The hypothesis that imidazole improves NCL reactivity by destabilizing the SE hairpin of the HM was tested using the mutant HM (W628A, W631R, D632A, Q652A) which was specifically designed to reduce the N-helix and C-helix binding in a trimer (**section 3.3.3**). It may be that imidazole is preventing hairpin formation; however, it is not fully understood how this may be the case with this protein. The PHI structure has been proposed to form after removal of the gp120s and has a fully extended structure like in step 2 of **Figure 1.11**. The existence of gp41 PHI is supported by functional studies. We proposed earlier that the HM needed to be monomer to be reactive toward NCL, but it may be required that the HM be in the PHI state as well. Extended PHI hairpin monomers are most clearly seen in **Figure 1.14** (2a,2b).

Three key results are that (1) imidazole alone does not act as a catalyst for the HM based on the absence of FPHM bands in the SDS-PAGE gel of the product (**Figure 3.10**); (2) MPAA alone does catalyze the NCL between mutant HM and FP (**Figure 3.10**); (3) highest yield using only MPAA on the mutant was ~40% like the non-mutant HM.

#### **3.4.4. Separation of HM and H<sub>6</sub>G<sub>6</sub>D<sub>4</sub>K\_FPHM by Affinity Chromatography**

The best conditions found for separation of HM from H<sub>6</sub>G<sub>6</sub>D<sub>4</sub>K\_FPHM were pH 3.5, 8 M guanidinium hydrochloride, 2.5 M Urea, and 0.5 M TCEP solution. The separation is a kinetic separation that will remove the HM protein from the column more quickly than the H<sub>6</sub>G<sub>6</sub>D<sub>4</sub>K\_FPHM. The difficulty in separating the HM from the H<sub>6</sub>G<sub>6</sub>D<sub>4</sub>K\_FPHM may be due to statistical mixed trimers that form between the two proteins. We hypothesized earlier that the NCL buffer allows reactive monomers to form, and we use a similar buffer for separation. Therefore, we also hypothesize that the separation buffer allows for the formation of HM monomers that can filter through the column more quickly than the H<sub>6</sub>G<sub>6</sub>D<sub>4</sub>K\_FPHM since the HM has no tag, monomer HM can be washed through the affinity column and removed. Monomer H<sub>6</sub>G<sub>6</sub>D<sub>4</sub>K\_FPHM may also form but will ideally bind to the Co<sup>2+</sup> resin, perhaps loosely bound due to low pH.

This approach is likely to be a generalized approach for producing high purity large proteins. In addition, this approach is applicable to systems where HPLC is not able to be utilized due to solubility or oligomerization issues.

#### **3.5. Summary**

This study reports the native chemical ligation synthesis of the gp41 ectodomain FPHM and describes a method of separation and purification of FPHM from HM protein. The SE hairpin of the HM protein is destabilized by imidazole to facilitate the native chemical ligation reaction. This destabilization is compared with mutations that also disrupt stabilizing interactions within the SE hairpin. In addition, this study also describes a protocol for producing bacterially expressed uniformly isotopically labeled HM with either <sup>13</sup>C or <sup>2</sup>H labels.

## REFERENCES

1. Dawson, P.E.; Muir, T.W.; Clark-Lewis, I.; Kent, S.B. *Science* **1994**, 266, 776-779.
2. Dawson, P.E.; Kent, S.B. *Annu. Rev. Biochem.* **2000**, 69, 923-960.
3. Grewe, C., Beck, A., and Gelderblom, H. R. (1990) HIV: early virus-cell interactions, *J. AIDS* 3, 965-974
4. Johnson, Erik., Kent, Stephen. (2006) Insights into the mechanism and Catalysis of the Native Chemical Ligation Reaction, *J. am. chem. soc.*, 128, 6640-6646
5. Wang, Chen.; Guo, Qing-Xiang.; Fu, Yao. (2011) Theoretical Analysis of the Detailed Mechanism of Native Chemical Ligation Reactions, *Chem. Asian, J.*, 6, 1241-1251
6. Freed, E. O., Delwart, E. L., Buchschacher, G. L., Jr., and Panganiban, A. T. (1992) A mutation in the human immunodeficiency virus type 1 transmembrane glycoprotein gp41 dominantly interferes with fusion and infectivity. *Proc. Natl. Acad. Sci. U. S. A.* 89,70-74.
7. Herold, N., Anders-Osswein, M., Glass, B., Eckhardt, M., Muller, B., and Krausslich, H. G. (2014) 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, 13956-13970.
8. Wan, Qian., Chen, Jin., Yuan, Yu., Danishefsky. (2008) Oxo-ester Mediated Native Chemical Ligation: Concept and Applications, *J. am. chem. soc.*, 130, 15814-15816
9. Fang, Ge-Min., Cui, Hong-Kui., Zheng, Ji-Shen., Liu, Lei. (2010) Chemoselective Ligation of Peptide Phenyl Esters with N-Terminal Cysteines, *ChemBioChem*, 11, 1061-1065
10. Muir, Tom., Sondhi, Dolan., Philip, Cole. (1998) Expressed Protein Ligation: A General Method for Protein Engineering, *Proc. Natl. Acad. Sci. USA*, 95, 6705-6710
11. Kent, Stephen. (2009) Total Chemical Synthesis of Proteins, *Chem. Soc. Rev.*, 38, 338-351
12. Liu, Chuan-Fa., Rao, Chang., Tam, James P. (1996) Orthogonal Ligation of Unprotected Peptide Segments Through Pseudoproline Formation for the Synthesis of HIV-1 Protease Analogs, *J. Am. Chem. Soc.*, 118, 307-312
13. Xie, Jianming., Schultz, Peter G. (2005) Adding Amino Acids to the Genetic Repertoire, *Current Opinion in Chemical Biology.*, 9, 548-554
14. Hofmann, Roseanne M., Muir, Tom W. (2002) Recent Advances in the Application of Expressed Protein Ligation to Protein Engineering, *Current Opinion in Biotechnology.*, 13, 297-303

15. Boerema, David J., Tereshko, Velentina A., Kent, Stephen B. H. (2007) Total Synthesis by Modern Chemical Ligation Methods and High Resolution (1.1 Å) X-ray Structure of Ribonuclease A, *Peptide Science.*, 90, 278-286
16. Miller, Maria. (2009) The Early Years of Retroviral Protease Crystal Structures, *Peptide Science.*, 94, 521-529
17. Torbeev, Vladimir Yu., Kent, Stephen B. H. (2007) Convergent Chemical Synthesis and Crystal Structure of a 203 Amino Acid "Covalent Dimer" HIV-1 Protease Enzyme Molecule, *Angew. Chem. Int. Ed.*, 46, 1667-1670
18. Durek, Thomas., Torbeev, Vladimir Yu., Kent, Stephen B. H. (2007) Convergent Chemical Synthesis and High-Resolution X-Ray Structure of Human Lysozyme, *PNAS*, 104, 4846-4851
19. Lee, Ji Yeon., Bang, Duhee. (2009) Challenges in the Chemical Synthesis of Average Sized Proteins: Sequential vs. Convergent Ligation of Multiple Peptide Fragments, *Peptide Science.*, 94, 441-447
20. Hackeng, Tilman M., Griffin, John H., Dawson, Phillip E. (1999) Protein Synthesis by Native Chemical Ligation: Extended Scope by Using Straightforward Methodology, *Proc. Natl. Acad. Sci. USA*, 96, 10068-10073
21. Cabezas, Edelmira., Wang, Meng., Parren, Paul W. H. I., Stanfield, Robyn L., Satterthwait, Arnold C. (2000) A Structure-Based Approach to a Synthetic Vaccine for HIV-1, *Biochemistry.*, 39, 14377-14391
22. Evans, Thomas C., Benner, Jack., Xu, Ming-Qun. (1999) The in Vitro Ligation of Bacterially Expressed Proteins Using an Intein from *Methanobacterium Thermoautotrophicum*, *J. Biol. Chem.*, 274, 3923-3926
23. Decostaire, Isidore E., Lelièvre, Dominique., Aucagne, Vincent., Delmas, Agnès F. (2014) Solid Phase Oxime Ligations for the Iterative Synthesis of Polypeptide Conjugates, *Org. Biomol. Chem.*, 12, 5536-5543
24. Wan, Qian., Danishefsky, Samuel J. (2007) Free-Radical-Based, Specific Desulfurization of Cysteine: A Powerful Advance in the Synthesis of Polypeptides and Glycopolypeptides, *Angew. Chem. Int. Ed.*, 46, 9248-9252
25. Lakomek, Nils-Alexander., Kaufman, Joshua D., Stahl, Stephen J., Louis, John M., Grishaev, Alexander., Wingfield, Paul T., Bax, Ad. (2013) Internal Dynamics of the Homotrimeric HIV-1 Viral Coat Protein gp41 on Multiple Time Scales, *Angew. Chem. Int. Ed.*, 52, 3911-3915
26. Rose, Keith. (1994) Facile Synthesis of Homogeneous Artificial Proteins, *J. Am. Chem. Soc.*, 116, 30-33

27. Shao, Jun., Tam, James P. (1995) Unprotected Peptides as Building Blocks for the Synthesis of peptide Dendrimers with Oxime, Hydrazone, and Thiazolidine Linkages, *J. Am. Chem. Soc.*, 117, 3893-3899
28. Low, Donald W., Hill, Michael G., Carrasco, Michael R., Kent, Stephen B.H., Botti, Paolo. (2001) Total Synthesis of Cytochrome b562 by Native Chemical Ligation Using a Removable Auxiliary, *PNAS*, 98, 6554-6559
29. Yan, Liang Z., Dawson, Phillip E. (2001) Synthesis of Peptides and Proteins Without Cysteine Residues by Native Chemical Ligation Combined with Desulfurization, *J. Am. Chem. Soc.*, 123, 526-533
30. Kent, Stephen. (2003) Total Chemical Synthesis of Enzymes, *J. Peptide Sci.*, 9, 574-593
31. Schnölzer, Martina., Kent, Stephen B. H. (1992) Constructing Proteins by Dove Tailing Unprotected Synthetic Peptides: Backbone-Engineered HIV Protease, *Science*, 256, 221-225
32. Cocchi, Fiorenza., DeVico, Anthony L., Garzino-Demo, Alfredo., Cara, Andrea., Gallo, Robert C., Lusso, Paolo. (1996) The V3 Domain of the HIV-1 gp120 Envelope Glycoprotein is Critical for Chemokine-Mediated Blockade of Infection, *Nature Medicine*, 2, 1244-1247
33. Thapa, Parashar., Zhang, Rui-Yang., Menon, Vinay., Bingham, Jon-Paul. (2014) Native Chemical Ligation: A Boon to Peptide Chemistry, *Molecules*, 19, 14461-14483
34. Torbeev, Vladimir Yu., Raghuramen, H., Hamelberg, Donald., Tonelli, Marco., Westler, William M., Perozo, Eduardo., Kent, Stephen B. H. (2011) Protein Conformational Dynamics in the Mechanism of HIV-1 Protease Catalysis, *PNAS*, 108, 20982-20987
35. Sackett, Kelly., TerBush, Allan., Weliky, David P. (2011) HIV gp41 Bundle Constructs Induce Rapid Vesicle Fusion at pH 3.5 and Little Fusion at pH 7.0: Understanding pH Dependence of Protein Aggregation, Membrane Binding, and Electrostatics, and Implications for HIV-host Cell Fusion, *Eur. Biophys. J.*, 40, 489-502
36. Gabrys, Charles M., Qiang, Wei., Sun, Yan., Xie, Li., Schmick, Scott D., Weliky, David P. (2013) Solid-State Nuclear Magnetic Resonance Measurements of HIV Fusion Peptide <sup>13</sup>CO to Lipid <sup>31</sup>P Proximities Support Similar Partially Inserted Membrane Locations of the  $\alpha$  Helical and  $\beta$  Sheet Peptide Structures, *J. Phys. Chem. A*, 117, 9848-9859
37. Xie, Li., Jia, Lihui., Liang, Shuang., Weliky, David P. (2015) Multiple Locations of Peptides in the Hydrocarbon Core of Gel-Phase membranes Revealed by Peptide <sup>13</sup>C to Lipid <sup>2</sup>H Rotational-Echo Double-Resonance Solid-State Nuclear Magnetic Resonance, *Biochemistry*, 54, 677-684



38. Jia, Lihui., Liang, Shuang., Sackett, Kelly., Xie, Li., Ghosh, Ujjayini., Weliky, David P. (2015) REDOR Solid-State NMR as a Probe of the Membrane Locations of Membrane-Associated Peptides and Proteins, *Journal of Magnetic Resonance.*, 253, 154-165
39. Hu, Jian., Qin, Huajun., Sharma, Mukesh., Cross, Timothy A., Gao, Fei Philip. (2008) Chemical Cleavage of Fusion Peptides for High-Level Production of Transmembrane Peptides and Protein Domains Containing Conserved Methionines, *Biochemica et Biophysica Acta.*, 1778, 1060-1066
40. Marley, Jonathan., Lu, Min., Bracken, Clay. (2001) A Method for Efficient Isotopic Labeling of Recombinant Proteins, *Journal of Biomolecular NMR.*, 20, 71-75
41. Gulion, Terry. (2008) Rotational-Echo, Double-Resonance NMR, *Modern Magnetic Resonance.*, 713-718
42. Gulion, Terry. (1998) Introduction to Rotational-Echo, Double-Resonance NMR, *Concepts in Magnetic Resonance.*, 10, 277-289
43. Dalgleish, Angus G., Beverley, Peter C. L., Clapham, Paul R., Crawford, Dorothy H., Greaves, Melvyn F., Weiss, Robin A. (1984) The CD4 (T4) Antigen is an Essential Component of the Receptor for the AIDS Retrovirus, *Nature.*, 312, 763-767
44. Chan, David C., Fass, Deborah., Berger, James M., Kim, Peter S. (1997) Core Structure of gp41 from the HIV Envelope Glycoprotein, *Cell.*, 89, 263-273
45. Aeffer, Sebastian., Reusch, Tobias., Weinhausen, Britta., Salditt, Tim. (2011) Energetics of Stalk Intermediates in Membrane Fusion are Controlled by Lipid Composition, *PNAS.*, E1609-E1618
46. Sackett, Kelly., Nethercott, Matthew J., Zheng, Zhaoxiong., Weliky, David P. (2014) Solid-State NMR Spectroscopy of the HIV gp41 Membrane Fusion Protein Supports Intermolecular Antiparallel  $\beta$  Sheet Fusion Peptide Structure in the Final Six-Helix Bundle State, *J. Mol. Biol.* 426, 1077-1094
47. Lakomek, Nils-Alexander., Kaufman, Joshua D., Stahl, Stephen J., Louis, John M., Grishaev, Alexander., Wingfield, Paul T., Bax, Ad. (2013) Internal Dynamics of the Homotrimeric HIV-1 Viral Coat Protein gp41 on Multiple Time Scales, *Angew. Chem. Int. Ed.*, 52, 3911-3915
48. Banerjee, Koyeli., Weliky, David P. (2014) 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.*, 53, 7184-7198

49. White, Judith M. Delos, Sue E., Brecher, Matthew., Schornberg, Kathryn. (2008) Structures and Mechanisms of Viral Membrane Fusion Proteins: Multiple Variations on a Common Theme, *Crit. Rev. Biochem. Mol. Biol.*, 43, 189-219
50. Melikyan, Gregory B. (2014) HIV Entry: a Game of Hide-and-Fuse?, *Current Opinion in Virology.*, 4, 1-7
51. Montero, Marinieve., Van Houten, Nienke E., Wang, Xin., Scott, Jamie K. (2008) The Membrane-Proximal External Region of the Human Immunodeficiency Virus Type 1 Envelope: Dominant Site of Antibody Neutralization and Target for Vaccine Design, *Microbiology and Molecular Biology Reviews.* 72, 54-84
52. Apellániz, Beatriz., Huarte, Nerea., Largo, Eneko., Nieva José L. (2014) The Three Lives of Viral Fusion Peptides, *Chemistry and Physics of Lipids.*, 181, 40-55
53. Liu, Jun., Bartesaghi, Alberto., Borgnia, Mario J., Sapiro, Guillermo., Subramaniam, Sriram. (2008) Molecular Architecture of Native HIV-1 gp120 Trimers, *Nature.*, 455, 109-114
54. Schülke, Norbert., Vesanen, Mika S., Sanders, Rogier W., Zhu, Ping., Lu, Min., Anselma, Deborah J., Villa, Anthony R., Parren, Paul W. H. I., Binley, James M., Roux, Kenneth H., Maddon, Paul J., Moore, John P., Olson, William C. (2002) Oligomeric and Conformational Properties of a Proteolytically Mature Disulfide-Stabilized Human Immunodeficiency Virus Type 1 gp140 Envelope Glycoprotein, *Journal of Virology.*, 76, 7760-7776
55. Sanders, Rogier W., Vesanen, Mika., Schuelke, Norbert., Master, Aditi., Schiffner, Linnea., Kalyanaraman, Roopa., Paluch, Maciej., Berkhout, Ben., Maddon, Paul J., Olson, William C., Lu, Min., Moore, John P. (2002) Stabilization of the Soluble, Cleaved Trimeric Form of the Envelope Glycoprotein Complex of Human Immunodeficiency Virus Type 1, *Journal of Virology.*, 76, 8875-8889
56. Bartesaghi, Alberto., Merk, Alan., Borgnia, Mario J., Milne, Jacqueline L. S., Subramaniam, Sriram. (2013) Pre-Fusion Structure of Trimeric HIV-1 Envelope Glycoprotein Determined by Cryo-Electron Microscopy, *Nat. Struct. Mol. Biol.*, 20, 1352-1357
57. Julien, Jean-Philippe et al. (2013) Crystal Structure of a Soluble Cleaved HIV-1 Envelope Trimer, *Science.*, 342, 1477-1483
58. Buzon, Victor., Natrajan, Ganesh., Schibli, David., Campelo, Felix., Kozlov, Michael M., Weissenhorn, Winfried. (2010) Crystal Structure of HIV-1 gp41 Including Both Fusion Peptide and Membrane Proximal External Regions, *PLoS Pathogens.*, 6, 1-7
59. Liang, S., Ratnayake, U., Keinath, C., Jia, L., Wolfe, R., Ranaweera, Weliky, D. P. Efficient Fusion at Neutral pH by Human Immunodeficiency Virus gp41 Trimers

Containing the Fusion Peptide and Transmembrane Domains. *Biochemistry* 2018, 57, 1219-1235.

60. Takeharu Mino, Seiji Sakamoto, Itaru Hamachi. Recent applications of N-acyl imidazole chemistry in chemical biology. *Bioscience, Biotechnology, and Biochemistry*, Volume 85, Issue 1, January 2021, Pages 53–60

61. Raffaello Verardi, Nathaniel J Traaseth, Larry R Masterson, Vitaly V Vostrikov, Gianluigi Veglia. Isotope labeling for solution and solid-state NMR spectroscopy of membrane proteins. *Adv. Exp. Med Biol.* **2012**; 992: 35-62.

**CHAPTER 4:**  
**APPLICATIONS OF NATIVE CHEMICAL LIGATION OF GP41 ECTODOMAIN TO**  
**STRUCTURAL ANALYSIS BY REDOR NMR**

#### 4.1. Introduction

Human immunodeficiency virus (HIV) is a membrane-enveloped virus whose initial infection of host cells begins with membrane fusion through a process initiated by gp160 (3,6,7). The gp160 glycoprotein complex is comprised of two noncovalently associated subunits, gp120 and gp41. The membrane of the virion has clusters of three gp160 comprised of three gp120 and three gp41 molecules (3,7). HIV gp41 has three major domains: the ectodomain, the transmembrane domain, and the cytoplasmic domain, which work together to mediate the membrane fusion functions of the gp41 protein (3,7). The gp41 is originally part of a larger noncovalent trimer of heterodimer complex with gp120. The gp120 recognizes target host cells by binding to CD4 and co-receptor proteins, which causes dissociation from the gp41 ectodomain and subsequent structural rearrangement of the ectodomain (3,7). The ectodomain itself is subdivided into different domains with defined structures and functions. **Figure 3.1** displays the protein ectodomain constructs we are investigating for this study. The domains include the fusion peptide (FP), and the hairpin + membrane proximal external region (MPER) previously described (**Section 3.1**).

FP includes 16 apolar residues at the N-terminus. Virus/cell fusion is impaired when there are deletions and mutations present in the FP. MPER is proposed to bind to the viral membrane and is adjacent to the transmembrane domain. Increased rates and extents of vesicle fusion have been experimentally observed in ectodomain constructs containing both the FP and the MPER however, there is little evidence of close contact between the FP and MPER. The synergistic effects of the FP and MPER together in increasing the final fusion extent in lipid vesicles indicates that there may be some interactions between the FP and the MPER (48).

The main benefit of the solid phase peptide synthesis (SPPS) reaction, relative to biosynthesis, is the addition of amino acids one at a time with the possibility of specific and selective labeling of individual residues. Any of these single amino acids may be specifically labeled at one or more atoms in the amino acid. The NCL reaction is then used in combination with the SPPS to synthesize larger protein constructs that retain the labeling selectivity of the SPPS fragment. In contrast, protein synthesis solely by bacteria is possible but require extensive molecular biology (59). Lack of site-specific isotopic labeling can create some ambiguity in the assignment of the signals in the solid state nuclear magnetic resonance (SSNMR) spectra. Multiple proteins have been successfully synthesized by NCL ranging across the different classes of proteins including redox proteins, intracellular proteins, membrane proteins, and enzymes (**Section 1.5**).

This study describes the application of REDOR NMR (37) to large protein constructs for the purpose of protein structural studies. REDOR in combination with site-specific isotopic labeling can provide several important insights into the structural biology of large protein constructs. Among these insights is the probing of proximity of protein domains relative to each other. In this study we will explore proximity of the FP to the HM. Site-specific labeling can also be used to probe membrane location of large constructs in lipid vesicles using  $^{13}\text{C}$  site-specific labels in the protein and  $^2\text{H}$  labels in the lipid. It is also possible to use REDOR to test the formation of monomers and trimers.

#### **4.2. Materials and Methods**

This section is like **Section 3.2** and the reader may wish to skip this section. Materials were purchased from the following companies: DNA – GenScript (Piscataway, NJ); Escherichia coli BL21(DE3) strain – Novagen (Gibbstown, NJ); Luria-Bertani (LB)

medium- Dot Scientific (Burton, MI); isopropyl  $\beta$ -D-thiogalactopyranoside (IPTG) and tris-(carboxyethyl) phosphine (TCEP) – Goldio (St. Louis, MO); Co<sup>2+</sup>-resin – Thermo Scientific (Waltham, MA); D-Glucose-1,2,3,4,5,6-<sup>13</sup>C<sub>6</sub> (98%) – Synthose (Ontario, Canada). Most other materials were obtained from Sigma-Aldrich (St. Louis, MO).

#### 4.2.1. Fusion Peptides (3.2.1)

FP peptides include FP  $\equiv$  gp41<sub>512-534(S534A)</sub> and non-native C-terminal (CO)S(CH<sub>2</sub>CH<sub>2</sub>COOH) for ligation with HM. The S534A mutation reduces sidechain volume and may increase ligation rate. In many cases, there was non-native N-terminal H<sub>6</sub>G<sub>6</sub>D<sub>4</sub>K which includes H<sub>6</sub> for FP\_HM binding to Co<sup>2+</sup>-resin, and D<sub>4</sub>K for enterokinase cleavage of FP\_HM from the resin. G<sub>6</sub> is an unstructured spacer that increases exposure of the H<sub>6</sub> and D<sub>4</sub>K segments. Increased exposure of the H<sub>6</sub> tag prevent the possibility of the tag being hidden from the Co<sup>2+</sup> affinity resin by other parts of the protein.

Solid-phase peptide synthesis was done manually using t-butoxycarbonyl (t-boc) chemistry and S-trityl- $\beta$ -mercaptopropinoyl-p-methyl-benzhydrylamine resin (230 mg, 0.88 meq/g). Sidechain protecting groups for amino acids include: His and Arg, tosyl; Asp, benzyl ester; Lys, carboxybenzyl; Ser and Thr, benzyl. A liquid or liquid solution with reagents was added to the resin in a 40 mL Teflon vessel with cap, filter, stopcock, and nozzle, followed by shaking the vessel, and then drainage of the liquid from the vessel. Synthesis began with resin-swelling in CH<sub>2</sub>Cl<sub>2</sub> (3 mL, 1 h) followed by trityl-group cleavage in 95:2.5:2.5 (v:v) TFA:H<sub>2</sub>O:triisopropylsilane (10 mL, 4 minutes, 2 $\times$ ).

The first cycle of coupling amino acid (Ala-534) to the resin began with resin washing with CH<sub>2</sub>Cl<sub>2</sub> (3 mL, 1 min, 5 $\times$ ) and then 5% N,N-diisopropylethylamine (DIEA) in CH<sub>2</sub>Cl<sub>2</sub> (3 mL, 1 min, 3 $\times$ ), with concurrent reaction in a flask between t-boc-Ala (6.8

mmole) and activator 3-(diethoxyphosphoryloxy)-1,2,3-benzotriazin-4(3H)-one (DEPBT) (6.4 mmole) in tetrahydrofuran (THF, 5.5 mL). DIEA (1.1 mL) was added, and the total solution used to couple Ala-534 to the resin (4 h), followed by washes with CH<sub>2</sub>Cl<sub>2</sub> (3 mL, 1 min, 5×).

Deprotection (t-boc cleavage) was done with 50:48:2 (v:v:v) TFA:CH<sub>2</sub>Cl<sub>2</sub>:anisole (3 mL) followed by coupling of Arg-533 to the resin using a similar procedure as above, but with 3.4 mmol amino acid and 3.2 mmol DEPBT in 2.8 mL THF. There were then sequential deprotection/coupling cycles to complete the synthesis. There was a final deprotection followed by washing with CH<sub>2</sub>Cl<sub>2</sub> (3 mL, 1 min, 5×) and 5% DIEA in CH<sub>2</sub>Cl<sub>2</sub> (3 mL, 1 min, 3×), and drying overnight in a vacuum desiccator. Peptide was cleaved from the resin using HF at Midwest Biotech Corporation (Fishers, IN) **Section 3.2.1**.

#### **4.2.2. Molecular Biology and Protein Expression and Separation (3.3.2)**

The FP\_HM amino acid sequence is shown in **Figure 3.1** and is based on the HXB2 laboratory strain of HIV. The FP sequence is gp41<sub>512-534(S534A)</sub> and the HM sequence is gp41<sub>535-581(M535C)</sub>-SGGRGG-gp41<sub>628-683</sub>. Like the gp41 SE, HM adopts helical hairpin structure that is hyperthermostable with  $T_m > 100$  °C. Replacement of residues 582-627 with SGGRGG improves solubility. The N-terminal Cys of HM is used for ligation with the C-terminal thioester of FP.

Protein was produced by expression in *E. coli* bacteria, BL21(DE3) strain, using vectors with inserts. Proteins included HM\_G6LEH<sub>6</sub> in the pET-24a(+) vector, which have been previously described. HM was produced in the pET-24a(+) vector after replacement of the first Gly codon with a stop codon.



Production of unlabeled protein began with addition of 1.5 mL *E. coli* glycerol stock to 50 mL LB medium followed by growth for 3 h. This and all other growths were done at 37 °C in shake flasks at 180 rpm, and with 50 µL of 50 mg/mL kanamycin in the medium. The culture was added to 1 L fresh LB medium, followed by: (1) growth for 2 h to OD<sub>600</sub> ≈ 0.8; (2) addition of 2 mM IPTG and induction of protein expression overnight at 37 °C; and (3) harvesting the cell pellet after centrifugation at 9000 *g* for 10 min.

Production of labeled protein began with growth in LB for 3 h followed by harvesting the cell pellet. Minimal medium was prepared by mixing autoclaved aqueous solutions that included: (1) 50 mL with M9 salts (0.34 g Na<sub>2</sub>HPO<sub>4</sub>, 0.15 g KH<sub>2</sub>PO<sub>4</sub>, 0.03 g NaCl, and 0.05 g NH<sub>4</sub>Cl); (2) 1 mL with 0.1 M CaCl<sub>2</sub>; and (3) 1 mL with 1 M MgSO<sub>4</sub>, as well as 0.5 mL MEM vitamin, and 200 mg glucose. The cell pellet was suspended in 10 mL minimal medium, and a 2 mL aliquot of this suspension was then added to the remaining minimal medium, followed by 4 h growth. Individual 5 mL aliquots of the suspension were then added to four separate flasks that each contained 50 mL fresh minimal medium, followed by 2 h growth, induction of expression, and harvesting of the cell pellet. Cells in this step of the growth process could be made into glycerol stock in an analogous manner to the LB glycerol stocks. The minimal media glycerol stocks could then be used to grow cell in minimal media to O.D. 0.8 followed by expression. Protein with fractional <sup>13</sup>C-labeling was produced using minimal medium that contained a mixture of unlabeled- and 1,2,3,4,5,6-<sup>13</sup>C D-glucose, and protein with fractional <sup>2</sup>H-labeling was produced using minimal medium that contained 1,2,3,4,5,6,6-<sup>2</sup>H D-glucose and a mixture of H<sub>2</sub>O and D<sub>2</sub>O.

Separation of inclusion body-rich material began with: (1) tip-sonication in ~30 mL PBS per 10g cells at pH 7.4 in a 50 mL beaker in an ice bath, (2) centrifugation at 27000*g*

for 30 min and harvesting the new pellet, and (3) 2× repetition of the sonication/centrifugation/harvesting steps with a resultant pellet 1. The next step was tip sonication of the pellet 1 in ~30 mL PBS at pH 7.4 with 6 M GuCl, followed by centrifugation. Much of pellet 1 was solubilized, as evidenced by a new pellet:pellet 1 volume ratio  $< \frac{1}{2}$ . The supernatant was dialyzed against deionized water overnight at 10 °C, using a 10 kDa cutoff membrane, with accompanying precipitation of HM-enriched material, followed by centrifugation at 11700 *g* for 40 min, and harvesting the pellet 2. There were two variants of the remaining procedure that are denoted A and B. For protocol A, much of pellet 2 was dissolved by vortexing in ~10 mL PBS at pH 8 with 8 M urea, followed by centrifugation at 11700 *g*, 40 minutes, and dialysis at 10 °C of the supernatant against water (one day total with three changes to fresh water) with accompanying precipitation by adding 0.2 g NaCl to initiate precipitation. Precipitation was usually performed after transferring protein in solution to a conical vial. The suspension was centrifuged and pellet IIIA was harvested. For protocol B, pellet II was desalted by vortexing in ~50 mL fresh deionized water, followed by centrifugation and harvesting pellet IIIB. Pellets IIIA and IIIB were sometimes lyophilized and stored at low temperature prior to using them in ligations or NMR experiments. Cells that expressed HM with a histidine tag were sometimes subjected to a previously published protocol which included sonication in PBS, centrifugation, and harvesting the pellet (3×), solubilization of the final pellet in PBS with 6 M GuHCl, and Co<sup>2+</sup>- affinity chromatography

### **Section 3.3.2.**

#### 4.2.3. Characterization of Proteins by Mass Spectrometry and Solid-State NMR (SSNMR) Spectroscopy (3.2.3)

MALDI-TOF mass spectrometry was done using a Kratos Analytical Axima-CFR Plus instrument. Protein (~0.1 mg) was vortexed in 1 mL of 98% formic acid, and a 2  $\mu$ L aliquot was then mixed by pipette with 4  $\mu$ L of a solution containing  $\alpha$ -cyano-4-hydrocinnamic acid (10 mg/mL) in 3:1 acetonitrile:0.1% TFA. A 2  $\mu$ L aliquot of the mixed solution was transferred to the MALDI plate, dried, and then subjected to MALDI-TOF in linear positive mode. There were often  $M^+$  and  $M^{2+}$  signals, with assignments to a single chemical species done using  $(m/z)_{M^+} \approx 2 \times (m/z)_{M^{2+}}$ .

SSNMR experiments were done using a 9.4 T Agilent Infinity Plus spectrometer and a magic angle spinning (MAS) probe equipped for a 4 mm diameter rotor and tuned simultaneously to  $^1\text{H}$ ,  $^{13}\text{C}$ , and  $^2\text{H}$  NMR frequencies. Rotational-echo double-resonance (REDOR) data were acquired using a pulse sequence: (1)  $^1\text{H}$   $\pi/2$  pulse, (2)  $^1\text{H}$ - $^{13}\text{C}$  cross polarization (CP), (3) dephasing period of duration  $\tau$ , and (4)  $^{13}\text{C}$  detection.  $S_0$  and  $S_1$  REDOR data were acquired alternately and differed in the pulses applied during the dephasing period. For both  $S_0$  and  $S_1$  there was a  $^{13}\text{C}$   $\pi$  pulse at the end of each rotor cycle except for the last one and for  $S_1$ , there was also a  $^2\text{H}$   $\pi$  pulse at the midpoint of each cycle. Typical parameters included: (1) 10 kHz MAS frequency and 1.5 ms CP contact time, (2) 50 kHz rf fields for  $^1\text{H}$   $\pi/2$  pulse and CP, (3) 55-66 kHz  $^{13}\text{C}$  CP ramp, (4) 60 kHz  $^{13}\text{C}$   $\pi$  pulses, 100 kHz  $^2\text{H}$   $\pi$  pulses with XY-8 phase cycling applied to all  $\pi$  pulses; and ~70 kHz two-pulse phase-modulated  $^1\text{H}$  decoupling during dephasing and acquisition. Typical recycle delays were 1 s ( $\tau = 2, 8, 16$  ms), 1.5 s ( $\tau = 24, 32$  ms), and 2 s ( $\tau = 40$  and 48 ms). Typical numbers of summed  $S_0$  or  $S_1$  scans were ~4000, 7000,

12000, 25000, 32000, 40000, and 50000 for  $\tau = 2, 8, 16, 24, 32, 40$  and  $48$  ms, respectively.  $^{13}\text{C}$  chemical shift referencing was done externally using the methylene peak of adamantane at 40.5 ppm **Section 3.2.3**.

#### **4.2.4. Solid State NMR Sample Preparation**

The lipid composition was DPPC (Dipalmitoylphosphatidylcholine), and DPPG (1,2-Dipalmitoyl-*sn*-glycero-3-phospho-(1'-*rac*-glycerol) (Sodium Salt)) with 4:1 ratio; and DPPC, DPPG, and cholesterol with 8:2:5 mole ratio. The cholesterol mole fraction in both compositions is close to that of the plasma membrane of host cells of HIV. This composition was chosen because: (1) the PC lipids are major fraction in HIV-1 virus host cell plasma membrane, and (2) the host cell plasma membrane is negatively charged, and the charge of the plasma membrane is like the 4:1 DPPC:DPPG charge.  $50\ \mu\text{mol}$  lipids were dissolved in 2 mL chloroform and methanol solution with a 9:1 volume ratio, and the solvent was removed by dry nitrogen gas flow and vacuum pumping overnight. 3mL of 10 mM 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES) and 5mM 2-(N-morpholino) ethane sulfonic acid (MES) buffer at pH 5.0 was used to hydrate the lipid film and followed by 10 times freeze-thaw cycles to make a homogenous suspension of unilamellar vesicles. The lipid-buffer suspension was extruded 10 times through a polycarbonate membrane with 100 nm pore size to get large unilamellar vesicles. Small unilamellar vesicles (SUV) are typically below 100 nm, large unilamellar vesicles (LUV) are between 100 nm and  $1\ \mu\text{m}$ , and giant unilamellar vesicles (GUV) are above  $1\ \mu\text{m}$ . The target protein ( $1\ \mu\text{mol}$ ) was dissolved in 10 mL of the HEPES/MES buffer and added to lipid vesicles drop by drop, then agitated overnight. The peptide-lipid complex was pelleted down by ultra-centrifugation at  $45000\ g$  for 4 hours. The proteo-

liposome complex pellet was lyophilized overnight. Lyophilization helps to reduce sample lost when packing the sample into the NMR rotor. The sample was packed in the NMR rotor and rehydrated with 10  $\mu$ L of the HEPES/MES buffer overnight at room temperature.

#### **4.2.5. Native Chemical Ligation and Purification (3.2.4)**

Ligation buffer was prepared with GuHCl (19.1 g), imidazole (4.3 g), 2.11 mL of 1 M  $\text{Na}_2\text{HPO}_4$ , 0.43 mL of 1 M  $\text{NaH}_2\text{PO}_4$ , and  $\text{H}_2\text{O}$  in ~25 mL total volume. The solution was heated to 100  $^\circ\text{C}$  for 1 minute and vortexed, cooled to room temperature, and ~5 mL water then added to achieve clarity. Typical solute concentrations were: GuHCl, 6.4 M, imidazole, 2.0 M, and phosphate, 1.4 mM. The FP solution was typically prepared with a FP variant like  $\text{H}_6\text{G}_6\text{D}_4\text{K\_FP}$  (~1.2 mg, ~0.3  $\mu\text{mol}$ ) and 4-mercaptophenylacetic acid (MPAA) catalyst (7.5 mg, 45  $\mu\text{mol}$ ) dissolved in 0.9 mL ligation buffer followed by constant stirring for 30 minutes. A HM construct in ligation buffer (~5 mg/mL) was prepared and ~0.2 mL (~0.07  $\mu\text{mol}$ ) transferred to the FP/MPAA solution. The reducing agent tris-(carboxyethyl) phosphine (TCEP, 0.50 M) solution was prepared in ligation buffer and 0.2 mL then transferred to the FP + HM solution, followed by addition of ligation buffer to achieve a total volume of 1.5 mL with pH adjusted to 6.8. The approximate solute concentrations were  $\text{H}_6\text{G}_6\text{D}_4\text{K\_FP}$ , 200  $\mu\text{M}$ ; HM, 50  $\mu\text{M}$ ; MPAA, 30 mM; and TCEP, 67 mM. The ligation reaction was done by stirring the solution overnight, and pH typically increase to ~7.0 during this time.

Separation of ligation product began with overnight dialysis of the reaction against 8 M urea using a membrane with 10 kDa cutoff, which removed  $\text{H}_6\text{G}_6\text{D}_4\text{K\_FP}$ . The  $\text{H}_6\text{G}_6\text{D}_4\text{K\_FP\_HM}$  and HM were then precipitated by dialysis against water for one day, with one water change. The dialysis suspension was centrifuged (48000g, 20 min), and

the solid pellet harvested and then solubilized by vortexing in 3 mL buffer that contained PBS, pH 7.4, and 6 M GuHCl. A 1 mL suspension of Co<sup>2+</sup>-resin in ethanol (HisPur, Thermo-Fisher) was added to a small plastic column, and excess ethanol removed by filtration. The resin was then washed with PBS, pH 7.4, with 8 M urea and 10 mM imidazole (1 mL, 3×) and then PBS, pH 7.4, with 6 M GuHCl (1 mL, 3×), followed by suspension in 3 mL of the latter buffer. The resin suspension was transferred to the protein solution, and the mixture then agitated overnight at 10 °C to achieve H<sub>6</sub>G<sub>6</sub>D<sub>4</sub>K\_FP\_HM binding to the resin. Excess liquid with unbound protein was removed by filtration, and the resin then washed with 6.7 M GuHCl, 2.1 M urea, and 0.5 M TCEP at pH 3.5 (1 mL, 1×), where the lower pH aided unbinding of HM from the resin. The resin was then washed with 50 mM Tris buffer, pH 8.0, with 10 mM CaCl<sub>2</sub> and 0.1% tween-20 (1 mL, 4×). The resin was suspended in 3 mL of the latter buffer, and transferred by Pasteur pipette to a larger vial, and buffer added to achieve 5 mL total volume. Enterokinase (50 units/50 µL; EKMax, Thermo-Fisher) was added to the resin suspension followed by overnight incubation at 37 °C without agitation. The cleavage between D<sub>4</sub>K and FP\_HM released FP\_HM into the solution. The resin suspension was filtered, followed by washes with PBS buffer, pH 8.0, with 8 M urea (1 mL, 3×). The filtrate and washes were combined and then dialyzed against water which precipitated the protein. The protein suspension was centrifuged (11600g, 40 min) and the protein pellet harvested

#### **Section 3.2.4.**

#### **4.2.6. Formation of HM Trimers with Mixed Labels**

Each protein was expressed separately in appropriate conditions for producing labeled HM (**Section 4.2.2**). The proteins were expressed and purified independently and

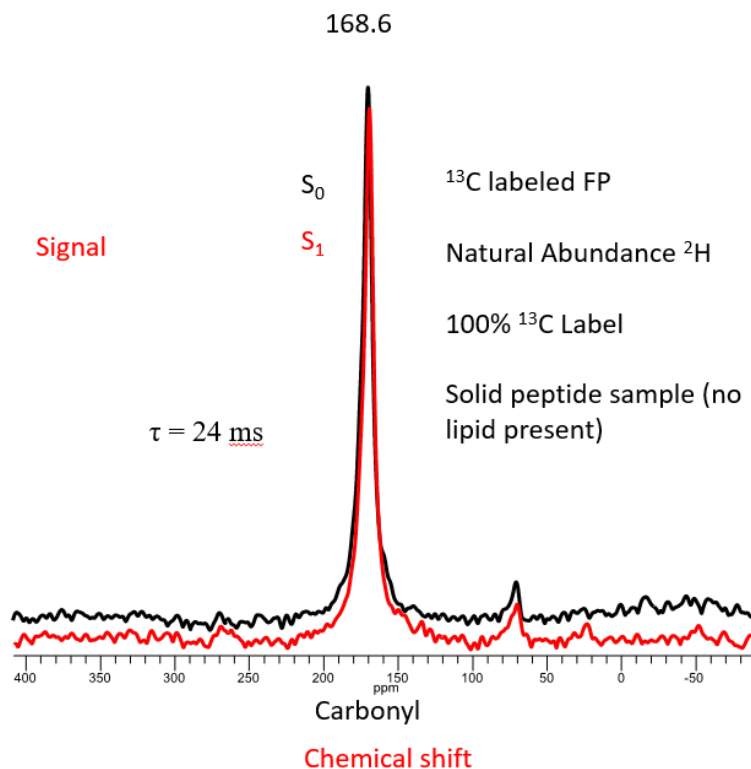
mixed under ligation conditions (**section 4.2.5**). We hypothesize that the HM forms monomers under ligation conditions due to the nature of NCL. The 1 mL of 50  $\mu$ M protein mixture was dialyzed against 50mM formate buffer to encourage formation of monomeric HM. The mixed protein in formate buffer was reconstituted in lipid vesicles overnight 8:2:5 (DPPC: DPPG: Cholesterol) vesicles in HEPES/MES pH 5.0 **Section 4.2.4**. The protein to lipid molar ratio was 1:100.

### **4.3. Results and Discussion**

Our strategy for REDOR studies has involved using site-specifically labeled FP. The FP was 100% labeled at the Gly-5, 10, and 16. HM proteins were produced with either  $^{13}\text{C}$  label,  $^2\text{H}$  label, or natural abundance label. Typically,  $^{13}\text{C}$  labeled HM was produced at approximately 13% label to prevent loss of signal from  $^{13}\text{C}$ - $^{13}\text{C}$  dipolar coupling.  $^2\text{H}$  HM was produced at approximately 54% label. Higher labeling fractions were produced but yield of HM protein decreased as concentration of  $^2\text{H}$  increased. HM protein with fractional  $^{13}\text{C}$  labeling was produced using expression in a minimal medium with a mixture of unlabeled and 1,2,3,4,5,6- $^{13}\text{C}$  D-glucose. HM protein with fractional  $^2\text{H}$  labeling was produced using 1,2,3,4,5,6,6- $^2\text{H}$  D-glucose and a mixture of  $\text{D}_2\text{O}$  and  $\text{H}_2\text{O}$ . FP peptides were synthesized with high purity using t-boc chemistry, as evidenced by mass spectrometry. The HM was expressed and purified from inclusion bodies, yielding protein of high purity as evidenced by MALDI MS. Additional data for HM is available (**Section 3, Figure 3.6**).

#### 4.3.1. Site-Specifically Labeled FP

$^{13}\text{C}$  labeled FP was synthesized by SPPS using t-boc chemistry. Glycine 5, 10, and 16 were replaced with a  $^{13}\text{CO}$  labeled glycine. The method for producing  $^{13}\text{C}$  labeled peptide followed the general procedure in (Section 4.2.1.)



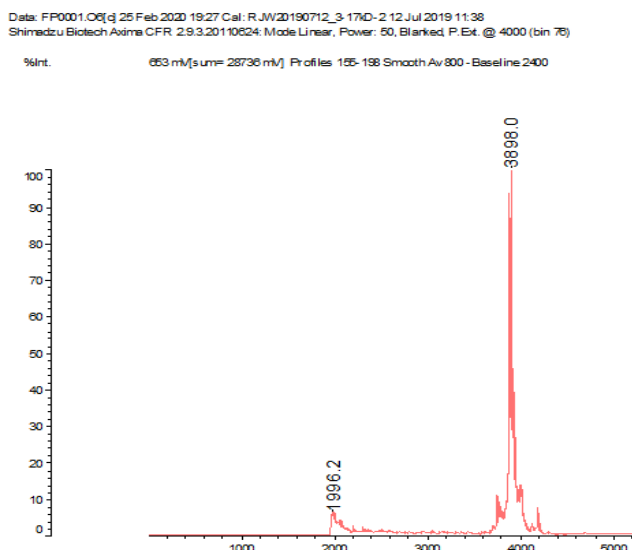
**Figure 4.1**  $^{13}\text{C}$  REDOR NMR of approximately 5 mg SPPS FP

(HHHHHHGGGGGGDDDDKAVGIGALFLGFLGAAGSTMGAAS). The  $S_0$  spectra shows a strong carbonyl peak at 168.6 ppm. The chemical shift corresponds to a likely  $\beta$  sheet structure. Glycine 5, 10, and 16 (AVGIGALFLGFLGAAGSTMGAAS) are the labeled amino acids in this protein construct.

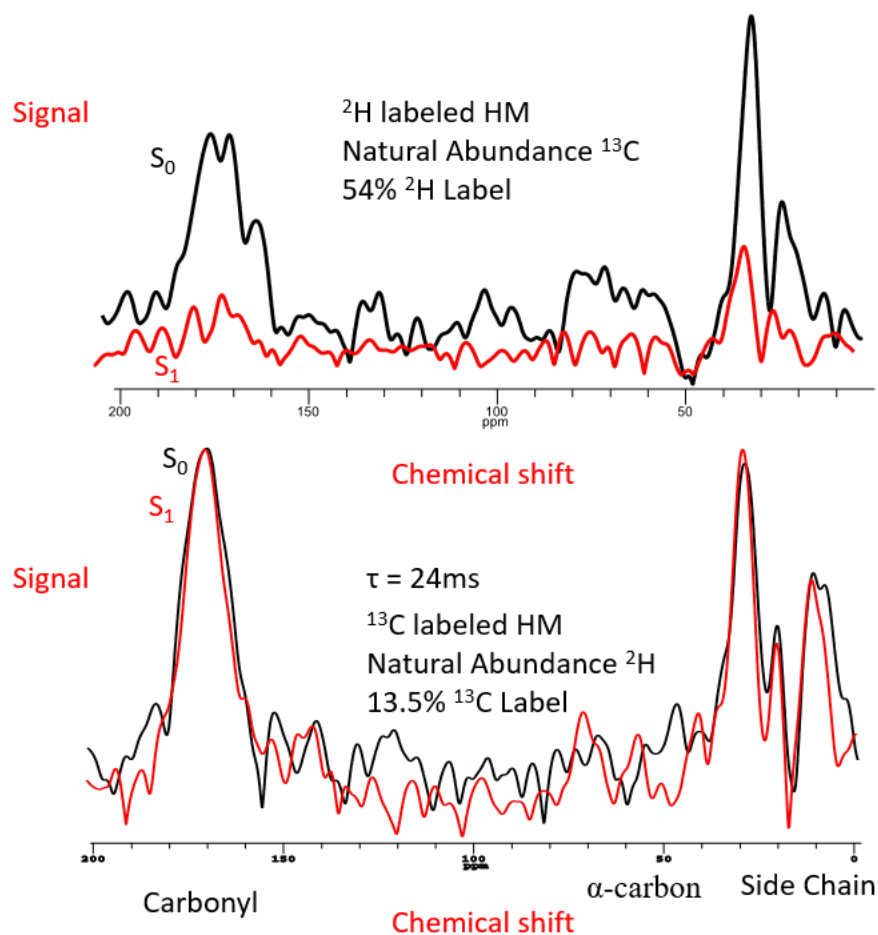
Isotopically labeled and lyophilized FP was probed by  $^{13}\text{C}$ - $^2\text{H}$  REDOR NMR (Figure 4.1). REDOR probes proximal ( $<8 \text{ \AA}$ )  $^{13}\text{C}$  and  $^2\text{H}$  nuclei, using attenuation of REDOR  $^{13}\text{C}$   $S_1$  vs.  $S_0$  signals. This attenuation is quantified as  $\Delta S/S_0$ , where  $\Delta S$  is the



difference between  $S_0$  and  $S_1$  signal intensities. For  $^{13}\text{C}$ 's proximal to  $^2\text{H}$ 's,  $\Delta S/S_0$  increases as a function of the experimental dephasing time  $\tau$ . (**Figure 4.1**) displays representative REDOR  $^{13}\text{C}$  spectra of FP with 100%  $^{13}\text{C}$  labeling based on mass spectroscopy (**Figure 4.3**). There is negligible attenuation of  $S_1$  signals, which correlates with only 0.01%  $^2\text{H}$  natural abundance. The MALDI mass spectra of the FP used in the NMR experiment is shown in (**Figure 4.2**).



**Figure 4.2** Representative MALDI mass spectra of  $^{13}\text{C}$  FP (three  $^{13}\text{C}$  labels) produced by SPPS (HHHHHHGGGGGGDDDDKAVGIGALFLGFLGAAGSTMGAAS). A single peak is observed at 3898. The calculated mass of FP is 3862. Sodium adducts may be the reason for the mass difference. The likely origin of the 1996 m/z peak is  $\text{M}^{2+}$  peak.



**Figure 4.3** Representative REDOR NMR spectra for  $^2\text{H}$  labeled HM protein (top) and  $^{13}\text{C}$  labeled HM protein (bottom). The representative protein spectra are the same as seen in **Figure 3.7**. The dephasing time is 24ms for both spectra. S1 only show strong dephasing in the  $^2\text{H}$  labeled sample and every observable region including side chain,  $\alpha$  carbons, and carbonyl carbons show dephasing.

#### 4.3.2. Preparation of Isotopically Labeled HM

Isotopically labeled and lyophilized HM was also probed by  $^{13}\text{C}$ - $^2\text{H}$  REDOR NMR (**Figure 4.3**). REDOR probes proximal ( $<8 \text{ \AA}$ )  $^{13}\text{C}$  and  $^2\text{H}$  nuclei, using attenuation of REDOR  $^{13}\text{C}$   $S_1$  vs.  $S_0$  signals. This attenuation is quantified as  $\Delta S/S_0$ , where  $\Delta S$  is the difference between  $S_0$  and  $S_1$  signal intensities. For  $^{13}\text{C}$ 's proximal to  $^2\text{H}$ 's,  $\Delta S/S_0$

increases as a function of the experimental dephasing time  $\tau$ . (**Figure 4.3**) displays representative REDOR  $^{13}\text{C}$  spectra of HM with 54%  $^2\text{H}$  labeling based on mass spectroscopy. The 1% naturally abundant  $^{13}\text{C}$  nuclei are randomly distributed throughout HM. The high attenuation of  $S_1$  intensity in all spectral regions evidence that most  $^{13}\text{C}$ 's have proximal  $^2\text{H}$ 's, i.e., the  $^2\text{H}$ 's are also randomly distributed throughout the non-exchangeable sites of HM. (**Figure 4.3**) also displays representative REDOR spectra of HM with 13.5%  $^{13}\text{C}$  labeling. There is negligible attenuation of  $S_1$  signals, which correlates with only 0.01%  $^2\text{H}$  natural abundance. REDOR spectra were also acquired for HM with 82%  $^{13}\text{C}$ -labeling, and there was also negligible attenuation of  $S_1$  vs.  $S_0$  signals. These samples exhibited reduced  $S_0$  signals at longer  $\tau$ , likely because of shorter  $^{13}\text{C}$ - $^{13}\text{C}$  distances and consequent larger  $^{13}\text{C}$ - $^{13}\text{C}$  dipolar couplings that are recoupled by the rotor-synchronized  $^{13}\text{C}$   $\pi$  pulses.

#### 4.3.3. REDOR NMR of $^{13}\text{CFP}$ - $^2\text{HHM}$ in Lipid Vesicles

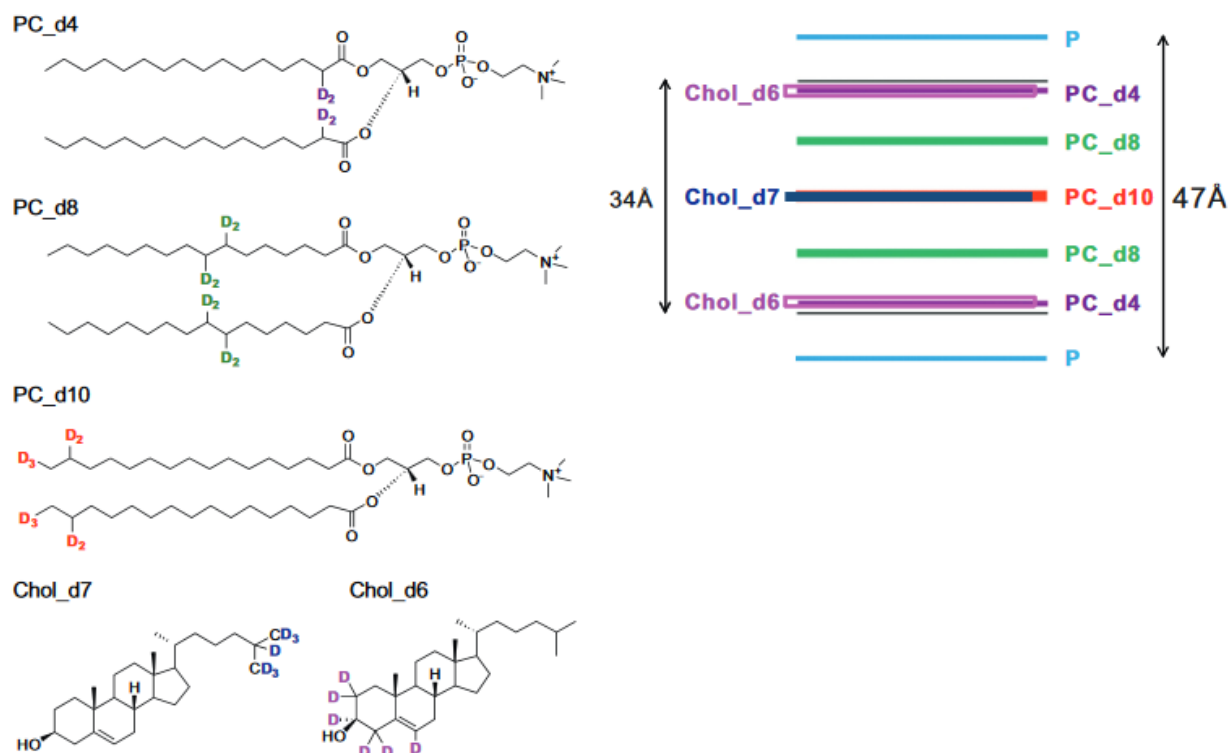
The full FP\_HM synthesis was done using the ligation and purification protocols presented in (**Section 3**). The FP\_HM was pure, and yield was ~10% relative to the HM limiting reactant. The  $G_6$  flexible linker in the FP N-terminal  $H_6G_6D_4K$  tag may be important for yield, both for  $H_6$  exposure for resin binding, and for  $D_4K$  exposure for enterokinase cleavage. This idea is supported by comparison with recombinant  $H_{10}S_2GHID_4K$ \_FP\_HM with  $S_2GHI$  rather than  $G_6$  linker. The (moles FP\_HM after cleavage)/ (moles  $H_{10}S_2GHID_4K$ \_FP\_HM initially added to the resin) is only ~10%.

Larger quantities of FP\_HM were produced using parallel reactions and purifications at the smaller scale. Reaction and purification (12 $\times$ ) were done using FP with  $^{13}\text{CO}$  labels at  $G_5$ ,  $G_{10}$ , and  $G_{16}$ , followed by pooling of the cleavage product solutions,

and precipitation of FP\_HM product by dialysis against water. This yielded ~1.5 mg FP\_HM which was then reconstituted in membrane using the protocol in (**Section 4.2.4**). **Figure 3.7C** shows  $^{13}\text{C}$  REDOR spectrum of this sample. For  $^{13}\text{C}$ 's proximal to  $^2\text{H}$ 's,  $\Delta S/S_0$  increases as a function of the experimental dephasing time  $\tau$  up to 24- 32ms. There is minor attenuation of  $S_1$  signals, which does not correlate with FP being in proximity with HM. This is consistent with other experimental results (60).

#### 4.3.4. Probing Membrane Depth of $^{13}\text{C}$ Labeled HM in $^2\text{H}$ Labeled Lipid Vesicles

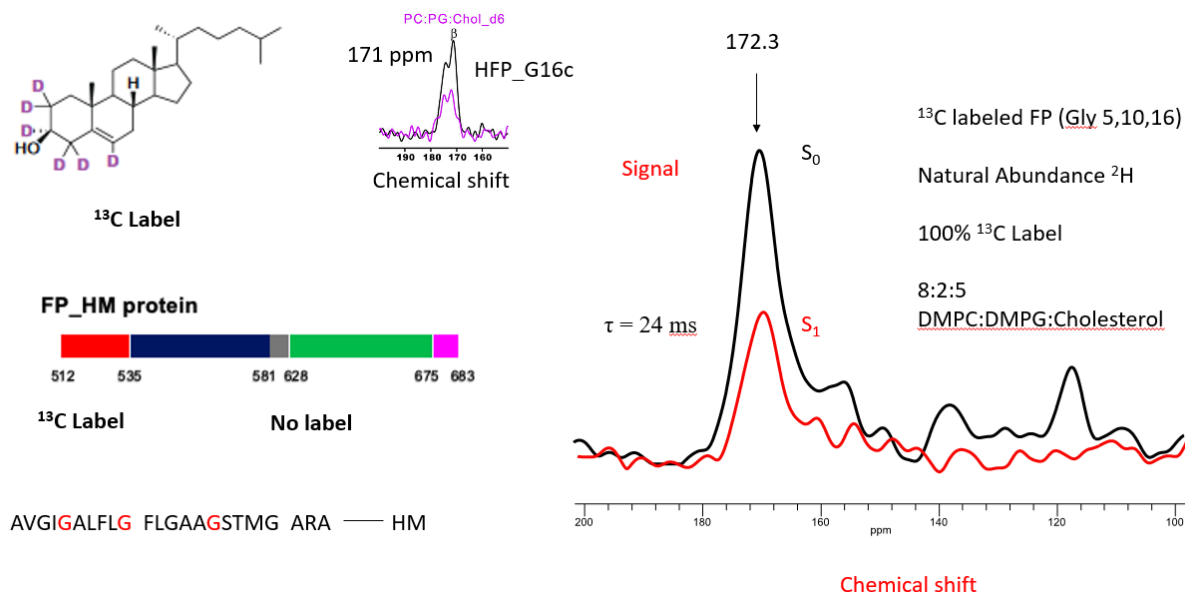
There are several lipids available for insertion experiments, examples of which are given in (**Figure 4.4**). For this work we have focused on Cholesterol d6 "chol\_d6". Cholesterol (Chol) is an important membrane component and represents ~0.25 mole fraction of the membrane of host cells of HIV and ~0.45 mole fraction of the HIV membrane (36-38). (**Figure 4.5**) displays NMR spectra for a sample containing FPHM with  $^{13}\text{C}$  Gly-5, 10,16 and membrane with lipid containing ~0.3 mole fraction "Chol\_d6". The  $^2\text{H}$ 's of Chol\_d6 are near the edge of the membrane. For FPHM, the most prominent spectral feature has  $\delta_{peak} \approx 172$  ppm which corresponds to Gly  $\beta$  sheet structure. The FPHM  $S_1$  dephasing supports insertion of either the G5, G10, or G16 residues within the membrane hydrocarbon core for a major fraction of FPHM molecules.



**Figure 4.4** Schematic diagram of potential lipids for probing protein membrane location.

On the right is a schematic diagram that gives an indication of where the deuterium labels of lipid are in the lipid vesicles.

Previous data (**Figure 4.5**) (38) shows that for either HFP\_G5c or HFP\_G16c, the most prominent spectral feature has  $\delta_{peak} \approx 171$  ppm which is also consistent with a Gly  $\beta$  sheet structure. HFP\_G16c also has a feature with  $\delta_{peak} \approx 174$  ppm which may correspond to HFP molecules with shorter  $\beta$  sheets that do not include G16.

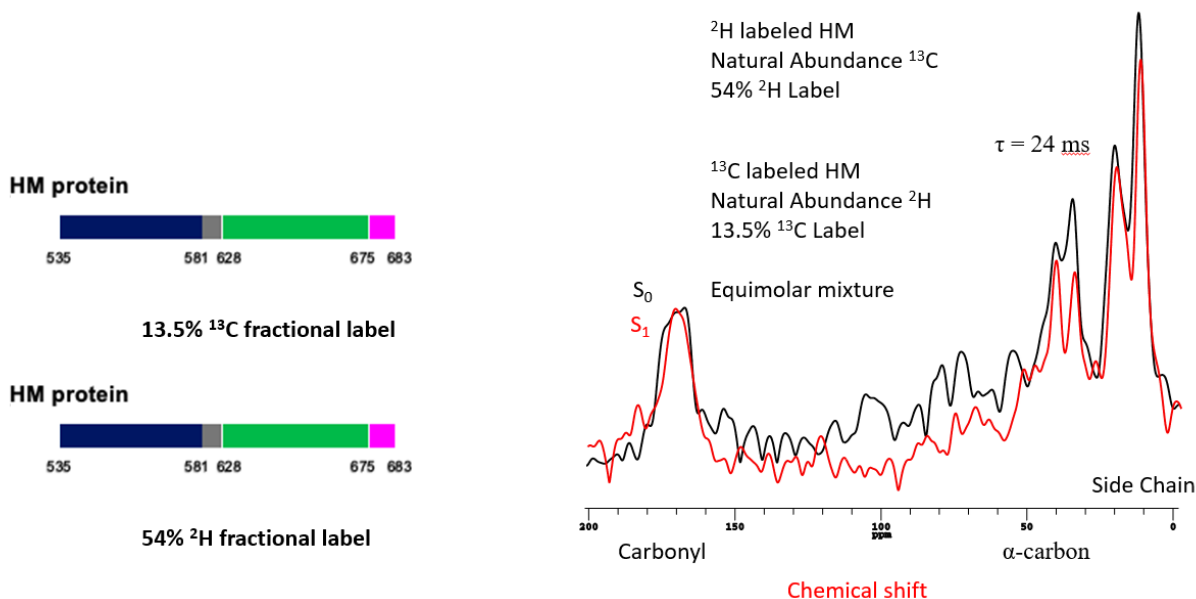


**Figure 4.5** (Top left) literature data for short FP peptide construct in cholesterol<sub>d6</sub>. (Bottom left) The protein construct in this work feature <sup>13</sup>C label at glycine 5, 10, and 16. The HM contain no label and has only natural abundance <sup>13</sup>C and <sup>2</sup>H. The center of the glycine peak is within 1 ppm of the literature value and similar dephasing is observed in the S<sub>1</sub> spectra. Protein: lipid molar ratio is approximately 1:100.

#### 4.3.5. Investigating the Equilibrium Process for the Formation of Monomer HM in Membranes

We were interested in determining if HM protein spontaneously forms monomers when inserted into lipid membranes. Previous SEC data has shown that gp41 monomers can exist in 50 mM sodium formate pH 3.2, 150 mM NaCl, and 0.2 mM TCEP (48). To investigate this, <sup>13</sup>C HM and <sup>2</sup>H HM are mixed in a state where we believe monomers are present and inserted into lipid vesicles. An equimolar mixture of <sup>13</sup>C labeled HM and <sup>2</sup>H HM is shown in (**Figure 4.6**). The process is detailed in (**Section 4.2.6**). The REDOR shows negligible dephasing in the carbonyl and side chain regions of the protein. It is possible that the small dephasing that we do see is due to dephasing in the <sup>2</sup>H labeled

HM. The  $^{13}\text{C}$  of the deuterium labeled protein only accounts for 7% of carbon so we would see little dephasing from the  $^2\text{H}$  HM. The lack of dephasing in the  $^{13}\text{C}$  carbonyl indicates that the  $^{13}\text{C}$  labeled HM and  $^2\text{H}$  HM are not in close contact with each other.



**Figure 4.6** REDOR spectra for  $^{13}\text{C}$  labeled HM and  $^2\text{H}$  labeled HM mixed. The two proteins mixed are the same except for the labeling (left). SSNMR spectra (right) has a protein: lipid molar ratio of 1:100. The lipid composition is 8:2:5 DMPC:DMPG:Chol.

#### 4.4. Discussion

This paper describes an approach to REDOR analysis of HM and FPHM constructs from bacterially expressed HM and FPHM produced from native chemical ligation. Key results include (1) REDOR NMR of  $^{13}\text{C}$  FP with  $^2\text{H}$  HM in a large FPHM construct; (2) probing of membrane insertion of FPHM protein with  $^2\text{H}$  labeled lipid vesicles; (3) Lack of formation of mixed SHB structure from a mixture of  $^2\text{H}$  HM and  $^{13}\text{C}$  HM when reconstituted in a lipid vesicle in formate buffer.

#### **4.4.1. REDOR NMR of $^{13}\text{C}$ FP with $^2\text{H}$ HM in a Large FPHM Construct**

There is not significant dephasing observed in the REDOR NMR spectra (**Figure 3.8C, Figure 3.12C**). This indicates that there is not close contact between the FP segment of the FPHM and any part of the HM segment. It would be interesting for future experiments to determine if FP encounters the TM (transmembrane) domain in FPHMTM.

#### **4.4.2. Probing of Membrane Insertion of FPHM Protein with $^2\text{H}$ Labeled Lipid**

##### **Vesicles**

$^{13}\text{C}$  labeled FPHM shows significant dephasing (**Figure 4.5**) when reconstituted into 8:2:5 DPPC:DPPG:Cholesterol vesicles. This methodology can be used for multiple applications including lipid insertion confirmation, and protein insertion depth into a membrane. The data observed in this study for FPHM is consistent with data seen for FP alone (38).

#### **4.5. Summary**

This study reports on the REDOR applications of FPHM and describes a new method of understanding the structural biology of gp41. These methods use site-specific labeling in combination with uniform labeling or labeled lipids to understand REDOR NMR in large protein constructs. Not only are these methods useful in understanding NMR of large proteins that have not been able to be analyzed by NMR previously, but these methods are also broadly applicable to many other membrane proteins.



## REFERENCES

1. Dawson, P.E.; Muir, T.W.; Clark-Lewis, I.; Kent, S.B. *Science* **1994**, 266, 776-779.
2. Dawson, P.E.; Kent, S.B. *Annu. Rev. Biochem.* **2000**, 69, 923-960.
3. Grewe, C., Beck, A., and Gelderblom, H. R. (1990) HIV: early virus-cell interactions, *J. AIDS* 3, 965-974
4. Johnson, Erik., Kent, Stephen. (2006) Insights into the mechanism and Catalysis of the Native Chemical Ligation Reaction, *J. AM. CHEM. SOC.*, 128, 6640-6646
5. Wang, Chen.; Guo, Qing-Xiang.; Fu, Yao. (2011) Theoretical Analysis of the Detailed Mechanism of Native Chemical Ligation Reactions, *Chem. Asian, J.*, 6, 1241-1251
6. Freed, E. O., Delwart, E. L., Buchschacher, G. L., Jr., and Panganiban, A. T. (1992) A mutation in the human immunodeficiency virus type 1 transmembrane glycoprotein gp41 dominantly interferes with fusion and infectivity. *Proc. Natl. Acad. Sci. U. S. A.* 89,70-74.
7. Herold, N., Anders-Osswein, M., Glass, B., Eckhardt, M., Muller, B., and Krausslich, H. G. (2014) 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, 13956-13970.
8. Wan, Qian., Chen, Jin., Yuan, Yu., Danishefsky. (2008) Oxo-ester Mediated Native Chemical Ligation: Concept and Applications, *J. AM. CHEM. SOC.*, 130, 15814-15816
9. Fang, Ge-Min., Cui, Hong-Kui., Zheng, Ji-Shen., Liu, Lei. (2010) Chemoselective Ligation of Peptide Phenyl Esters with N-Terminal Cysteines, *ChemBioChem*, 11, 1061-1065
10. Muir, Tom., Sondhi, Dolan., Philip, Cole. (1998) Expressed Protein Ligation: A General Method for Protein Engineering, *Proc. Natl. Acad. Sci. USA*, 95, 6705-6710
11. Kent, Stephen. (2009) Total Chemical Synthesis of Proteins, *Chem. Soc. Rev.*, 38, 338-351
12. Liu, Chuan-Fa., Rao, Chang., Tam, James P. (1996) Orthogonal Ligation of Unprotected Peptide Segments Through Pseudoproline Formation for the Synthesis of HIV-1 Protease Analogs, *J. Am. Chem. Soc.*, 118, 307-312
13. Xie, Jianming., Schultz, Peter G. (2005) Adding Amino Acids to the Genetic Repertoire, *Current Opinion in Chemical Biology.*, 9, 548-554
14. Hofmann, Roseanne M., Muir, Tom W. (2002) Recent Advances in the Application of Expressed Protein Ligation to Protein Engineering, *Current Opinion in Biotechnology.*, 13, 297-303

15. Boerema, David J., Tereshko, Velentina A., Kent, Stephen B. H. (2007) Total Synthesis by Modern Chemical Ligation Methods and High Resolution (1.1 Å) X-ray Structure of Ribonuclease A, *Peptide Science.*, 90, 278-286
16. Miller, Maria. (2009) The Early Years of Retroviral Protease Crystal Structures, *Peptide Science.*, 94, 521-529
17. Torbeev, Vladimir Yu., Kent, Stephen B. H. (2007) Convergent Chemical Synthesis and Crystal Structure of a 203 Amino Acid "Covalent Dimer" HIV-1 Protease Enzyme Molecule, *Angew. Chem. Int. Ed.*, 46, 1667-1670
18. Durek, Thomas., Torbeev, Vladimir Yu., Kent, Stephen B. H. (2007) Convergent Chemical Synthesis and High-Resolution X-Ray Structure of Human Lysozyme, *PNAS*, 104, 4846-4851
19. Lee, Ji Yeon., Bang, Duhee. (2009) Challenges in the Chemical Synthesis of Average Sized Proteins: Sequential vs. Convergent Ligation of Multiple Peptide Fragments, *Peptide Science.*, 94, 441-447
20. Hackeng, Tilman M., Griffin, John H., Dawson, Phillip E. (1999) Protein Synthesis by Native Chemical Ligation: Extended Scope by Using Straightforward Methodology, *Proc. Natl. Acad. Sci. USA*, 96, 10068-10073
21. Cabezas, Edelmira., Wang, Meng., Parren, Paul W. H. I., Stanfield, Robyn L., Satterthwait, Arnold C. (2000) A Structure-Based Approach to a Synthetic Vaccine for HIV-1, *Biochemistry.*, 39, 14377-14391
22. Evans, Thomas C., Benner, Jack., Xu, Ming-Qun. (1999) The in Vitro Ligation of Bacterially Expressed Proteins Using an Intein from *Methanobacterium Thermoautotrophicum*, *J. Biol. Chem.*, 274, 3923-3926
23. Decostaire, Isidore E., Lelièvre, Dominique., Aucagne, Vincent., Delmas, Agnès F. (2014) Solid Phase Oxime Ligations for the Iterative Synthesis of Polypeptide Conjugates, *Org. Biomol. Chem.*, 12, 5536-5543
24. Wan, Qian., Danishefsky, Samuel J. (2007) Free-Radical-Based, Specific Desulfurization of Cysteine: A Powerful Advance in the Synthesis of Polypeptides and Glycopolypeptides, *Angew. Chem. Int. Ed.*, 46, 9248-9252
25. Lakomek, Nils-Alexander., Kaufman, Joshua D., Stahl, Stephen J., Louis, John M., Grishaev, Alexander., Wingfield, Paul T., Bax, Ad. (2013) Internal Dynamics of the Homotrimeric HIV-1 Viral Coat Protein gp41 on Multiple Time Scales, *Angew. Chem. Int. Ed.*, 52, 3911-3915
26. Rose, Keith. (1994) Facile Synthesis of Homogeneous Artificial Proteins, *J. Am. Chem. Soc.*, 116, 30-33

27. Shao, Jun., Tam, James P. (1995) Unprotected Peptides as Building Blocks for the Synthesis of peptide Dendrimers with Oxime, Hydrazone, and Thiazolidine Linkages, *J. Am. Chem. Soc.*, 117, 3893-3899
28. Low, Donald W., Hill, Michael G., Carrasco, Michael R., Kent, Stephen B.H., Botti, Paolo. (2001) Total Synthesis of Cytochrome b562 by Native Chemical Ligation Using a Removable Auxiliary, *PNAS*, 98, 6554-6559
29. Yan, Liang Z., Dawson, Phillip E. (2001) Synthesis of Peptides and Proteins Without Cysteine Residues by Native Chemical Ligation Combined with Desulfurization, *J. Am. Chem. Soc.*, 123, 526-533
30. Kent, Stephen. (2003) Total Chemical Synthesis of Enzymes, *J. Peptide Sci.*, 9, 574-593
31. Schnölzer, Martina., Kent, Stephen B. H. (1992) Constructing Proteins by C-Terminal Tailing of Unprotected Synthetic Peptides: Backbone-Engineered HIV Protease, *Science*, 256, 221-225
32. Cocchi, Fiorenza., DeVico, Anthony L., Garzino-Demo, Alfredo., Cara, Andrea., Gallo, Robert C., Lusso, Paolo. (1996) The V3 Domain of the HIV-1 gp120 Envelope Glycoprotein is Critical for Chemokine-Mediated Blockade of Infection, *Nature Medicine*, 2, 1244-1247
33. Thapa, Parashar., Zhang, Rui-Yang., Menon, Vinay., Bingham, Jon-Paul. (2014) Native Chemical Ligation: A Boon to Peptide Chemistry, *Molecules*, 19, 14461-14483
34. Torbeev, Vladimir Yu., Raghuraman, H., Hamelberg, Donald., Tonelli, Marco., Westler, William M., Perozo, Eduardo., Kent, Stephen B. H. (2011) Protein Conformational Dynamics in the Mechanism of HIV-1 Protease Catalysis, *PNAS*, 108, 20982-20987
35. Sackett, Kelly., TerBush, Allan., Weliky, David P. (2011) HIV gp41 Bundle Constructs Induce Rapid Vesicle Fusion at pH 3.5 and Little Fusion at pH 7.0: Understanding pH Dependence of Protein Aggregation, Membrane Binding, and Electrostatics, and Implications for HIV-host Cell Fusion, *Eur. Biophys. J.*, 40, 489-502
36. Gabrys, Charles M., Qiang, Wei., Sun, Yan., Xie, Li., Schmick, Scott D., Weliky, David P. (2013) Solid-State Nuclear Magnetic Resonance Measurements of HIV Fusion Peptide <sup>13</sup>CO to Lipid <sup>31</sup>P Proximities Support Similar Partially Inserted Membrane Locations of the  $\alpha$  Helical and  $\beta$  Sheet Peptide Structures, *J. Phys. Chem. A*, 117, 9848-9859
37. Xie, Li., Jia, Lihui., Liang, Shuang., Weliky, David P. (2015) Multiple Locations of Peptides in the Hydrocarbon Core of Gel-Phase membranes Revealed by Peptide <sup>13</sup>C to Lipid <sup>2</sup>H Rotational-Echo Double-Resonance Solid-State Nuclear Magnetic Resonance, *Biochemistry*, 54, 677-684

38. Jia, Lihui., Liang, Shuang., Sackett, Kelly., Xie, Li., Ghosh, Ujjayini., Weliky, David P. (2015) REDOR Solid-State NMR as a Probe of the Membrane Locations of Membrane-Associated Peptides and Proteins, *Journal of Magnetic Resonance.*, 253, 154-165
39. Hu, Jian., Qin, Huajun., Sharma, Mukesh., Cross, Timothy A., Gao, Fei Philip. (2008) Chemical Cleavage of Fusion Peptides for High-Level Production of Transmembrane Peptides and Protein Domains Containing Conserved Methionines, *Biochemica et Biophysica Acta.*, 1778, 1060-1066
40. Marley, Jonathan., Lu, Min., Bracken, Clay. (2001) A Method for Efficient Isotopic Labeling of Recombinant Proteins, *Journal of Biomolecular NMR.*, 20, 71-75
41. Gulion, Terry. (2008) Rotational-Echo, Double-Resonance NMR, *Modern Magnetic Resonance.*, 713-718
42. Gulion, Terry. (1998) Introduction to Rotational-Echo, Double-Resonance NMR, *Concepts in Magnetic Resonance.*, 10, 277-289
43. Dalgleish, Angus G., Beverley, Peter C. L., Clapham, Paul R., Crawford, Dorothy H., Greaves, Melvyn F., Weiss, Robin A. (1984) The CD4 (T4) Antigen is an Essential Component of the Receptor for the AIDS Retrovirus, *Nature.*, 312, 763-767
44. Chan, David C., Fass, Deborah., Berger, James M., Kim, Peter S. (1997) Core Structure of gp41 from the HIV Envelope Glycoprotein, *Cell.*, 89, 263-273
45. Aeffer, Sebastian., Reusch, Tobias., Weinhausen, Britta., Salditt, Tim. (2011) Energetics of Stalk Intermediates in Membrane Fusion are Controlled by Lipid Composition, *PNAS.*, E1609-E1618
46. Sackett, Kelly., Nethercott, Matthew J., Zheng, Zhaoxiong., Weliky, David P. (2014) Solid-State NMR Spectroscopy of the HIV gp41 Membrane Fusion Protein Supports Intermolecular Antiparallel  $\beta$  Sheet Fusion Peptide Structure in the Final Six-Helix Bundle State, *J. Mol. Biol.* 426, 1077-1094
47. Lakomek, Nils-Alexander., Kaufman, Joshua D., Stahl, Stephen J., Louis, John M., Grishaev, Alexander., Wingfield, Paul T., Bax, Ad. (2013) Internal Dynamics of the Homotrimeric HIV-1 Viral Coat Protein gp41 on Multiple Time Scales, *Angew. Chem. Int. Ed.*, 52, 3911-3915
48. Banerjee, Koyeli., Weliky, David P. (2014) 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.*, 53, 7184-7198

49. White, Judith M. Delos, Sue E., Brecher, Matthew., Schornberg, Kathryn. (2008) Structures and Mechanisms of Viral Membrane Fusion Proteins: Multiple Variations on a Common Theme, *Crit. Rev. Biochem. Mol. Biol.*, 43, 189-219
50. Melikyan, Gregory B. (2014) HIV Entry: a Game of Hide-and-Fuse?, *Current Opinion in Virology.*, 4, 1-7
51. Montero, Marinieve., Van Houten, Nienke E., Wang, Xin., Scott, Jamie K. (2008) The Membrane-Proximal External Region of the Human Immunodeficiency Virus Type 1 Envelope: Dominant Site of Antibody Neutralization and Target for Vaccine Design, *Microbiology and Molecular Biology Reviews.* 72, 54-84
52. Apellániz, Beatriz., Huarte, Nerea., Largo, Eneko., Nieva José L. (2014) The Three Lives of Viral Fusion Peptides, *Chemistry and Physics of Lipids.*, 181, 40-55
53. Liu, Jun., Bartesaghi, Alberto., Borgnia, Mario J., Sapiro, Guillermo., Subramaniam, Sriram. (2008) Molecular Architecture of Native HIV-1 gp120 Trimers, *Nature.*, 455, 109-114
54. Schülke, Norbert., Vesanen, Mika S., Sanders, Rogier W., Zhu, Ping., Lu, Min., Anselma, Deborah J., Villa, Anthony R., Parren, Paul W. H. I., Binley, James M., Roux, Kenneth H., Maddon, Paul J., Moore, John P., Olson, William C. (2002) Oligomeric and Conformational Properties of a Proteolytically Mature Disulfide-Stabilized Human Immunodeficiency Virus Type 1 gp140 Envelope Glycoprotein, *Journal of Virology.*, 76, 7760-7776
55. Sanders, Rogier W., Vesanen, Mika., Schuelke, Norbert., Master, Aditi., Schiffner, Linnea., Kalyanaraman, Roopa., Paluch, Maciej., Berkhout, Ben., Maddon, Paul J., Olson, William C., Lu, Min., Moore, John P. (2002) Stabilization of the Soluble, Cleaved Trimeric Form of the Envelope Glycoprotein Complex of Human Immunodeficiency Virus Type 1, *Journal of Virology.*, 76, 8875-8889
56. Bartesaghi, Alberto., Merk, Alan., Borgnia, Mario J., Milne, Jacqueline L. S., Subramaniam, Sriram. (2013) Pre-Fusion Structure of Trimeric HIV-1 Envelope Glycoprotein Determined by Cryo-Electron Microscopy, *Nat. Struct. Mol. Biol.*, 20, 1352-1357
57. Julien, Jean-Philippe et al. (2013) Crystal Structure of a Soluble Cleaved HIV-1 Envelope Trimer, *Science.*, 342, 1477-1483
58. Buzon, Victor., Natrajan, Ganesh., Schibli, David., Campelo, Felix., Kozlov, Michael M., Weissenhorn, Winfried. (2010) Crystal Structure of HIV-1 gp41 Including Both Fusion Peptide and Membrane Proximal External Regions, *PLoS Pathogens.*, 6, 1-7

59. Purushottam, L., Adusumalli, S., Singh, U., Unnikrishnan, V. B., Rawale, D., Gujrati, M., Mishra, R., Rai, V. Single-site glycine-specific Labeling of Proteins. *Nature Communications* volume 10, Article number: 2539 (2019)
60. Liang, S., Ratnayake, U., Keinath, C., Jia, L., Wolfe, R., Ranaweera, Weliky, D. P. Efficient Fusion at Neutral pH by Human Immunodeficiency Virus gp41 Trimers Containing the Fusion Peptide and Transmembrane Domains. *Biochemistry* 2018, 57, 1219-1235.

**CHAPTER 5:**  
**EXPRESSION, PURIFICATION, SOLUBILIZATION, AND CHARACTERIZATION OF**  
**SARS-COV-2 PROTEIN CONSTRUCTS PRODUCED FROM *E.COLI***

## 5.1. Introduction

The global pandemic caused by the novel coronavirus (2019-nCoV) has created a public health emergency resulting in almost 600k deaths in the United States, and over 3M deaths worldwide (1,2). The S2 subunit of the of the CoV spike (S) glycoprotein is a key protein facilitating membrane fusion and thus infection by 2019-nCoV (3-6). The S2 subunit is an attractive target for study of therapeutics that may inhibit the process of 2019-nCoV infection. To investigate the structure and function of the S2 subunit more in depth we have produced and purified several protein constructs by recombinant means from *E. coli*. Constructs were purified and characterized by a combination of analytical methods including SDS-PAGE, western blot, and proteomics. This study also examines CD spectra of folded CoV spike protein and lipid vesicle fusion assays of protein constructs with and without the fusion peptide.

SARS-CoV-2 (SARS2) is a zoonotic virus that acts as the pathogen of the COVID-19 pandemic. SARS2 is enveloped by a membrane that is obtained during viral budding from an infected host cell (3-6). Infection of a new cell requires fusion of the virus membrane with the target cell membrane and subsequent deposition of the viral nucleocapsid in the cytoplasm (12,39). This process is catalyzed by the Spike (S) protein subunit 2 (S2). There is a single gene that codes for the S protein subunit 1 (S1) and S2 with respective residues 1-685 and 686-1273 (5). The S1 subunit and the large S2 ectodomain (Ed, residues 687-1207) are outside the virus, followed by the viral transmembrane domain (TM, 1208-1234) and then endodomain (1235-1273) inside the virus (5). Target cells are identified by binding between S1 and the extracellular domain of angiotensin-converting enzyme 2 (ACE2) (7). If there has also been proteolytic cleavage between S1 and S2, the S1 separates from the S2. This separation is followed



by a large structural change of the Ed to form a final hairpin structure (5,6). For initial infection of respiratory epithelial cells by SARS2, cellular proteases may do S1/S2 cleavage so that fusion occurs with the plasma membrane (11-16). For systemic infection of cells in other tissues, there may be endocytosis after S1/ACE2 binding followed by endosome maturation with reduction of pH < 6, activation of cathepsin L proteases at the low pH, S1/S2 cleavage, and then fusion with the endosome membrane (11-16). Both fusion pathways result in deposition of the nucleocapsid in the cytoplasm (11-16). The initial trimeric S state and final hairpin structure of the S2 protein are the basis for identifying S2 as a “class I” fusion protein. There are ~225 residues N-terminal of the hairpin structure and at least five distinct ~20-residue segments have been proposed as a FP (3-5). See (**Section 1.2**) for further details on the structure of S2. The epitopes of many neutralizing antibodies may be in the S2 region that is N-terminal of the hairpin, as this was observed for antibodies from convalescent patients of the 2002-2004 SARS epidemic. The S2 sequence of SARS2 has ~90% sequence identity with S2 of the SARS1 viral pathogen of the earlier epidemic (60,63-64).

## **5.2. Materials and Methods**

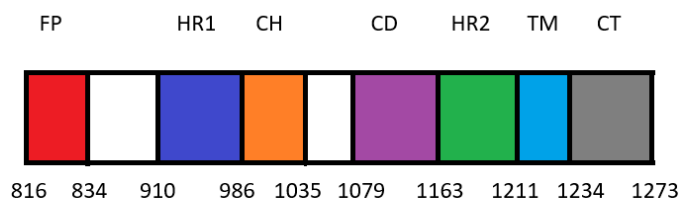
Materials were purchased from the following companies: DNA – GenScript (Picataway, NJ); Escherichia coli BL21(DE3)pLysS strain – competent cells grown in house; Luria-Bertani (LB) medium- Dot Scientific (Burton, MI); isopropyl  $\beta$ -D-thiogalactopyranoside (IPTG) and tris-(carboxyethyl) phosphine (TCEP) – GoldBio (St. Louis, MO); Co<sup>2+</sup>-resin and SnakeSkin dialysis tubing– Thermo Scientific (Waltham, MA); Most other materials were obtained from Sigma-Aldrich (St. Louis, MO).

### 5.2.1. Making Calcium Competent Cells

See (Section 2.3.3.)

### 5.2.2. Molecular Biology, Protein Expression, Separation

The S2 amino acid sequence is shown in (Figures 5.1-5.4) and is based on the 2019-nCoV strain (3). One S2 sequence begins at S2' and is designated as S2\_816-1273 (Figure 5.1) and some of the S2 sequence used in this paper is S2\_903-998-SGGRGG-1153-1207 (Figure 5.3) or S2\_903-998-SGGRGG-1163-1207 (Figure 5.4). We replaced residues 999-1152 with the flexible linker SGGRGG. In addition, all S2 constructs have the C-terminal purification tag G6LEH6.

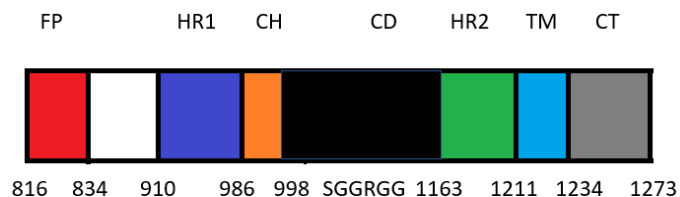


```

AS S816FIEDLLFNK VTLADAGFI834K QYGDSLGDIA ARDLISAQKF NGLTVLPPLL TDEMIAQYTS
ALLAGTITSG WTFGAGAAALQ IPFAMQMAYR FNGIG910VTQNV LYENQKLIAN QFNSAIGKIQ DSLSSTASAL
GKLQDVVNQN AQALNTLVKQ LSSNFGAISS VLNDILSRDL K986VEAEVQIDR LITGRLQSLQ TYVTQQLIRA
AEIRASANLA ATKMSSEVLG1035 QSKRVDFSGK GYHLMSFPQS APHGVVFLHV TYVPAQEKNF
TTAP1079AISHDG KAHFDREGVF VSNGTHWFVT QRNFYEPQII TTDNTFVSGN SDVVIGIVNN TVYDPLQPEL
DSFKEELDKY FKNHTSPD1163VD LGDISGINAS VVNIQKEIDR LNEVAKNLNE SLIDLQELGK
YEQYIK1211WPWY IWLGFIAGLI AIVMTIML1234S SMTSSSSSLK GSSSSGSSSK FDEDDSEPVL
KGVKLHYT1273GG GGGGLEHHHHHH

```

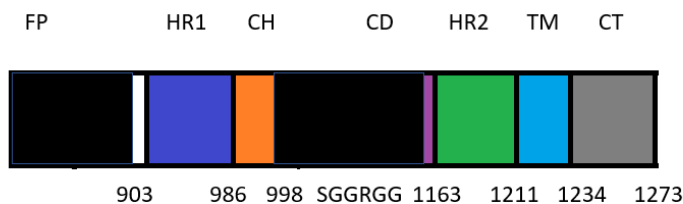
**Figure 5.1** Schematic diagram and amino acid sequence for S2\_816-1273. The construct contains both the fusion peptide and the transmembrane domain.



AS S<sub>816</sub>FIEDLLFNK VTLADAGFI<sub>834</sub>K QYGDSLGDIA ARDLISAKKF NGLTVLPPLL TDEMIAQYTS  
 ALLAGTITSG WTFGAGAALQ IPFAMQMA<sub>903</sub>YR FNGIG<sub>910</sub>VTQNV LYENQKLIAN QFNSAIGKIQ DSLSSTASAL  
 GKLQDVVNQN AQUALNTLVKQ LSSNFGAISS VLNDILSRD K<sub>986</sub>VEAEVQIDR LIT<sub>998</sub> SGRGG D<sub>1163</sub>VD  
 LGDISGINAS VVNIQKEIDR LNEVAKNLNE SLIDLQELGK YE<sub>1207</sub>GG GGGGLEHHHHHH

**Figure 5.2** Schematic diagram and amino acid sequence for S2\_816-

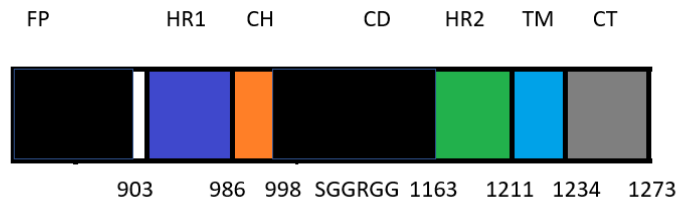
998SGGRGG1163-1207. This construct features the FP but lacks the transmembrane domain and features a non-native SGRGG loop that replaces residues 999-1152.



AS A<sub>903</sub>YR FNGIG<sub>910</sub>VTQNV LYENQKLIAN QFNSAIGKIQ DSLSSTASAL GKLQDVVNQN AQUALNTLVKQ  
 LSSNFGAISS VLNDILSRD K<sub>986</sub>VEAEVQIDR LIT<sub>998</sub> SGRGG D<sub>1153</sub>KY FKNHTSPDD<sub>1163</sub>VD LGDISGINAS  
 VVNIQKEIDR LNEVAKNLNE SLIDLQELGK YE<sub>1207</sub>GG GGGGLEHHHHHH

**Figure 5.3** Schematic diagram and amino acid sequence for S2\_903-

998SGGRGG1153-1207. This construct lacks the FP and transmembrane domain and features a non-native SGRGG loop that replaces residues 999-1152.



AS A<sub>903</sub>YR FNGIG<sub>910</sub>VTQNV LYENQKLIAN QFNSAIGKIQ DSLSTASAL GKLQDVVNQN AQA<sub>1273</sub>N<sub>1273</sub>TLVKQ  
LSSNFGA<sub>903</sub>ISS VLNDILSR<sub>986</sub>LD K<sub>986</sub>VEAEVQIDR LIT<sub>998</sub> SGGRGG D<sub>1163</sub>VD LGDISGINAS VVNIQKEIDR  
LNEVAKNLNE SLIDLQELGK YE<sub>1207</sub>GG GGGGLEHHHHHH

**Figure 5.4** Schematic diagram and amino acid sequence for S2\_903-998SGGRGG1163-1207. This construct lacks the FP and transmembrane domain and features a non-native SGGRGG loop that replaces residues 999-1162.

Several related proteins were produced by expression in *E. coli* bacteria, BL21(DE3) pLysS strain, using vectors with inserts. Proteins included:

S2\_816-1273 (MW = 51717),

S2\_903-998SGGRGG1163-1207 (MW = 17413),

S2\_903-998SGGRGG1153-1207 (MW = 18516), and

S2\_816-998SGGRGG1153-1207 (MW = 28020)

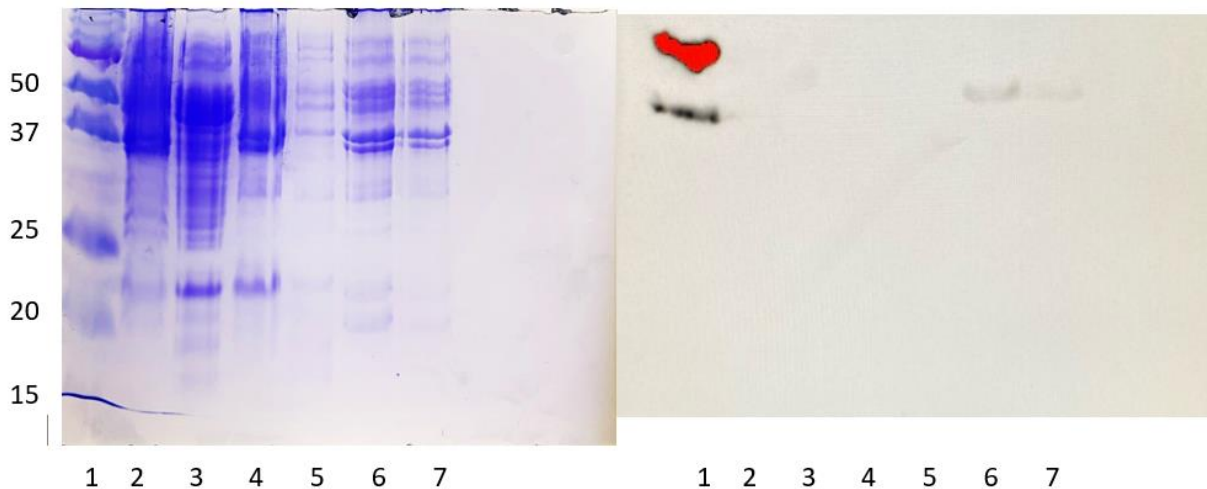
The DNA coding for each protein construct was sub-cloned into the pET-24a(+) vector by Genscript company. Transformation into BL21(DE3)pLysS began with addition of 5 ng of the plasmid provided by Genscript for S2 to a 50  $\mu$ L suspension of *E. coli* competent-cells, incubation on ice for 30 min, heat shock in a 42 °C bath without shaking for 50 s, and then ice for 2 min. LB medium (450  $\mu$ L) was added followed by incubation at 37 °C for 1 h. The cell suspension (20-200  $\mu$ L) was added to kanamycin/chloramphenicol-resistant selective plates at 37 °C, followed by incubation overnight. A single colony was selected from the plate and added to a flask containing 25 mL LB medium augmented

with 50 mg/mL kanamycin and 50 mg/mL chloramphenicol. Cells grew to an OD<sub>600</sub> of approximately 0.8. The plasmid DNA was isolated and purified from an aliquot of the cell suspension with a Wizard Plus Minipreps kit (Promega – Madison, WI), with subsequent sequencing that confirmed the insert. The remaining suspension was divided into 1 mL aliquots with subsequent addition of 0.6 mL 50% glycerol, and stored at -80 °C.

Protein expression began with the addition of 1.5 mL E. coli glycerol stock to 1L LB medium followed by growth for 3 h. This and all other growths were done at 37 °C in shake flasks at 180 rpm, and with 1 mL of 50 mg/mL kanamycin and 1 mL of 50 mg/mL chloramphenicol added to the growth medium. (1) Growth phase was approximately 8 h to OD<sub>600</sub> of approximately 0.8; (2) protein expression was induced by 2 mM IPTG and occurred overnight at 19 °C; and (3) the cell pellet was harvested after centrifugation at 9000g for 30 min. TB medium was later used to grow the E. coli cells. 1 L of TB contained 24 g yeast, 12 g tryptone, 4 mL glycerol, and 900 mL DI H<sub>2</sub>O. 100 mL pH adjusting buffer was also prepared with 2.31 and 12.54 g monobasic and dibasic sodium phosphate, respectively. The ~pH 7.0 buffering solution was added to the 1 L TB at the time of beginning cell growth.

Separation of desired protein material began with: (1) tip-sonication of cell pellet in ~30 mL PBS at pH 7.4 in a 50 mL beaker on ice; (2) centrifugation at 48000g for 20 min and harvesting of the new pellet. The supernatant was kept and analyzed with binding of soluble proteins to cobalt affinity resin; and (3) 1x repetition of the sonication/centrifugation/harvesting steps with resultant pellet 1 and retained washes. The next step was tip sonication of the pellet 1 in ~30 mL PBS at pH 7.4 with 6 M GuCl, Followed by centrifugation. Much of pellet 1 was solubilized, as evidenced by a new pellet:

pellet 1 volume ratio  $<1/2$ . The supernatant was then subjected to  $\text{Co}^{2+}$  affinity chromatography. The protein was analyzed following affinity chromatography (**Figures 5.5-5.7**). Based on this solubility test the general solubilities of the proteins can be determined.



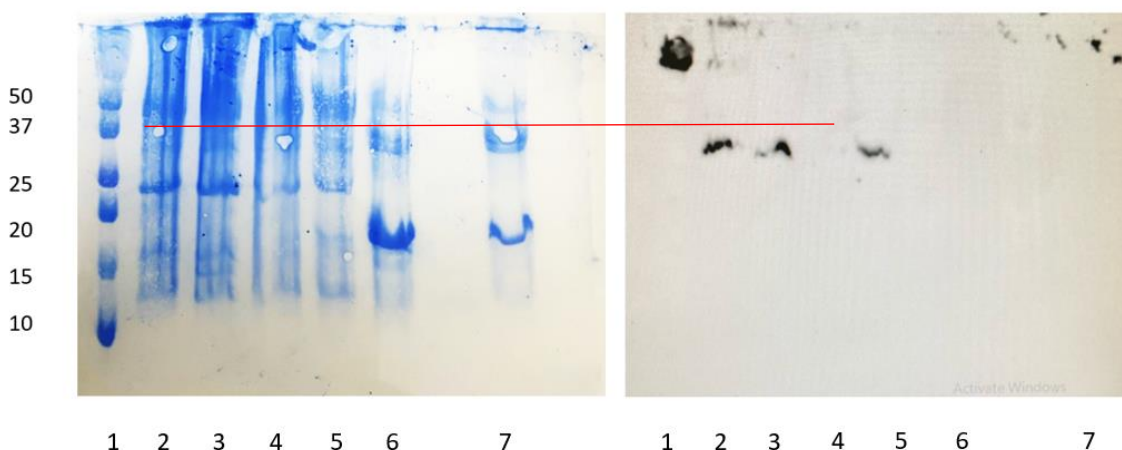
Lane 1: MW standards  
 Lane 2: PBS wash 1, retained by  $\text{Co}^{2+}$  resin  
 Lane 3: PBS wash 1, flow through from  $\text{Co}^{2+}$  resin  
 Lane 4: PBS wash 2, retained by  $\text{Co}^{2+}$  resin  
 Lane 5: PBS wash 2, flow through from  $\text{Co}^{2+}$  resin  
 Lane 6: 6M GuCl retained by  $\text{Co}^{2+}$  resin  
 Lane 7: 6M GuCl flow through from  $\text{Co}^{2+}$  resin

Western blot of SDS-PAGE on left. S2\_816-1273 pET-24a(+) appears to be insoluble in PBS and soluble in GuCl and is retained by affinity resin.

**Figure 5.5** Lysate of S2\_816-1273. Each PBS wash is done with 30 mL PBS buffer or 30 mL 6 M GuCl buffer. PBS wash contains only proteins that are soluble in PBS. The

**Figure 5.5 (cont'd)**

target protein bands in the western blot correspond to approximate molecular weight (51.7kDa) of monomer S2\_816-1273. Proteins “retained by”  $\text{Co}^{2+}$  resin are eluted by addition of 250mM imidazole in PBS or 6M GuCl respectively. The S2\_816-1273 is present in both lanes 6 & 7 indicating the poor binding to affinity resin.



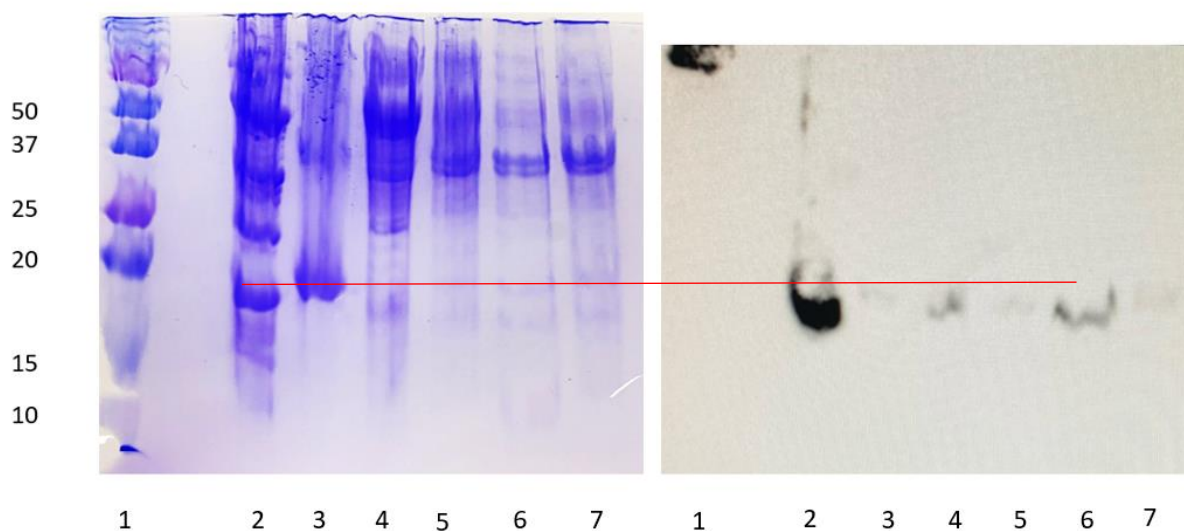
Lane 1: MW standards  
Lane 2: PBS wash 1, retained by  $\text{Co}^{2+}$  resin  
Lane 3: PBS wash 1, flow through from  $\text{Co}^{2+}$  resin  
Lane 4: PBS wash 2, retained by  $\text{Co}^{2+}$  resin  
Lane 5: PBS wash 2, flow through from  $\text{Co}^{2+}$  resin  
Lane 6: 6M GuCl retained by  $\text{Co}^{2+}$  resin  
Lane 7: 6M GuCl flow through from  $\text{Co}^{2+}$  resin

Western blot of SDS-PAGE on left. Soluble fractions appear to be monomers in PBS according to approximate molecular weight (28kDa).

**Figure 5.6** SDS-PAGE of lysate transfer to nitrocellulose paper and accompanying western blot of S2\_816-998SGGRGG1153-1207 (MW = 28020). The wash contains proteins soluble in PBS buffer. Western blot shows monomer fraction is soluble in PBS

### Figure 5.6 (cont'd)

by presence of His-tagged protein in the PBS soluble lanes (wash 1 and wash 2). Higher order oligomers appear to be soluble only in GuCl as there is some signal in the western blot of protein with higher than 50 kDa for the GuCl soluble fractions. These possible oligomers have not however been purified and confirmed. The western blot shows protein signal in 2,3, and 5.



Lane 1: MW standards  
Lane 2: PBS wash 1, retained by  $\text{Co}^{2+}$  resin  
Lane 3: PBS wash 1, flow through from  $\text{Co}^{2+}$  resin  
Lane 4: PBS wash 2, retained by  $\text{Co}^{2+}$  resin  
Lane 5: PBS wash 2, flow through from  $\text{Co}^{2+}$  resin  
Lane 6: 6M GuCl retained by  $\text{Co}^{2+}$  resin  
Lane 7: 6M GuCl flow through from  $\text{Co}^{2+}$  resin  
Lane 8 GuCl non-binding

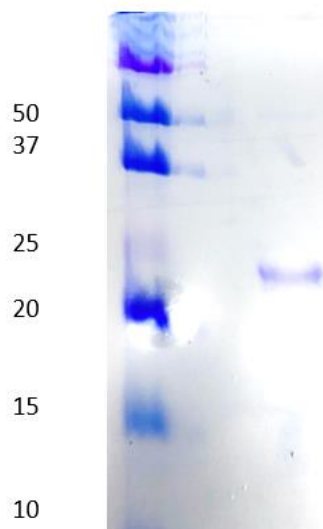
Western blot of SDS-PAGE on left.  
Largest fraction of S2\_903-998SGGRGG1163-1207 appears to be soluble in PBS

**Figure 5.7** Lysate of S2\_903-998SGGRGG1163-1207. Each PBS wash is done with 30 mL PBS buffer or 30 mL 6 M GnCl buffer. The largest portion of target protein appears to be soluble as evidenced by the western blot at approximately (17.4 kDa).



The larger protein construct S2\_816-1273, containing the FP segment was much less soluble in PBS (**Figure 5.5**) and so separation of desired protein material for S2\_816-1273 began with: (1) tip-sonication in ~50 mL 6M GuHCl at pH 7.4 in a 200 mL beaker in an ice bath; (2) centrifugation at 48000g for 30 min and harvesting the new pellet. (3) The new pellet is sonicated once more in 6M GuCl and separated from the insoluble pellet by centrifugation. (4) The total supernatant liquid is dialyzed against H<sub>2</sub>O twice. The desired protein precipitates and soluble proteins are removed by centrifugation and discarding the supernatant solution which show a yellow color. (5) That pellet of protein precipitate is then solubilized in 8M urea. (6) The resulting supernatant contains mostly large S2 protein construct but can be further purified by His-tag affinity chromatography and size exclusion chromatography.

Protein constructs S2\_903-998SGGRGG1163-1207 (**Figure 5.7**), and S2\_903-998SGGRGG1153-1207 contained no FP and were soluble in PBS. Separation of protein material began with: (1) tip-sonication in ~100 mL 0.05% CHAPS in PBS at pH 7.4 in a 200 mL beaker in an ice bath; (2) centrifugation at 48000g for 30 min and harvesting the new pellet. (3) The new pellet was then sonicated once more in 0.05% CHAPS and separated from the insoluble pellet by centrifugation. (4) The total supernatant liquid was then mixed with affinity resin and purified by affinity chromatography. (5) SEC can then be used following affinity chromatography to further purify protein. (**Figure 5.8**)



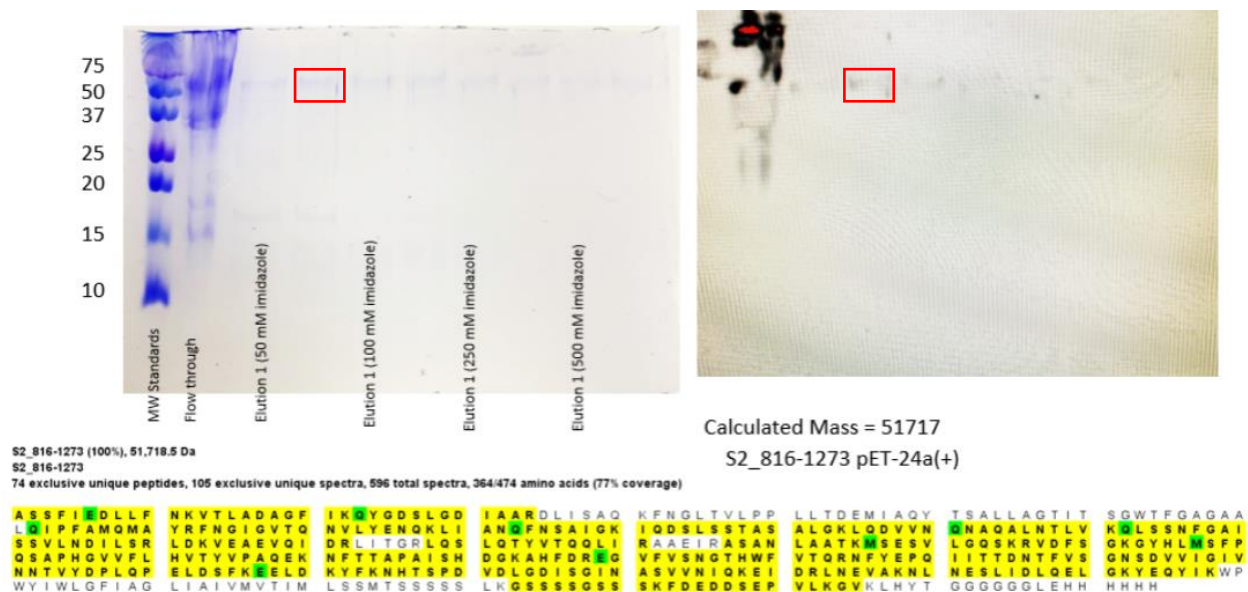
**Figure 5.8** SDS-PAGE of S2\_903-998SGGRGG1153-1207 following purification by affinity chromatography and FPLC. FPLC fractions were confirmed by Bradford assay. A single band is seen in the SDS-PAGE at around 22kDa gel shifted from the calculated mass of S2\_903-998SGGRGG1153-1207 (18.5kDa). The first lane is the molecular weight standards, and the second lane is the protein.

### 5.2.3. Characterization of Proteins by Proteomics Data and Western Blot

A cassette was prepared with a gel from SDS-PAGE, nitrocellulose membrane, foam inserts, and casing. Buffer (100 mL) was prepared with 250 mM Tris at pH 8.3, 1.94 M glycine, and 1% SDS (w/v) and then diluted with 700 mL H<sub>2</sub>O and 200 mL MeOH. The cassette was placed in a tub with this solution, with cooling achieved by placing a plastic bottle filled with ice in the solution and placing the tub in a larger ice bath. Protein was transferred to the nitrocellulose membrane by applying 100 V for 1 h. The nitrocellulose membrane was shaken for 1 h in a solution made from mixing: (1) 200 mM Tris buffer at pH 8.0 with 1.37 M NaCl, 2.0 mL; (2) H<sub>2</sub>O, 18 mL; (3) 50% Tween-20, 0.04 mL; and (4) nonfat dried milk, 1.0 g. The membrane was then washed

with the (1) + (2) + (3) solution without milk. The membrane was shaken for 1 h in 10 mL of the latter solution with 10  $\mu$ L anti-H5-HRP-conjugate antibody, and then washed for 5 min with the solution without antibody (4x). The membrane was then immersed in 5 mL Clarity ECL chemiluminescence substrate and analyzed by digital imaging software.

Proteolysis followed by mass spectrometric detection of peptides expected from S2 sequences were used to confirm the identity of select bands cut from SDS-PAGE gels. (**Figure 5.9**)



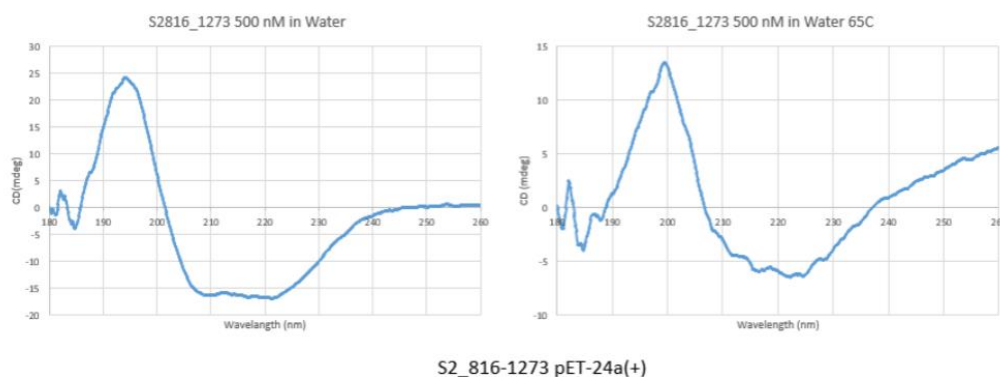
**Figure 5.9** Affinity purification of S2\_816-1273. The affinity purified protein shows only the monomer S2\_816-1273. Mass spectroscopy data following trypsin digestion show a 77% sequence coverage of the S2\_816-1273 confirming the presence of monomer (51.7 kDa).

### **Figure 5.9 (cont'd)**

Proteomics data was obtained from the same protein sample that produced this gel but ran on a different gel. The elution is of increasing imidazole concentration and approximately 3 mL of buffer is added per band. The lanes with labels are at the start of the new 3 mL addition of imidazole and the lanes without labels are at the end of that same 3 mL.

#### **5.2.4. Circular Dichroism and Lipid Vesicle Assay**

Circular dichroism (CD) spectra were collected to characterize protein folding. Samples were prepared by pipetting 1 mL of protein in buffer. Typical concentrations of protein were in the range of 10 $\mu$ M to 500nM. Each spectra taken was the (protein +buffer) – buffer difference. Protein was prepared from solubilization in 8M urea in the case of S2\_816-1273 and refolded by slow dialysis against pure water twice. The dialysis was performed at a ratio of 20mL solubilized protein (~0.6-0.8 mg/mL) to 2L water. CD experiments in other solutions were performed using an aliquot of the protein in water diluted into different solutions. S2\_903-998SGGRGG1163-1207, and S2\_903-998SGGRGG1153-1207 were solubilized in CHAPS buffer and dialyzed against water in an analogous method. Snakeskin dialysis tubing from Thermo-Fisher with 10kDa MW cutoff was used in all cases. CD spectra at ambient temperature was recorded in a quartz cuvette with a 190-260 nm wavelength range scanned in 0.5 nm steps. For some samples, a temperature series of CD spectra were acquired over a 25-65 °C range. There was no visible precipitation at any temperature measured, or after cooling. There was some change in the CD spectra of S2\_816-1273 as the temperature was increased (**Figure 5.10**).



**Figure 5.10** CD spectra of S2\_816-1273 taken at ambient temperature on the left and 65 °C on the right. The protein appears to be less thermally stable than other S2 constructs, as well as other similar enveloped membrane proteins from HIV. Molar ellipticity at 222 nm is related to mdeg at 222 nm by  $\theta = m^0M/LC$  where  $\theta$  = molar ellipticity,  $m^0$  is circular dichroism in mdeg,  $M$  = molar mass,  $L$  = path length of the cell, and  $C$  = concentration. For this experiment  $M = 51717$  g/mol,  $L = 10$  mm,  $C = .51717$  g/mL. Molar ellipticity =  $15000 \text{ deg}(\text{cm}^2)/\text{dmol}$  (left) and  $5300 \text{ deg}(\text{cm}^2)/\text{dmol}$  (right) compared to total helical value of  $33000 \text{ deg}(\text{cm}^2)/\text{dmol}$ .

Fusion activity of the different S2 constructs was assayed by protein-induced lipid mixing between unilamellar vesicles. Assays were performed for vesicles with the lipid composition: 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine (POPC), 1-palmitoyl-2-oleoyl-sn-glycero-3-phospho(1'-rac-glycerol) (POPG) at a 4:1 molar ratio. The composition has a net negative charge due to the -1 charge on POPG. Vesicle preparation began with dissolution of POPC, and POPG (1.6, and 0.4  $\mu\text{mol}$  respectively) in 1 mL of chloroform, followed by solvent removal using dry nitrogen gas and then overnight vacuum. The dry lipid films were then suspended in 2 mL of aqueous buffer and subjected to 10 freeze/thaw cycles to create unilamellar vesicles.

Subsequent extrusion (10-fold) of the suspension through a polycarbonate membrane with 100 nm diameter pores resulted in vesicles with typical diameters of 200-300 nm, as observed by electron microscopy. A set of companion “labeled” vesicles were also prepared that contained additional 2 mol % fluorescent lipid N-(7-nitro-2,1,3-benzoxadiazol-4-yl) phosphatidylethanolamine and 2 mol % quenching lipid N-(lissamine rhodamine B sulfonyl) phosphatidylethanolamine. Systematic differences among vesicle compositions were minimized by weighing all lipids at the same time using the same scale. The final vesicle solution contained POPC, and POPG at a total concentration of about 225  $\mu$ M and a 1:9 labeled to unlabeled vesicle ratio. The solution was transferred to a quartz cuvette in a fluorimeter and subjected to constant stirring at 37 °C. Fluorescence was monitored using 467 nm excitation, 530 nm detection, and a 1 s time increment. The initial baseline fluorescence,  $F_0$ , was determined, and an aliquot of a protein stock solution was then added ( $t=0$ ). The stock solution contained 40  $\mu$ M protein in water. The time-dependent fluorescence increase [ $\Delta F(t) = F(t) - F_0$ ] was diagnostic of protein-mediated fusion between a labeled and unlabeled vesicle (40-43). Relative to the initial labeled vesicle, the fused vesicle is expected to display a higher fluorescence because of the longer average fluorophore-quencher distance (40-43). The dead time after protein addition was  $\sim 5$  s, and the final asymptotic fluorescence change was usually achieved by 600 s. A 12  $\mu$ L aliquot of 10% Triton X-100 was then added to solubilize the vesicles, with a corresponding maximum fluorescence change ( $\Delta F_{\text{max}}$ ). The percent fusion parameter was calculated as  $[\Delta F(t)/\Delta F_{\text{max}}] \times 100$ . Assignment of protein as the cause of fusion is supported by negligible fusion (close to 0% fusion) after addition of an aliquot of a stock solution without protein.

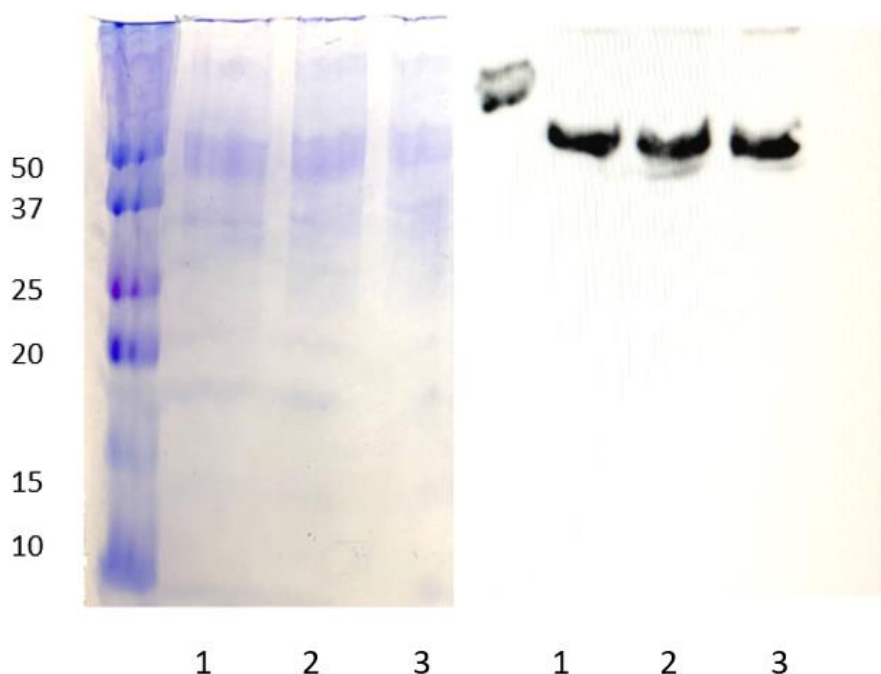
### 5.3. Results and Discussion

Our strategy for producing protein constructs based on the S2 involved the use of a non-native C-terminal His-tag to purify the protein from *E. coli* proteins. In addition to affinity purification, solubility purification of the protein in inclusion bodies was incorporated into the purification process. The different protein constructs displayed variable solubilities in water or PBS buffer.

#### 5.3.1. Initial Bacterial Growth and Expression and Solubility Check

S2 protein constructs were first tested for solubility to determine which fraction the target protein would be present in. Following bacterial growth and IPTG controlled expression, cells were lysed by tip-sonication in PBS buffer. The soluble proteins were separated from insoluble precipitate by centrifugation at 15000 rpm. The soluble proteins in PBS were then mixed with Co<sup>2+</sup> affinity resin. The insoluble protein fraction was solubilized in 6M GuCl and the soluble portion of the GuCl solubilization was mixed with Co<sup>2+</sup> affinity resin. Fractions were sorted by solubility in PBS or GuCl and then further divided into affinity resin binding and non-affinity resin binding proteins. Non-affinity resin binding proteins were obtained by washing the Co<sup>2+</sup> resin with PBS. Affinity binding protein was obtained by washing the Co<sup>2+</sup> resin with PBS and 250mM imidazole. The fractions were subject to SDS-PAGE analysis. The SDS-PAGE by itself is not useful because it contains too many different proteins; however, the accompanying western blot only shows a signal where the anti-Penta His antibody binds to the His-6 tag. The western blot shows that the S2\_816-1273 is insoluble in PBS buffer and is only soluble in 6M GuCl and binds to the affinity resin as seen in (**Figure 5.5**). The bands here are due to the low protein concentration in the lysate. The bands are shown more clearly in (**Figure 5.9**).

The S2\_903-998SGGRGG1163-1207, and S2\_903-998SGGRGG1153-1207 were both shown to be soluble in PBS and were shown to bind to the affinity resin (**Figure 5.6-5.7**). The differences in solubility result from the differences between the proteins, as well as, the increased hydrophobicity of adding the FP region and the transmembrane domain region from 1211-1234.



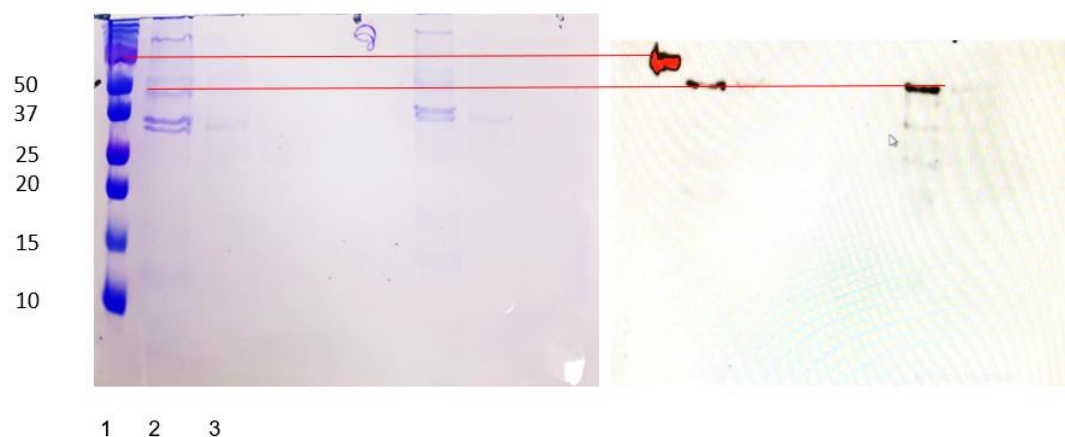
**Figure 5.11** Product of solubility purification of S2\_816-1273. The strongest band in the SDS-PAGE corresponds to monomer S2\_816-1273 as evidenced by the accompanying western blot.

### 5.3.2. Affinity Purification of S2 Constructs

All protein constructs featured a C-terminal GGGGGGLEHHHHHHH affinity tag for purification by Co<sup>2+</sup> affinity chromatography and for identification of target protein by western blot. Various purification protocols are shown for S2\_816-1273 in (**Figure 5.8-5.12**). Three different purification protocols were attempted, including:



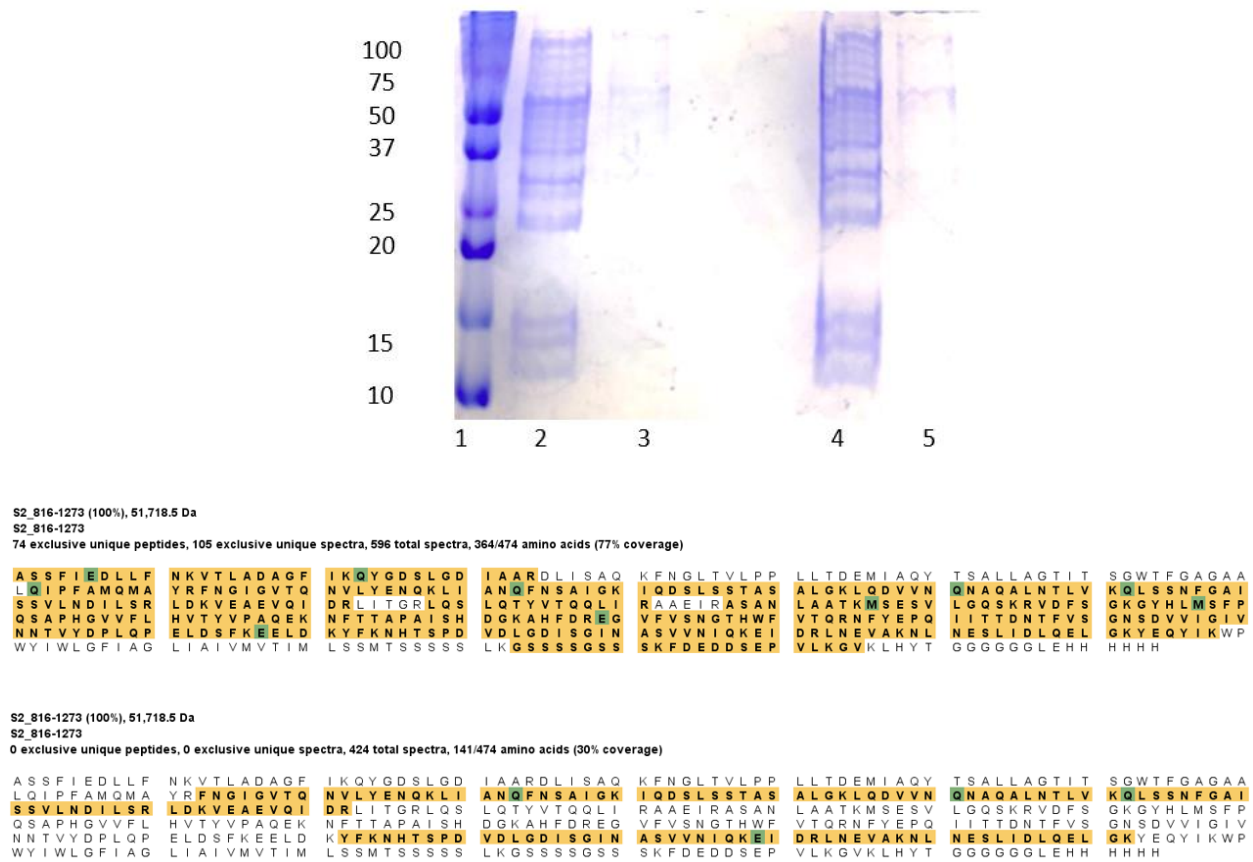
1) Solubility purification (**Figure 5.10**) of S2\_816-1273 described previously for large protein constructs (**Section 3.2.2.**). Separation of inclusion body-rich material began with tip-sonication in ~30 mL PBS per 10g cells at pH 7.4 in a 50 mL beaker in an ice bath then centrifugation at 27000g for 30 min and harvesting the new pellet and 2x repetition of the sonication/centrifugation/harvesting steps with a resultant pellet. The next step was tip sonication of the pellet 1 in ~30 mL PBS at pH 7.4 with 6 M GuCl, followed by centrifugation. The supernatant was dialyzed against deionized water overnight at 10 °C, using a 10 kDa cutoff membrane, with accompanying precipitation of S2 enriched material. The precipitated material is then solubilized in 8M urea and centrifuged retaining the supernatant. Solubility purification gives high yields of S2\_816-1273 approximately 12mg/L culture based on A280 of S2\_816-1273. However, the protein is not purified to any reasonable extent as seen with some impurity bands.



**Figure 5.12** SDS-PAGE and western blot of S2\_816-1273. The lanes are MW standards (lane 1) followed by the purified protein from the solubility purification (lane 2) and another lane run at half quantity (lane 3). There are lingering impurities seen on the SDS-PAGE. The lanes are repeated on the right side of the gel. Western blot shows a signal at approximately the same location as the mass of monomer S2\_816-1273.

2) Affinity purification (**Figure 5.9**) without solubility purification. Purification began with 3x PBS wash, like protocol 1, followed by solubility of protein in 6M GuHCl the protein is then mixed Co<sup>2+</sup> affinity resin overnight. Protein on the resin is washed with 10x the volume of resin with 6M GuHCl and then eluted with increasing concentrations of imidazole from 10mM to 500mM. The GuHCl is removed by dialysis against pure water but does not precipitate. The protein is solubilized in 8M urea prior to running the SDS-PAGE gel. The yield was too low to be useful for further experimentation (approximately 0.4mg/L culture).

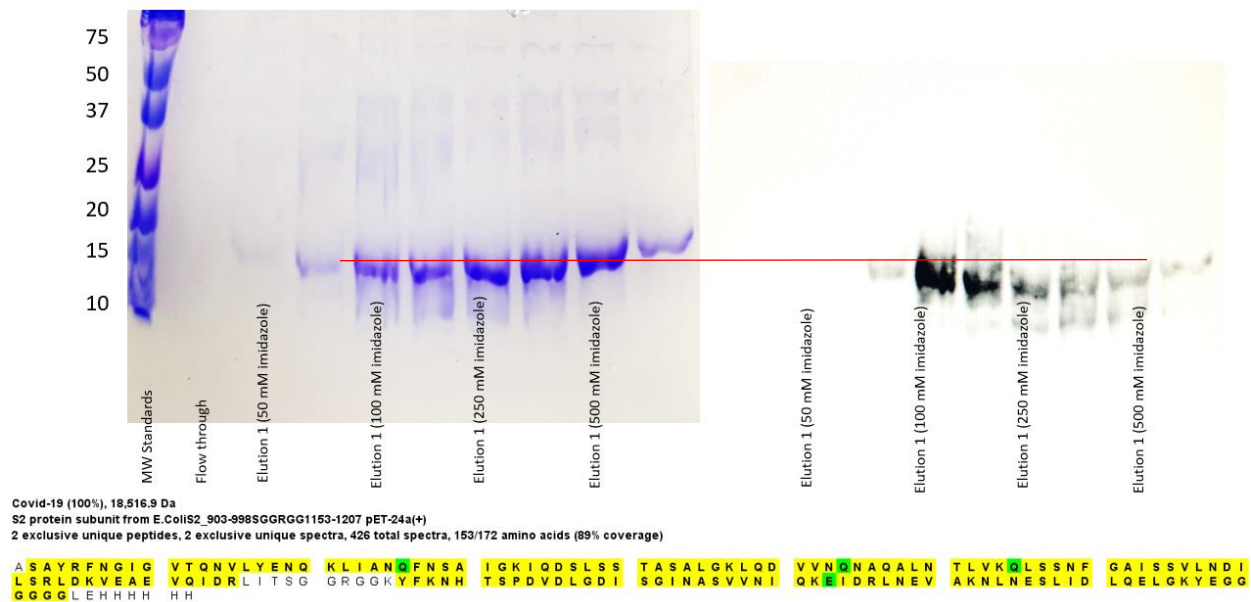
3) The third option was to combine the previous two purification protocols by using solubility purification to remove the bulk of non-target protein followed by affinity chromatography (**Figure 5.12**). Protocol 3 is protocol 2 with an additional precipitation, and solubilization in urea prior to affinity chromatography. The flow through of the solubility purification contains the proteins that do not bind to Co<sup>2+</sup> affinity resin. The elution is the affinity purified S2\_816-1273 removed from the Co<sup>2+</sup> resin by elution with 250 mM imidazole and 8 M urea. The idea was to increase binding of His-tag by removing other proteins that could interfere with the Co<sup>2+</sup> resin. The protein had a high degree of purity as well as a higher yield of ~4 mg/L culture. Bands may be less intense than expected on the SDS-PAGE gel due to formation of aggregates (above 100kDa), and general these protein constructs do not produce intense bands even when concentrations more than 1 mg/mL protein is used.



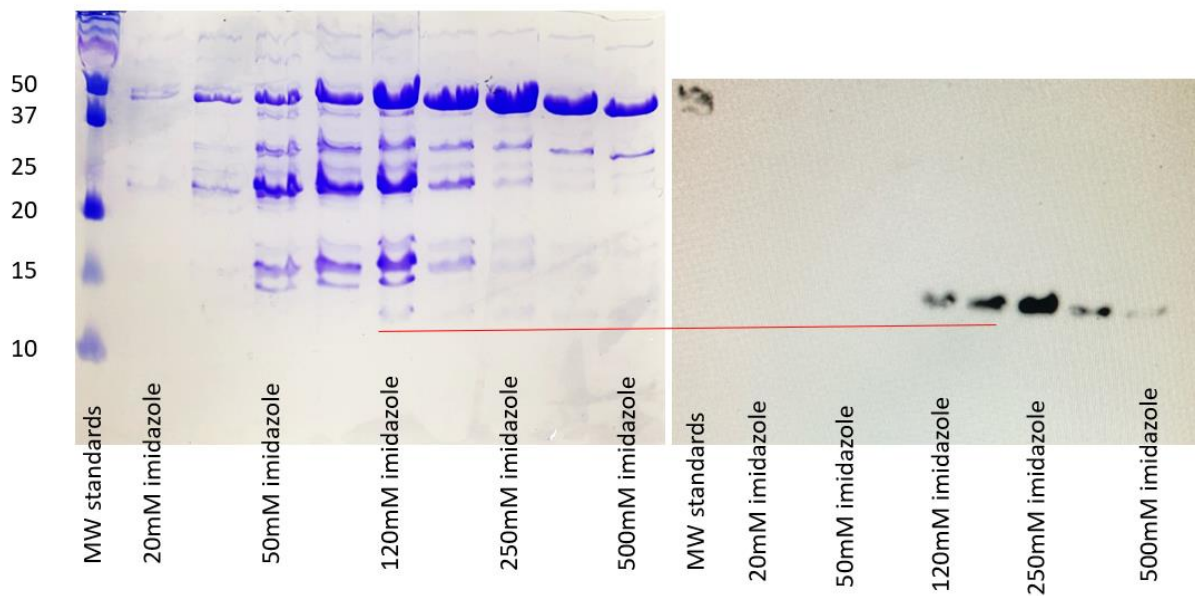
**Figure 5.13** SDS-PAGE of S2\_816-1273 purified by solubility followed by affinity chromatography. The first lane is the molecular weight standards. The second lane is the flow through from the affinity purification. High amounts of S2\_816-1273 (51.7 kDa) are still present in the flow through following solubility purification. The third lane is the affinity purified S2\_816-1273. Only the monomer band (51.7kDa) and a larger band dimer (100kDa) are observed. Both bands are confirmed to be S2\_816-1273 based on proteomics data. Lanes 4 & 5 are repeats of 2 & 3.

Purification of S2\_903-998SGGRGG1153-1207, and S2\_903-998SGGRGG1163-1207 was much more straightforward due to the increased solubility in PBS of the target proteins. Purification began with lysing cells in 0.05% CHAPS buffer followed by centrifugation to remove the supernatant containing the protein from the insoluble cell

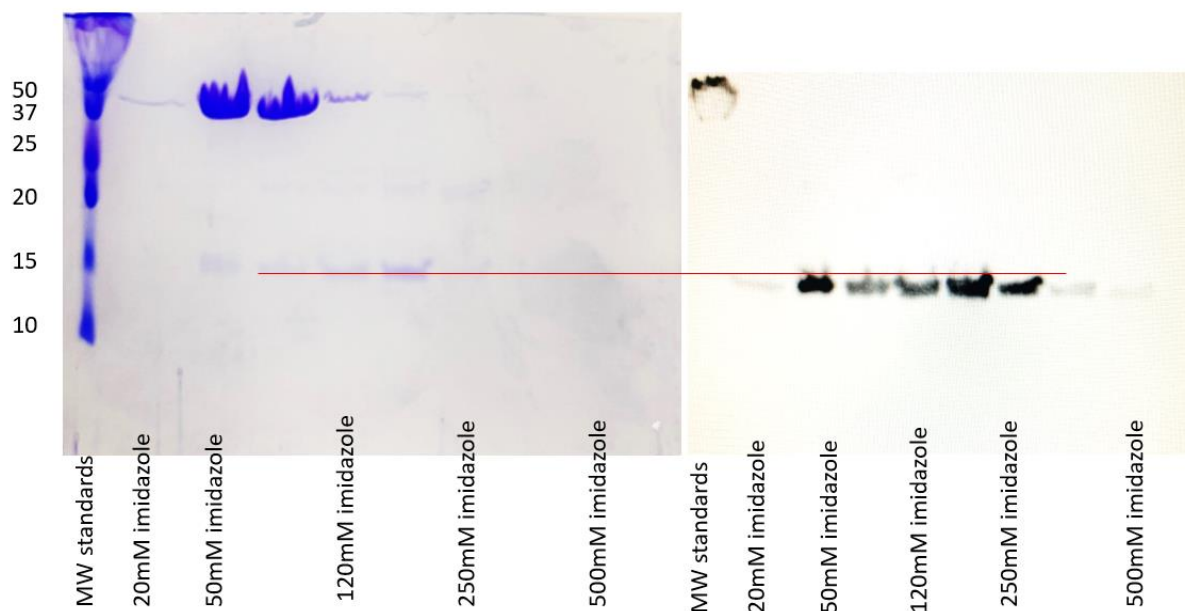
debris. The supernatant is then combined with  $\text{Co}^{2+}$  affinity resin overnight. The resin is washed with 10x bed volume of CHAPS buffer and then eluted with imidazole concentrations ranging from 10mM to 500mM. The target proteins come out pure as seen in (Figure 5.13-5.16). All protein constructs were confirmed by western blot and proteomics data. Some gels displayed prominent trimer bands while others showed no trimer bands. Western blots do not show signals for trimer bands, but trimer presence was confirmed by proteomics data.



**Figure 5.14** Affinity purification of S2\_903-998SGGRGG1153-1207. Little target protein is lost in the flow through. The affinity purified protein shows only the monomer S2\_903-998SGGRGG1153-1207 with high purity. Mass spectroscopy data following trypsin digestion show an 89% sequence coverage of the S2\_903-998SGGRGG1153-1207 confirming the presence of monomer. The lanes with labels are at the start of the new 3 mL addition of imidazole and the lanes without labels are at the end of that same 3 mL.

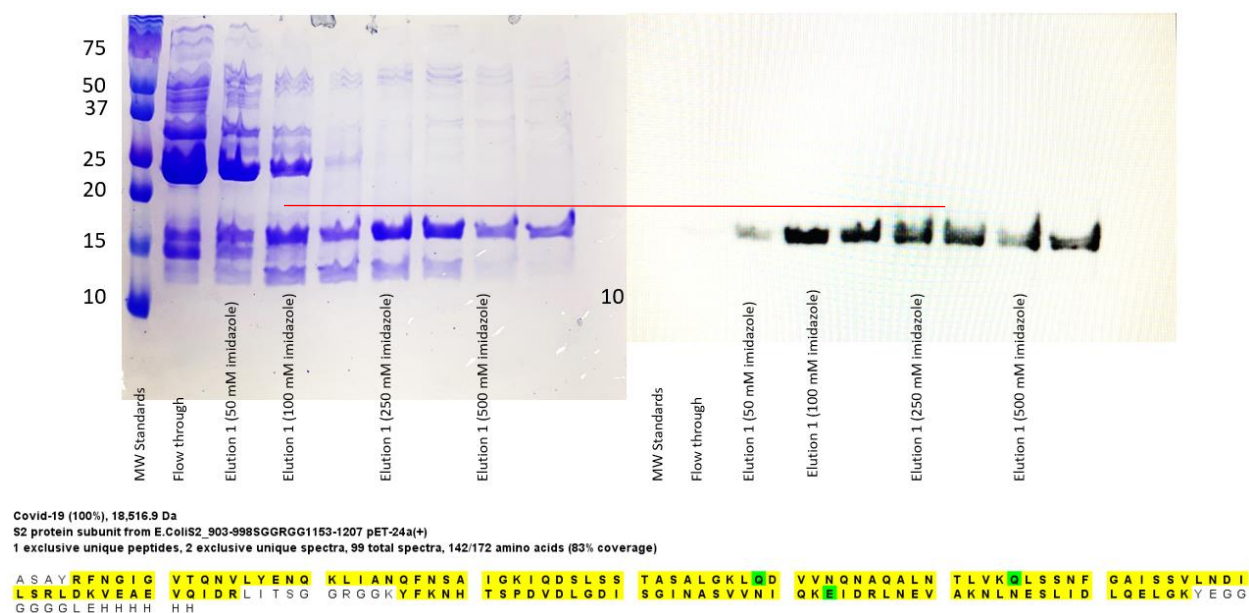


**Figure 5.15** Cell Lysis of S2\_903-998SGGRGG1163-1207. The cells are lysed by tip sonication in 0.05% CHAPS buffer and mixed with cobalt affinity resin overnight. The lanes on the SDS-PAGE are MW standards followed by imidazole elution gradient of 20mM, 50mM, 120mM, 250mM, and 500mM imidazole. The western blot on the right shows the presence of the S2\_903-998SGGRGG1163-1207 monomer (17.4kDa) is not a major product in this purification. However, the strong band at about 50kDa is identified as S2\_903-998SGGRGG1163-1207 trimer by proteomics data. The lanes with labels are at the start of the new 3 mL addition of imidazole and the lanes without labels are at the end of that same 3 mL.



**Figure 5.16** SDS-PAGE of affinity chromatography of S2\_903-998SGGRGG1163-1207.

The accompanying western blot confirms the presence of monomer S2\_903-998SGGRGG1163-1207. The S2\_903-998SGGRGG1163-1207 monomer shows strong signal in the western blot the more intense SDS-PAGE bands at 50kDa do not produce signals in the western blot. We do not know why trimer elutes at 50 mM imidazole in this purification. The lanes with labels are at the start of the new 3 mL addition of imidazole and the lanes without labels are at the end of that same 3 mL.



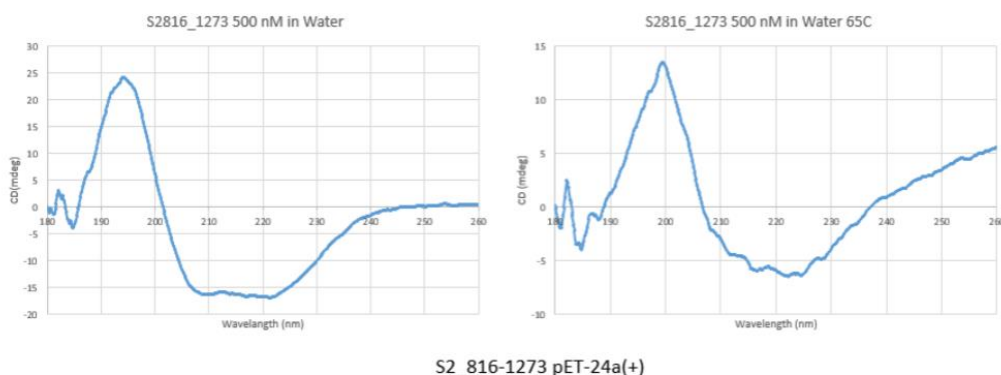
**Figure 5.17** Affinity purification of S2\_903-998SGGRGG1163-1207. The affinity purified protein shows only the monomer S2\_903-998SGGRGG1163-1207 with high purity. Mass spectroscopy data following trypsin digestion show an 83% sequence coverage of the S2\_903-998SGGRGG1163-1207 confirming the presence of monomer. The elution is a continuous flow of increasing imidazole concentration and approximately 3mL of buffer is added per band. The lanes with labels are at the start of the new imidazole concentration and the lanes without labels are at the end.

### 5.3.3. Circular Dichroism Spectroscopy of S2 Proteins

CD spectra were obtained for S2\_816-1273, S2\_903-998SGGRGG1153-1207, and S2\_903-998SGGRGG1163-1207. The CD spectra were measured to determine proper refolding of the protein after solubilization. Spectra were examined both at ambient temperature and at 65 °C to determine thermal stability. The CD spectra of S2\_816-1273 are given in (**Figure 5.17**). The room temperature CD data show a classically alpha helical spectra in water based on absorbance at 208 nm and 222 nm (**Figure 2.10**). Alpha helicity was also measured in various buffers to determine whether the structure remains alpha

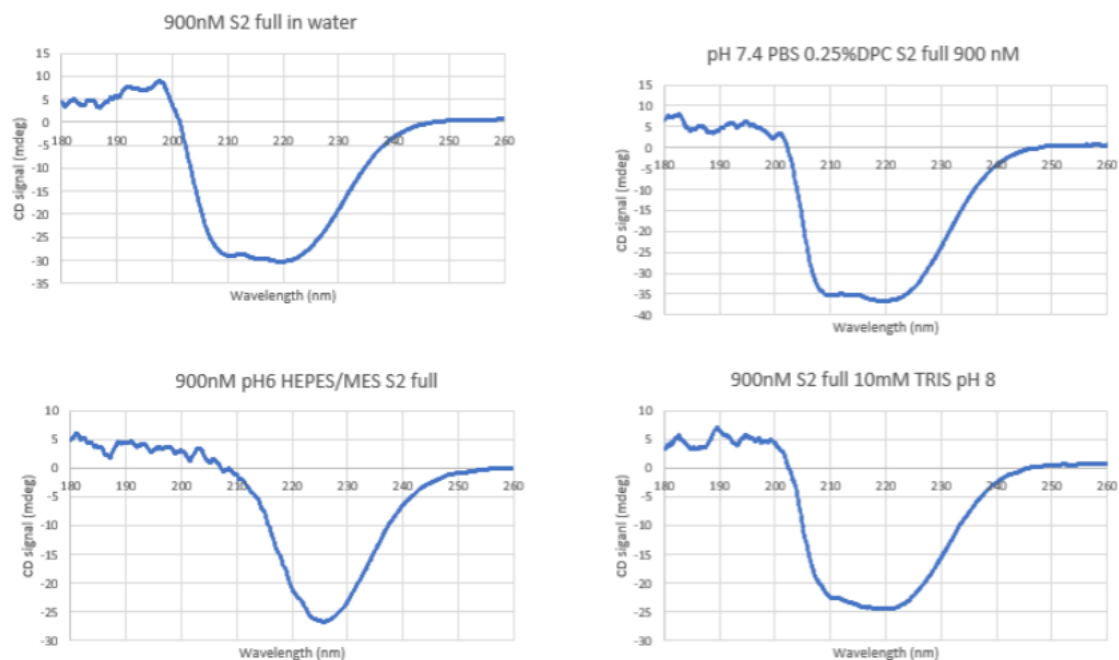


helical in both buffer and detergent. Alpha helical structure is retained in 0.25% DPC as well as in 10mM TRIS at pH 8 (**Figure 5.18**). Helicity is lost however if HEPES/MES buffer at pH 6 is used. The CD of protein in HEPES/MES appears to take on a  $\beta$ -sheet structure. S2\_903-998SGGRGG1153-1207 does not show any obvious change in the CD spectrum when transitioning from room temperature to 65 °C as shown in (**Figure 5.19**), although the CD spectra do not show alpha helical structure regardless. S2\_903-998SGGRGG1163-1207 shown in (**Figure 5.20**) exhibits both strong alpha helical structure as well as high thermal stability.

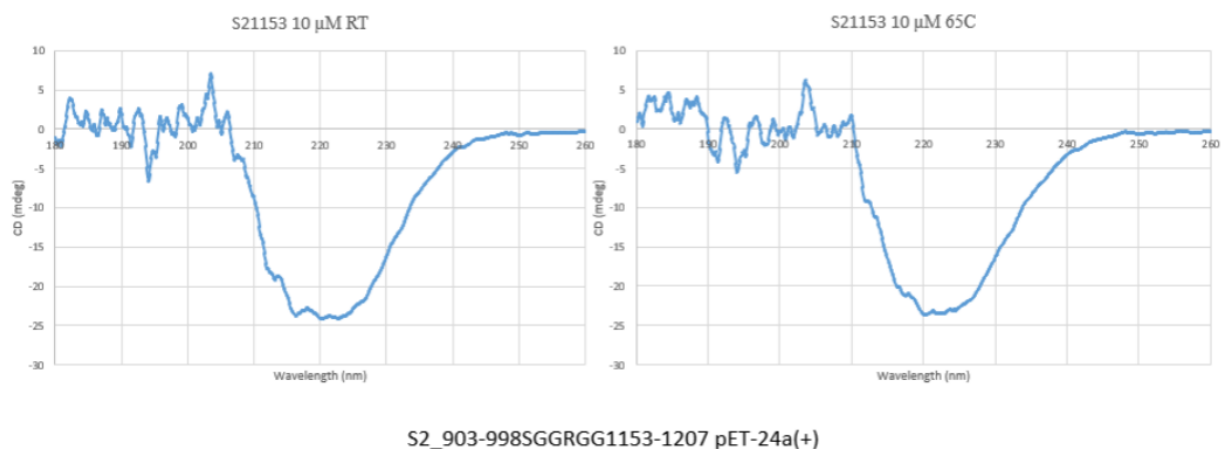


**Figure 5.18** CD spectra of S2\_816-1273 taken at ambient temperature on the left and 65 °C on the right. The protein appears to be less thermally stable than other S2 constructs, as well as other similar enveloped membrane proteins from HIV. Molar ellipticity at 222 nm is related to mdeg at 222 nm by  $\theta = m^0M/LC$  where  $\theta$  = molar ellipticity,  $m^0$  is circular dichroism in mdeg,  $M$  = molar mass,  $L$  = path length of the cell, and  $C$  = concentration. For this experiment  $M = 51717$  g/mol,  $L = 10$ mm,  $C = .51717$  g/mL. Molar ellipticity = 15000 deg(cm<sup>2</sup>)/dmol (left) and 5300 deg(cm<sup>2</sup>)/dmol (right) compared to total helical value of 33000 deg(cm<sup>2</sup>)/dmol.

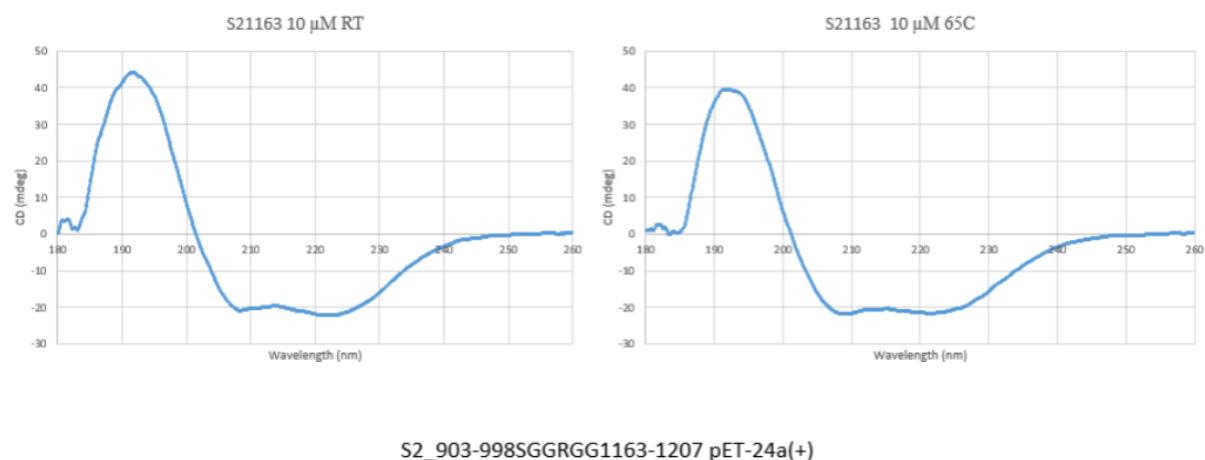




**Figure 5.19** CD spectra of S2\_816-1273 in detergent and buffers. Spectra measured for PBS/ 0.25%DPC, and 10mM TRIS are like spectra observed for water. HEPES/MES buffer appears to give a different CD spectrum from the others. Helicity at top left is 15100 deg(cm<sup>2</sup>)/dmol, top right is 18200 deg(cm<sup>2</sup>)/dmol, bottom right is 12000 deg(cm<sup>2</sup>)/dmol. Helicity is compared to total helical value of 33000 deg(cm<sup>2</sup>)/dmol.



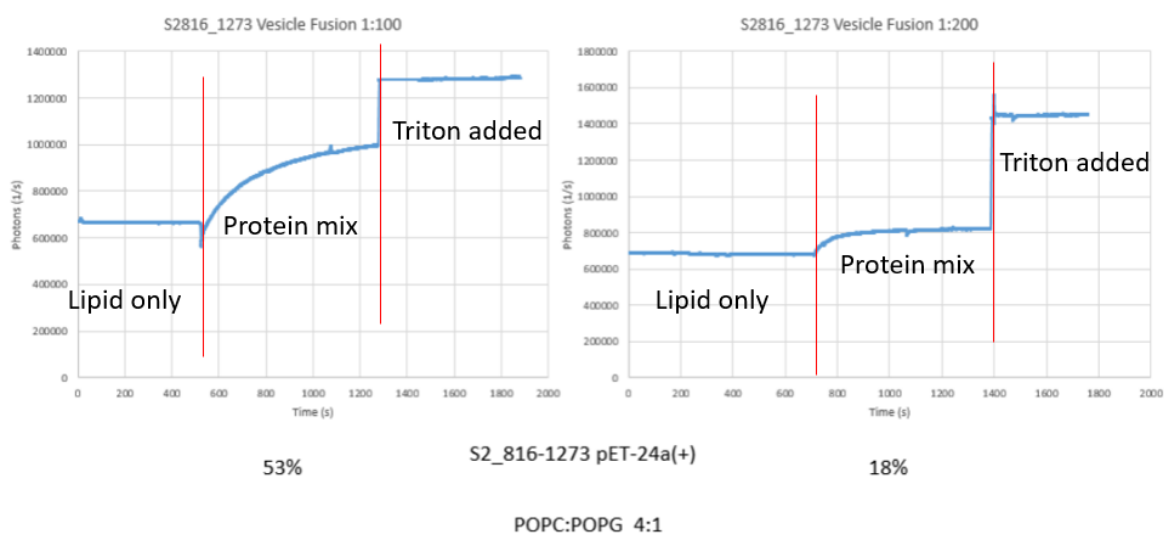
**Figure 5.20** CD spectra for S2\_903-998SGGRGG1153-1207 in water. The protein is less distinctly alpha helical from either S2\_816-1273 or S2\_903-998SGGRGG1163-1207.



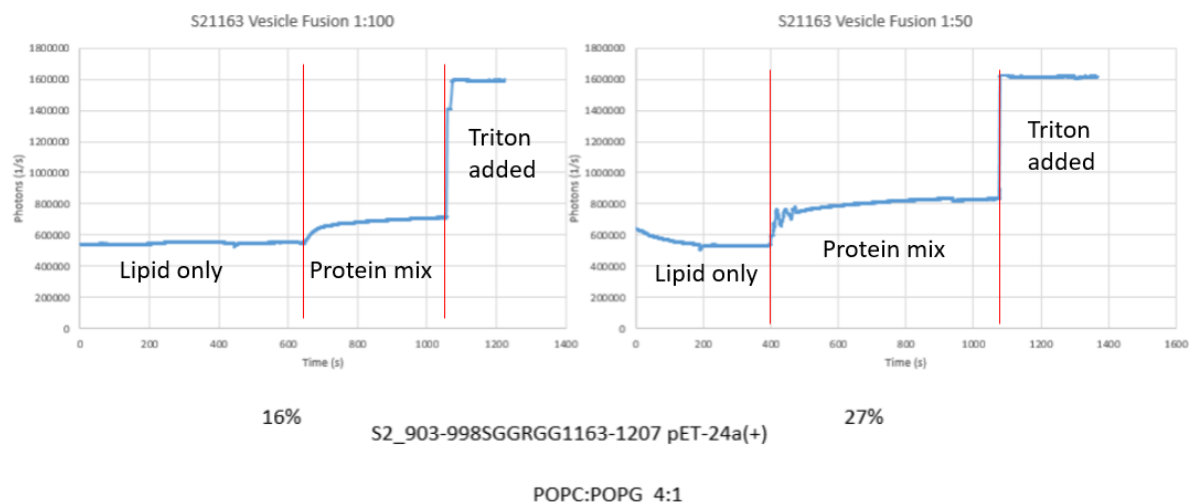
**Figure 5.21** CD spectra for S2\_903-998SGGRGG1163-1207 in water. The protein appears to be more thermally stable than S2\_816-1273 as evidenced by the similarity of CD spectra at ambient temperature on the left and 65 °C on the right. The molar ellipticity is 36400 deg(cm<sup>2</sup>)/dmol.

### 5.3.4. Vesicle Fusion Assay Shows Protein is Fusogenic

The vesicle fusion assay performed on POPC:POPG 4:1 lipid vesicles (**Figure 5.21**) shows a strong ability for alpha helical S2816-1273 to fuse with lipid vesicles. Vesicle fusion of the full length S2' construct is >50% for a protein to lipid ratio of 1:100, which is similar in magnitude to vesicle fusion observed by other proteins including those of HIV-1 (26,30). Measured vesicle fusion in S2\_816-1273 is also much higher than vesicle fusion observed for S2\_903-998SGGRGG1163-1207 (**Figure 5.22**). The S2' construct S2\_816-1273 contains both the FP and the TM and shows a greater than 3x increase in vesicle fusion at protein:lipid ratio 1:100.



**Figure 5.22** Lipid vesicle assay for S2\_816-1273 in POPC: POPG 4:1. The S2 protein in stock solution 3mg/mL in water pH 7 was added to the lipid vesicles in pH 5 HEPES/MES buffer (10mM MES, 5mM HEPES, 0.01% NaN<sub>3</sub>). The protein exhibits 53% fusion of vesicles at 37 °C. for a 1:100 protein to lipid ratio.



**Figure 5.23** Lipid vesicle assay for S2\_903-998SGGRGG1163-1207 in POPC: POPG 4:1. The protein exhibits 16% fusion of vesicles at 37 °C. Vesicles are the same as (Figure 5.21) for a 1:100 protein to lipid ratio. The S2\_816-1273 exhibits a more than 3x greater ability to fusion vesicles.

#### 5.4. Summary

This study reports expression, purification, solubilization, and purification of multiple S2' protein constructs in the pET-24a(+) vector. Procedures are described for acquiring mg quantities of pure protein which is demonstrated to be folded properly and induce lipid vesicle fusion.

## REFERENCES

1. Gisanddata. *ArcGIS*, John Hopkins University & Medicine <https://gisanddata.maps.arcgis.com>. (2021)
2. U.S. Department of Health and Human Services. Centers for Disease Control and Prevention. [https://www.cdc.gov/nchs/nvss/vsrr/covid19/excess\\_deaths.htm](https://www.cdc.gov/nchs/nvss/vsrr/covid19/excess_deaths.htm). (2021)
3. Wu, F., Zhao, S., Yu, B., Chen, Y. M., Wang, W., Song, Z. G., . . . Zhang, Y. Z. (2020) A new coronavirus associated with human respiratory disease in China, *Nature* 579, 265-269.
4. Lu, R. J., Zhao, X., Li, J., Niu, P. H., Yang, B., Wu, H. L., . . . Tan, W. J. (2020) Genomic characterization and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding, *Lancet* 395, 565-574.
5. Wrapp, D., Wang, N., Corbett, K. S., Goldsmith, J. A., Hsieh, C.-L., Abiona, O., . . . McLellan, J. S. (2020) Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation, *Science* 367, 1260-1263.
6. Ng, M. L., Tan, S. H., See, E. E., Ooi, E. E., and Ling, A. E. (2003) Early events of SARS coronavirus infection in Vero cells, *J. Med. Virol.* 71, 323-331.
7. Zhang, H., Wang, G. W., Li, H., Nie, Y. C., Shi, X. L., Lian, G. W., . . . Deng, H. K. (2004) Identification of an antigenic determinant on the S2 domain of the severe acute respiratory syndrome coronavirus spike glycoprotein capable of inducing neutralizing antibodies, *J. Virol.* 78, 6938-6945.
8. Zhong, X. F., Yang, H. H., Guo, Z. F., Sin, W. Y. F., Chen, W., Xu, J. J., . . . Guo, Z. H. (2005) B-cell responses in patients who have recovered from severe acute respiratory syndrome target a dominant site in the S2 domain of the surface spike glycoprotein, *J. Virol.* 79, 3401-3408.
9. Xia, S., Liu, M. Q., Wang, C., Xu, W., Lan, Q. S., Feng, S. L., . . . Lu, L. (2020) Inhibition of SARS-CoV-2 (previously 2019-nCoV) infection by a highly potent pan-coronavirus fusion inhibitor targeting its spike protein that harbors a high capacity to mediate membrane fusion, *Cell Research* 30, 343-355.
10. White, J. M., Delos, S. E., Brecher, M., and Schornberg, K. (2008) Structures and mechanisms of viral membrane fusion proteins: Multiple variations on a common theme, *Crit. Rev. Biochem. Mol. Biol.* 43, 189-219.
11. Heald-Sargent, T., and Gallagher, T. (2012) Ready, Set, Fuse! The Coronavirus Spike Protein and Acquisition of Fusion Competence, *Viruses-Basel* 4, 557-580.

12. Kielian, M. (2014) Mechanisms of virus membrane fusion proteins, *Annual Rev. Virol.* 1, 171-189.
13. Harrison, S. C. (2015) Viral membrane fusion, *Virology* 479, 498-507.
14. White, J. M., and Whittaker, G. R. (2016) Fusion of enveloped viruses in endosomes, *Traffic* 17, 593-614.
15. Tang, T., Bidon, M., Jaimes, J. A., Whittaker, G. R., and Daniel, S. (2020) Coronavirus membrane fusion mechanism offers a potential target for antiviral development, *Antiviral Research* 178, 104792.
16. Grewe, C., Beck, A., and Gelderblom, H. R. (1990) HIV: early virus-cell interactions, *J. AIDS* 3, 965-974.
17. Herold, N., Anders-Osswein, M., Glass, B., Eckhardt, M., Muller, B., and Krausslich, H. G. (2014) 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, 13956-13970.
18. Melikyan, G. B. (2014) HIV entry: a game of hide-and-fuse?, *Curr. Opin. Virol.* 4, 1-7.
19. Kemble, G. W., Danieli, T., and White, J. M. (1994) Lipid-anchored influenza hemagglutinin promotes hemifusion, not complete fusion, *Cell* 76, 383-391.
20. Blumenthal, R., Sarkar, D. P., Durell, S., Howard, D. E., and Morris, S. J. (1996) Dilation of the influenza hemagglutinin fusion pore revealed by the kinetics of individual cell-cell fusion events, *Journal of Cell Biology* 135, 63-71.
21. Markosyan, R. M., Cohen, F. S., and Melikyan, G. B. (2000) The lipid-anchored ectodomain of influenza virus hemagglutinin (GPI-HA) is capable of inducing nonenlarging fusion pores, *Mol. Biol. Cell* 11, 1143-1152.
22. Leikina, E., LeDuc, D. L., Macosko, J. C., Epand, R., Shin, Y. K., and Chernomordik, L. V. (2001) The 1-127 HA2 construct of influenza virus hemagglutinin induces cell-cell hemifusion, *Biochemistry* 40, 8378-8386.
23. Munoz-Barroso, I., Durell, S., Sakaguchi, K., Appella, E., and Blumenthal, R. (1998) Dilation of the human immunodeficiency virus-1 envelope glycoprotein fusion pore revealed by the inhibitory action of a synthetic peptide from gp41, *J. Cell Biol.* 140, 315-323.
24. Markosyan, R. M., Cohen, F. S., and Melikyan, G. B. (2003) HIV-1 envelope proteins complete their folding into six-helix bundles immediately after fusion pore formation, *Mol. Biol. Cell* 14, 926-938.

25. Freed, E. O., Delwart, E. L., Buchsacher, G. L., Jr., and Panganiban, A. T. (1992) A mutation in the human immunodeficiency virus type 1 transmembrane glycoprotein gp41 dominantly interferes with fusion and infectivity, *Proc. Natl. Acad. Sci. U.S.A.* 89, 70-74.
26. Liang, S., Ratnayake, P. U., Keinath, C., Jia, L., Wolfe, R., Ranaweera, A., and Weliky, D. P. (2018) Efficient fusion at neutral pH by Human Immunodeficiency Virus gp41 trimers containing the fusion peptide and transmembrane domains, *Biochemistry* 57, 1219-1235. PMC6151270 Contact PD/PI: Weliky, David References Cited Page 37
27. Ranaweera, A., Ratnayake, P. U., and Weliky, D. P. (2018) 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, 5480-5493. PMC6433127
28. Ratnayake, P. U., Ekanayaka, E. A. P., Komanduru, S. S., and Weliky, D. P. (2016) 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 Purif.* 117, 6-16. PMC4684446
29. 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* 1848, 289-298. PMC4258546
30. 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. PMC4245979
31. Roche, J., Louis, J. M., Grishaev, A., Ying, J. F., and Bax, A. (2014) Dissociation of the trimeric gp41 ectodomain at the lipid-water interface suggests an active role in HIV-1 Env-mediated membrane fusion, *Proc. Natl. Acad. Sci. U.S.A.* 111, 3425-3430.
32. Pabis, A., Rawle, R. J., and Kasson, P. M. (2020) Influenza hemagglutinin drives viral entry via two sequential intramembrane mechanisms, *Proc. Natl. Acad. Sci. U.S.A.* 117, 7200-7207.
33. Kim, C. S., Epand, R. F., Leikina, E., Epand, R. M., and Chernomordik, L. V. (2011) The final conformation of the complete ectodomain of the HA2 subunit of Influenza Hemagglutinin can by itself drive low pH-dependent fusion, *J. Biol. Chem.* 286, 13226-13234.

34. Boonstra, S., Blijleven, J. S., Roos, W. H., Onck, P. R., van der Giessen, E., and van Oijen, A. M. (2018) Hemagglutinin-mediated membrane fusion: A biophysical perspective, *Ann. Revs. Biophys.* 47, 153-173.
35. Ranaweera, A., Ratnayake, P. U., Ekanayaka, E. A. P., Declercq, R., and Weliky, D. P. (2019) 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, 2432-2446. PMC6536117
36. Qiao, H., Armstrong, R. T., Melikyan, G. B., Cohen, F. S., and White, J. M. (1999) A specific point mutant at position 1 of the influenza hemagglutinin fusion peptide displays a hemifusion phenotype, *Mol. Biol. Cell* 10, 2759-2769.
37. Borrego-Diaz, E., Peeples, M. E., Markosyan, R. M., Melikyan, G. B., and Cohen, F. S. (2003) Completion of trimeric hairpin formation of influenza virus hemagglutinin promotes fusion pore opening and enlargement, *Virology* 316, 234-244.
38. 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. PMC2830311
39. Aydin, H., Al-Khooly, D., and Lee, J. E. (2014) Influence of hydrophobic and electrostatic residues on SARS coronavirus S2 protein stability: Insights into mechanisms of general viral fusion and inhibitor design, *Prot. Sci.* 23, 603-617.
40. Yao, H. W., and Hong, M. (2013) Membrane-dependent conformation, dynamics, and lipid Interactions of the fusion peptide of the paramyxovirus PIV5 from solid-state NMR, *J. Mol. Biol.* 425, 563-576.
41. Yang, S. T., Kiessling, V., and Tamm, L. K. (2016) Line tension at lipid phase boundaries as driving force for HIV fusion peptide-mediated fusion, *Nature Communications* 7, 1-9.
42. Ghosh, U., and Weliky, D. P. (2020) <sup>2</sup>H 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, 183404.
43. Lai, A. L., Millet, J. K., Daniel, S., Freed, J. H., and Whittaker, G. R. (2017) The SARS-CoV Fusion Peptide Forms an Extended Bipartite Fusion Platform that Perturbs Membrane Order in a Calcium-Dependent Manner, *J. Mol. Biol.* 429, 3875-3892.



44. Durrer, P., Galli, C., Hoenke, S., Corti, C., Gluck, R., Vorherr, T., and Brunner, J. (1996) H<sup>+</sup>-induced membrane insertion of influenza virus hemagglutinin involves the HA2 amino-terminal fusion peptide but not the coiled coil region, *J. Biol. Chem.* 271, 13417-13421.
45. 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. Contact PD/PI: Weliky, David References Cited Page 38
46. Qiang, W., Bodner, M. L., and Weliky, D. P. (2008) Solid-state NMR spectroscopy of human immunodeficiency virus fusion peptides associated with host-cell-like membranes: 2D correlation spectra and distance measurements support a fully extended conformation and models for specific antiparallel strand registries, *J. Am. Chem. Soc.* 130, 5459-5471. PMC4487652
47. Lorieau, J. L., Louis, J. M., and Bax, A. (2010) 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, 11341-11346.
48. Sackett, K., Nethercott, M. J., Zheng, Z. X., and Weliky, D. P. (2014) Solid-state NMR spectroscopy of the HIV gp41 membrane fusion protein supports intermolecular antiparallel beta sheet fusion peptide structure in the final six-helix bundle state, *J. Mol. Biol.* 426, 1077-1094. PMC3944376
49. Ghosh, U., Xie, L., Jia, L. H., Liang, S., and Weliky, D. P. (2015) 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, 7548-7551. PMC4481145
50. Caffrey, M., Cai, M., Kaufman, J., Stahl, S. J., Wingfield, P. T., Covell, D. G., . . . Clore, G. M. (1998) Threedimensional solution structure of the 44 kDa ectodomain of SIV gp41, *EMBO J.* 17, 4572-4584.
51. Yang, Z. N., Mueser, T. C., Kaufman, J., Stahl, S. J., Wingfield, P. T., and Hyde, C. C. (1999) The crystal structure of the SIV gp41 ectodomain at 1.47 Å resolution, *J. Struct. Biol.* 126, 131-144.
52. Chen, J., Skehel, J. J., and Wiley, D. C. (1999) 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, 8967-8972.
53. Buzon, V., Natrajan, G., Schibli, D., Campelo, F., Kozlov, M. M., and Weissenhorn, W. (2010) Crystal structure of HIV-1 gp41 including both fusion peptide and membrane proximal external regions, *PLoS Pathog.* 6, e1000880.

54. Han, X., and Tamm, L. K. (2000) A host-guest system to study structure-function relationships of membrane fusion peptides, *Proc. Natl. Acad. Sci. U.S.A.* 97, 13097-13102.
55. Belouzard, S., Chu, V. C., and Whittaker, G. R. (2009) Activation of the SARS coronavirus spike protein via sequential proteolytic cleavage at two distinct sites, *Proc. Natl. Acad. Sci. U.S.A.* 106, 5871-5876.
56. Duquerroy, S., Vigouroux, A. N., Rottier, P. J. M., Rey, F. A., and Bosch, B. J. (2005) Central ions and lateral asparagine/glutamine zippers stabilize the post-fusion hairpin conformation of the SARS coronavirus spike glycoprotein, *Virology* 335, 276-285.
57. Guillen, J., Kinnunen, P. K. J., and Villalain, J. (2008) Membrane insertion of the three main membranotropic sequences from SARS-CoV S2 glycoprotein, *Biochim. Biophys. Acta* 1778, 2765-2774.
58. Guillen, J., Perez-Berna, A. J., Moreno, M. R., and Villalain, J. (2008) A second SARS-CoV S2 glycoprotein internal membrane-active peptide. Biophysical characterization and membrane interaction, *Biochemistry* 47, 8214-8224.
59. Guillen, J., de Almeida, R. F. M., Prieto, M., and Villalain, J. (2008) Structural and dynamic characterization of the interaction of the putative fusion peptide of the S2SARS-CoV virus protein with lipid membranes, *J. Phys. Chem. B* 112, 6997-7007.
60. Madu, I. G., Roth, S. L., Belouzard, S., and Whittaker, G. R. (2009) Characterization of a highly conserved domain within the Severe Acute Respiratory Syndrome coronavirus spike protein S2 domain with characteristics of a viral fusion peptide, *J. Virol.* 83, 7411-7421.
61. Madu, I. G., Belouzard, S., and Whittaker, G. R. (2009) SARS-coronavirus spike S2 domain flanked by cysteine residues C822 and C833 is important for activation of membrane fusion, *Virology* 393, 265-271.
62. Mahajan, M., and Bhattacharjya, S. (2015) NMR structures and localization of the potential fusion peptides and the pre-transmembrane region of SARS-CoV: Implications in membrane fusion, *Biochim. Biophys. Acta* 1848, 721- 730.
63. Ou, X., Zheng, W., Shan, Y., Mu, Z., Dominguez, S. R., Holmes, K. V., and Qian, Z. (2016) Identification of the fusion peptide-containing region in betacoronavirus spike glycoproteins, *J. Virol.* 90, 5586-5600.
64. Basso, L. G. M., Vicente, E. F., Crusca, E., Jr., Cilli, E. M., and Costa-Filho, A. J. (2016) SARS-CoV fusion peptides induce membrane surface ordering and curvature, *Sci. Rep.* 6.

65. Millet, J. K., and Whittaker, G. R. (2018) Physiological and molecular triggers for SARS-CoV membrane fusion and entry into host cells, *Virology* 517, 3-8.
66. Mahajan, M., Chatterjee, D., Bhuvaneswari, K., Pillay, S., and Bhattacharjya, S. (2018) NMR structure and localization of a large fragment of the SARS-CoV fusion protein: Implications in viral cell fusion, *Biochim. Biophys. Acta* 1860, 407-415.
67. Meher, G., Bhattacharjya, S., and Chakraborty, H. (2019) Membrane cholesterol modulates oligomeric status and peptide-membrane interaction of Severe Acute Respiratory Syndrome coronavirus fusion peptide, *J. Phys. Chem. B* 123, 10654-10662.
68. Yang, J., Prorok, M., Castellino, F. J., and Weliky, D. P. (2004) Oligomeric  $\beta$ -structure of the membrane-bound HIV-1 fusion peptide formed from soluble monomers, *Biophys. J.* 87, 1951-1963. PMC1304598 Contact PD/PI: Weliky, David References Cited Page 39
69. Duan, J. Z., Yan, X. Y., Guo, X. M., Cao, W. C., Han, W., Qi, C., . . . Jin, G. (2005) A human SARS-CoV neutralizing antibody against epitope on S2 protein, *Biochem. Biophys. Res. Comm.* 333, 186-193.
70. Keng, C. T., Zhang, A., Shen, S., Lip, K. M., Fielding, B. C., Tan, T. H. P., . . . Tan, Y. J. (2005) Amino acids 1055 to 1192 in the S2 region of severe acute respiratory syndrome coronavirus S protein induce neutralizing antibodies: Implications for the development of vaccines and antiviral agents, *J. Virol.* 79, 3289-3296.
71. Lip, K. M., Shen, S., Yang, X. M., Keng, C. T., Zhang, A. H., Oh, H. L. J., . . . Tan, Y. J. (2006) Monoclonal antibodies targeting the HR2 domain and the region immediately upstream of the HR2 of the S protein neutralize in vitro infection of severe acute respiratory syndrome coronavirus, *J. Virol.* 80, 941-950.
72. Tripet, B., Kao, D. J., Jeffers, S. A., Holmes, K. V., and Hodges, R. S. (2006) Template-based coiled-coil antigens elicit neutralizing antibodies to the SARS-coronavirus, *J. Struct. Biol.* 155, 176-194.
73. Elshabrawy, H. A., Coughlin, M. M., Baker, S. C., and Prabhakar, B. S. (2012) Human monoclonal antibodies against highly conserved HR1 and HR2 comains of the SARS-CoV spike protein are more broadly neutralizing, *Plos One* 7.
74. Kong, R., Xu, K., Zhou, T. Q., Acharya, P., Lemmin, T., Liu, K., . . . Mascola, J. R. (2016) Fusion peptide of HIV-1 as a site of vulnerability to neutralizing antibody, *Science* 352, 828-833.
75. Xu, K., Acharya, P., Kong, R., Cheng, C., Chuang, G.-Y., Liu, K., . . . Kwong, P. D. (2018) Epitope-based vaccine design yields fusion peptide-directed antibodies that neutralize diverse strains of HIV-1, *Nature Medicine* 24, 857- 867.

76. Montero, M., van Houten, N. E., Wang, X., and Seott, J. K. (2008) The membrane-proximal external region of the human immunodeficiency virus type 1 envelope: Dominant site of antibody neutralization and target for vaccine design, *Microbiol. Mol. Biol. Rev.* 72, 54-84.
77. Ni, L., Zhu, J. Q., Zhang, J. J., Yan, M., Gao, G. F., and Tien, P. (2005) Design of recombinant protein-based SARS-CoV entry inhibitors targeting the heptad-repeat regions of the spike protein S2 domain, *Biochem. Biophys. Res. Comm.* 330, 39-45.
78. Park, J.-E., and Gallagher, T. (2017) Lipidation increases antiviral activities of coronavirus fusion-inhibiting peptides, *Virology* 511, 9-18.
79. Xia, S., Yan, L., Xu, W., Agrawal, A. S., Algaissi, A., Tseng, C.-T. K., . . . Lu, L. (2019) A pan-coronavirus fusion inhibitor targeting the HR1 domain of human coronavirus spike, *Science. Advances* 5.
80. Louis, J. M., Baber, J. L., and Clore, G. M. (2015) The C34 peptide fusion inhibitor binds to the six-helix bundle core domain of HIV-1 gp41 by displacement of the C-terminal helical repeat region, *Biochemistry* 54, 6796-6805.
81. Walls, A. C., Tortorici, M. A., Snijder, J., Xiong, X., Bosch, B.-J., Rey, F. A., and Veasler, D. (2017) Tectonic conformational changes of a coronavirus spike glycoprotein promote membrane fusion, *Proc. Natl. Acad. Sci. U.S.A.* 114, 11157-11162.
82. 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. PMC2782608
83. Pereira, F. B., Valpuesta, J. M., Basanez, G., Goni, F. M., and Nieva, J. L. (1999) Interbilayer lipid mixing induced by the human immunodeficiency virus type-1 fusion peptide on large unilamellar vesicles: the nature of the nonlamellar intermediates, *Chem. Phys. Lipids* 103, 11-20.
84. Worman, H. J., Brasitus, T. A., Dudeja, P. K., Fozzard, H. A., and Field, M. (1986) Relationship between lipid fluidity and water permeability of bovine tracheal epithelial cell apical membranes, *Biochemistry* 25, 1549-1555.
85. Kobayashi, T., Beuchat, M. H., Chevallier, J., Makino, A., Mayran, N., Escola, J. M., . . . Gruenberg, J. (2002) Separation and characterization of late endosomal membrane domains, *J. Biol. Chem.* 277, 32157-32164.
86. Guha, S., Rajani, M., and Padh, H. (2007) Identification and characterization of lipids from endosomes purified by electromagnetic chromatography, *Indian J. Biochem. Biophys.* 44, 443-449.

87. Petit, C. M., Melancon, J. M., Chouljenko, V. N., Colgrove, R., Farzan, M., Knipe, D. M., and Kousoulas, K. G. (2005) Genetic analysis of the SARS-coronavirus spike glycoprotein functional domains involved in cell-surface expression and cell-to-cell fusion, *Virology* 341, 215-230.
88. Petit, C. M., Chouljenko, V. N., Iyer, A., Colgrove, R., Farzan, M., Knipe, D. M., and Kousoulas, K. G. (2007) Palmitoylation of the cysteine-rich endodomain of the SARS-coronavirus spike glycoprotein is important for spikemediated cell fusion, *Virology* 360, 264-274.
89. Shulla, A., and Gallagher, T. (2009) Role of spike protein endodomains in regulating coronavirus entry, *J. Biol. Chem.* 284, 32725-32734.

**CHAPTER 6:**  
**SUMMARY AND FUTURE DIRECTIONS**

My research during the past few years has been focused on HIV gp41 and SARS CoV2 (S2). These two proteins are responsible for the initial step of virus infection by catalyzing the fusion between viral and host cell membrane. HIV infection happens on the surface of immune cells. After the HIV glycoprotein gp120 binds to the host cell, gp41 is exposed and fusion between host cell membrane and viral membrane starts. The fusion peptide (FP) and transmembrane (TM) domains are the segments that binds to membrane and are critical for fusion. As for SARS CoV2, the spike protein is proteolytically cleaved which triggers a conformational change in the S2. The ~18 N-terminal residues of S2' is fusion peptide domain (FP). The FP binds to host cells in the lungs and is necessary for fusion. However, the structure of gp41 and S2 as well as the infection mechanism are still not fully understood. The overall goal is to understand the gp41 and S2 structure and their interaction with the host cell membrane.

Chapter 3 mainly discusses the native chemical ligation synthesis and selective isotopic labeling of HIV gp41 fusion protein. A gp41 construct including the FP and MPER were synthesized by a combination of solid phase peptide synthesis, expression in *E. Coli*, and native chemical ligation. Both  $^{13}\text{C}$  and  $^2\text{H}$  labeling of the HM protein construct was explored; labeling was demonstrated by both REDOR experiments and MALDI mass spectroscopy. Selective isotopic labeling was also accomplished by solid phase peptide synthesis. The native chemical ligation reaction then allowed for synthesis of large protein constructs with singularly labeled amino acids. The resulting constructs were utilized in various experiments outlined in Chapter 4. Among these experiments were probing the insertion depth of the FP into lipid bilayer in a large protein construct. We also examined the formation of monomer in the lipid environment and used differential labeling to check

for close contact between FP and MPER. We proposed that there is no contact between FP and MPER domains in lipid according to the REDOR NMR.

Future studies can be done to determine the proximity of FPHMTM incorporated in membrane using  $^{13}\text{C}$ - $^{15}\text{N}$  REDOR solid-state NMR. The FPHMTM can be obtained by native chemical ligation of  $^{13}\text{C}$  labeled FP peptide and  $^{15}\text{N}$  labeled HMTM. The FP peptide can be synthesized by solid-phase peptide synthesis with site-specific labeling, while the HMTM can be labeled with  $^{15}\text{N}$  Phe residue by bacterial expression. There are three Phe in HMTM: one of them in MPER and two in the TM. The Phe in MPER could be mutated to other residues to avoid the FP and MPER interaction being observed. The result of a big dephasing buildup of the  $^{13}\text{C}$ - $^{15}\text{N}$  REDOR indicates close contact between FP and TM, and small dephasing buildup indicates no contact. A possible issue lies in the expression of  $^{15}\text{N}$  Phe labeled HMTM, as other residues may also be labeled due to *E. coli* scrambling. If the yield of ligation experiment is high, it is worth trying to ligate FP and TM to HM. The methodology developed in Chapter 3 provides a means of producing the ligation product between FP and HMTM. HMTM could possibly be made by ligation between HM which has been chemically modified at the C-terminus to have a thioester and the TM with and N-terminal Cys. A similar study could be performed with deuterium labeled TM. Multiple other experiments could be proposed utilizing the methodology we have developed here across many membrane proteins including influenza and SARS-CoV2 proteins.

In Chapter 5, we laid the foundation for further study of SARS-CoV2 S2 membrane protein by expression, solubilization, and purification of protein constructs with and without the FP and TM. We showed that the full length S2 was properly folded and



exhibited the ability to fuse lipid vesicles. We were also able to express a protein construct of S2 that did not have the fusion peptide. This construct could be used in an analogous manner to the method described in Chapter 3 to produce S2 protein from native chemical ligation that has site-specific isotopic labeling. This methodology could be used then to study S2 in the lipid bilayer.

## REFERENCES

1. Levy, J. A. (1993) Pathogenesis of Human-Immunodeficiency-Virus Infection, *Microbiol. Rev.* 57, 183-289.
2. Webster, R. G., Bean, W. J., Gorman, O. T., Chambers, T. M., and Kawaoka, Y. (1992) Evolution and Ecology of Influenza-A Viruses, *Microbiol. Rev.* 56, 152-179.
3. Engen, J. R. (2009) Analysis of Protein Conformation and Dynamics by Hydrogen/Deuterium Exchange MS, *Anal. Chem.* 81, 7870-7875.
4. Curtis-Fisk, J., Spencer, R. M., and Weliky, D. P. (2008) Isotopically labeled expression in E-coli, purification, and refolding of the full ectodomain of the influenza virus membrane fusion protein, *Protein Expr. Purif.* 61, 212-219.
5. Ghosh, U., Xie, L., Jia, L. H., Liang, S., and Weliky, D. P. (2015) 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, 7548-7551.

## APPENDIX A: Tables of NMR Values

**Table A.1. REDOR NMR Values for  $^{13}\text{C}$  Labeled FP**

$^{13}\text{CFP}$	$S_0$	$S_1$	$\Delta S$	$\Delta S/S_0$	# scans
16 ms	527.28	515.57	11.71	2.2%	30195
24 ms	80.5	79.4	1.1	1.3%	4958

**Table A.2. REDOR NMR Values for  $^{13}\text{C}$  Labeled Carbonyl Region of HM**

$^{13}\text{CHM (CO)}$	$S_0$	$S_1$	$\Delta S$	$\Delta S/S_0$	# scans
2 ms	6.5409	6.2545	0.2864	4.4%	65
8 ms	4.2247	3.6576	0.5671	13.4%	428
16 ms	1.6629	1.6728	0.0099	0.6%	65
24 ms	14.3119	15.7534	1.4415	10.1%	1754
48 ms	8.1266	7.5575	0.5691	7.0%	14942

**Table A.3. REDOR NMR Values for  $^{13}\text{C}$  Labeled Alpha Carbon Region of HM**

$^{13}\text{CHM } (\alpha\text{C})$	$S_0$	$S_1$	$\Delta S$	$\Delta S/S_0$	# scans
2 ms	3.3723	3.3603	0.012	3.5%	65
8 ms	1.2372	0.7372	0.5	40.4%	428
16 ms	0.1222	0.1261	0.0039	3.2%	65
24 ms	3.0456	3.9940	0.9484	31%	1754
48 ms	3.5809	2.1224	1.4585	40.1%	14942

**Table A.4. REDOR NMR Values for  $^{13}\text{C}$  Labeled Sidechain Region of HM**

$^{13}\text{CHM (sidechain)}$	$S_0$	$S_1$	$\Delta S$	$\Delta S/S_0$	# scans
2 ms	12.1197	11.7847	0.335	2.7%	65
8 ms	9.1906	8.7672	0.4234	4.6%	428
16 ms	0.4636	0.6776	0.215	46.4%	65
24 ms	13.0326	13.2290	0.1964	1.5%	1754
48 ms	22.7065	21.5026	1.2039	5.3%	14942

**Table A.5. REDOR NMR Values for  $^2\text{H}$  Labeled Carbonyl Region of HM**

$^2\text{HHM (CO)}$	$S_0$	$S_1$	$\Delta S$	$\Delta S/S_0$	# scans
2 ms	129.2078	80.0482	49.1596	38%	35355
32 ms	33.2123	3.5515	29.6608	89%	32635
40 ms	13.7150	2.6106	11.1044	81%	17235
48 ms	13.2293	1.0510	12.1783	92%	22132

**Table A.6. REDOR NMR Values for  $^2\text{H}$  Labeled Alpha Carbon Region of HM**

$^2\text{HHM } (\alpha\text{C})$	$S_0$	$S_1$	$\Delta S$	$\Delta S/S_0$	# scans
2 ms	33.1308	5.9037	27.2271	82%	35355
32 ms	8.3840	0.2574	8.1266	97%	32635
40 ms	4.5642	0.3221	4.2421	93%	17235
48 ms	3.3650	0.3308	3.0342	90%	22132

**Table A.7. REDOR NMR Values for  $^2\text{H}$  Labeled Sidechain Region of HM**

$^2\text{HHM}$ (sidechain)	$S_0$	$S_1$	$\Delta S$	$\Delta S/S_0$	# scans
2 ms	105.56	43.8747	61.6853	58.4%	35355
32 ms	12.1834	2.2955	9.8879	81%	32635
40 ms	2.0283	1.1715	0.8568	42.2%	17235
48 ms	5.8237	1.2429	4.5808	78.7%	22132

**Table A.8. REDOR NMR Carbonyl Values for  $^{13}\text{CFP}$  Chemically Ligated to  $^2\text{H}$  Labeled HM**

$^{13}\text{CFP-}^2\text{HHM (CO)}$	$S_0$	$S_1$	$\Delta S$	$\Delta S/S_0$	# scans
2 ms	57.3805	55.5108	1.8697	3.2%	10000
8 ms	5.0805	4.3203	0.7602	15%	902
16 ms	43.7525	36.9136	6.8389	15.6%	10000
24 ms	41.9143	33.5451	8.3692	20%	10000
32 ms	35.0007	26.5674	8.9994	26%	10000
40 ms	13.5668	11.9381	1.6287	12%	5524
48 ms	19.7004	15.4688	4.2316	21.5%	10000

**Table A.9. REDOR NMR Alpha Carbon Values for  $^{13}\text{CFP}$  Chemically Ligated to  $^2\text{H}$  Labeled HM**

$^{13}\text{CFP-}^2\text{HHM } (\alpha\text{C})$	$S_0$	$S_1$	$\Delta S$	$\Delta S/S_0$	# scans
2 ms	10.1301	7.8396	2.2905	23%	10000
8 ms	N/A	N/A	N/A	N/A	902
16 ms	2.9179	1.7706	1.1473	39.3%	10000
24 ms	N/A	N/A	N/A	N/A	10000
32 ms	N/A	N/A	N/A	N/A	10000
40 ms	N/A	N/A	N/A	N/A	5524
48 ms	N/A	N/A	N/A	N/A	10000

**Table A.10. REDOR NMR Sidechain Values for  $^{13}\text{CFP}$  Chemically Ligated to  $^2\text{H}$  Labeled HM**

$^{13}\text{CFP-}^2\text{HHM}$ (sidechain)	$S_0$	$S_1$	$\Delta S$	$\Delta S/S_0$	# scans
2 ms	60.7748	20.4884	40.2864	66.3%	10000
8 ms	N/A	N/A	N/A	N/A	902
16 ms	6.2984	2.4786	3.8198	60.1%	10000
24 ms	N/A	N/A	N/A	N/A	10000
32 ms	N/A	N/A	N/A	N/A	10000
40 ms	N/A	N/A	N/A	N/A	5524
48 ms	N/A	N/A	N/A	N/A	10000

**Table A.11. REDOR NMR Carbonyl Values for  $^{13}\text{CFP}$  Chemically Ligated to HM in  $^2\text{H}$  Labeled Lipids**

$^{13}\text{CFP}$ -HM in $^2\text{H}$ lipid (CO)	$S_0$	$S_1$	$\Delta S$	$\Delta S/S_0$	# scans
2 ms	1.3637	0.4021	0.9616	70.5%	2089
8 ms	2.6898	1.6698	1.02	38%	18177
16 ms	6.7312	3.7244	3.0068	44.6%	41889
24 ms	35.9107	16.6622	19.2485	53.6%	24261

**Table A.12. REDOR NMR Carbonyl Values for  $^{13}\text{CFP}$  mixed with  $^2\text{H}$  Labeled HM**

$^{13}\text{CHM}$ $^2\text{HHM}$ mixed trimers (CO)	$S_0$	$S_1$	$\Delta S$	$\Delta S/S_0$	# scans
2 ms	4.3543	1.8901	2.4642	56.6%	14214
8 ms	6.5007	4.7588	1.7419	26.8%	995
16 ms	2.6126	1.5616	1.051	40.2%	923
24 ms	1.7547	0.6828	1.0719	61.1%	1037
40 ms	1.8023	1.4491	0.3532	19.6%	3864
48 ms	4.4005	1.9003	2.5002	57%	14214

**Table A.13. REDOR NMR Alpha Carbon Values for  $^{13}\text{CFP}$  mixed with  $^2\text{H}$  Labeled HM**

$^{13}\text{CHM}$ $^2\text{HHM}$ mixed trimers ( $\alpha\text{C}$ )	$S_0$	$S_1$	$\Delta S$	$\Delta S/S_0$	# scans
2 ms	5.4453	0.9867	4.4586	82%	14214
8 ms	N/A	N/A	N/A	N/A	995
16 ms	N/A	N/A	N/A	N/A	923
24 ms	N/A	N/A	N/A	N/A	1037
40 ms	N/A	N/A	N/A	N/A	3864
48 ms	N/A	N/A	N/A	N/A	14214

**Table A.14. REDOR NMR Carbonyl Values for  $^{13}\text{CFP}$  mixed with  $^2\text{H}$  Labeled HM**

$^{13}\text{CHM}$ $^2\text{HHM}$ mixed trimers (sidechain)	$S_0$	$S_1$	$\Delta S$	$\Delta S/S_0$	# scans
2 ms	38.9773	27.5980	11.3793	29.2%	14214
8 ms	9.0486	7.9888	1.0598	11.2%	995
16 ms	2.5922	1.1149	1.4773	57%	923
24 ms	7.1371	5.6632	1.4739	20.1%	1037
40 ms	7.1061	6.5258	0.5803	8.1%	3864
48 ms	38.9894	27.6350	11.3544	29%	14214