

EXPRESSION, PURIFICATION AND FUNCTIONAL CHARACTERIZATION OF  
INFLUENZA HEMAGGLUTININ (HA2) MEMBRANE FUSION PROTEIN

By

Punsisi Upeka Ratnayake

A DISSERTATION

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

Chemistry-Doctor of Philosophy

2015

## ABSTRACT

### EXPRESSION, PURIFICATION AND FUNCTIONAL CHARACTERIZATION OF INFLUENZA HEMAGGLUTININ (HA2) MEMBRANE FUSION PROTEIN

By

Punsisi Upeka Ratnayake

Influenza virus causes substantial public health risk worldwide. Influenza is an enveloped virus and hemagglutinin membrane protein present in the viral membrane plays an important role in the viral infection process. Hemagglutinin protein is composed of two subunits called HA1 and HA2. Binding of the virus to the host cell is governed by HA1 subunit, and HA2 subunit is responsible for the fusion of viral membrane and endosomal membrane.

My research has focused on production and characterization of several protein constructs containing different domains (with or without FP or with or without TM) and the characterization of the full length HA2. Biophysical comparisons between full-length HA2 and shorter constructs including SHA2, FHA2, and SHA2-TM were performed. Biophysical characterization techniques such as CD spectroscopy, SEC, crosslinking experiments, mass spectrometry and vesicle fusion assays were used in this study. Physiologically relevant oligomeric states of these hemagglutinin constructs were identified. At pH 7.4, the physiological pH, these constructs are trimeric helical molecules in detergents. The melting temperature for full length HA2 is  $> 90$  °C in decylmaltoside (DM) representing a highly thermostable structure. All constructs are positively charged at pH 5.0 and induce vesicle fusion with negatively-charged vesicles. However, fusion with negatively charged vesicles at pH 7.4 was negligible. With positively charged vesicles, pH-dependence was reversed leading to conclude that attractive protein/vesicle electrostatics play a role in fusion between vesicles and hemagglutinin constructs.

My work will help future scientists set up crystallography experiments since this is the first time that full length HA2 has been produced in mg quantities using bacterial expression system. Furthermore, it is hypothesized that in the final stage of the fusion of full length HA2 constructs as well as full length HIV gp41 constructs there is a formation of complex in between FP and TM. But, still there is no experimental evidence for support this hypothesis. HDX experiment completed and proposed in this dissertation will help to investigate this hypothesis.

Copyright by  
PUNSISI UPEKA RATNAYAKE  
2015

Dedicated to my loving mother Mrs.S.A.K Perera, father Mr.Gamini Ratnayake and my husband  
Dr.E.A Prabodha Ekanayaka,

## 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 without showing any hesitation what I am going to do. I highly appreciated his patience and encouragement.

I would like to acknowledge Dr. A. Daniel Jones who guided me to join with Dr. Weliky's research group at the start of my graduate life at MSU. He has given me numerous support and guidance throughout my PhD research. In addition I thank my PhD guidance committee members including Dr. Jones, Dr. Dana Spence and Dr. Kevin D. Walker for their support.

I am grateful for the past and present members of the Weliky group. Dr. Kelly Sackett helped me to adapt to the lab during my first year of studies and I would like to appreciate his kindness and helpfulness. I would like to thank Dr. Koyeli Banerjee for teaching me basic biochemical and molecular biology techniques in the lab. I am grateful to Dr. Erica Vogel, Dr. Li Xie and Dr. Matthew J. Nethercott for useful discussion regarding our research. I would like to thank present group members of the lab, Dr. Charles Gabrys, Ujjayini Ghosh, Lihui Jia, Shuang Liang, Robert Wolfe and Ahinsa Ranaweera for being so friendly and helpful. They all made my graduate life with a journey fulfilled with fun and laughter.

My husband Prabodha's constant encouragement and support motivate me to be optimistic. His guidance, love and kindness help me to jump out of difficult situations. I had a great family that was always a great source of encouragement. My mother and father's dedication, good wishes, blessings and encouragement helped me move forward in challenging circumstances.

I would like to thank Dr. Jian Hu in biochemistry for useful discussions regarding crystallography. And I would like to thank his postdoctoral associate Dr. Tuo Zhang for helping me with the setting up crystallography trays. I would like to extend my thank for Dr. Lisa Lapidus, Dr. Kinshuk Raj Srivastava and Mr. Terry Ball for allowing to use the CD instrument and fluorimeter. They never restricted my usage of either of the instruments. They were willing to help me whenever I need them.

I would like to thank Dr. Sandhya Payankulam and Dr. Ramya Rajagopalan for helping me with PCR techniques. I would like to acknowledge Dr. Sundari Chodavarapu for showing me how to set up HDX experiment. I am grateful to Dr. Dilini Ratnayake for teaching me cloning experiments. I was blessed to have you all in MSU, whom become great source of knowledge too.

## TABLE OF CONTENTS

LIST OF TABLES.....	xii
LIST OF FIGURES.....	xiii
KEY TO ABBREVIATIONS.....	xxi
Chapter 1 Introduction to influenza hemagglutinin membrane fusion protein (HA2) .....	1
1-1 Introduction.....	<b>Error! Bookmark not defined.</b>
1-2 Flu and flu vaccine.....	3
1-3 Influenza virus .....	4
1-3-1 Influenza virus life cycle.....	6
1-3-2 Endocytosis path way .....	7
1-3-3 Structural studies of hemagglutinin .....	9
1-3-4 Proposed mechanism for influenza membrane fusion.....	12
1-4 Membrane proteins .....	16
1-5 Influenza fusion peptide.....	17
1-6 Transmembrane domain.....	19
1-7 Antiviral drugs .....	21
1-7-1 Neuraminidase Inhibitors.....	21
1-7-2 M2 Inhibitors .....	22
1-7-3 Fusion inhibitors .....	23
REFERENCES.....	24
Chapter 2 Materials and Methods.....	30
2-1 Introduction.....	31
2-2 Molecular subcloning.....	33
2-2-1 Basics steps involved in subcloning .....	35
2-3 Protein expression.....	41
2-3-1 Lac operon .....	42
2-3-2 pet vector system.....	45
2-3-3 Inclusion bodies .....	45
2-4 Solubilization and purification of expressed protein .....	48
2-4-1 Immobilized metal affinity chromatography (IMAC).....	50
2-4-2 Ion exchange chromatography.....	51
2-5 SDS PAGE (Sodium Dodecyl Sulphate Poly Acrylamide Gel Electrophoresis) .....	52
2-6 Crosslinking experiment .....	53
2-7 Gel filtration chromatography (SEC-Size Exclusion Chromatography) .....	54
2-8 CD spectroscopy .....	58
2-9 Western blots .....	60
2-10 Lipid mixing assays .....	61

APPENDIX.....	64
REFERENCES.....	66
Chapter 3 Expression, purification and functional characterization of full length HA2 and shorter constructs with or without fusion peptide or transmembrane domain.....	69
3-1 Introduction.....	70
3-2 Materials and Methods.....	73
3-2-1 Abbreviations.....	73
3-2-2 Constructs, bacterial culture, and protein expression .....	75
3-2-3 Protein expression.....	79
3-2-4 RP solubilization and purification .....	79
3-2-5 IMAC purification of RP .....	80
3-2-6 RP Identification using western blot.....	81
3-2-7 RP Identification using Mass spectrometry .....	81
3-2-8 Detergent exchange and refolding .....	82
3-2-9 Circular dichroism (CD) .....	82
3-2-10 Cross-linking.....	83
3-2-11 Size-exclusion chromatography (SEC).....	83
3-2-12 Protein-induced vesicle fusion.....	84
3-3 Results.....	85
3-3-1RP solubilization and purification .....	85
3-3-2 RP confirmation by mass spectrometry.....	92
3-3-4 Hyperthermostable $\alpha$ helical structure .....	96
3-3-5 Trimers and higher-order oligomers at pH 7.4 .....	101
3-3-6 Vesicle fusion.....	107
3-4 Discussion.....	112
3-4-1 Significant findings.....	112
3-4-2 Full-length HA2 purified from E. coli .....	113
3-4-3 Dominant thermostable trimer supported by earlier biophysical studies.....	114
3-4-4 SHA2 aggregation in DM .....	115
3-4-5 Viral fusion overview .....	116
3-4-6 Common fusion features.....	118
3-4-7 HIV vs influenza fusion.....	121
3-5 Conclusions.....	122
REFERENCES.....	123
Chapter 4 Investigation of pH-dependent vesicle fusion induced by the ectodomain constructs of human immunodeficiency virus membrane fusion protein gp41 .....	129
4-1 Introduction.....	130
4-2 Materials and Methods.....	135

4-2-1 Materials. ....	135
4-2-2 Protein-induced vesicle fusion.....	136
4-3 Results.....	137
4-3-1 Vesicle fusion requires attractive electrostatic energy.....	137
4-3-2 Fast rate is strongly positively correlated with lipid charge .....	140
4-3-3 Slow extent is inversely dependent on lipid charge.....	140
4-3-4 Higher fusion for FP-HP.....	141
4-5 Discussion.....	145
4-5-1 Fast and slow fusion processes .....	145
4-5-2 Slow vesicle fusion as a model for HIV hemifusion .....	147
REFERENCES.....	150
Chapter 5 Characterization of full length HA2 using hydrogen deuterium exchange mass spectrometry (HDX) .....	154
5-1 Introduction.....	155
5-2 Experimental design.....	156
5-3 Reagent and Chemicals.....	158
5-4 LC-MS analysis of peptides generated from the pepsin digestion of FP, HA2 and FHA2 protein.....	159
5-4-1 Off-column digestion of fusion peptide and the proteins followed by LC-MS analysis .....	159
5-4-2 On-column digestion of the proteins followed by LC-MS analysis .....	159
5-4-3 Hydrogen – Deuterium exchange experiment .....	161
5-5 Results .....	163
5-5-1 Calculating Mass accuracy in ppm .....	166
5-7 Discussion.....	173
REFERENCES.....	175
Chapter 6 Sample preparation and preliminary studies for the crystallization of influenza full length HA2 and mini projects completed .....	177
6-1 Introduction.....	178
6-1-1 Stability test for the HA2 .....	179
6-1-2 Setting up the crystallographic trays.....	180
6-1-3 Purification of cell palette with other detergents .....	181
6-1-4 Extracting HA2 from membrane fraction .....	182
6-1-5 CD with protein purified from membrane fraction.....	183
6-2 Results and Discussion .....	183
6-3 Future work.....	187
6-4 Mini projects completed .....	188
6-4-1 Subcloning of FPHM, HMTM and FPHMTM.....	188

6-4-2 Purification of HM and FPHM .....	191
6-5 Results and Discussion .....	193
6-6 Future Work.....	193
REFERENCES.....	195

## LIST OF TABLES

Table 1-1: DNA size resolution based on agarose content in the gel. Depending on the size of the vector/gene (insert) user has to decide which percentage of agarose gel has to be used for the preparative agarose gels.....	36
Table 4-1: Best-fit parameters of vesicle fusion.....	141
Table 5-1: Solvent gradient used for the separation of pepsin digestion products of proteins on the C18 analytical column. ....	161
Table 5-2: Table showing the sequence of the predicted pepsin digestion products for FP and their accurate masses. Yellow color highlighted columns showing the peptides that were identified by the experiment. ....	164
Table 5-3: Preliminary results of HD exchange experiment (ND – Not Determined). Due to instrument unavailability, 0.5 min, 3 min and 24 hour time points for HA2 could not be performed.....	173
Table 6-1: Highlighted slots show the spots that showed the crystals.....	181

## LIST OF FIGURES

Figure 1-1 Basic structure of influenza virus.....	5
Figure 1-2 Influenza viral life cycle.....	6
Figure 1-3 An X-ray crystallography structure of trimeric hemagglutinin containing HA1 and HA2 ectodomain at pH 7.5 (PDB entry 1rd8.) This structure includes residues 4-328 of HA1 and 1- 175 residues of HA2. HA1 is in yellow color and HA2 is colored in blue [2] .....	10
Figure1-4 (A) displays a schematic of the HA2 constructs with domains colored: fusion peptide (FP), red; soluble ectodomain (SE), blue; transmembrane (TM) domain, green; and endodomain (Endo), yellow. (B) The amino acid sequence of the full length HA2 .....	10
Figure 1-5 (A) X-ray crystallography structure of monomeric hemagglutinin HA2 ectodomain at pH 7.5. This structure includes residues 1- 175 of HA2. In this structure fusion peptide is buried inside the HA2. HA1 was omitted for the clarity. (B) X-ray crystallography structure of monomeric HA2 at ~ pH 5. Dramatic structural reorganization of HA2 molecules moves N-terminal part from the trimeric core towards the opposite end and relocates the C-terminal part to the same end of the molecule [2, 6] .....	11
<p><b>(a)</b> Hemagglutinin protein exists in viral membrane. "rb" in (a) is receptor binding domain (HA1) and it is not shown in diagrams beyond figure (1-6(a)). Initially HA1 domain of the hemagglutinin interacts with sialic acid present in host cell and endocytosis of the virus happens. <b>(b)</b> Decrease of endosomal pH to the 5-6 range causes conformational changes to trimeric HA2 protein and movement of FP from trimeric core towards endosomal membrane. HA2 forms a long extended structure which is known as perhairpin intermediate (PHI). <b>(c)</b> Multiple trimers come close and initiate fusion of endosomal membrane and viral membrane. <b>(d)</b> Protein begins to fold allowing N-and C-terminal regions to come close together <b>(e)</b> Formation of six-helix bundle brings both membranes to close proximity and mixing of outer leaflets (hemifusion). <b>(f)</b> Finally, the fusion pore formation followed by release of viral genome to host cell happens. Even though mechanism is proposed as above, high resolution crystal structures are available only for the structures in (a) and (f) [3]. .....</p>	15
Figure 1-7 Backbone structural models of (A) open (boomerang structure) of IFP20 of HA3 (B) closed IFP23 of HA1, and (C) semiclosed IFP23 of HA1. C, N, O atoms are respectively represented by green, blue, and red verticals. The dash lines between 2 chains are between N atom of F9 and CO of G16 with the distances $r_{\text{opened}} = 11.5\text{\AA}$ , $r_{\text{closed}} = 3.9\text{\AA}$ and $r_{\text{semiclosed}} = 5.5\text{\AA}$ . .....	19
Figure 1-8 Diagrams showing the activity of the Neuraminidase inhibitors. NA inhibitors bind with NA protein inhibiting its activity. Therefore the daughter viruses continue to attach to the host cell since they cannot be released. ....	22

Figure-2-1 Flow chart showing basic steps involved in characterization of a protein. These steps are briefly discussed below.....	32
Figure 2-2-Diagram showing the steps of sub-cloning.....	39
Figure 2-3 (A) Picture of 0.5% preparative agarose gel run for the separation of cut pet24a+ vector. (B) Picture of 1% preparative agarose gel run for the separation of insert (gene of interest) from pUC57-Kan vector. ....	40
Figure 2-4 Diagram showing the lac operator, ribosomal binding site (rbs), multiple cloning region (Nde1-Xho1), and His tag of the pet24a+ vector. ....	40
Figure 2-5-The graph showing the four phases of <i>E.coli.</i> growth .....	42
Figure 2-6 Diagram showing the functionality of <i>Lac</i> operon (a) In the absence of lactose (b) In the presence of lactose [1].....	44
Figure 2-7 (A) Transmission electron microscopy pictures of uninduced bacterial cells (B) Transmission electron microscopy pictures of bacterial cells induced to produce FHA2 protein (After 5 hours of induction time). ....	47
Figure 2-8 (A) Structure of Bis (sulfosuccinimidyl) suberate (B) NHS ester reaction scheme for chemical conjugation to a primary amine in protein. (R) Represents one end of a crosslinker having the NHS ester reactive group; (P) represents a protein that contains the target functional group (primary amine). The reaction forms amide bond with protein and cross linker releasing N-hydroxysuccinimide (NHS). ....	54
Figure 2-9 Mechanism of gel filtration.....	56
Figure 2-10 The plot showing SEC of $A_{280}$ of MW standards loaded on Superdex 200 columns. The protein standards used were Thyroglobulin (669 kDa), Ferritin (440 kDa), Aldolase (158 kDa), Conalbumin (75 kDa), Ovalbumin (44 kDa), Carbonic Anhydrase (29 kDa), Ribonuclease A (13.7 kDa) and Aprotinin (6.5 kDa). The void volume is represented by the $V_0$ . ....	57
Figure 2-11 The plot of $K_{av}$ Vs $\log_{10}$ (MW). $K_{av} = (V_e - V_0)/(V_c - V_0)$ where $V_e$ , $V_c$ , $V_0$ are the elution volume, column volume and void volume respectively. ....	57
Figure 2-12 Characteristic CD curves for different secondary structures of protein molecules [2]. ....	60
Figure 2-13 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. ....	62

Figure 2-14 Principle of fluorescence quenching assay. NBD excitation happens around 467 nm and emits around 530nm. When donor (NBD) and acceptor (Rhodamine) are in close 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. Therefore, increased fluorescence signal being observed..... 63

Figure 3-1 (A) Displays a schematic of the HA2 constructs with domains colored: fusion peptide (FP), red; soluble ectodomain (SE), blue; transmembrane (TM) domain, green; and endodomain (Endo), yellow. The SHA2-TM<sub>7</sub>, SHA2-TM<sub>14</sub>, SHA2-TM<sub>21</sub>, and SHA2-TM<sub>26</sub> constructs respectively include 7, 14, 21, and 26 residues of the transmembrane domain. (B) displays the amino acid sequence of the HA2 construct with color-coding matching (A). The sequence has a non-native C-terminal region in black which includes a H<sub>6</sub> tag for affinity chromatography. (C) DNA sequence of full length HA2 corresponding to amino acid sequence in figure 3-1(B) ..... 77

Figure 3-2 Amino acid sequences of (A) FHA2 (B) SHA2 (C) SHA2-TM<sub>26</sub> constructs with color coding matching Figure 3-1 (A). The non-native C-terminal regions are shown in black and include a H<sub>6</sub> tag for affinity chromatography. For constructs that include the TM region G<sub>6</sub> is added as a flexible region to increase H<sub>6</sub> exposure during chromatography. .... 78

Figure 3-4 HA2 affinity purification from soluble cell lysate using 0.5% SRC..... 87

Figure 3-5 SHA2 affinity purification with PBS buffer. .... 87

Figure 3-6 SDS-PAGE of purified fractions for SHA2-TM<sub>21</sub>.The cells were first lysed in PBS followed by centrifugation to separate the insoluble material enriched in RP from soluble fraction enriched in bacterial proteins. The procedure was repeated three times and the pellet was then solubilized with the buffer containing either (A) Urea; or (B) 0.5% (w/v) sarkosyl (SRC) with subsequent Co<sup>2+</sup> affinity purification..... 89

Figure 3-7 SDS-PAGE of affinity purified HA2 . The cells were first lysed in PBS followed by centrifugation to separate the insoluble material enriched in RP from soluble fraction enriched in bacterial proteins.The procedure was repeated three times and the pellet was then solubilized with the buffer containing either (A) 0.17% (B) 0.1% DDM (C) 0.1% DPC with subsequent Co<sup>2+</sup> affinity purification..... 90

Figure 3-8 SDS-PAGE of purification eluents for SHA2-TM<sub>26</sub> with either a (A) LEH<sub>6</sub> or (B) G<sub>6</sub>LEH<sub>6</sub> terminal tag. Sequential 0.5 mL elutions are shown in (A) lanes 2 and 3 and (B) lanes 2-5. Intermediate lanes in gel A were cropped for clarity..... 92

Figure 3-9 (A) Peptide mapping via trypsin digestion and mass spectrometry; and (B) analysis of liquid chromatography-electrospray ionization mass spectra. The mapping in panel A was done for the most intense monomer band of the SDS-PAGE of the purification eluent (Figure 3-3).

The underlined regions of a sequence correspond to detected peptides and the percent sequence coverages are: HA2, 81%, SHA2-TM<sub>26</sub>, 55%; FHA2, 77%; and SHA2, 65%. The color coding of the sequences matches Figure 3-1. The experimental peak masses in panel B and expected masses are: HA2, 26678 and 26676 Da; SHA2-TM<sub>26</sub>, 23663 and 23663 Da; FHA2, 22379 and 22378 Da; and SHA2, 20444 and 20444 Da. The ESI mass spectra are presented in figure 3-10.

..... 94

Figure 3-10 Mass Spectra of (A) HA2 (B) SHA2-TM<sub>26</sub> (C) FHA2 (D) SHA2 after liquid chromatography and electrospray ionization. .... 95

Figure 3-11 Circular dichroism spectra of samples containing ~20 μM protein in buffer with 10 mM Tris-HCl, 0.17% DM detergent, and 1 mM DTT at pH 7.4. The spectra were obtained at ambient temperature. .... 97

Figure 3-12 Circular dichroism spectra of samples containing ~ 20 μM protein in (A) 0.17% DM at pH 9.0 and (B) 0.17% and 0.30% DM at pH 7.4. .... 97

Figure 3-13 Thermal melts of samples containing ~20 μM protein in buffer with 10 mM Tris-HCl, 0.17% DM detergent, and 1 mM DTT at pH 7.4. Temperature is plotted Vs the mean-residue molar ellipticity at 222 nm. .... 99

Figure 3-14 Circular dichroism spectra at pH 7.4 in 0.17% DM detergent at temperature 20 °C-85 °C. .... 100

Figure 3-15 (A) SDS-PAGE after chemical cross-linking and (B) SEC of the proteins in the absence of cross-linking. Solutions contained either (top) 0.17% DM or (bottom) 0.10% SRC detergent at pH 7.4. Cross-linking was done with pure proteins (Figure 3-3) using bis (sulfosuccinimidyl) suberate and the displayed SDS-PAGE are 8% gels with band diffuseness at low MW's. The trimer region is marked. SEC was done with *A*<sub>280</sub> detection and ~0.8 mg protein/mL loading. The vertical arrows in the plots mark the elution times of the MW standards with *V*<sub>o</sub> ≈ column void volume. The SEC of the MW standards is provided in Figure 3-16. A fraction of protein molecules are typically in oligomeric states larger than a trimer as evidenced by diffuse high MW intensity in SDS-PAGE and by *A*<sub>280</sub> intensity at shorter SEC elution times. .... 104

Figure 3-16 SEC (A) plot and (B) analysis of MW standards. The SEC was performed with a Superdex 200 column with *A*<sub>280</sub> detection. For each standard, *K*<sub>av</sub> was calculated as  $[(V_e - V_o)/(V_c - V_o)]$  using *V*<sub>e</sub> ≡ elution volume of the standard, *V*<sub>o</sub> ≡ column void volume, and *V*<sub>c</sub> ≡ column volume. The standards include the thyroglobulin (669 kDa), ferritin (440 kDa), aldolase (158 kDa), conalbumin (75 kDa), Ovalbumin (44 kDa), carbonic anhydrase (29 kDa), ribonuclease A (13.7 kDa), and aprotinin (6.5 kDa). *V*<sub>o</sub> corresponded to blue dextran with MW = 2 MDa. .... 105

Figure 3-17 SDS PAGE of protein after one hour of cross-linking. The faint marked bands are assigned as protein monomer. Cross-linked oligomers are represented by the diffuse bands with MW's typically >60kDa. The bands of oligomers of different MW's are not resolved. This is a 15% gel. .... 106

Figure 3-18 Protein-induced fusion of negatively-charged vesicles at pH 5.0. The protein:lipid mole ratio = 1:280. The vesicle composition is POPC:POPG (4:1). For each trace, protein is added a few seconds before the first displayed point. There is typically < ±1% variation in final extent of vesicle fusion among replicate assays and no fusion is observed with addition of detergent solution without protein ..... 108

Figure 3-19 Protein-induced fusion of negatively-charged vesicles at pH 5.0 (A) different SHA2-TM constructs (B) dose response for FHA2. The protein: lipid mole ratio for figure (A) = 1:280. The vesicle composition is POPC:POPG (4:1)..... 108

Figure 3-20 Protein-induced fusion of (A) negatively-charged vesicles at pH 7.4 and (B) positively charged vesicles at pH 5.0 and pH 7.4. (C) dose response for SHA2 at pH 5.0. The protein:lipid mole ratio = 1:280 in panel(A) and (B). The vesicle composition is (A) POPC : POPG (4:1) and (B) POPC : DOTAP (4:1). For each trace, protein is added a few seconds before the first displayed point. There is typically < ±1% variation in final extent of vesicle fusion among replicate assays and no fusion is observed with addition of detergent solution without protein. .... 110

Figure 3-21 HA2-induced fusion for POPC-only and POPC:POPG (4:1) vesicles with protein:lipid = 1:140. Replicate assays consistently showed at least twice the fusion for PC:PG vesicles relative to PC-only vesicles at 50 s after protein addition. Because of instrument breakage, this vesicle fusion assays were done with a different fluorimeter than the rest of the assays of this study. Because this fluorimeter had lower sensitivity, we needed to increase the integration and sampling times to 3 s. .... 112

Figure 3-22 Model of trimeric hairpin HA2 prior to fusion with the FP and TM regions in (A) the same membrane or (B) different membranes. The two membranes would correspond to: (1) two vesicles; (2) HA-cell and RBC; or (3) virus and endosome. For cases (2) and (3), the TM is respectively in the HA-cell and the viral membrane. The color coding is the same as Figure 3-1 with FP (red), SE (blue), TM (green), and Endo (yellow). A single HA2 trimer is shown with structural elements: FP, helix-(tight-turn)-helix; SE, trimer-of-hairpins; and TM, helix. For either fusion type at pH ~5, the SE binds to both membranes in part because of attractive electrostatic energy between the positively-charged protein and negatively-charged membrane. The resultant membrane apposition and membrane perturbation aid catalysis of the subsequent membrane fusion..... 119

Figure 4-1 (A) The diagram showing the ectodomain of the HIV gp41 protein including fusion peptide (FP), N-heptad repeat (NHR), loop, C-heptad repeat (CHR), and membrane-proximal

external region (MPER). (B,C) Amino acid sequences of the hairpin (HP) and FP-hairpin (FP-HP) proteins with color-matching with panel A. The sequences of HP and FP-HP are from the HXB2 laboratory strain of HIV-1. According to the gp160 numbering HP includes NHR [(535-581(M535C))-[SGGRGG]-CHR-[628-666]. FP-HP includes [512-581 (M535C))-[SGGRGG]-[628-666]. The first bracket for HP includes much of the NHR for HP and the first bracket for FP-HP includes the full FP and much of the NHR. For both constructs, the third bracket includes much of the CHR. The intervening nonnative loop (SGGRGG) is shorter than ~20 residue native loop but still allows formation of stable hairpin with antiparallel NHR and CHR helices. [1-3]

..... 131

Figure 4-2 Representative time-courses of protein-induced vesicle fusion for protein:total lipid mole ratio = 1:700 at pH (A) 3.5 and (B) 4.0. Chol is not considered in the total lipid quantity. Best-fits are displayed in black. .... 139

Figure 4-3 (A) Buildup rate and (B) final extent best-fit parameters at pH 3.5. Each displayed parameter value and uncertainty is respectively the average and standard deviation among three replicates with separate fitting of each trial. The fitting uncertainty of a parameter for an individual trial is typically <1% of the best-fit value. For PC:PG:Chol = 5:5:5 and 8:2:5 and PC:PS:Chol = 5:5:5, the fitting model is a sum of a fast and a slow exponential-buildup while for other compositions, the model is a single slow buildup. Choice of fitting model was based on visual inspection of goodness-of-fit. .... 143

Figure 4-4 Plots of (A)  $k_f$  vs  $|(\text{charge/protein}) \times (\text{charge/lipid})|$  and (B)  $M_s$  vs  $(\text{charge/lipid})^{-2}$ . The positive correlation of plot A supports the fast process resulting from release of electrostatic energy from protein/vesicle binding. Plot B supports an inverse correlation between the extent of slow fusion and inter-vesicle repulsion. Plot A displays data for HP at pH 3.5 and 4.0 and plot B displays data for HP at pH 3.5. .... 144

Figure 4-5 Qualitative model of FP-HP-induced fusion of anionic vesicles with FP-HP color coding matched to Fig. 4-1. Monomers from solution bind to the vesicles with release of electrostatic energy that results in fast fusion. The bound monomers also diffuse on the vesicle surface and self-associate as small oligomers with formation of an intermolecular antiparallel FP  $\beta$  sheet which is associated with slow fusion [5]. .... 149

Figure 5-1 Diagram showing the basic procedure used for the HD exchange experiment [4]... 158

Figure 5-2 H-D exchange apparatus used for the experiments [6]. .... 162

Figure 5-3 Amino acid sequence (27 residues) of the fusion peptide used for the preliminary identification of peptides resulted from FP region of the proteins. Nonnative G<sub>2</sub>K<sub>4</sub>G<sub>1</sub> tag present in the C-terminus was added for the purpose of solubilization and purification with HPLC..... 163

Figure 5-4 (A) Chromatogram obtained by separating the pepsin digestion peptides of FP on the C18 analytical column (B) Mass spectra corresponding to peaks a, b and c.....	165
Figure 5-5 (A) Extracted ion chromatogram for $m/z$ 852.46 from the chromatographic profile of pepsin digestion products of FHA2. (B) Spectrum corresponding to the chromatographic peak “ $c_{FHA2}$ ” observed in A. $m/z$ 852.4637 represents $[m+H]^+$ the signature peptide selected for FP in FHA2.....	167
Figure 5-6 (A) Extracted ion chromatogram for $m/z$ 852.46 from the LC/MS analysis of pepsin digestion products of HA2. (B) Spectrum corresponding to the chromatographic peak “ $c_{HA2}$ ” observed in A. The peak at $m/z$ 852.4614 observed here represents the $[M+H]^+$ of the signature peptide present in HA2. 54 Since the signature peptide for FP was observed in FHA2 and HA2, the next step was to perform the deuteration and characterize the HD exchange rates of the different proteins under different time points. ....	168
Figure 5-7: Deuterium incorporation into the signature peptide for FP in FHA2 after 0.5 min of exchange .....	169
Figure 5-8: Deuterium incorporation into the signature peptide for FP in FHA2 after 1.5 min of exchange .....	169
Figure 5-9: Deuterium incorporation into the signature peptide for FP in FHA2 after 3.0 min of exchange .....	170
Figure 5-10 Deuterium incorporation into the signature peptide for FP in FHA2 after 10.0 min of exchange. ....	170
Figure 5-11: Deuterium incorporation into the signature peptide of FP in FHA2 after 24 hour of exchange. ....	171
Figure 5-12: Deuterium incorporation into the signature peptide for FP in HA2 after 1.5 min of exchange. ....	172
Figure 5-13: Deuterium incorporation into the signature peptide for FP in HA2 after 10 min of exchange. ....	172
Figure 6-1 SEC $A_{280}$ profile of concentrated ( $\sim 8$ mg/mL) full length HA2 protein after keeping 3 days in room temperature.....	179
Figure 6-3 SDS PAGE showing HA2 purification from membrane fraction using 0.5% sarkosyl solubilization followed by affinity purification with 0.5% sarkosyl .....	184
Figure 6-4 SDS PAGE showing HA2 purification from membrane fraction using 0.17% DM solubilization followed by affinity purification with 0.17% DM .....	185

Figure 6-5 SDS PAGE showing HA2 purification from membrane fraction using 0.1% DDM solubilization followed by affinity purification with 0.1% DDM .....	185
Figure 6-6 SDS PAGE showing HA2 purification from membrane fraction using 0.04% LDAO solubilization followed by affinity purification with 0.04% LDAO.....	186
Figure 6-7 Circular dichroism of full length HA2 (~20μM) extracted from membrane fraction using 0.5% sarkosyl followed by dialysis in to buffer containing 10mM Tris-HCl, 0.17% DM at pH 7.4.....	187
Figure 6-8 Schematic diagram of the HIV-gp41 where FP; fusion peptide, NHR; N-heptad repeat, CHR; C-heptad repeat, MPER; MPER, TM; Transmembrane domain. Amino acid sequences given below color coded according to the different domains [2]. .....	188
Figure 6-9 SDS PAGE showing HM purification from cell palette using 0.8% sarkosyl+ 8 M urea solubilization followed by affinity purification with 0.8% sarkosyl+8 M urea.....	192
Figure 6-10 SDS PAGE showing FPHM purification from cell palette using 0.8% sarkosyl+urea solubilization followed by affinity purification with 0.8% sarkosyl+8 M urea.....	193

## KEY TO ABBREVIATIONS

HA	Hemagglutinin protein
NA	Neuraminidase protein
TM	Transmembrane domain
FP	Fusion peptide
PHI	Pre Hairpin Intermediate
CD	Circular Dichroism
GuHCl	Guanidine Hydrochloride
IB	Inclusion Bodies
SHB	Six-Helix Bundle
CHR	C-heptad repeat
NHR	N-heptad repeat
RP	Recombinant Protein
SEC	Size Exclusion chromatography
SSNMR	Solid State Nuclear magnetic resonance
HIV	Human Immunodeficiency Virus
LB	Lauria-Bertani
CMC	Critical Micelle Concentration
SDS	Sodium dodecyl sulfate
PAGE	Poly Acrylamide Gel Electrophoresis
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
MES	2-(N-morpholino) ethanesulfonic acid
POPC	1-palmitoyl-2-oleoyl- <i>sn</i> -glycero-3-phosphocholine
POPS	1-palmitoyl-2-oleoyl- <i>sn</i> -glycero-3-[phospho-L-serine]

POPG	1-palmitoyl-2-oleoyl- <i>sn</i> -glycero-3-[phosphor-rac-(1-glycerol)]
HDX	Hydrogen Deuterium Exchange
LCP	Lipidic Cubic Phase
DM	Decylmaltoside
DOTAP	1,2-dioleoyl-3-trimethylammonium-propane (chloride salt)
DTT	Dithiothreitol
SRC	<i>N</i> -lauroylsarcosine
IPTG	Isopropyl $\beta$ -D-1-thiogalactopyranoside
IMAC	Immobilized Metal Affinity Chromatography
SE	Soluble Ectodomain
Chol	Cholesterol
DDM	Dodecylmaltoside
DPC	Dodecylphosphocholine
<i>N</i> -NBD-DPPE	N-(7-nitro-2, 1, 3-benzoxadiazol-4-yl) (ammonium salt) dipalmitoylphosphatidylethanolamine
<i>N</i> -Rh-DPPE	N-(lissamine rhodamine B sulfonyl) (ammonium salt) dipalmitoylphosphatidylethanolamine

**Chapter 1 Introduction to influenza hemagglutinin membrane  
fusion protein (HA2)**

## 1-1 Introduction

Influenza virus poses a significant threat to human health since this is the virus that causes seasonal flu and various severe forms of it such as swine flu. According to WHO (world Health Organization) influenza infects 3-5 million people causing 250000-500000 deaths each year worldwide [1]. According to CDC (Center for Disease Control) in United States there are about 200000 hospitalization and 36,000 deaths reported annually. In USA for example, recent estimates put the cost of influenza epidemics to the economy at US\$ 71-167 billion per year [5].

There were four major influenza pandemics in history. In 1918 flu pandemic is known as Spanish flu and 500 million people were infected from it causing 50 million deaths across the world. Flu pandemics were recorded in 1957, 1968 and 2009 years too. The influenza virus mutates rapidly and as a result each year, people get infected by a new strain of the virus. Therefore, despite having a vaccine for influenza, it continues to pose a substantial health risk to public worldwide. Influenza virus is an “enveloped virus” meaning that it has a viral membrane. Investigating the mechanism by which the influenza virus membrane fuses with host cell membrane aid the development of a model that explains the viral fusion with cellular membranes. Research discussed in this dissertation provides the information regarding the characterization of the full length influenza viral fusion protein – hemagglutinin (discussed in detail under section 1-3-3 and chapter 3). This research lays the foundation for further structural and functional studies about this protein that may lead to the development of influenza antiviral drugs. Having the structural details of the full protein (drug target) allows more specific drugs to be developed successfully as opposed to trying to develop drugs based on partial structural details. Obtaining structural details for such a successful drug target requires that the protein is purified in its physiologically relevant form with the physiologically relevant oligomeric state.

Research presented in this dissertation has accomplished both these requirements for influenza full length HA2 protein and therefore lays the foundation for future structure based anti-influenza drug design.

## **1-2 Flu and flu vaccine**

CDC identified children younger than the age of 5 years, adults of 65 year of age and older, pregnant women and people with certain medical conditions such as lung, heart, kidney disease and weakened immune system, as being the most vulnerable to influenza infections. During flu season 90% of deaths occur among people of age 65 years or older. Flu season in United States starts from early October and lasts until late May. Flu vaccine protects against two types of influenza A viruses (H1N1 and H3N2) and influenza B virus. Influenza A viruses subdivides into H1N1/H3N2 etc. based on the two membrane proteins present in the viral envelope namely hemagglutinin (HA) and neuraminidase (NA). To date, there are 18 different H (H1 to H18) and 11 different known N antigens (N1 to N11). This nomenclature of naming influenza viruses is based on the results of immune-double-diffusion tests (Schild et al., 1980). In immune-double-diffusion tests antibodies that are specific for hemagglutinin or neuraminidase are placed in the wells of an agarose gel. Then hemagglutinin and neuraminidase proteins are placed in another set of wells in the same agarose gel. Then these were left for 2 days. Diffusion of protein and the antibodies occurs during this period. When the two diffusion fronts meet, if the antibodies recognize and form an immune complex with their respective protein, a precipitate forms. Depending on the shape of the precipitate front nomenclature of Influenza A virus was assigned.

### 1-3 Influenza virus

Influenza virus belongs to the family of orthomyxovirus. It is an enveloped virus which obtains the envelope from the host cell by viral budding. Budding is a method for releasing new viruses from infected host cells. This process is used by enveloped viruses to acquire a host-derived membrane. According to mass spectrometry based lipid analysis of MDCK cells (Madin-Darby-canine kidney cells) infected with influenza virus showed that lipid envelop of influenza virus consisted of 52% cholesterol, 19% sphingolipids and 29% phospholipids based on mole percentages [7]. There are evidences suggesting “lipid rafts” (subdomains of plasma membrane that contains high concentrations of cholesterol and glycosphingolipids) act as microdomains where influenza virus budding takes place [8, 9]

Influenza virus can be filamentous, spherical or oval in shape. Spherical or oval shaped viruses have a diameter of 80-120 nm [10]. Viral envelope contains three proteins. These are hemagglutinin (HA), neuraminidase (NA) and, M2 channel. Hemagglutinin is a trimer and there are about 400 trimers per virus [2, 10]. Hemagglutinin is responsible for binding of virus to host cell and is involved in membrane fusion (This will be briefly discussed under section 1-3-4) [11]. Neuraminidase (NA) acts in later stage of the viral life cycle. Neuraminidase cleaves terminal sialic acids present in the host cell glycoproteins and glycolipids and helps daughter viruses to escape from host cell, allowing virus to spread [12]. Newly formed viruses are bound to the host cell through HA1 and sialic acid residues present in the host cell glycoprotein and glycolipids (figure 1-8). Therefore without the cleavage by NA, newly formed viruses will not be able to escape the host cell. M2 protein is a proton channel involved in the uncoating of the virus (uncoating is explained under section 1-3-2). Matrix protein (M1) lies underneath the viral envelope and forms a shell around the viral genetic material. Influenza virus has eight separate

RNA molecules and code for 11 proteins: hemagglutinin (HA), neuraminidase (NA), nucleoprotein (NP), M1 (matrix protein), M2, NS1, NEP (nuclear export protein also referred to as NS2), PA, PB1, PB1-F2 and PB2. RNA molecules are surrounded by ribonucleoprotein (RNP). Viral RNA polymerase proteins (PA, PB1, PB1-F2 and PB2) encoded by RNA segments are known to be active during mRNA transcription in the host cell. NEP protein in the interior of the virus is known to export viral RNAs from host cell nucleus and pack viral RNAs into newly formed viruses [13].

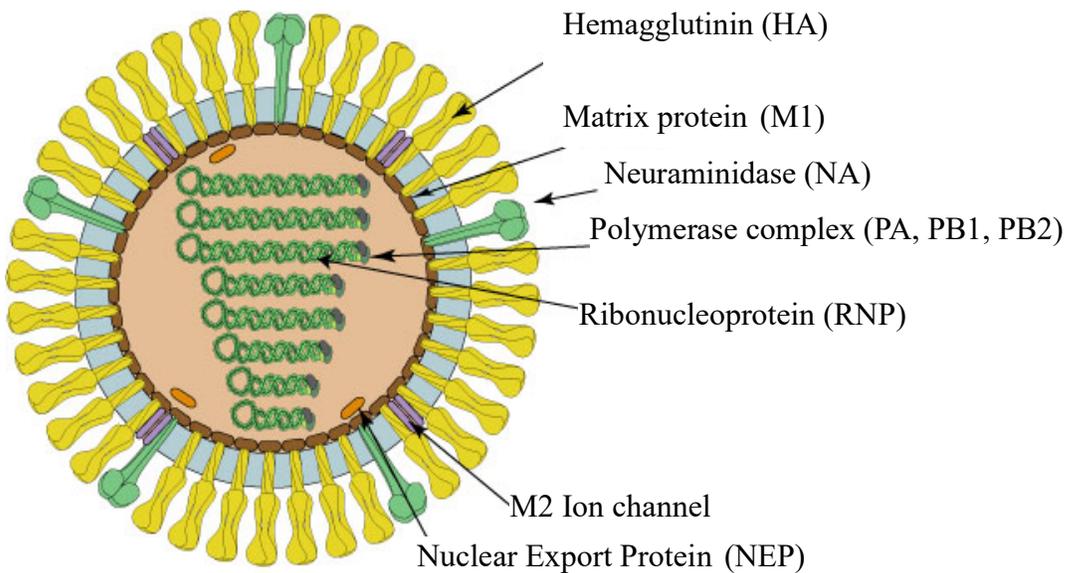
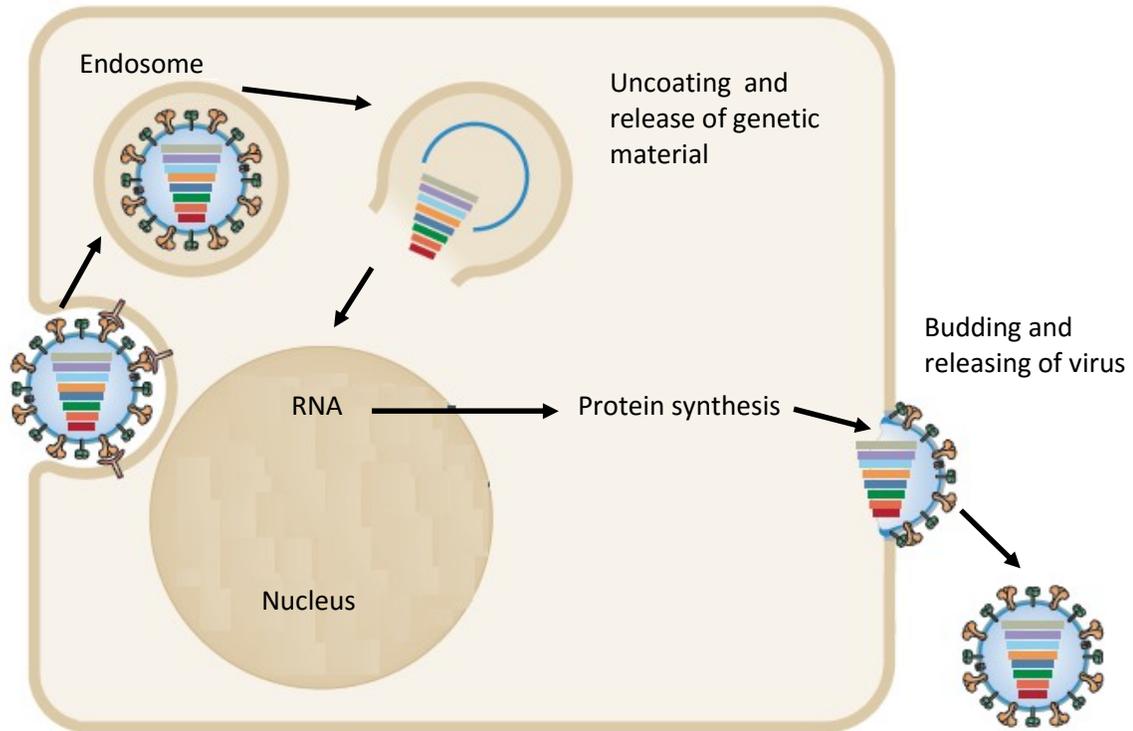


Figure 1-1 Basic structure of influenza virus.

### 1-3-1 Influenza virus life cycle



Binding of the virus to a sialic acid-containing glycolipids or glycoproteins present in the host cell epithelial cells results in the virus being endocytosed (figure 1-2 (a)). This happens through the interaction between HA1 subunit of hemagglutinin with the sialic acid present in glycolipids or glycoproteins of the host cell membrane [14]. Decrease of endosomal pH to the 5-6 range causes conformational changes to trimeric HA2 protein followed by the joining of viral and endosomal membranes (fusion). Through this fusion process viral RNA enters the host cell cytosol (figure 1-2 (c)). These viral RNA molecules are “negative sense” RNA molecules. The negative sense RNAs cannot be translated directly into protein. First it is transcribed in to positive sense RNA that can act as mRNA. This process is governed by viral RNA polymerase

enzyme. These mRNA molecules leave the nucleus (figure 1-2 (1d)) and translation of the mRNA into viral proteins happens in the host cell cytosol. Replication of viral RNAs required for the progeny virus happens in host cell nucleus (figure 1-2 (2d)). These viral RNA produced in the nucleus as well as the viral proteins produced in the cell cytosol (figure 1-2(1e), (2e)), are transported towards host cell plasma membrane. Finally, budding of new virus happens (figure 1-2 (f)) where host cell plasma membrane becomes the viral envelope.

### ***1-3-2 Endocytosis path way***

Endocytic pathway plays an important role in cholesterol homeostasis. LDL-derived cholesterol (low density lipoprotein derived cholesterol) enters the cells through endocytic pathway. Transportation of cholesterol from their entry point towards the endoplasmic reticulum also occurs through endocytic pathway [15].

Influenza virus uses endocytic pathway for the viral entry. Infection is initiated by binding of the HA1 subunit to the cellular sialic acid containing glycoproteins or glycolipids present in the host cell [14, 16]. The virus is then endocytosed. The early endosome forms after initial endocytosis has the pH ~ 6. Early endosome starts to migrate from plasma membrane towards the host cell nucleus. When viral particle continues its journey in the endocytic pathway, lowering of the pH to ~5.5 happens. These endosomes are now known as late endosomes. Late endosomes can fuse with the lysosome which contain hydrolase enzymes and have pH < 5 [17, 18].

This low pH of the late endosomes triggers conformational changes of HA2 protein. These conformational (discussed in section 1-3-4) changes initiate viral fusion with endosomal

membrane. But it is not clear whether the virus first fuses with the membrane of the late endosome or it is fuse with the membrane of the vesicle with in the endosomal lumen. After the binding it takes about 8 minutes until the virus reaches the late endosomal stage [18]. After fusion, viral RNPs are transported to host cell nucleus where transcription and replication happens. RNPs are not transported into the host cell nucleus when it is associated with M1 protein. Therefore dissociation of M1 and RNPs is crucial. M1 protein interacts with RNP predominantly by electrostatic forces. Lowering the pH weakens the electrostatic attraction between M1 protein and RNPs followed be by the dissociation of M1 from RNPs. This dissociation results the removal of M1 shell around the virus genetic material and this process is known as “uncoating”. After the uncoating viral RNPs leaves the virus and enters the host cell nucleus. [19-21].

To our knowledge there has been little detailed study of virus endosomal fusion and there are some studies with virus/vesicle fusion. Most of the studies are based on fusion between cells expressing HA and cells containing sialic acid such as red blood cells [18, 22, 23]

### ***1-3-3 Structural studies of hemagglutinin***

In influenza, mediation of viral fusion is performed by glycoprotein hemagglutinin (HA) (Skehel and Wiley, 2000). Hemagglutinin is expressed as a single polypeptide HA<sub>0</sub> and is composed of disulfide linked subunits HA1 and HA2 (Wiley and Skehel, 1987). There is a crystal structure of HA1 and HA2 ectodomain at pH 7.5 (figure 1-3) depicting a trimeric oligomerization of the HA protein [2]. HA1 has 328 amino acids and HA2 has 221 residues. HA2 contains (figure 1-4) ~25 residues N-terminal fusion peptide (FP), f~160 soluble ectodomain (SE), ~25 residue transmembrane domain (TM) and a ~10 residue C-terminus endodomain [24]. In this 1981 crystal structure, HA1 subunit is located at the top of the long coil-coiled HA2 subunit. HA1 consists of antiparallel  $\beta$ -sheets. The stem region (HA2 subunit) consists of three alpha helices. In this pre fusion structure N-terminal fusion peptide is buried inside the HA2 trimer (figure 1-5 (A)). The total distance from the top to the bottom of the crystal structure of the HA1/HA2 trimer is 135 Å [2].

The crystal structure of soluble ectodomain (SE) (residues 23-185) without HA1 domain at pH ~5 displays a trimeric hairpin structure. These trimeric hairpin molecules forms a bundle including residue 38-105 of N-terminal residues followed by an 180° turn with C-terminal residues which is packed antiparallell to the N-terminal region (figure 1-5 (B)).

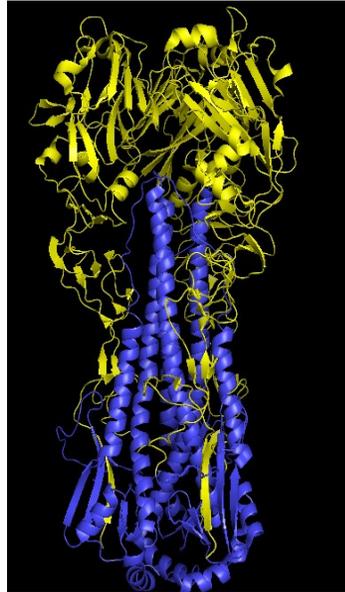


Figure 1-3 An X-ray crystallography structure of trimeric hemagglutinin containing HA1 and HA2 ectodomain at pH 7.5 (PDB entry 1rd8.) This structure includes residues 4-328 of HA1 and 1- 175 residues of HA2. HA1 is in yellow color and HA2 is colored in blue [2]

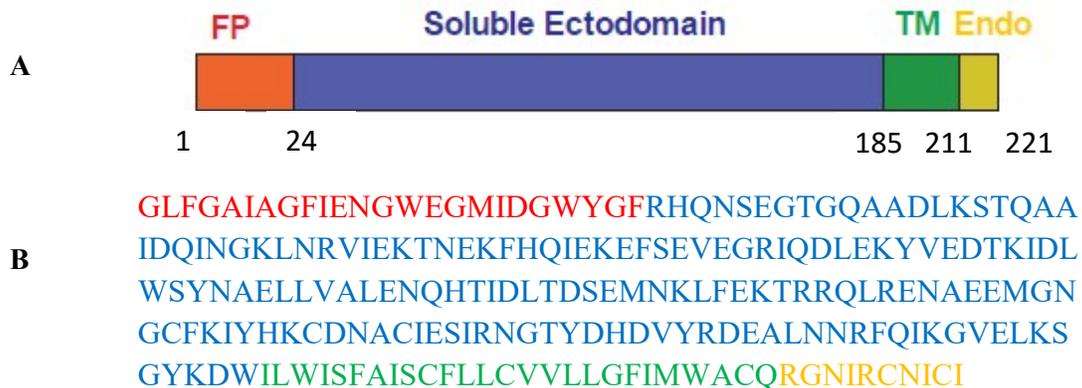


Figure1-4 (A) displays a schematic of the HA2 constructs with domains colored: fusion peptide (FP), red; soluble ectodomain (SE), blue; transmembrane (TM) domain, green; and endodomain (Endo), yellow. (B) The amino acid sequence of the full length HA2

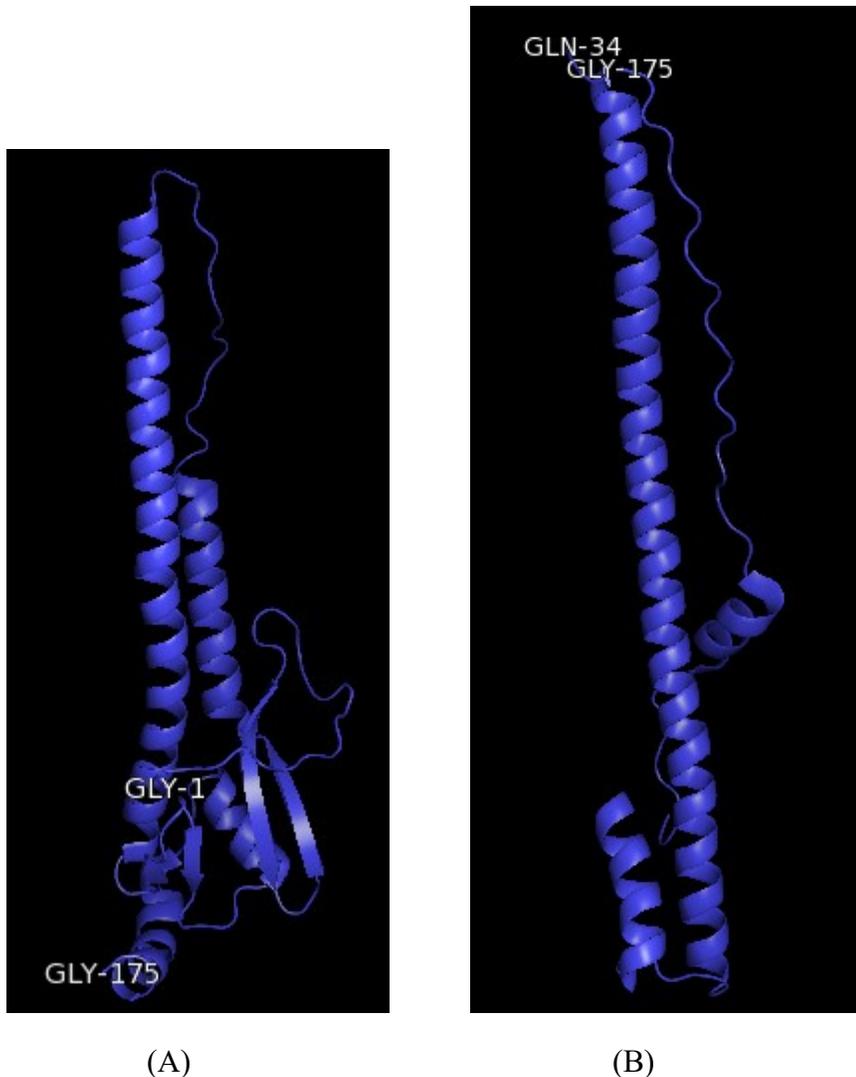


Figure 1-5 (A) X-ray crystallography structure of monomeric hemagglutinin HA2 ectodomain at pH 7.5. This structure includes residues 1- 175 of HA2. In this structure fusion peptide is buried inside the HA2. HA1 was omitted for the clarity. (B) X-ray crystallography structure of monomeric HA2 at ~ pH 5. Dramatic structural reorganization of HA2 molecules moves N-terminal part from the trimeric core towards the opposite end and relocates the C-terminal part to the same end of the molecule [2, 6]

### ***1-3-4 Proposed mechanism for influenza membrane fusion***

Figure 1-6 (a) depicts that the hemagglutinin protein attached to the viral membrane. In the figure “rb” is HA1 and three cylinders represent trimeric HA2 domain. The top membrane is the host cell membrane and bottom is the viral membrane to which hemagglutinin HA2 attaches. This 1981 crystal structure resembles this pre fusion structure of hemagglutinin, but that structure does not have the transmembrane domain and endodomain [2]. Influenza virus HA1 subunit binds to cell surface glycoproteins or glycolipids containing terminal sialic acids. Oligosaccharides ending with sialic acid are found on many cell surface glycoproteins and glycolipids. It has been demonstrated that the removal of sialic acid from these glycoproteins and glycolipids by NA treatment destroys the binding of the protein to the host cell. [16, 25]. Binding of HA1 subunit with sialic acid present in host cell glycoproteins or glycolipids causes endocytosis of the virus. Influenza virus that infects humans preferentially binds with sialic acids attached to galactose by  $\alpha$  (2, 6) linkage in host cell glycoproteins or glycolipids. This means carbon atom at position 2 of the sialic acid is joined via an oxygen atom to the carbon at position 6 of the hexose sugar. These 2, 6 linked sialic acids are most commonly found in upper respiratory tract of humans where viral replication starts. [4]. In this prefusion structure the fusion peptide is buried inside the core of the trimeric HA2 near the viral envelope (figure 1-5 (A)).

Decrease of endosomal pH to the 5-6 range causes conformational changes to trimeric HA2 protein and insertion of fusion peptide to endosomal membrane. This structure is known as prehairpin (PHI) intermediate (figure 1-6 (b)). It is believed structural rearrangement of HA2 propels the fusion peptide from the trimeric core to the opposite end of the molecule, allowing it to interact with the endosomal membrane (described in section 1-3-2).

PHI intermediate was studied using paramyxovirus PIV5 and lipid coated Nano beads. To trap the PHI a peptide called C-1 which binds to the PIV5 F protein was used. Binding of the C-1 peptide to the protein PIV5 F prevents the formation of the final hairpin structure and trap the PIV5 F protein in its PHI state. In the presence and absence of the C-1 peptide, the distance between the viral and Nano bead coated with bilayer was measured. In the absence of C-1, peptide fusion was observed between PIV5 virus and nano particle and virus/nano bead-supported bilayer was separated by a very short (~1 nm) distance. In the presence of the C-1, interparticle distance between the virus and the nano bead was 20 nm. In the prefusion structure of PIV5 F protein, it extended up to 12 nm from the viral membrane from top to bottom. Computer modeling studies predicted that bilayer separation (viral membrane and host cell membrane) is 21 nm revealing that PIV 5 F protein has to extend 21 nm if it forms the PHI structure. Since experimental results closely matches with the modeling studies, it is considered that this study was able to capture PHI state of virial fusion.

For gp 41 protein (fusion protein for HIV virus) the presence of PHI state is supported only by some functional studies. HIV infection is inhibited by NHR or CHR+MPER peptides. It is believed that these peptides bind with the CHR or NHR regions of PHI respectively inhibiting the formation of SHB [26, 27]. The epitope exposure studies have shown that dissociation of the three HA1 domains from each other can happen upon the lowering of pH. In this study, addition of several antibodies to the hemagglutinin protein at neutral pH and pH~5 was performed. Certain epitopes which were not accessible at neutral pH (epitopes which buried in the trimer interface such as HA1 98-106, 305-328) was exposed after low pH treatment [28-30].

In the next step, multiple trimers come close (figure 1-6 (c)) and protein begins to fold, allowing N- and C- terminal regions (figure 1-6 (d)) to come close together. This folding causes formation of a six-helix bundle which in turn brings two target membranes near each other. There is a crystal structure of SHB including residues 23-185. This crystal structure consists of three molecules each with hairpin structure. The six helix bundle formed by these hairpin molecules consists of an N-terminal residue 38-105 and is followed by a 180° turn with antiparallel packed C-terminus [6].

During the merging, if only the outer lipid layers of the two membranes mix, keeping the inner lipid layers apart, it is known as hemifusion. During hemifusion the contents of the two compartments do not mix (figure 1-6 (e)). Hemifusion barrier is next broken forming a fusion pore which results in the formation of a single lipid bilayer. The genetic material of the virus gets introduced into the host cell through the fusion pore.

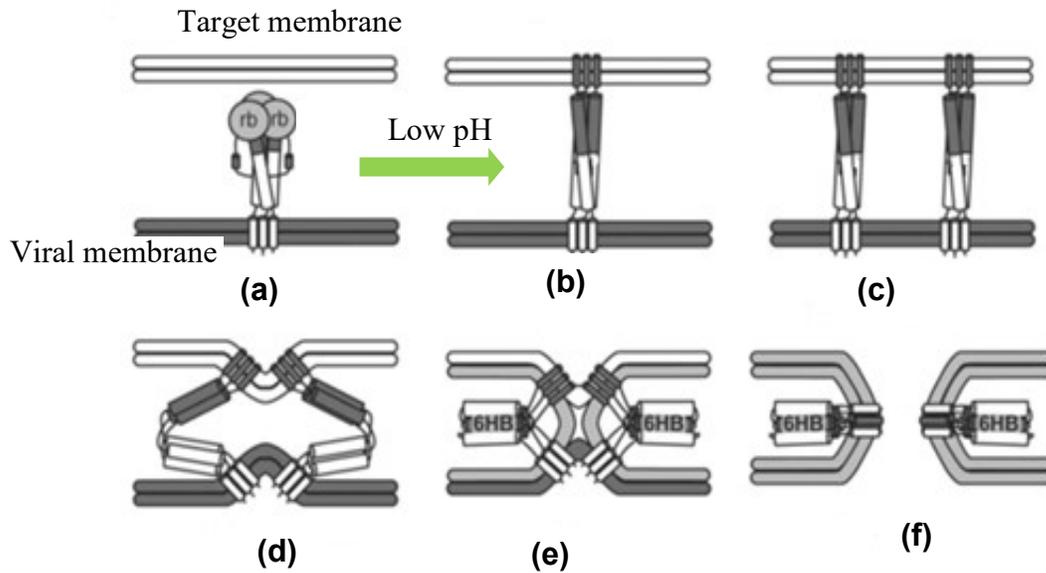


Figure 1-6 Diagram showing the proposed membrane fusion mechanism for hemagglutinin protein. **(a)** Hemagglutinin protein exists in viral membrane. "rb" in (a) is receptor binding domain (HA1) and it is not shown in diagrams beyond figure (1-6(a)). Initially HA1 domain of the hemagglutinin interacts with sialic acid present in host cell and endocytosis of the virus happens. **(b)** Decrease of endosomal pH to the 5-6 range causes conformational changes to trimeric HA2 protein and movement of FP from trimeric core towards endosomal membrane. HA2 forms a long extended structure which is known as perhairpin intermediate (PHI). **(c)** Multiple trimers come close and initiate fusion of endosomal membrane and viral membrane. **(d)** Protein begins to fold allowing N- and C-terminal regions to come close together **(e)** Formation of six-helix bundle brings both membranes to close proximity and mixing of outer leaflets (hemifusion). **(f)** Finally, the fusion pore formation followed by release of viral genome to host cell happens. Even though mechanism is proposed as above, high resolution crystal structures are available only for the structures in (a) and (f) [3].

## 1-4 Membrane proteins

HA2 is a membrane protein and this dissertation provides methods to perform cloning, expression and purification of this protein. Being a membrane protein poses several challenges on developing methods to characterize this protein as discussed below and in the following chapters.

In any organism, membrane proteins represent 20-30% of cell proteomes [31]. Membrane proteins perform many of functions in our body such as signal transduction, transporting substances and fusion of membranes [32-34]. Even though 40% of drugs target membrane proteins [35], only 2% of total crystal structures found in Protein Data Bank (PDB) represent crystal structures of membrane proteins [36, 37]. This is due to the inherent challenges involved in membrane protein characterization and production. (1) First, membrane proteins are of low abundance in their native biological systems. Recombinant protein expression techniques are used, to obtain a large quantity of a low abundance protein (discussed further in chapter 2). However, recombinant expression of membrane proteins in systems such as *E.coli* leads to their aggregation in cytoplasm resulting accumulation of incorrectly folded protein [38]. Such incorrectly folded protein cannot be used for structural studies since it does not represent the physiologically relevant structure. However, some of such proteins can be refolded into their biologically active forms [39]. (2) Membrane proteins are found embedded in cell membranes and therefore these proteins are constituted of domains that are inherently hydrophobic. Therefore, aqueous solutions cannot be used to isolate these proteins in the absence of detergents or denaturing agents. Use of ionic detergents (sarkosyl, sodium dodecyl sulfate) or denaturants (8M urea, 6M guanidine hydrochloride) cause change of the native protein folding [36].

## 1-5 Influenza fusion peptide

About 25 residues present in N-terminus of HA2 is known as influenza fusion peptide (IFP) and is vital in membrane fusion. FP is the most highly conserved region in the HA2 sequence [40]. Introducing mutations into IFP sequence resulted decrease in the fusion activity as well as complete loss of fusion, indicating the vital role of the FP region. These studies were based on the fusion between HA-cell and RBCs [41].

The structure of IFP in lipid membranes either can be  $\alpha$  helical monomer or  $\beta$  sheet structure depending on the composition of the membrane. Addition of cholesterol results in higher fractions of  $\beta$  sheet structure compared to  $\alpha$  helical structure [52]. The helical structure is predominant in membranes lacking cholesterol [53, 54].

Several IFP structural studies have been performed in the presence of detergent micelles. However, due to the presence of detergents, these studies are considered to be less physiologically relevant. Structures obtained from the membrane associated IFP are considered to be the more physiologically relevant. In a structural study using liquid state NMR with detergent micelles, it was reported 20 residue fusion peptide (IFP20) at pH 5.0 forms a N-terminal helix from residue 2-10 / turn (residue 11-12) / C-terminal helix from residue 13-18. At pH 7.4 the structure is N-terminal helix from residue 2-9/ turn (residue 11-12) / C-terminal extended structure from 14-20. At pH 5.0, the fusion peptide structure is considered to be an open boomerang due to the oblique interhelical angles. Using EPR measurements it was reported that, at pH 5.0, N- and C-terminal portions are able to penetrate more deeply into the membrane than at pH 7.4. At pH 5.0 spin labelled F3 residue of FP was  $\sim 4\text{\AA}$  from the bilayer center and at

pH 7.4 F3 residue was 10Å from the bilayer center. Han *et al.*, suggested that this deeper insertion probably perturbs the lipid molecules and facilitates the fusion [42].

In another structural study performed in the presence of detergents, using 23 residues N-terminal fusion peptide (IFP23) showed tightly packed anti-parallel N- and C- terminal helices at pH 4.0 and pH 7.4. Structure of IFP23 in detergent was a ‘closed hairpin’ structure with N-helix/turn/C-helix at both the pHs [43].

The structural studies performed by the Weliky group using SSNMR with IFP20 and IFP23 with lipid membranes showed a mixture of closed and semi-closed structures containing N-helix/turn/C-helix for both the pHs. IFP20 and IFP23 showed higher lipid mixing at pH 5.0 compared to pH 7.0. This is related to a larger fraction of semiclosed structure compared to closed structure at pH 5.0. Open interhelical geometry of semiclosed structure resulted in higher hydrophobic area and consequently increased of membrane perturbation and membrane fusion. Since IFP induces fusion between vesicles rather than the micelles, structures obtained using lipid membranes are more physiologically relevant compared to structures obtained using micelles.

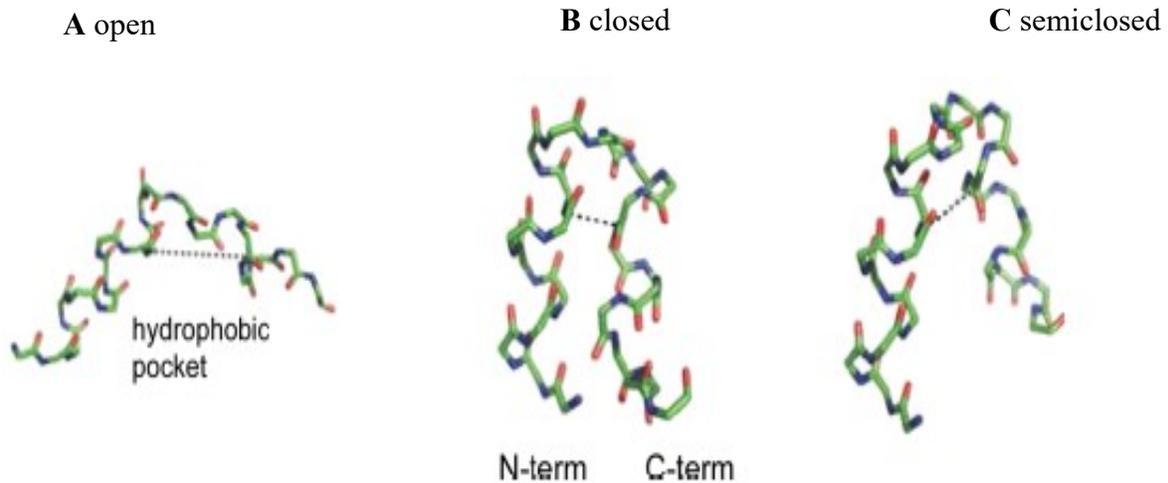


Figure 1-7 Backbone structural models of (A) open (boomerang structure) of IFP20 of HA3 (B) closed IFP23 of HA1, and (C) semiclosed IFP23 of HA1. C, N, O atoms are respectively represented by green, blue, and red verticals. The dash lines between 2 chains are between N atom of F9 and CO of G16 with the distances  $r_{\text{opened}} = 11.5\text{\AA}$ ,  $r_{\text{closed}} = 3.9\text{\AA}$  and  $r_{\text{semiclosed}} = 5.5\text{\AA}$  [57].

### 1-6 Transmembrane domain

Transmembrane domain anchors the HA2 fusion protein to the viral membrane. It is composed of ~25 amino acids and is considered as a single span transmembrane domain. Except for the second amino acid residue, leucine in the transmembrane domain, other amino acid residues are not highly conserved among different sero types of hemagglutinin [40] It is reported that synthetic peptide which corresponds to the TM domain of X31 strain of influenza virus adopted  $\alpha$  helical shape in detergent micelles and phospholipid bilayers [6, 44].

The role of transmembrane domain in viral fusion has been discussed and investigated to some extent. Replacing the TM and endo domain of HA with glycosylphosphatidylinositol (GPI) anchor could only promote hemifusion indicating that presence of TM contributes toward full fusion. This study was based on the cell-cell fusion of GPI-anchored HA cells and RBC. It was observed that GPI-anchored HA promotes only mixing of lipids but not the soluble content of the two cells. Mixing of only the lipids where no mixing of the soluble content occurs is known as hemi fusion. When both lipid membranes and the contents of the cells are mixed it is called full fusion. [22].

In another study, fusion activity of wild-type HA expressing cells with RBC was compared with that of modified HA expressing cells with RBC s. The modified HA consisted of a TM and an endodomain from a non-viral polyimmunoglobulin protein. The ectodomains of both proteins were the same. This study revealed that when TM and endodomain are derived from the same protein they are more effective in promoting fusion [45].

On the other hand, another group reported that sequences of the transmembrane domain or endo domain or their length is not crucial for membrane fusion activity [46]. Another group concluded that there is a length requirement for TM to display fusion activity, and to promote full fusion, TM must span the lipid bilayer [47].

## **1-7 Antiviral drugs**

### ***1-7-1 Neuraminidase Inhibitors***

In addition to HA, influenza viral envelope contains another protein known as Neuraminidase (NA). The function of NA proteins is the cleavage of sialic acid from glycolipids and glycoproteins present in the host cell and allowing daughter virus to bud out from the infected cell. Therefore NA is required for spreading the infection throughout the body and it is an effective drug target. Commercially available zanamivir and oseltamivir are two kinds of drugs active as reversible competitive inhibitors and interact with NA active site, reducing the viral spreading to other cells [55]. Tamiflu (oseltamivir phosphate), Relenza (zanamivir) and Rapivab (peramivir) are the three FDA-approved influenza antiviral drugs recommended by CDC for use against recently circulating influenza viruses (Figure 1-8).

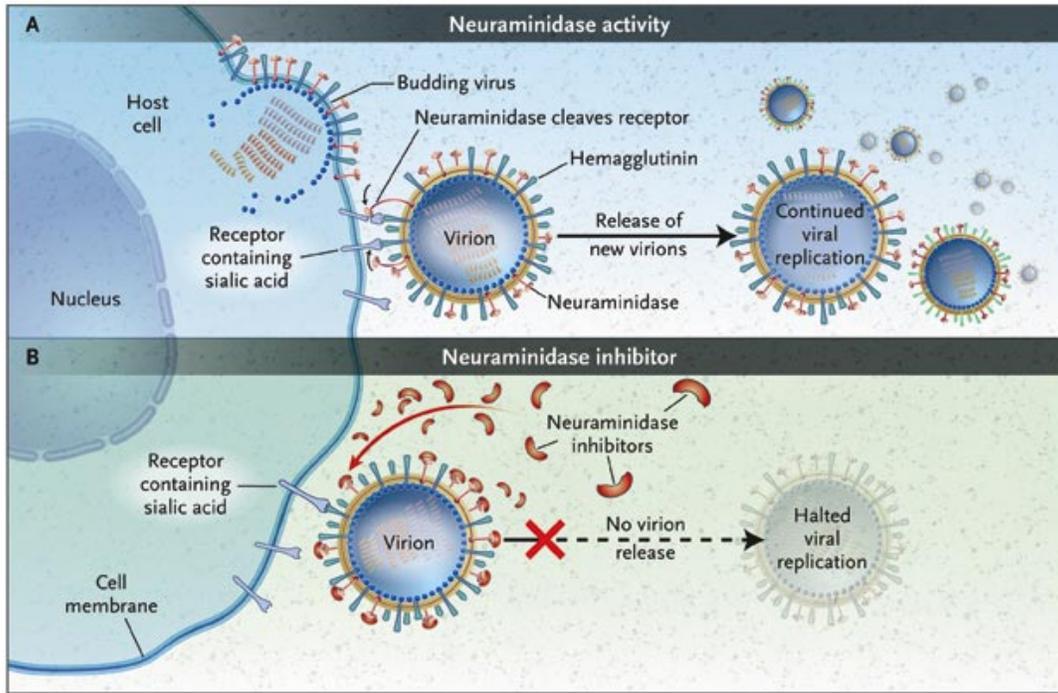


Figure 1-8 Diagrams showing the activity of the Neuraminidase inhibitors. NA inhibitors bind with NA protein inhibiting its activity. Therefore the daughter viruses continue to attach to the host cell since they cannot be released [56].

### 1-7-2 M2 Inhibitors

M2 protein is another membrane protein which present in low abundance and acts as an ion channel in the influenza virus envelope. This ion channel allows conduction of protons across the viral envelope resulting in uncoating (described in section 1-3-2) of the virus and release of viral content into the cytoplasm of the host cell. Commercially available M2 inhibitor drugs such as amantadine and rimantadine bind with transmembrane region of the M2 channel and sterically block the channel. Binding of the drug prevents passage of protons to the interior of the virus and

thereby inhibits the uncoating of the virus [48]. But certain strains of influenza such as the 2009 H1N1 influenza are now resistant to these two drugs. Therefore, CDC has not recommended the use of these two drugs for recently circulating influenza viruses.

### ***1-7-3 Fusion inhibitors***

Activation of M2 inhibitors and Neuraminidase inhibitors occurs after the fusion of viral and endosomal membrane. Since conformational changes of the HA are necessary for infection to proceed, blocking the conformational changes using a drug could be effective in the prevention of the disease. Benzoquinones and hydro-quinones act as fusion inhibitors [49, 50]. In one such study using ELISA and monoclonal antibody showing that the drug BMY-27709 acts upon inhibition of low pH induced conformational changes of HA protein. After exposing the virus to BMY-27709 it was confirmed that some amino acids present in HA had mutated, implying the new resistance to BMY-27709. The exact mechanism of the inhibition was not clear but, it is believed that BMY-27709 acts as the glue and inhibits the movement of fusion peptide or other regions of the HA protein. Crystal structure of another fusion inhibitor drug TBHQ (tert-butyl hydroquinone) and HA shows that TBHQ binds between two monomers of the HA trimer. According to the crystal structure, binding pocket of TBHQ lies between the interface of highly conserved  $\beta$  hairpin of HA1 and long  $\alpha$ -helix of HA2 [51]. Earlier studies have shown that occurrence of mutated HA and inhibition of fusion activity between RBCs and HA-expressing CHO cells in the presence of TBHQ molecules. Modifications of TBHQ structure could promote the development of potential anti-influenza viral drugs [49].

## **REFERENCES**

## REFERENCES

1. WHO, W. *Influenza (Seasonal)*. 2015.
2. Wilson, I.A., J.J. Skehel, and D.C. Wiley, *Structure of the haemagglutinin membrane glycoprotein of influenza virus at 3 Å resolution*. *Nature* 1981. **289** p. 366- 373.
3. White, J.M., et al., *Structures and mechanisms of viral membrane fusion proteins: Multiple variations on a common theme*. *Critical Reviews in Biochemistry and Molecular Biology*, 2008. **43**(3): p. 189-219.
4. Shi, Y., et al., *Enabling the 'host jump': structural determinants of receptor-binding specificity in influenza A viruses*. *Nature Reviews Microbiology*, 2014. **12**(12): p. 822-831.
5. WHO, W. *Influenza*. 2003.
6. Chen, J., J.J. Skehel, and D.C. Wiley, *N- and C-terminal residues combine in the fusion-pH influenza hemagglutinin HA(2) subunit to form an N cap that terminates the triple-stranded coiled coil*. *Proceedings of the National Academy of Sciences of the United States of America*, 1999. **96**(16): p. 8967-8972.
7. Gerl, M.J., et al., *Quantitative analysis of the lipidomes of the influenza virus envelope and MDCK cell apical membrane*. *Journal of Cell Biology*, 2012. **196**(2): p. 213-221.
8. Zhang, J., A. Pekosz, and R.A. Lamb, *Influenza virus assembly and lipid raft microdomains: a role for the cytoplasmic tails of the spike glycoproteins*. *Journal of Virology*, 2000. **74**(10): p. 4634-4644.
9. Nayak, D.P., E.K.W. Hui, and S. Barman, *Assembly and budding of influenza virus*. *Virus Research*, 2004. **106**(2): p. 147-165.
10. Harris, A., et al., *Influenza virus pleiomorphy characterized by cryoelectron tomography*. *Proceedings of the National Academy of Sciences of the United States of America*, 2006. **103**(50): p. 19123-19127.
11. Skehel, J.J. and D.C. Wiley, *Receptor binding and membrane fusion in virus entry: The influenza hemagglutinin*. *Annual Review of Biochemistry*, 2000. **69**: p. 531-569.
12. Gubareva, L.V., L. Kaiser, and F.G. Hayden, *Influenza virus neuraminidase inhibitors*. *Lancet*, 2000. **355**(9206): p. 827-835.
13. O'Neill, R.E., J. Talon, and P. Palese, *The influenza virus NEP (NS2 protein) mediates the nuclear export of viral ribonucleoproteins*. *Embo Journal*, 1998. **17**(1): p. 288-296.

14. Gamblin, S.J., et al., *The structure and receptor binding properties of the 1918 influenza hemagglutinin*. Science 2004. **303** p. 1838-1842
15. Mobius, W., et al., *Recycling compartments and the internal vesicles of multivesicular bodies harbor most of the cholesterol found in the endocytic pathway*. Traffic, 2003. **4**(4): p. 222-231.
16. Wiley, D.C. and J.J. Skehel, *The structure and functions of the hemagglutinin membrane glycoprotein of influenza virus*. Annual Review of Biochemistry, 1987. **56**: p. 365-394.
17. Kobayashi, T., et al., *Separation and characterization of late endosomal membrane domains*. Journal of Biological Chemistry, 2002. **277**(35): p. 32157-32164.
18. Lakadamyali, M., et al., *Visualizing infection of individual influenza viruses*. Proceedings of the National Academy of Sciences of the United States of America, 2003. **100**(16): p. 9280-9285.
19. Bui, M., G. Whittaker, and A. Helenius, *Effect of M1 protein and low pH on nuclear transport of influenza virus ribonucleoproteins*. Journal of Virology, 1996. **70**(12): p. 8391-8401.
20. Zhirnov, O.P., *Solubilization of matrix protein M1/M from virions occurs at different pH for orthomyxoviruses and paramyxoviruses*. Virology, 1990. **176**(1): p. 274-279.
21. Martin, K. and A. Helenius, *Nuclear transport of influenza-virus ribonucleoproteins-The viral matrix protein (M1) promotes export and inhibits import*. Cell, 1991. **67**(1): p. 117-130.
22. Kemble, G.W., T. Danieli, and T.M. White, *Lipid anchored influenza hemagglutinin promotes hemi fusion not complete fusion*. Cell 1994. **76**: p. 383-391.
23. Park, H.E., J.A. Gruenke, and J.M. White, *Leash in the groove mechanism of membrane fusion*. Nature Structural Biology, 2003. **10**(12): p. 1048-1053.
24. Chen, J., J.J. Skehel, and D.C. Wiley, *A polar octapeptide fused to the N-terminal fusion peptide solubilizes the influenza virus HA(2) subunit ectodomain*. Biochemistry, 1998. **37**(39): p. 13643-13649.
25. Bumet, F.V.M. and W.V.M. Stanley, *The Viruses*. 3 ed. 1959.
26. Matthews, T., et al., *Enfuvirtide: The first therapy to inhibit the entry of HIV-1 into host CD4 lymphocytes*. Nature Reviews Drug Discovery, 2004. **3**(3): p. 215-225.
27. Wild, C.T., et al., *Peptides corresponding to a predictive alpha helical domain of human immunodeficiency virus type-1 GP41 are potent inhibitors of virus infection*.

- Proceedings of the National Academy of Sciences of the United States of America, 1994. **91**(21): p. 9770-9774.
28. White, J.M. and I.A. Wilson, *Anti-peptide antibodies detect steps in a protein conformational change-low pH activation of the influenza virus-hemagglutinin*. Journal of Cell Biology, 1987. **105**(6): p. 2887-2896.
  29. Godley, L., et al., *Introduction of intersubunit disulfide bonds in the membrane distal region of influenza hemagglutinin abolishes membrane fusion activity*. Cell, 1992. **68**(4): p. 635-645.
  30. White, J.M., *Membrane-Fusion*. Science, 1992. **258**(5084): p. 917-924.
  31. Krogh, A., et al., *Predicting transmembrane protein topology with a hidden Markov model: Application to complete genomes*. Journal of Molecular Biology, 2001. **305**(3): p. 567-580.
  32. Bartfai, T., et al., *The state of GPCR research in 2004*. Nature Reviews Drug Discovery, 2004. **3**(7): p. 574-626.
  33. Nikaido, H., *Molecular basis of bacterial outer membrane permeability revisited*. Microbiology and Molecular Biology Reviews, 2003. **67**(4): p. 593-656.
  34. Sollner, T.H., *Intracellular and viral membrane fusion: a uniting mechanism*. Current Opinion in Cell Biology, 2004. **16**(4): p. 429-435.
  35. Overington, J.P., B. Al-Lazikani, and A.L. Hopkins, *Opinion - How many drug targets are there?* Nature Reviews Drug Discovery, 2006. **5**(12): p. 993-996.
  36. Carpenter, E.P., et al., *Overcoming the challenges of membrane protein crystallography*. Current Opinion in Structural Biology, 2008. **18**(5): p. 581-586.
  37. Smith, S., *Strategies for the Purification of Membrane Proteins*, in *Protein Chromatography*, D. Walls and S.T. Loughran, Editors. 2011, Humana Press. p. 485-496.
  38. Link, A.J. and G. Georgiou, *Advances and challenges in membrane protein expression*. Aiche Journal, 2007. **53**(4): p. 752-756.
  39. 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.
  40. Nobusawa, E., et al., *Comparison of complete amino acid sequences and receptor binding properties among 13 serotypes of hemagglutinins of influenza virus A- viruses*. Virology, 1991. **182**(2): p. 475-485.

41. Gething, M.J., et al., *Studies on the mechanism of membrane- fusion –site specific mutagenesis of the hemagglutinin of influenza –virus*. Journal of Cell Biology, 1986. **102**(1): p. 11-23.
42. Han, X., et al., *Membrane structure and fusion-triggering conformational change of the fusion domain from influenza hemagglutinin*. Nature Structural Biology, 2001. **8**(8): p. 715-720.
43. Lorieau, J.L., J.M. Louis, and A. Bax, *The complete influenza hemagglutinin fusion domain adopts a tight helical hairpin arrangement at the lipid:water interface*. Proceedings of the National Academy of Sciences of the United States of America, 2010. **107**(25): p. 11341-11346.
44. Tatulian, S.A. and L.K. Tamm, *Secondary structure, orientation, oligomerization, and lipid interactions of the transmembrane domain of influenza hemagglutinin*. Biochemistry, 2000. **39**(3): p. 496-507.
45. Melikyan, G.B., et al., *Amino acid requirement of transmembrane and cytoplasmic domain of influenza virus hemagglutinin for viable membrane fusion*. . Molecular biology of the cell, 1999. **10**: p. 1821-1836.
46. SchrothDiez, B., et al., *Fusion activity of transmembrane and cytoplasmic domain chimeras of the influenza virus glycoprotein hemagglutinin*. Journal of Virology, 1998. **72**(1): p. 133-141.
47. Armstrong, R.T., A.S. Kushnir, and J.M. White, *The transmembrane domain of influenza hemagglutinin exhibits a stringent length requirement to support the hemifusion to fusion transition*. J. Cell Biology 2000. **151**: p. 425-437.
48. Pielak, R.M. and J.J. Chou, *Influenza M2 proton channels*. Biochimica Et Biophysica Acta-Biomembranes, 2011. **1808**(2): p. 522-529.
49. Hoffman, L.R., I.D. Kuntz, and J.M. White, *Structure-based identification of an inducer of the low-pH conformational change in the influenza virus hemagglutinin: Irreversible inhibition of infectivity*. Journal of Virology, 1997. **71**(11): p. 8808-8820.
50. Combrink, K.D., et al., *Salicylamide inhibitors of influenza virus fusion*. Bioorganic & Medicinal Chemistry Letters, 2000. **10**(15): p. 1649-1652.
51. Russell, R.J., et al., *Structure of influenza hemagglutinin in complex with an inhibitor of membrane fusion*. Proceedings of the National Academy of Sciences of the United States of America, 2008. **105**(46): p. 17736-17741.
52. Wasniewski, C.M., et al., *Solid-state nuclear magnetic resonance studies of HIV and influenza fusion peptide orientations in membrane bilayers using stacked glass plate samples*. Chemistry and Physics of Lipids, 2004. **132**(1): p. 89-100.

53. Bodner, M.L., et al., *Temperature dependence and resonance assignment of C-13 NMR spectra of selectively and uniformly labeled fusion peptides associated with membranes*. *Magnetic Resonance in Chemistry*, 2004. 42(2): p. 187-194.
54. Bodner, M.L., et al., *(13)C-(13)C and (15)N-(13)C correlation spectroscopy of membrane-associated and uniformly labeled human immunodeficiency virus and influenza fusion peptides: Amino acid-type assignments and evidence for multiple conformations*. *Journal of Chemical Physics*, 2008. 128(5).
55. Gubareva, L.V., L. Kaiser, and F.G. Hayden, *Influenza virus neuraminidase inhibitors*. *Lancet*, 2000. 355(9206): p. 827-835.
56. Moscona, A., *Neuraminidase Inhibitors for Influenza*. *New England Journal of Medicine*, 2005. 353(13): p. 1363-1373.
57. Ghosh, U., Xie, L., Jia, L., Liang, S., 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. *Journal of the American Chemical Society* 137, 7548-7551.

## **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. This is due to the challenges in investigating the structures of membrane proteins as discussed in chapter 1. Obtaining a sufficient amount of folded protein, which is in non-aggregated oligomeric state, is necessary in order 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  $\mu\text{L}$  of a protein sample that is 40  $\mu\text{M}$  in concentration per experiment (i.e. about 0.4 mg/mL for a protein with molecular mass of 10.7 kDa) [4, 5]. For a structural study such as a crystallography experiment, 10 – 20 mg of protein is required and the volume needed depends on how many conditions are being investigated. [6]. Crystallographic studies discussed in this dissertation required 60  $\mu\text{L}$  of sample that is 10 mg/ml in concentration investigate 384 different crystallographic conditions. Therefore exploring methods that can express, solubilize and purify the proteins in large quantities is an important endeavor in characterization of proteins. Obtaining protein for various characterization experiments requires the successful completion of multiple steps as depicted by the flow chart given below. In this chapter, basic methods that were used for my experiments are discussed and specific experimental procedures are discussed in each chapter.

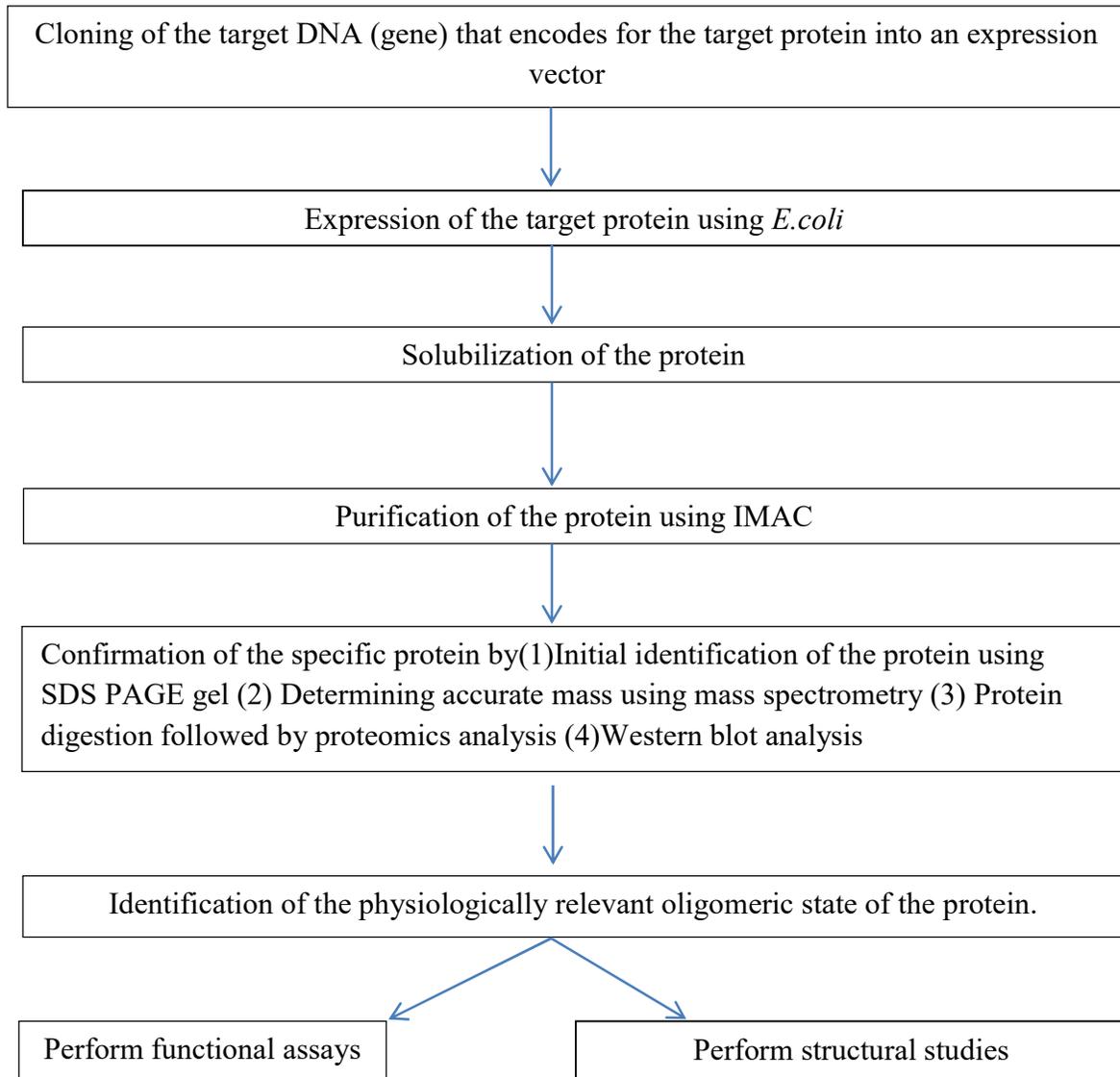


Figure-2-1 Flow chart showing basic steps involved in characterization of a protein. These steps are briefly discussed below.

## 2-2 Molecular subcloning

Getting a protein expressed using *E.coli* requires the insertion of the targeted DNA that encodes for the protein of interest into the expression vector. The gene was ordered from Genscript Inc. (NJ). Genscript provides the gene in their vector, named pUC57-Kan. The objective of the subcloning discussed here is to cut the gene (insert) which is originally in pUC57-Kan vector and ligate it with cut pet24a+ expression vector (Figure 2-2). The term ligation means the joining of two DNA fragments in this case gene (insert) with the cut pet24a+ vector through the action of an enzyme called T4 DNA Ligase.

Before ordering the gene from the company, the restriction sites to be used for the subcloning need to be decided. I requested Genscript Inc. to put NdeI restriction site 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 over using other restriction sites to prevent the occurrence of non-native sequences in the expressed proteins.

- (a) Use of BamHI instead of NdeI 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, BamHI site can be used instead of the NdeI restriction site.

(b) Selection of the second restriction site XhoI at 3' end depends on the position of His-tag. Since it was planned to purify the expressed protein using IMAC (discussed briefly under section 2-4(a)) which requires the presence of His tag in the protein being expressed, this restriction site was used. Usage of XhoI results only L (Leucine) E (Glutamate) amino acids to appear before the tag after the protein is expressed. Use of SalI instead of XhoI will cause unnecessary amino acid sequence to appear on C-terminal of the protein. Use of SalI will cause ArgGlnAlaCysGlyArgThrArgAla to appear in the C-terminus of the protein before the tag (Figure 2-4). Therefore Usage of XhoI 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.

Therefore, usage of XhoI and NdeI 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 NdeI is CATATG and that for XhoI is CTCGAG.

### ***2-2-1 Basic steps involved in subcloning***

- 1) Transformation of gene (insert) which is present in pUC57-Kan vector and empty (pet24a+) expression vector separately.
- 2) Extraction of the plasmid (pUC57-Kan) that contains the gene of interest (insert) and extraction of the empty expression vector (pet24a+) from the overnight grown cell cultures is performed. This process is commonly referred as miniprep and commercially available plasmid extraction kit was used in this 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.

In vectors, commonly used restriction sites are localized into one area known as “multiple cloning site”. In pet24 a+ vector multiple cloning site is present in between Nde1 site and Xho1 site (Figure 2-4). Due to the availability of several restriction sites, that region is known as “multiple cloning site”. Depending on the users requirement, as discussed earlier user have freedom to use any of these restriction sites. Restriction sites present in the multiple cloning site, do not appear anywhere else in the vector sequence (Figure 2-4). Therefore usage of restriction

sites present in the multiple cloning site guarantees that vector will not cut into several different pieces.

4) Run two preparative DNA gels (agarose gels) to isolate cut pet24a+ vector from uncut pet24a+ vector and to isolate gene (insert) from pUC57-Kan vector (Figure 2-3).

Depending on the size of the vector and insert, percentage of the agarose needed for the DNA gel preparation varies. 0.5% agarose gel was used for the separation of cut pet24a+ vector from uncut pet24a+ vector (~5.3 kbp) (Figure 2-3 (A)). 1% agarose gel was used for the isolation of gene of interest (insert) (~500 bp) from pUC57-Kan vector (~2.5 kbp) (Figure 2-3 (B)). Presence of low amount of agarose results in large pore size in gel and is useful in the isolation of constructs having large number 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 (table 1).

Table 1-1: DNA size resolution based on agarose content in the gel. Depending on the size of the vector/gene (insert) user has to decide which percentage of agarose gel has to be used for the preparative agarose gels.

Percentage Agarose Gel (w/v)	DNA Size Resolution
0.50%	1 kbp to 30 kbp
0.70%	800 bp to 12 kbp
1.00%	500 bp to 10 kbp
1.20%	400 bp to 7 kbp
1.50%	200 bp to 3 kbp
2.00%	50 bp to 2 kbp

6) Setting up a ligation reaction.

Ligation reaction has to be set up by considering the molar ratio of the empty vector and the insert. The equation used to calculate the quantity of insert that needed to be added to the reaction mixture is given below (Equation 2-1). Typically the amount of empty vector used for the ligation reaction is 80-200 ng. Generally three ligation reactions are set up using insert: vector molar ratio as 1:1, 3:1 and 1:3. First, the inset and vector are mixed at these ratios. Next, this mixture is combined with T4 DNA ligase enzyme and its buffer and stored at 16 C<sup>0</sup> overnight for ligation to take place. (T4 DNA ligase enzyme 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) was purchased from New England Biolabs Inc.)

$$\frac{\text{Length of the Insert (kb)}}{\text{Length of the empty Vector (kb)}} \times \text{ng of the empty Vector} = \text{ng of the Insert} \quad - \text{Equation 2-1}$$

7) The next day, transformation (explained below) of the ligation product was performed and the transformed cells were plated in the agar plates which contain the antibiotic (kanamycin) that the pet24a+ vector displays resistance to. Transformation was performed as described below.

The competent cells were taken out from -80°C freezer and thawed on ice. A volume of 1 to 5µL of ligation product was added into microcentrifuge tube which contains the competent cells. Ligation product and competent cells were gently mixed by flicking the bottom of the microcentrifuge tube. Then competent cell/ligation product mixture was placed on ice for 20-30 min. After that, tubes were placed in a 42°C water bath for 45 seconds (This step is known as heat shock). Then tubes were put back in the ice bath again for 2 minutes. 450 µL of LB (Lauria Broth) was added to the competent cell/ ligation product mixture. Then the tube was placed in

the incubator at 37 °C for 1 hour. And the incubator was shaken at 225 rpm speed. After 1 hour, competent cell/ligation cell mixture was plated on agar plate which contains kanamycin antibiotic. Agar plates were kept overnight in the incubator at 37°C.

8) Once the initial colonies appeared on the agar plate, some colonies were grown overnight at 37 °C in LB medium in the presence of desired antibiotic (kanamycin) and plasmids were extracted the following day.

9) Finally extracted plasmids were submitted to DNA sequencing facility to confirm the presence of desired gene in the pet24a+ vector.

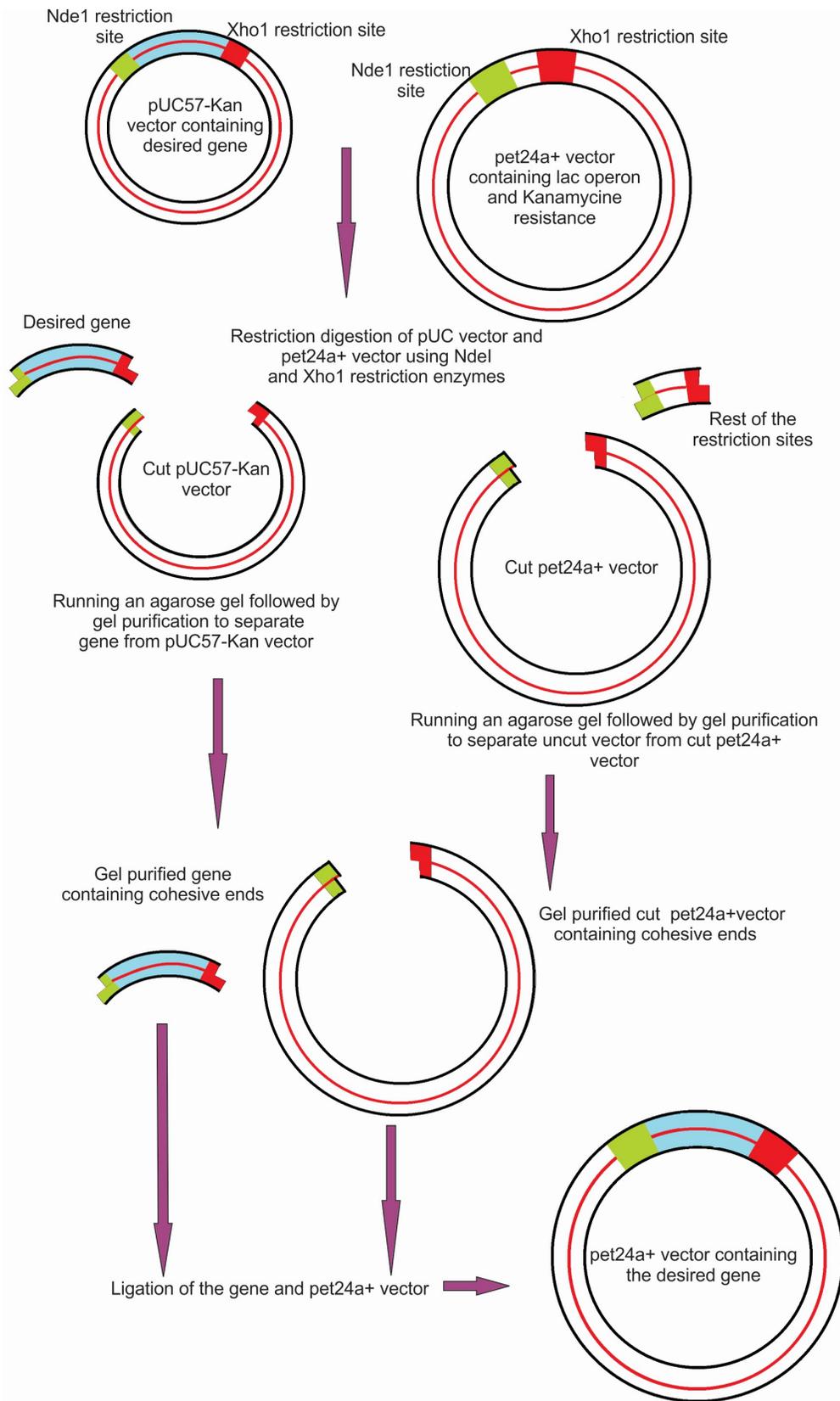


Figure 2-2-Diagram showing the steps of sub-cloning

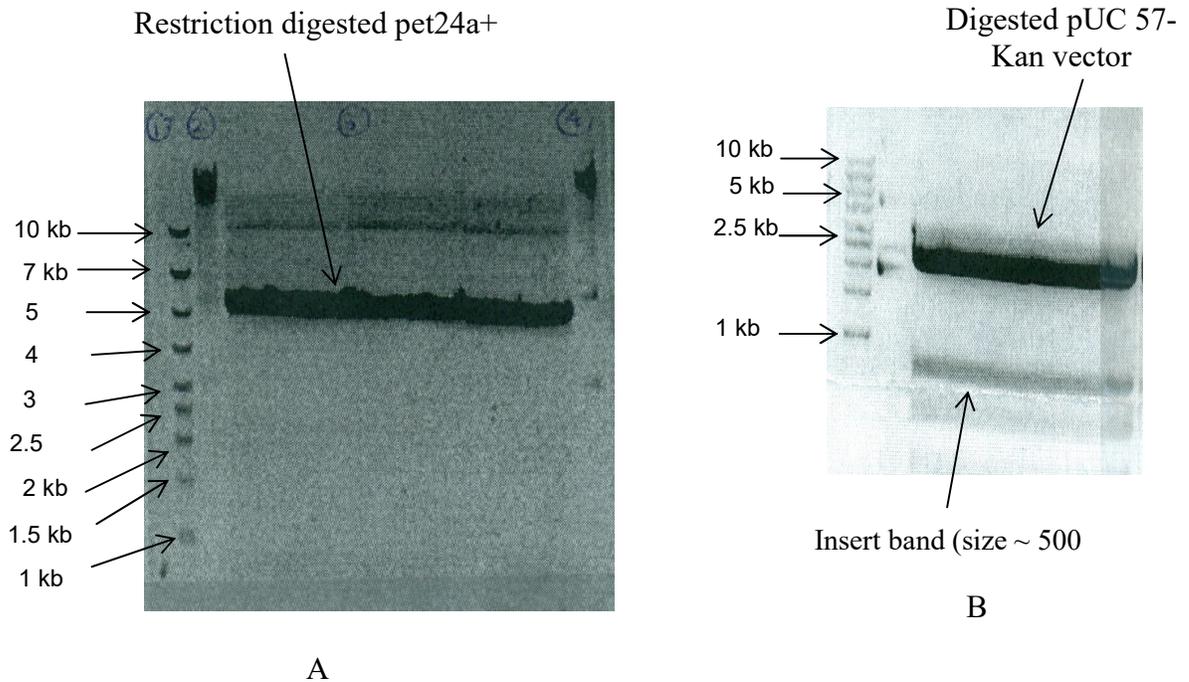


Figure 2-3 (A) Picture of 0.5% preparative agarose gel run for the separation of cut pet24a+ vector. (B) Picture of 1% preparative agarose gel run for the separation of insert (gene of interest) from pUC57-Kan vector.

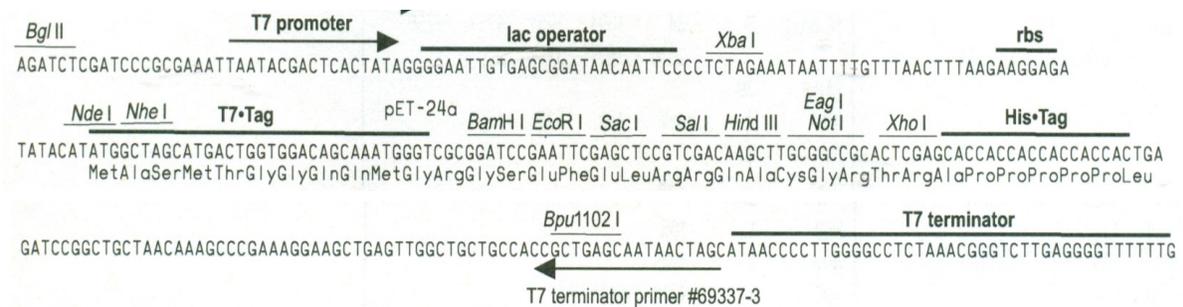


Figure 2-4 Diagram showing the lac operator, ribosomal binding site (rbs), multiple cloning region (NdeI-XhoI), and His tag of the pet24a+ vector.

## 2-3 Protein expression

Bacterial expression systems represent the most commonly 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 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 of *E.coli* consists of 4 phases namely lag phase, log phase, stationary phase and death phase (Figure 2-5). At the lag phase there is only a limited number of cells. During log phase, *E.coli* cells grow very rapidly due to the presence of plenty of nutrients in the culture medium. Therefore, they are in the optimum state to produce proteins. In the stationary phase, nutrients in the medium become limited (due to being consumed in 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* cell growth is in log phase. For an example, to reach OD<sub>600</sub>= 0.5 HA2 constructs needs ~1 hour and addition of IPTG is done after 1 hour.

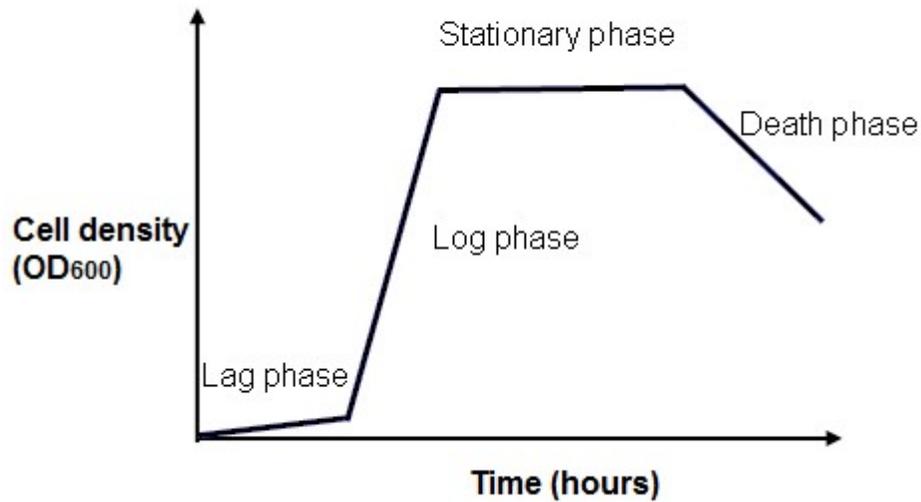


Figure 2-5-The graph showing the four phases of *E.coli.* growth

### 2-3-1 *Lac operon*

An operon is a cluster of bacterial genes which is controlled under single promoter (Figure 2-6). 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. The *lac* operon consists of promoter, operator, terminator and three other genes known as *lacZ*, *lacY* and *lacA* (Figure 2-6). *LacZ* 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 in to the cells. *LacA* encodes galactoside O-acetyltransferase, an enzyme that transfers an acetyl group from acetyl-CoA to  $\beta$ -galactosidase.

In the absence of lactose, the regulatory gene *lacI* produces *lac* repressor protein and it can bind with operator (Figure 2-6 (a)). Once repressor protein binds to the operator it will prevent the transcription of 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 protein to operator (Figure 2-6 (b)). This allows RNA polymerase to transcribe *lac* genes.

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

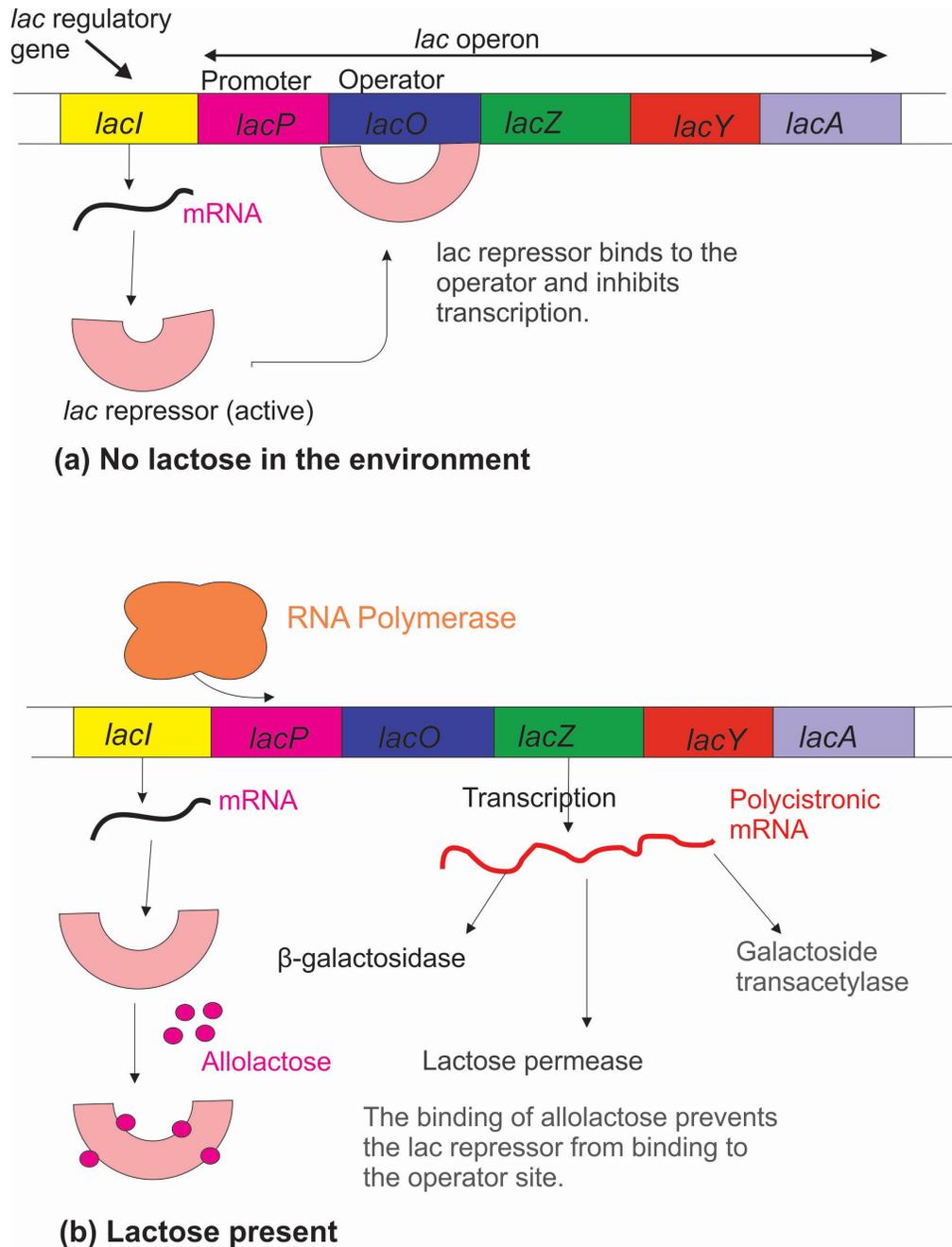


Figure 2-6 Diagram showing the functionality of *Lac* operon (a) In the absence of lactose (b) In the presence of lactose [1].

### ***2-3-2 pet vector system***

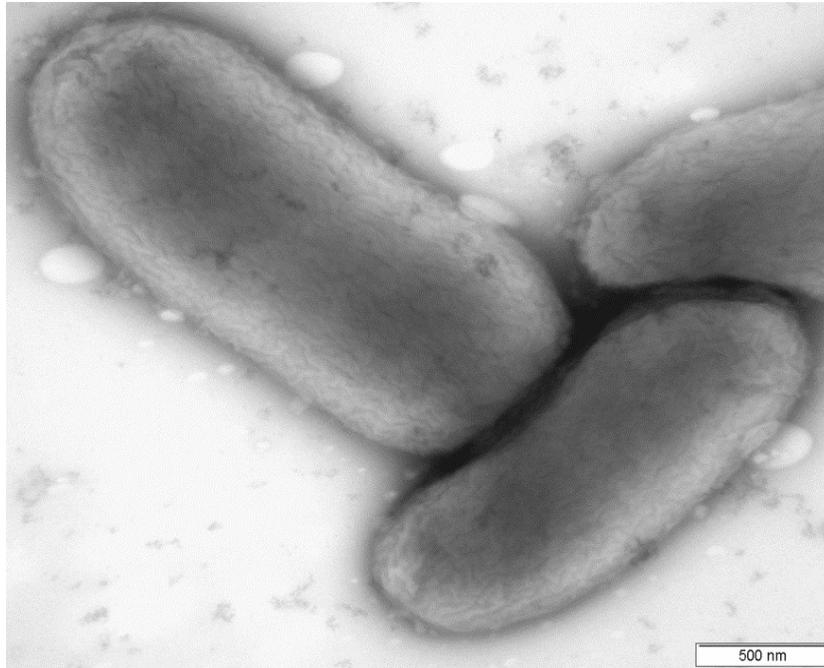
The pet24a+ vector was used in this project to take the advantage of lac operator. This vector basically consists of LacI gene, T7 promoter, lac operator, T7 and terminator (Figure 2.4). Bacteriophage T7 promoter is nonnative to bacterial genome and it can only recognize T7 RNA polymerase but not the 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 inducer molecule IPTG, T7 RNA polymerase is not produced and therefore gene of interest is not transcribed. Upon addition of IPTG, it induces both the expression of RNA polymerase and thereby the target protein. As discussed above, pet vectors expression of targeted protein is tightly regulated by T7 promoter and T7 RNA polymerase [3].

### ***2-3-3 Inclusion bodies***

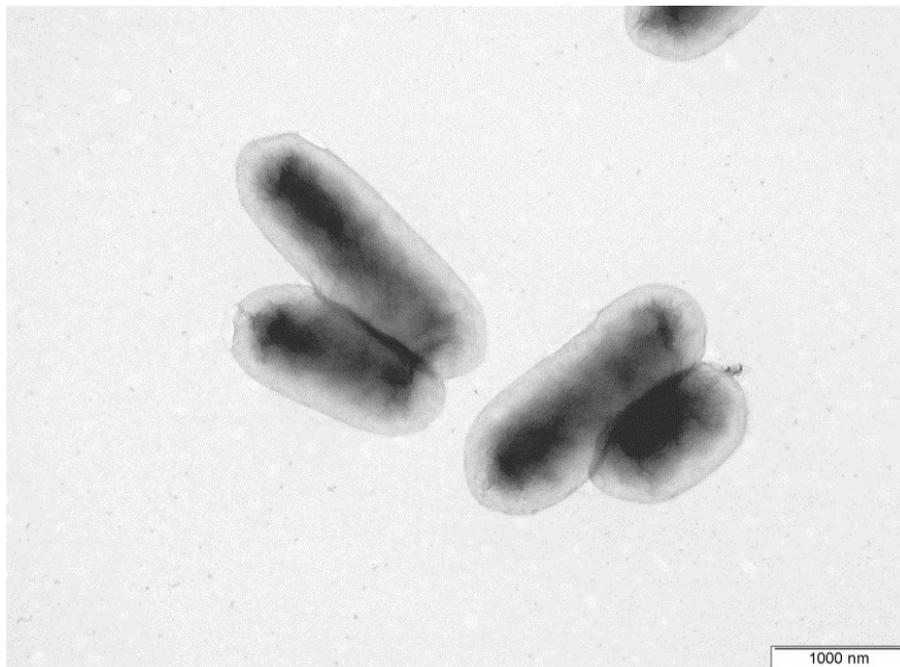
Overexpression of many recombinant proteins using *E.coli* results in the accumulation of proteins in inclusion bodies. In the inclusion bodies, in addition to the overexpressed protein, phospholipids from the *E.coli* membrane, other *E.coli* proteins and RNA are present [9]. Inclusion bodies produced in bacteria have different shapes such as spherical, ellipsoidal, cylindrical and even tear-shaped with the size ranging from 50-700 nm [10]. Proteins present in

inclusion body have either amorphous (devoid of any structural regularity) or paracrystalline (short to medium term ordering in the structure) nature [11, 12]. The presence of inclusion bodies in bacterial cells can be easily identified using transmission electron microscopy. To identify the presence of inclusion bodies induced and uninduced cells that encode FHA2 protein were analyzed using transmission electron microscopy. Uninduced cells (Figure 2-7 (a)) exhibited no visible regions of densely aggregated portions in the *E.coli* cells while induced cells consisted of dense regions inside the cell (Figure 2-7 (b)).

The inclusion bodies can be solubilized using denaturants and detergents such as 8 M urea, 6 M guanidinium hydrochloride or ionic detergents (0.5% sarkosyl, 0.5% SDS) [13, 14]. There are some advantages of using protein purified from inclusion bodies for structural and functional studies. (1) The amount of protein obtained from inclusion bodies is up to 15-25% of the total cell mass. (2) Those are easy to isolate from cells since their density (~1.3 mg/ml) is higher compared with cell debris. Therefore using high speed centrifugation (20000 rpm, 20 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. [15]. In chapter 3, brief description of methods that I have developed in our lab for the isolation and purification of inclusion bodies is provided.



(A)



(B)

Figure 2-7 (A) Transmission electron microscopy pictures of uninduced bacterial cells (B) Transmission electron microscopy pictures of bacterial cells induced to produce FHA2 protein (After 5 hours of induction time).

Figure 2-7 (A) and (B) pictures were generated at MSU Electron Microscopy Facility using the following protocol. First, *E. coli* cells were grown for five hours with and without induction using IPTG. 5  $\mu$ L of sample was added into the formvar coated grid (a special grid used for this experiment provided by the Electron Microscopy Facility). Next, the sample was air dried. About 100  $\mu$ L of negative stain 1% uranyl acetate was added and air dried for few minutes. Next, the image was taken using JEOL 100CX transmission electron microscope, (Japan Electron Optics Laboratory, Japan).

#### **2-4 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 as discussed above. First, the cells are sonicated in the presence of PBS. This breaks the cells and releases the soluble proteins into the medium. Next, this lysate is centrifuged at 20000 rpm for 20 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 for 1 hour at 4 °C) of this supernatant results the membrane fraction of the *E. coli* being separated as a pellet. Membrane fraction pellet formed has a transparent brownish color. Membrane proteins are more likely to be present in this fraction.

For membrane protein isolation, detergents and chaotropic agents are used as the solubilizing reagents [13, 14, 16]. 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 capacity to store these in their correctly folded state. (2) Expressed protein is often stored as insoluble aggregates known as inclusion bodies and detergents (Ex: 0.5% sarkosyl, 0.5% SDS) or denaturing reagents (Ex: 8 M urea, 6 M guanidinium hydrochloride) has to be used for solubilization. Furthermore, proteins extracted from inclusion bodies have to be refolded before, they used for any structural or functional experiments. (3) The downstream protein purification procedures including IMAC chromatography, ion exchange chromatography and gel filtration chromatography techniques can fail due to a number of reasons discussed below. (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 oftentimes limited/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 tends 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 moves along with the hydrophobic domain making it also unexposed to the solvent. This causes the above phenomena and makes His tag based purification of membrane proteins challenging. Usage of longer linker region (addition of glycine tag) in between the protein sequence and his tag sequence can solve this problem to some extent. (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 inside the gel filtration columns if the concentration of detergent being used is less than the CMC level of that detergent. Different detergents must be screened to identify the ones in which no aggregation occurs. The clogging of the gel filtration column due to the protein aggregation will cause increased back pressure. Denaturing solutions like of 8 M urea or 6 M guanidine hydrochloride has to be used to clean the column or if the protein cannot be removed, gel filtration column may need replacement.

#### ***2-4-1 Immobilized metal affinity chromatography (IMAC)***

In immobilized metal affinity chromatography (IMAC), histidine residues present in the surface of the protein molecule and immobilized metal ions form weak coordinate bonds. Immobilized metal such as Ni (II), Cu (II), and Co (II) 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 histidines present in the His-tag protein [17]. Elution of the protein from affinity matrix can be achieved by (1) using a competitive displacement agent such as imidazole (2) using a low pH buffer (3) using a strong chelating agent as EDTA.

It is easy to elute the His-tag protein (polyhistidine tag with six histidine residues) attached to metal ion with a small molecule such as imidazole since imidazole also is capable of displaying the same interaction shown by the His tag. Therefore His tag protein is detached and elutes from the metal while imidazole binds with the metal. Using 250-300 mM imidazole helps in eluting out the protein from the affinity resin. pKa of imidazole ring of histidine residues is ~6.0. Under low pH, histidines become protonated and cannot form coordinate bonds with metal ions since

its lone pairs are donated to the  $H^+$ . Therefore, His tagged protein it will not stay bound with the resin in low pH conditions and will elute off. Addition of strong chelating agent like EDTA form a complex with metal ion and disrupt the interaction with metal ion and his tag proteins, causing the his tag protein to elute out.

#### ***2-4-2 Ion exchange chromatography***

The basis for the ion exchange chromatography is the ionic interaction between the protein molecules and the ion exchange resin. The separation is driven by 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 ionizable sidechains of amino acid residues in the protein. Depending on the pH and pI value proteins can achieve either positive or negative charge and proteins can be separated using ion exchange resin. Anion exchange is performed at pH values above the pI value of the protein (where protein is negatively charged) using positively charged resin. Cation exchange is performed at pH values below the pI value (where protein is positively charged) using a negatively charged resin. To elute the proteins from the ion exchange resin, salt gradient or manipulation of the pH can be used. The molecules that are weakly bound start to elute at low concentration of salt and higher salt concentration is necessary to elute strongly bound proteins. Raising the pH of 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 of 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 of 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.

## **2-5 SDS PAGE (Sodium Dodecyl Sulphate Poly Acrylamide Gel Electrophoresis)**

SDS PAGE allows mass selective separation of protein molecules. SDS (Sodium dodecyl sulphate) is an anionic detergent that coats over the proteins resulting in 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). This causes proteins to attain intrinsic negative charge. Therefore, SDS treated proteins have similar charge-to mass ratio.

SDS-bound protein molecules migrate in the SDS PAGE gel depending primarily on their masses, since all protein molecules attain 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 we can change the pore size of the SDS PAGE. Higher percentage gels are used for separation of the small proteins since the pore size is small. Molecules with larger mass will experience more resistance to migration when the pore size of the gel is smaller and *vice versa*.

## 2-6 Crosslinking experiment

Running the SDS PAGE (Sodium Dodecyl Sulfate Poly Acrylamide Gel Electrophoresis) allows for the determination of the monomer mass of a protein. Since SDS denatures the protein and non-covalent interactions are disrupted by heating before running SDS PAGE, the molecular mass of the native oligomeric form of the protein cannot be determined by direct SDS PAGE. Thus, to find oligomeric state of the protein we can first crosslink the protein with BS3 Bis (sulfosuccinimidyl) suberate, followed by the SDS PAGE experiment. The protein molecules in the oligomer are covalently attached to each other using the crosslinker. Therefore, denaturing the protein sample after crosslinking to run the SDS PAGE does not break the covalently attached crosslinked protein molecules. Therefore this allows for the determination of the molecular mass of native oligomeric state of the protein.

Bis (sulfosuccinimidyl) suberate (BS3) contains N-hydroxysulfosuccinimide (NHS) esters at both ends (Figure 2-6 (A)). This NHS esters are reactive towards amines at pH 7-9. After reacting with amines these can form stable amide bonds leaving N-hydroxysulfosuccinimide (Figure 2-6 (B)). Primary amine of the side chain of the lysine residues and N-terminus of the protein are the expected reactive species towards crosslinking with BS3.

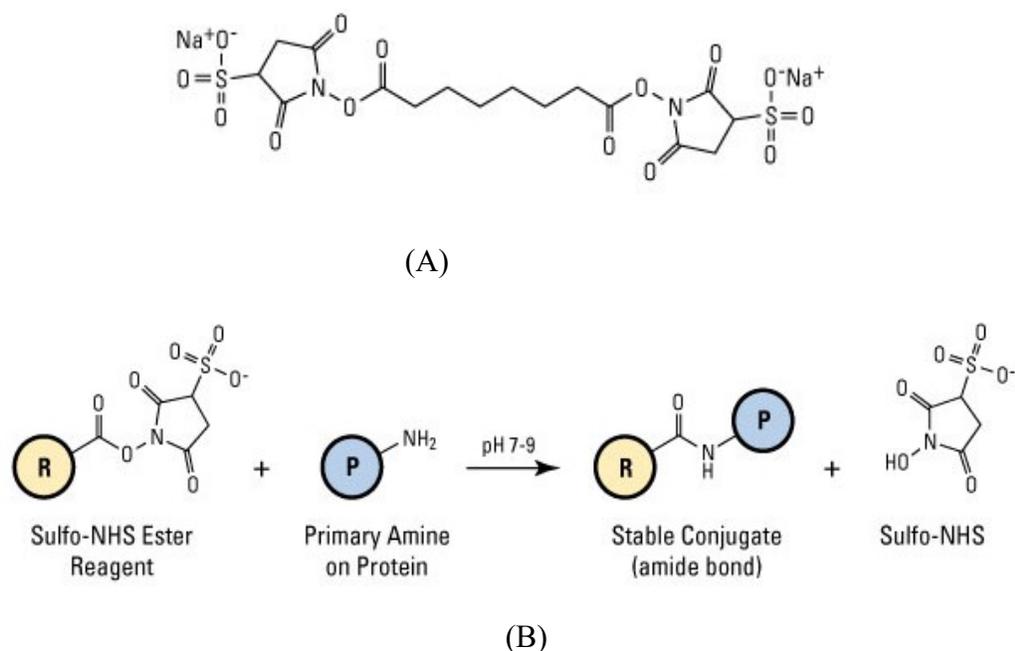


Figure 2-8 (A) Structure of Bis (sulfosuccinimidyl) suberate (B) NHS ester reaction scheme for chemical conjugation to a primary amine in protein. (R) Represents one end of a crosslinker having the NHS ester reactive group; (P) represents a protein that contains the target functional group (primary amine). The reaction forms amide bond with protein and cross linker releasing N-hydroxysuccinimide (NHS).

## 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 passes through the stationary phase (column matrix) allowing molecules to diffuse into the pores. Larger molecules do not enter into 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 into

the column pores and move through the column slowly (Figure 2-9). In gel filtration chromatography the exclusion limit means 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 that was used in our lab has the exclusion limit of 1.3 MDa and permeation limit of 6.5 kDa. Using gel filtration chromatography molecular weight of the proteins in their native states can be found. In membrane protein studies usage of detergent is common. They form micelles with these proteins and therefore when the protein elutes off of the column the observed mass for the protein is often greater from the actual molecular mass of the protein. Therefore, the observed mass has the mass contributed from detergent molecules as well as that from the protein molecule. [18]. For the determination of the molecular mass of an unknown protein, first, a set of protein calibration standards of known molecular weight are run (Figure 2-10). Then the unknown protein sample can be run in the column and its partition coefficient ( $K_{av}$ ) can be calculated using following equation (Equation 2-2). Void volume,  $V_o$  refers to the elution volume of a molecule that does not go into the pores of the gel filtration matrix. Column volume,  $V_c$  refers to the volume of the packed gel filtration bed.

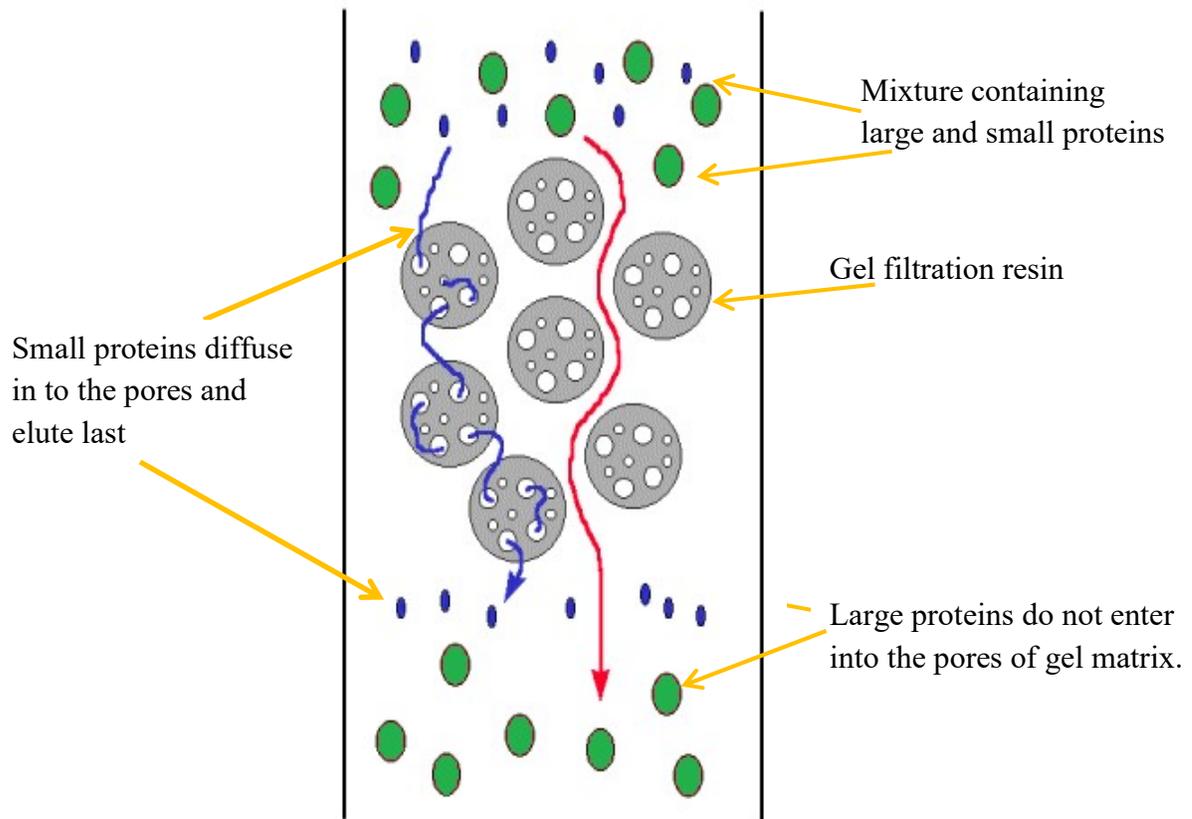


Figure 2-9 Mechanism of gel filtration

*Partition coefficient*  $K_{av} = (V_e - V_o) / (V_c - V_o)$  -----Equation 2-2

where  $V_e$ ,  $V_c$ ,  $V_o$  are the elution volume and column volume and void volume respectively. In superdex 200 column  $V_c = 24$  ml and  $V_o = 9$  ml

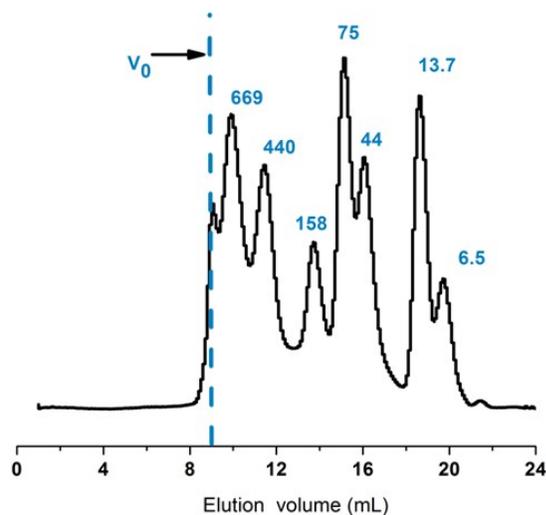


Figure 2-10 The plot showing SEC of  $A_{280}$  of MW standards loaded on Superdex 200 columns. The protein standards used were Thyroglobulin (669 kDa), Ferritin (440 kDa), Aldolase (158 kDa), Conalbumin (75 kDa), Ovalbumin (44 kDa), Carbonic Anhydrase (29 kDa), Ribonuclease A (13.7 kDa) and Aprotinin (6.5 kDa). The void volume is represented by the  $V_0$ .

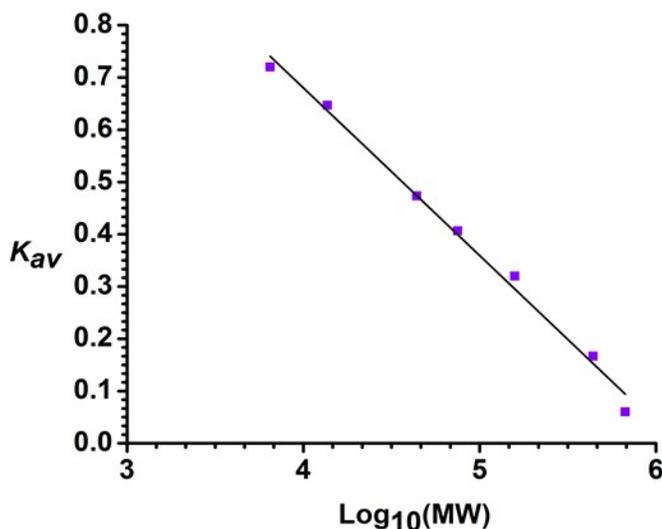


Figure 2-11 The plot of  $K_{av}$  Vs  $\log_{10}(\text{MW})$ .  $K_{av} = (V_e - V_0)/(V_c - V_0)$  where  $V_e$ ,  $V_c$ ,  $V_0$  are the elution volume, column volume and void volume respectively.

## 2-8 CD spectroscopy

CD spectroscopy 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. In a CD plot, mean residue molar ellipticity ( $\theta_{MR}$ ) vs wavelength (nm) is plotted. Alpha-helix, beta-sheet, and random coil structures give characteristic shapes of CD spectrum in the region of 190-250 nm (Figure 2-12).

Circular dichroism (CD) is measured as the differential absorbance of left ( $A_{LCP}$ ) and right circularly polarized ( $A_{RCP}$ ) light, and can be expressed as

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

According to Beer-Lambert's law

$$A = \epsilon C l$$

$$\Delta A = \Delta \epsilon C l$$

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

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

The concentration of protein ( $C$ ) in molar is multiplied by the number of amino acids ( $N$ ) in the protein to provide the mean residue concentration  $C_{MR}$ ,

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

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

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

Differential absorption of the left and right circularly polarized light is detected by the instrument using a photomultiplier and converted into ellipticity,  $\theta$  which has units of millidegrees.

Ellipticity can be related to differential absorption by

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

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

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

$\theta_{MR}$  - Mean residue molar ellipticity

Mean residue molar ellipticity ( $\theta_{MR}$ ) of  $-33,000 \text{ deg cm}^2\text{dmol}^{-1}$  at wavelength 222 nm is considered as 100% helicity. This value was determined using different reference structures consisting of poly-L-lysine with the secondary structures of completely helical, random coiled and beta sheet. Mean molar residue ellipticity was calculated after scanning over the wavelength region of 190nm-250nm of these poly-L-lysine structures. At  $\theta_{222 \text{ nm}}$  random coiled and beta sheet structures had the very low value for the mean residue molar ellipticity and alpha helix had the maximum value of  $-33,000$ . Depending on that value, percent helicity of a protein molecule can be calculated [2].

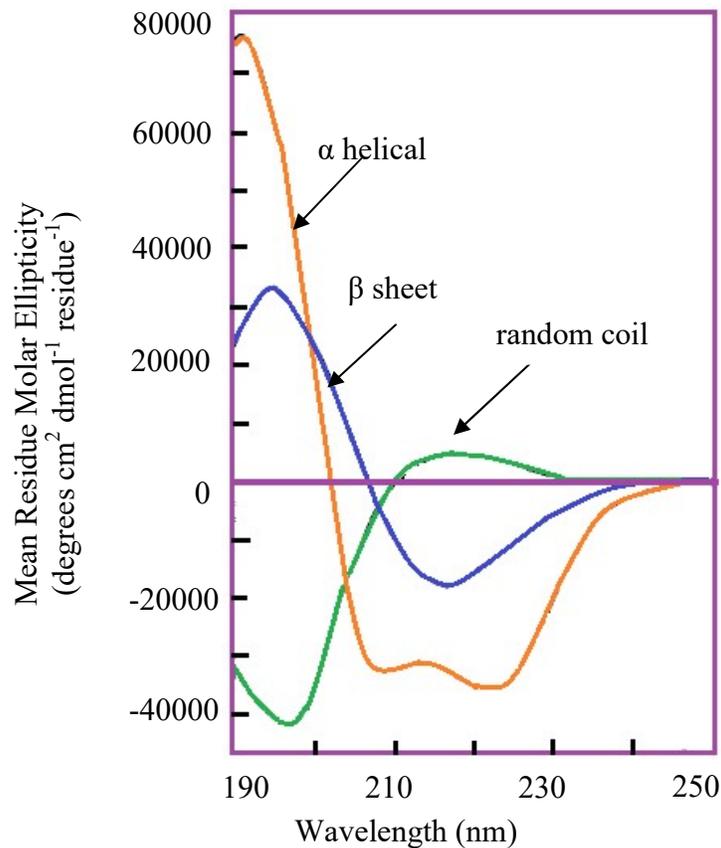


Figure 2-12 Characteristic CD curves for different secondary structures of protein molecules [2].

## 2-9 Western blots

Western blot is a useful technique to identify 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 transferred to a membrane. Next step is known as blocking and milk solution is added to the membrane. Milk contains casein protein which binds with membrane and prevents the nonspecific binding of the antibody with membrane. Then the 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. Secondary antibody binds to the primary antibody and produces a specific color in the presence of specific substrate.

The expressed proteins in our lab have C-terminal his tag which aids in the affinity purification. For the detection of these proteins, an anti-His antibody was used. This anti-His antibody is conjugated to HRP enzyme which reacts with the chemiluminescence substrate. This reaction results in a brown color, allowing the detection of the proteins that has his tag.

## **2-10 Lipid mixing assays**

Fluorescence quenching assay was used to compare the fusogenicity among different protein constructs [19]. This assay depends on the interaction between a donor molecule and an acceptor molecule when emission band of the donor molecule overlaps with the excitation band of the acceptor molecule and when they exist in close proximity (Figure 2-14).

The efficiency of this energy transfer is inversely proportional to the sixth power of the distance between donor and acceptor. *N*-NBD-PE {*N*-(7-nitro-2, 1, 3-benzoxadiazol-4-yl) (ammonium salt) dipalmitoylphosphatidylethanolamine} is the energy donor and *N*-Rh-PE {*N*-(lissamine rhodamine B sulfonyl) (ammonium salt) dipalmitoylphosphatidyl- ethanolamine} is the energy acceptor [19]. Head group of the phosphatidylethanolamine lipid is modified to add NBD fluorophore and rhodamine quencher.

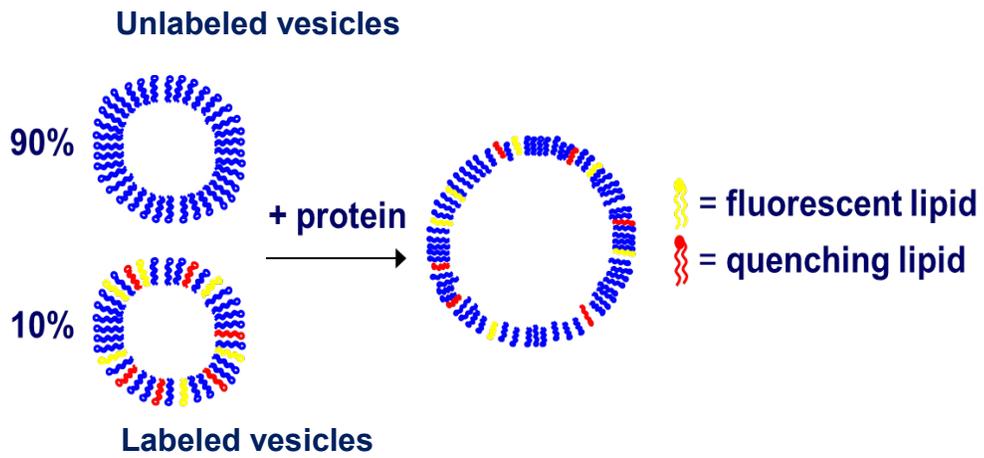


Figure 2-13 Principle of lipid mixing. A small fraction of labeled vesicles containing both quenching lipid (acceptor) and fluorescent lipid (donor) is mixed with unlabeled vesicles. Fusion associated dilution results in increased fluorescence signal .

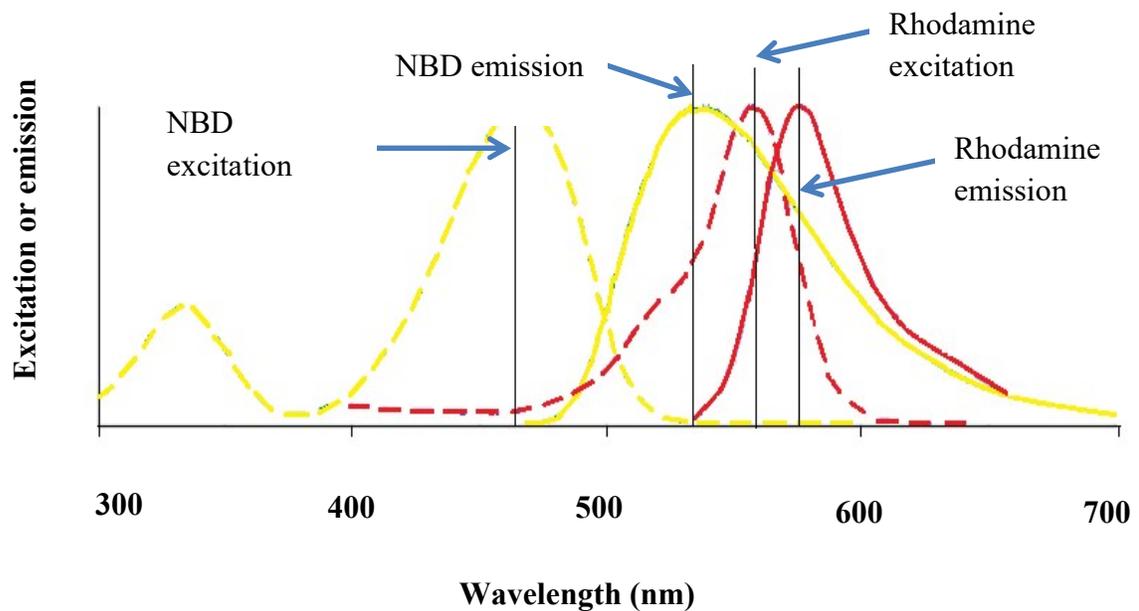


Figure 2-14 Principle of fluorescence quenching assay. NBD excitation happens around 467 nm and emits around 530nm. When donor (NBD) and acceptor (Rhodamine) are in close proximity emission signal of donor is effectively absorbed by acceptor molecule. In the presence of protein, vesicular fusion happens, resulting in increased donor acceptor distance. This increase of distance results in a decrease of quenching by the acceptor. Therefore, increased fluorescence signal being observed.

## **APPENDIX**

## APPENDIX - GLOSSARY OF TERMS

### **Plasmid**

A segment of DNA independent of the chromosomes and capable of replicating, which is present in bacteria and yeast. Plasmids are used in recombinant DNA procedures to transfer genetic material from one cell to another.

### **Expression vector**

A plasmid containing the required regulatory sequences that in turn manipulates the cell's mechanism to produce the protein encoded by the gene which is present in the plasmid. Expression vector consists of a transcriptional promoter, ribosomal binding site, start codon, stop codon and a selective marker (Ex: antibiotic resistance). Ex: pet24a+ vector containing T7 promoter, ribosomal binding site (rbs), start codon (ATG), stop codon (TGA) and kanamycin resistance (Figure 2-4).[3]

### **Ligation**

Process of attaching two DNA fragments through an enzymatic reaction is ligation. It is performed when a foreign DNA fragment is inserted into a plasmid. Formation of phosphodiester bonds between the 3'-hydroxyl of one DNA terminus with the 5'-phosphoryl of another DNA molecule joins the two DNA fragments.

### **Molecular cloning**

A set of techniques used to insert recombinant DNA from a prokaryotic or eukaryotic source into a replicating vehicle such as plasmids or viral vectors is known as "molecular cloning."

### **Subcloning**

Subcloning refers to moving of a particular gene of interest from a parent vector to a destination vector for further characterization.

### **Restriction site**

A specific location on a DNA molecule containing specific sequence of nucleotides which are recognized by restriction enzymes. Ex: XhoI restriction site corresponds to the CTCGAG nucleotide sequence.

### **Cohesive ends (Sticky ends)**

A single-stranded extension on each end of double stranded DNA molecule having a nucleotide sequence complementary to that of another DNA molecule's single stranded extension. This enables the two strands to be connected by base pairing. Cohesive ends are produced in the laboratory by using restriction enzymes.

## **REFERENCES**

## REFERENCES

1. Alberts, B., et al., *Molecular Biology of the Cell*. Third Edition ed. 1994.
2. Greenfield, N and G.D. Fasman, *Computed circular dichroism spectra for evaluation of protein conformation*. *Biochemistry*, 1969. **8**(10): p. 4108-4116.
3. Novagen.Inc, *pET-24a-d(+) Vectors*, in *pET-24a-d(+) Technical Bulletin*
4. 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 conformation*. *Biochimica et Biophysica Acta (BBA) - Biomembranes*, 2015. **1848**(1, Part B): p. 289-298.
5. 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.
6. Benvenuti, M. and S. Mangani, *Crystallization of soluble proteins in vapor diffusion for x-ray crystallography*. *Nature Protocols*, 2007. **2**(7): p. 1633-1651.
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-Polumerase travels far ahead of ribosomes invivo*. *Journal of Bacteriology*, 1992. **174**(2): p. 619-622.
9. Wang, L., *Towards revealing the structure of bacterial inclusion bodies*. *Prion*, 2009. **3**(3): p. 139-145.
10. Peternel, S. and R. Komel, *Active Protein Aggregates Produced in Escherichia coli*. *International Journal of Molecular Sciences*, 2011. **12**(11): p. 8275-8287
11. Taylor, G., et al., *Size and density of protein inclusion-bodies*. *Bio-Technology*, 1986. **4**(6): p. 553-557.
12. Singh, A., et al., *Protein recovery from inclusion bodies of Escherichia coli using mild solubilization process*. *Microbial Cell Factories*, 2015. **14**.
13. 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.

14. 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
15. 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.
16. 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.
17. Bornhorst, J.A. and J.J. Falke, *[16] Purification of Proteins Using Polyhistidine Affinity Tags*. Methods in Enzymology, 2000. **326**: p. 245-254.
18. 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.
19. Struck, D.K., D. Hoekstra, and R.E. Pagano, *Use of Resonance energy transfer to monitor membrane fusion*. Biochemistry, 1981. **20**: p. 4093-4099.

**Chapter 3 Expression, purification and functional characterization  
of full length HA2 and shorter constructs with or without fusion  
peptide or transmembrane domain**

### 3-1 Introduction

Influenza is a class I enveloped virus surrounded by a membrane obtained during budding from an infected host cell. The membrane contains hemagglutinin (HA) protein which assembles as three HA1 and three HA2 subunits with ~400 HA trimers per virion [1]. HA2 is a monotopic integral membrane protein with fusion peptide (FP), soluble ectodomain (SE), transmembrane (TM), and endodomain regions with respective lengths of ~25, 160, 25, and 10 residues (Figure 3-1 (A) and (B)). The ectodomain is outside the virus as is the ~330-residue HA1 subunit. Infection of respiratory epithelial cells is initiated by binding of the HA1 subunits to cellular sialic acid receptors. The virus is then endocytosed and subsequent reduction of the endosomal pH to the 5-6 range leads to a large structural rearrangement of HA2 subunits followed by fusion (joining) of the viral and endosomal membranes. The timescale of fusion is several minutes [2]. To our knowledge, it is not known whether influenza fuses initially with the membrane of the endosome or with the membrane of a vesicle in the interior of the endosome [3-5].

Much of our understanding of fusion of the influenza virus is based on HA-expressing cells with other cells or with vesicles [6-10]. Influenza fusion occurs within the endosome whereas in these studies, fusion is between the membrane of the virus or cell and the membrane of another cell or vesicle. To our knowledge, the significance of this difference, if any, is not known. There is likely some difference in fusion topology because the virus is enclosed within the endosomal body with which it fuses and successful fusion requires release of viral contents (nucleocapsid) outside of this body. In the model system, the fusing bodies do not enclose one another and successful fusion is considered to be mutual release of contents between them. This is more analogous to fusion of a virus with the plasma membrane which can occur for other viruses like HIV[11].

There is significant sequence conservation of HA2 across viral strains, particularly in the FP[12]. The FP and TM are important in fusion of influenza or HA-expressing cells with other cells or vesicles and are the only HA2 regions which are deeply inserted in the fused membrane [10, 13-16]. Exogenous addition of many HA2 constructs to solutions of vesicles (often negatively-charged) leads to vesicle fusion with greater fusion at pH 5.0 than pH 7.4. [17-21]. Fusion is typically greater for a construct containing both the FP and the SE than for either the FP or SE alone. Fusion is impaired for the G1E mutant in the FP region which matches the effect of this mutant on HA catalyzed cell/cell fusion. There is also leakage of small molecules out of the vesicles and the rates of leakage and fusion are comparable [19]. Leakage has also been observed in virus/vesicle fusion and is likely not problematic for intracellular influenza fusion so there probably is not evolutionary bias against it [7]. Such bias is expected for other viruses which fuse directly with the plasma membrane because leakage could disrupt cell homeostasis and result in cell death.

There is a crystal structure of the HA2 ectodomain (FP+SE) in its initial complex with HA1 [22]. Three HA2 molecules form a trimer in complex with three HA1 molecules. A structure of the SE without HA1 also shows three molecules each with hairpin structure with *N*-terminal regions (residues 38-105) forming a helical bundle followed by a 180° turn and *C*-terminal regions on the outside of the bundle and antiparallel to the *N*-terminal regions [23]. This is probably the final HA2 state in fusion. Although the FP and TM are not in the construct, the structure implies that they both on the same end of the hairpin which is opposite the hairpin turn.

There are also several structures of the monomeric FP in detergent-rich media with significant differences among these structures [24-26]. In particular, a peptide composed of the 20 *N*-terminal HA2 residues has (*N*-helix)-(open turn)-(C-helix) structure at pH 5 and (*N*-helix)-

(open turn)-(C-coil) structure at pH 7. At pH 5, the structure is an “open boomerang” with an oblique angle between the *N*- terminal and *C*-terminal helices. In some contrast, a peptide composed of the 23 *N*-terminal HA2 residues has (*N*-helix)-(tight turn)-(C-helix) “closed hairpin” structure in detergent at both pH 5 and 7 [27, 28] The helices are antiparallel with close contact including hydrogen bonds between the *N*- and *C*- helices. The hydrophobic surfaces of the open and closed structures are also different and are located in the interhelical pocket (open structure) or on a shared face of the two helices (closed structure). These different hydrophobic surfaces are the basis for different models of membrane binding, either membrane insertion of the *N*-helix (open structure) or membrane surface location of the hydrophobic face (closed structure).

To our knowledge, neither peptide catalyzes fusion between detergent-rich micelles or bicelles but both peptides catalyze inter-vesicle fusion for peptide: lipid mole ratio of ~1:50 [51]. Fusion is moderately higher for the 23-residue relative to the 20-residue peptide and for pH 5 relative to pH 7. Fusion for both peptides at both pHs is difficult to understand based on the very different structures reported in detergent but is well-correlated to their structures in membrane. Both peptides adopt (*N*-helix)-(tight turn)-(C-helix) structure in membrane at both pH 5 and 7 [27, 28]. The helices can either pack tightly in the closed structure or a little less tightly in a semi-closed structure. For both peptides at both pHs, there are significant populations of both structures with greater semi-closed population at pH 5. The moderate differences in vesicle fusion correlate well with moderate differences in hydrophobic surface areas, with greater surface area for the 23-residue relative to the 20-residue peptide and for the semi-closed relative to the closed structure.

The present work enhances our understanding of HA2-mediated fusion. Purified full length HA2 is produced in mg quantities after expression in bacteria as well as the shorter constructs SHA2 (SE), FHA2 (FP + SE), and SHA2-TM (SE + TM). All constructs can form folded trimers in detergent. At pH 5.0, the constructs are positively-charged and induce significant fusion of negatively-charged vesicles whereas at pH 7.4, the constructs are negatively-charged and induce negligible fusion. The contribution of attractive protein-vesicle electrostatics is further supported by significant HA2-induced fusion of positively-charged vesicles at pH 7.4 with much less fusion at pH 5.0. Comparison of fusion among constructs reveals that the SE is the most important region for fusion with moderate effect of the FP and little effect of the TM.

## **3-2 Materials and Methods**

### ***3-2-1 Abbreviations***

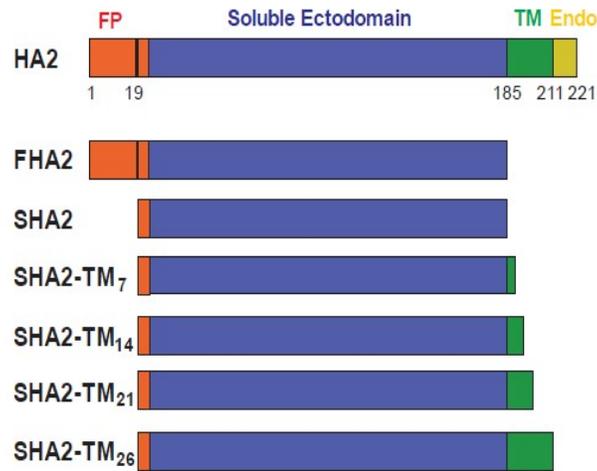
Buffer A, 50 mM phosphate, 0.5% SRC, 50 mM sodium phosphate, and 300 mM NaCl at pH 8.0; Buffer B, 10 mM Tris-HCl, 0.17% DM, and 1 mM DTT at pH 7.4; Buffer C, 10 mM Tris-HCl, 0.10% SRC, and 1 mM DTT at pH 7.4; CD, circular dichroism; DM, decylmaltoside; DOTAP, 1, 2-dioleoyl-3-trimethylammonium-propane (chloride salt); DTT, dithiothreitol; Endo, endodomain; FP, fusion peptide; HA, hemagglutinin; HA-cell, HA-expressing cell; HEPES, 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid; IB, inclusion body; IPTG, isopropyl  $\beta$ -D-1-thiogalactopyranoside; LB, Luria-Bertani; MES, 2-(N-morpholino) ethanesulfonic acid; N-NBD-DPPE, N-(7-nitro-2, 1, 3-benzoxadiazol-4-yl) (ammonium salt) dipalmitoylphosphatidylethanolamine; N-Rh-DPPE, N-(lissamine rhodamine B sulfonyl) (ammonium salt) dipalmitoylphosphat-

idylethanolamine; POPC, 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine; POPG, 1-palmitoyl-2-oleoyl-sn-glycero-3-[phospho-rac-(1-glycerol)] (sodium salt); RBC, red blood cell; RP, recombinant protein; SE, soluble ectodomain; SEC, size exclusion chromatography; SRC, N-lauroylsarcosine; SSNMR, solid-state nuclear magnetic resonance; TM, transmembrane; Tris-HCl, tris (hydroxymethyl)aminomethane.

### ***3-2-2 Constructs, bacterial culture, and protein expression***

Figure 3-1 (A) displays a schematic of the HA2 constructs of the present study with color coding of different regions. Figure 3-1(B) displays the amino acid sequence of full-length HA2 (residues 1-221) with color coding matching Figure 3-1(A). The sequence corresponds to the X31 strain of influenza virus except that all native cysteines are mutated to other amino acid types, either serine (residue 137) or alanine (residues 144, 148, 195, 199, 210, 217, and 220). Non-native residues at the C-terminus are color-coded black in the sequence and include a H6 tag for affinity chromatography and preceding G<sub>6</sub> to increase H<sub>6</sub> exposure during this chromatography.

Other constructs include FHA2 (residues 1-185)  $\equiv$  full ectodomain  $\equiv$  FP+SE, SHA2 (residues 20-185)  $\equiv$  primarily SE, and SHA2-TM<sub>7</sub>, SHA2-TM<sub>14</sub>, SHA2-TM<sub>21</sub>, and SHA2-TM<sub>26</sub>  $\equiv$  SE + 7, 14, 21, and 26 TM residues, respectively. The term SHA2-TM refers to these latter four constructs. The amino acid sequences of all constructs are given in Figure 3-2.



(A)

```

1  GLFGAIAGFIENGWEGMIDGWYGFRHQNSE
31 GTGQAADLKSTQAAIDQINGKLNRVIEKTN
61 EKFHQIEKEFSEVEGRIQDLEKYVEDTKID
91 LWSYNAELLVALENQHTIDLTDSEMNKLFE
121 KTRRQLRENAEEMGNGSFKIYHKADNAAIE
151 SIRNGTYDHDVYRDEALNRRFQIKGVELKS
181 GYKDWILWISFAISAFLLAVVLLGFIMWAA
211 QRGNIRANIAIGGGGGGLEHHHHHH

```

(B)

```

ggctgttcgggtgctatcgctggctttattgaaaacgggtgggaaggcatgatcgacggc
tggtagcggcttcccatcaaaaactcagaaggcaccggtcaggcggcgatctgaaaagc
acgcaggcagctattgaccaaactcaacggcaaactgaatcgtgtgatgaaaaaaccaac
gaaaaattccatcagatcgaaaagaattttctgaagtcgaaggtcgattcaagatctg
gaaaaatgtggaagatacgaaaatcgacctgtggcatacaacgcggaactgctggtt
gccctggaatcagcacaccattgatctgacggactcggaatgaacaaactgttcgaa
aaaaccgctgccaaactcgtgaaaacgcagaagaatgggcaacggtagtttcaaatc
taccataagctgataacgcggcattgaatccatccgcaatggcacgtatgatcacgac
gttaccgtgacgaagcgtgaacaatcgcttcagattaaaggcgtcgaactgaaatcc
ggttcaaaagattggattctgtggatcagcttgcatttctgcttctgctggcctg
gctctgctgggttcatcatgtggcgccgagcgtggcaacattcgtgcaaacatcgca
atcggtagcgggtggcgccgctcagaccaccaccaccac

```

(C)

Figure 3-1 (A) Displays a schematic of the HA2 constructs with domains colored: fusion peptide (FP), red; soluble ectodomain (SE), blue; transmembrane (TM) domain, green; and endodomain (Endo), yellow. The SHA2-TM<sub>7</sub>, SHA2-TM<sub>14</sub>, SHA2-TM<sub>21</sub>, and SHA2-TM<sub>26</sub> constructs respectively include 7,

Figure 3-1 (cont'd) 14, 21, and 26 residues of the transmembrane domain. (B) displays the amino acid sequence of the HA2 construct with color-coding matching (A). The sequence has a non-native C-terminal region in black which includes a H<sub>6</sub> tag for affinity chromatography. (C) DNA sequence of full length HA2 corresponding to amino acid sequence in figure 3-1(B) .

GLFGAIAGFIENGWEGMIDGWYGRHRQNSEG TGQAADLKSTQ  
AAIDQINGKLN RVIEKTNEKFHQIEKEFSEVEGRIQDLEKYVEDT  
KIDLWSYNAELLVALENQHTIDLTDSEMKNLFEKTRRQLRENAE  
EMGNGSFKIYHKADNAAIESIRNGTYDHDVYRDEALNNRFQIKG  
VELKSGYKDWLEHHHHHH

(A) FHA2 amino acid sequence

CWYGRHRQNSEG TGQAADLKSTQAAIDQINGKLN RVIEKTNEKFHQ  
IEKEFSEVEGRIQDLEKYVEDTKIDLWSYNAELLVALENQHTIDLTD  
SEMKNLFEKTRRQLRENAEEMGNGSFKIYHKADNAAIESIRNGTYD  
HDVYRDEALNNRFQIKGVELKSGYKDWLEHHHHHH

(B) SHA2 amino acid sequence

CWYGRHRQNSEG TGQAADLKSTQAAIDQINGKLN RVIEKTNEKFHQ  
IEKEFSEVEGRIQDLEKYVEDTKIDLWSYNAELLVALENQHTIDLTD  
SEMKNLFEKTRRQLRENAEEMGNGSFKIYHKADNAAIESIRNGTYD  
HDVYRDEALNNRFQIKGVELKSGYKDWILWISFAISAFLLAVVLLGFI  
MWAAQGGGGGGGLEHHHHHH

(C) SHA2-TM<sub>26</sub> amino acid sequence

Figure 3-2 Amino acid sequences of (A) FHA2 (B) SHA2 (C) SHA2-TM<sub>26</sub> constructs with color coding matching Figure 3-1 (A). The non-native C-terminal regions are shown in black and include a H<sub>6</sub> tag for affinity chromatography. For constructs that include the TM region G<sub>6</sub> is added as a flexible region to increase H<sub>6</sub> exposure during chromatography.

### ***3-2-3 Protein expression***

The DNA coding for each construct was inserted in a pUC57-Kan plasmid with codon optimization for *E.coli*. The sequence of the construct was then subcloned as discussed in chapter 1 via the NdeI and XhoI restriction sites into a pet24a(+) plasmid that contained the Lac operon and kanamycin resistance. The DNA sequence of the HA2 insert is given in Figure 3-1(C). The plasmid was transformed into *E. coli* cells, BL21(DE3) strain, followed by preparation of 1 mL stock cultures with 50% glycerol that were then stored at  $-80\text{ }^{\circ}\text{C}$ . Culture growth was initiated by adding bacterial stock to a flask containing 50 mL LB broth and kanamycin antibiotic (50 mg/L). After growth overnight at  $37\text{ }^{\circ}\text{C}$  with stirring at 180 rpm, the culture was added to a baffled flask containing 1 L of fresh LB medium with kanamycin. After growth to  $\text{OD}_{600} \approx 0.5$ , recombinant protein (RP) expression was induced with addition of IPTG to 1 mM final concentration. Expression continued for five hours at  $37\text{ }^{\circ}\text{C}$  followed by centrifugation at  $9000g$  for 10 minutes at  $4\text{ }^{\circ}\text{C}$ . The harvested cell pellet was stored at  $-20\text{ }^{\circ}\text{C}$ .

### ***3-2-4 RP solubilization and purification***

For HA2 and SHA2-TM, much purer RP was obtained by first removing much of the native cellular material before performing RP solubilization. The separation began with sonication of  $\sim 5$  g wet cell mass in 40 mL buffer (10 mM  $\text{Na}_2\text{HPO}_4$ , 2 mM  $\text{KH}_2\text{PO}_4$ , 140 mM NaCl, and 3 mM KCl at pH 7.4). Sonication conditions included  $4 \times 1$  minute rounds separated by 1 minute rests. Each round was consist of 80% amplitude for 0.8 s followed by 0.2 s rest. After sonication, the insoluble material was pelleted by centrifugation ( $48000\text{ }g$  for 20 min at  $4\text{ }^{\circ}\text{C}$ ). Relative to whole cells, the pellet was enriched in RP both from the bacterial membrane and

from inclusion bodies. There were two subsequent courses of sonication/centrifugation of the insoluble material.

For all RP's, solubilization was achieved by sonication in 40 mL of buffer A (50 mM phosphate, 0.5% *N*-lauroylsarcosine (SRC) detergent, 50 mM sodium phosphate, and 300 mM NaCl at pH 8.0) with 10 mM imidazole. Sonication of either the HA2- or SHA2-TM-enriched pellet resulted in a clear solution which was then stirred for one hour. No solid was visible after centrifugation which supports complete RP solubilization. RP separation was not used for SHA2 and FHA2 because purified yields of >10 mg/L culture were obtained from sonication of whole cells in buffer A with 10 mM imidazole followed by centrifugation and purification of the soluble lysate.

### ***3-2-5 IMAC purification of RP***

The purification procedure began with addition of 1 mL of Co<sup>2+</sup> affinity resin to the ~40 mL solution containing RP. Binding of RP to the resin was achieved during one hour of agitation at ambient temperature. The RP-coated beads were separated by pouring the suspension through a fritted column. Weakly-bound proteins were removed from the beads by addition of buffer A containing 10 mM imidazole (0.75 mL, 3×). RP was then eluted from the beads by addition of buffer A containing 250 mM imidazole (0.50 mL, 4×). RP was quantified using  $A_{280}$  as well as the Bradford assay.

### ***3-2-6 RP Identification using western blot***

The RP was initially identified by band corresponds to the expected molecular weight in the SDS-PAGE by running the elutions. Further identification of proteins was carried out by running a western blot. After running SDS-PAGE with purified proteins, it was transfer to the nitrocellulose membrane. Membrane was incubated in 10 ml of TBST at pH 7.6 containing anti His antibody and 5% (w/v) nonfat dry milk. The membrane was developed using Supersignal West Pico chemiluminescent substrate.

### ***3-2-7 RP Identification using Mass spectrometry***

RP identity was also investigated by subjecting the band from SDS-PAGE to trypsin digestion, sequencing the resultant peptides by tandem mass spectrometry, and matching with continuous regions of the RP sequence. This was done by submitting the corresponding RP band present in the SDS page to the MSU Proteomics Facility.

Additional accurate mass analysis was done using reverse-phase HPLC coupled to a time-of-flight mass spectrometer (Waters, Xevo G2-S QTOF). Sample preparation included protein precipitation from the eluent using acetonitrile followed by centrifugation. The protein pellet was dissolved in lab stock formic acid and the solution was injected into the LC-MS instrument. Instrument parameters included a C18 analytical column and electrospray ionization in positive ion and continuum modes. The protein masses were derived from the spectrum using the maximum entropy algorithm, MaxEnt1.

### ***3-2-8 Detergent exchange and refolding***

CD spectra could not be obtained from the RP eluent because of ultraviolet absorption by imidazole and SRC detergent. Imidazole was removed and SRC exchanged for decylmaltoside (DM) detergent by dialysis against buffer B (10 mM Tris-HCl, 0.17% DM, and 1 mM DTT at pH 7.4). This dialysis and subsequent dialyses conducted done at 4 °C over two days with four buffer changes. Precipitation was not observed during dialyses. The CD spectra supported folding for SHA2 and FHA2 but not for HA2 and SHA2-TM. The latter eluents were therefore subjected to a refolding protocol in which eluent was added to two-fold excess ice-cold buffer (10 mM Tris-HCl, 0.17% DM, 2 mM EDTA, and 1M *L*-arginine at pH 8.0) [29, 30] After overnight agitation at 4 °C, the arginine and EDTA were removed by dialysis against buffer B.

### ***3-2-9 Circular dichroism (CD)***

Spectra were obtained with a Chirascan instrument (AppliedPhotophysics). Parameters included: (1) ~20 μM protein concentration; (2) 1 mm path length; (3) 260-190 nm spectral range with 0.5 nm wavelength increments and 1.5 s averaging time; and (4) summing of three scans. The final spectrum was the (protein + buffer) – (buffer only) difference. For each of the four constructs (SHA2, FHA2, SHA2-TM26, and HA2), the “textbook” α helix shape of the CD spectrum was reproducible among replicate samples. There was some variability in the magnitude of the absolute molar ellipticity among replicate samples, probably because of  $A_{280}$  measurement error with the NanoDrop instrument. Comparative spectra of the four constructs at ambient temperature were obtained by: (1) simultaneous purification and sample preparation; and (2) successive spectral acquisition within a single day. This approach resulted in similar molar ellipticities among the constructs. For each construct, a different sample was prepared for

which spectra were obtained at a series of increasing temperatures. These samples were prepared on different days and there was greater variability of absolute molar ellipticities between samples.

### ***3-2-10 Cross-linking***

Protein in buffer B was dialyzed into buffer containing 20 mM HEPES and either 0.17% DM or 0.1% SRC at pH 7.4. Cross-linking between lysine  $-NH_2$  groups of the RP molecules was done with bis (sulfosuccinimidyl) suberate. Cross-linking was done with  $[RP] = 0.5$  mg/mL and with 50-fold molar excess cross-linking agent. The reaction was allowed to proceed at room temperature for one hour and then quenched by adding Tris-HCl at pH 6.8 with final  $[Tris-HCl] = 50$  mM. The oligomer sizes were analyzed by using SDS-PAGE gel.

### ***3-2-11 Size-exclusion chromatography (SEC)***

The SEC columns require  $[NaCl] \geq 150$  mM in the running buffer to inhibit electrostatic interaction between the protein and the column. The protein solution in buffer B was therefore dialyzed against buffer B with 150 mM NaCl or against buffer C (10 mM Tris-HCl, 0.10% SRC detergent, and 1 mM DTT at pH 7.4) with 150 mM NaCl. The chromatography was performed with a DuoFlow Pathfinder 20 instrument (Bio-Rad) with Tricorn Superdex 200 semi-preparative column (General Electric). The column was equilibrated with dialysis buffer before each run. Parameters included 0.8 mg protein/mL loading concentration, 0.3 mL/min flow rate, and  $A_{280}$  detection.

### 3-2-12 Protein-induced vesicle fusion

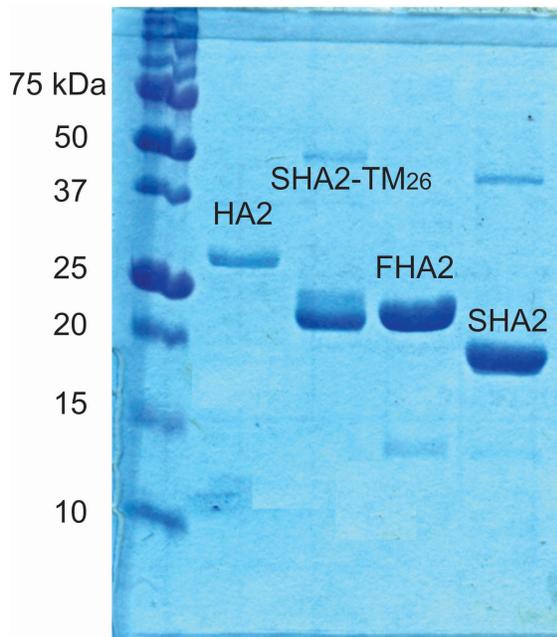
Lipid was dissolved in chloroform followed by chloroform removal with nitrogen gas and vacuum pumping. The film was homogenized by freeze-thaw cycles in ~1 mL buffer and extruded through 100 nm diameter pores to form unilamellar vesicles. The buffer contained 5 mM HEPES/10 mM MES at either pH 5.0 or 7.4. Fluorescently labeled vesicles were similarly prepared except that the mixture also contained 2 mole% of the fluorescent lipid *N*-NBD-PE and 2 mole% of the quenching lipid *N*-Rh-PE. Labeled and unlabeled vesicles were mixed in 1:9 ratio with total [lipid]  $\approx$  150  $\mu$ M. Fluorescence of the stirring vesicle solution was measured at 37  $^{\circ}$ C with 467 nm excitation, 530 nm detection, and 1 s time increment. After measurement of the baseline fluorescence  $F_0$ , a protein aliquot from stock was added and marked time  $t = 0$ . Vesicle fusion was reflected in the increased fluorescence  $\Delta F(t) = F(t) - F_0$  due to longer distances between fluorescent and quenching lipids in a fused (labeled+unlabeled) vesicle relative to the initial labeled vesicle. The dead-time in the assay was  $\sim$ 5 s and asymptotic fluorescence ( $\Delta F_f$ ) was usually reached by  $\sim$ 600 s. The maximum fluorescence change ( $\Delta F_{max}$ ) was detected after addition of 12  $\mu$ L 10% Triton X-100 which solubilized the vesicles. Percent fusion was  $M(t) = [\Delta F(t)/\Delta F_{max}] \times 100$ . Comparison among assay replicates showed  $\delta(M_f)/M_f \approx 0.02$ . The protein stock contained 40  $\mu$ M protein in 10 mM Tris-HCl at pH 7.4 with 1 mM DTT and either 0.17% DM (HA2 and SHA2-TM) or 0.10% SRC (FHA2 and SHA2). For these detergent choices, the trimer is the dominant protein oligomeric species. No fusion was detected after addition of either detergent solution without protein.

### 3-3 Results

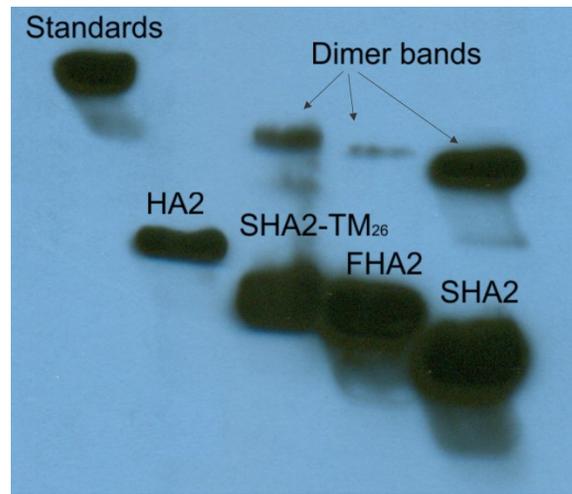
#### *3-3-1 RP solubilization and purification*

The first approach was sonication of the cell mass in Buffer A containing 0.5% SRC detergent followed by centrifugation and  $\text{Co}^{2+}$  affinity chromatography of the supernatant. For SHA2 and FHA2, SDS-PAGE of the purification eluent showed highly pure RP with respective yields of ~15 and ~10 mg/L culture (Figure 3-3). SHA2 was solubilized even with PBS buffer but purification resulted significant amount of impurities (Figure 3-5). In some contrast, for SHA2-TM or full-length HA2, there were significant impurities after purification (Figure 3-4). Low expression was likely not the problem because analysis of solid-state NMR spectra of RP-expressing cells has always  $\geq$  shown 100 mg RP/L culture including for RP $\equiv$  FHA2 [31].

Many native proteins are soluble in PBS whereas there is low solubility for SHA2-TM and HA2. Separation based on this solubility difference was done by sonication of the cell mass in PBS followed by centrifugation and discarding the supernatant enriched in cell protein. The procedure was repeated three times and followed by complete solubilization of the RP-enriched pellet in buffer containing SRC, and subsequent  $\text{Co}^{2+}$  affinity chromatography. This protocol resulted in high-purity SHA2-TM and HA2 with purified yield of ~2 mg HA2/L culture (Figure 3-3(A)). The proteins are primarily monomeric in SDS-PAGE with a small fraction of dimer.

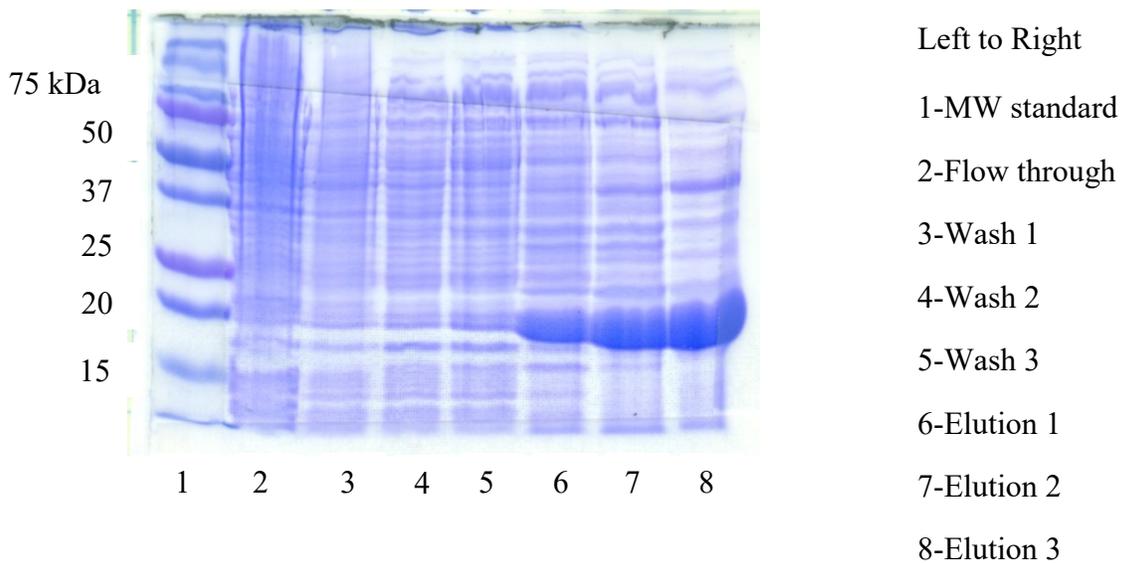
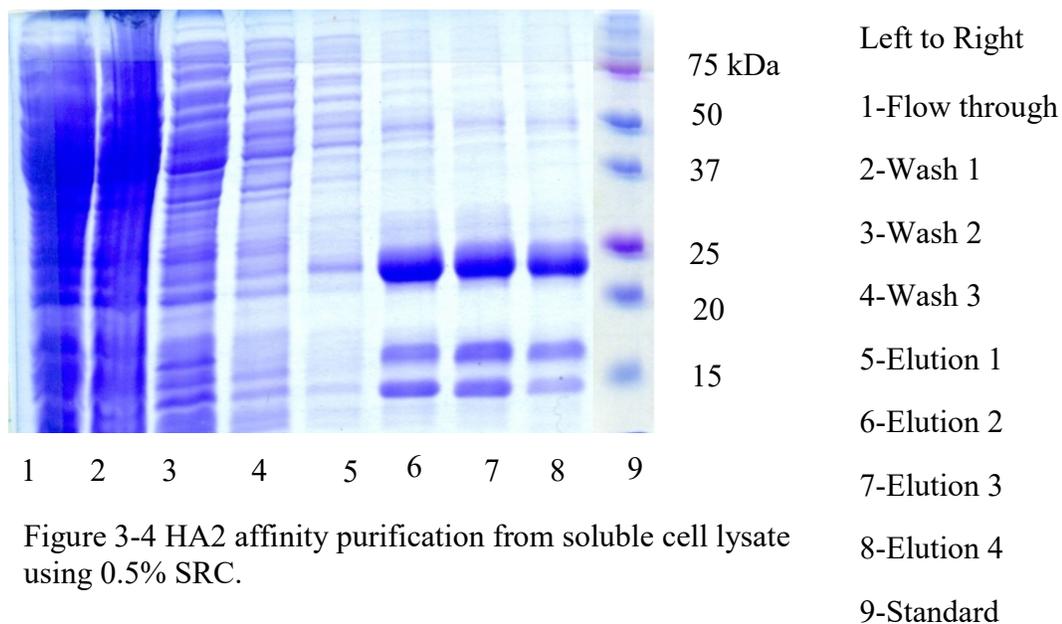


A



B

Figure 3-3 Panel (A) displays SDS-PAGE of the purification eluents and Panel (B) is the corresponding anti-H<sub>6</sub> Western blot. The expected MW's are: HA2, 26.7 kDa; SHA2-TM<sub>26</sub>, 23.7 kDa; FHA2, 22.4 kDa, and SHA2, 20.4 kDa. Dominant monomer and minor dimer bands are observed where the latter is evidenced by binding to the anti-H<sub>6</sub> antibody and by qualitatively displays SDS-PAGE (panel A) of the purification eluents and panel B is the corresponding anti-H<sub>6</sub> Western blot.



Different solubilizing agents such as 8 M urea and 6 M GuHCl were also tried but the highest yields were obtained with 0.5% SRC (Figure 3-6 (A) and (B)). Higher purified yields were obtained for SHA2-TM with a G<sub>6</sub>LEH<sub>6</sub> C-terminal tag than with a LEH<sub>6</sub> tag (Figure 3-8). Inclusion of the G<sub>6</sub> spacer may result in greater H<sub>6</sub> exposure and consequent better RP binding to the Co<sup>2+</sup> resin.

Solubilization and purification of HA2 was also tried using 0.17% DM, 0.10% dodecylmaltoside (DDM), or 0.10% dodecylphosphocholine (DPC) detergents. There was poorer solubilization as well as much lower final purities and yields with these detergents than with 0.5% SRC (Figure 3-7). Sodium dodecyl sulfate detergent had previously been successfully used for solubilization and purification of the related “Fgp41” full ectodomain construct of the HIV gp41 membrane fusion protein [32]. There was better solubilization of Fgp41 with SDS than with SRC but SDS is poorly exchangeable and we have not yet tried SDS with the HA2 constructs.

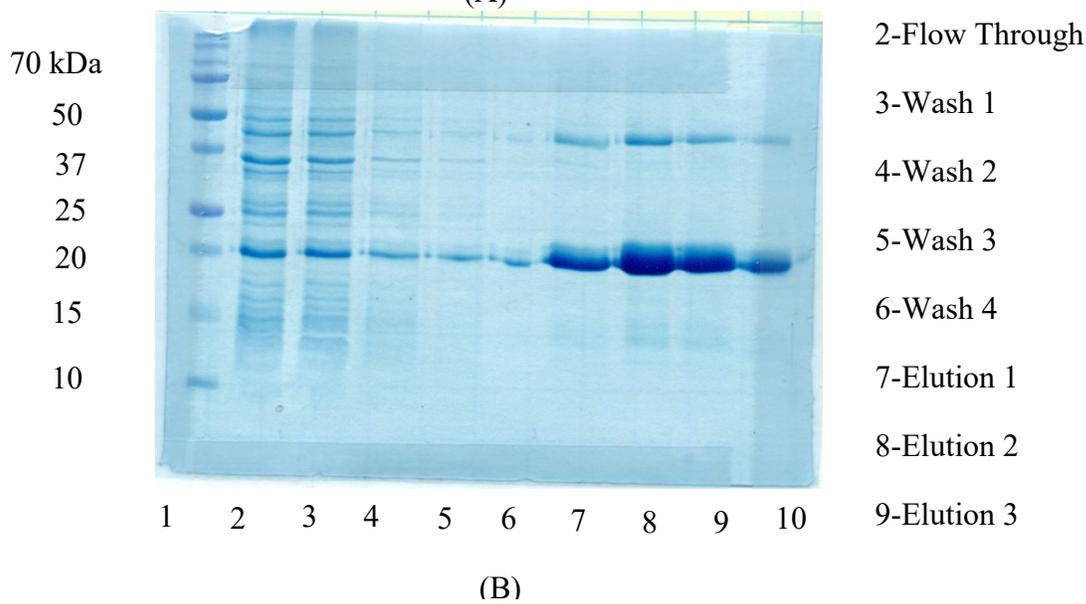
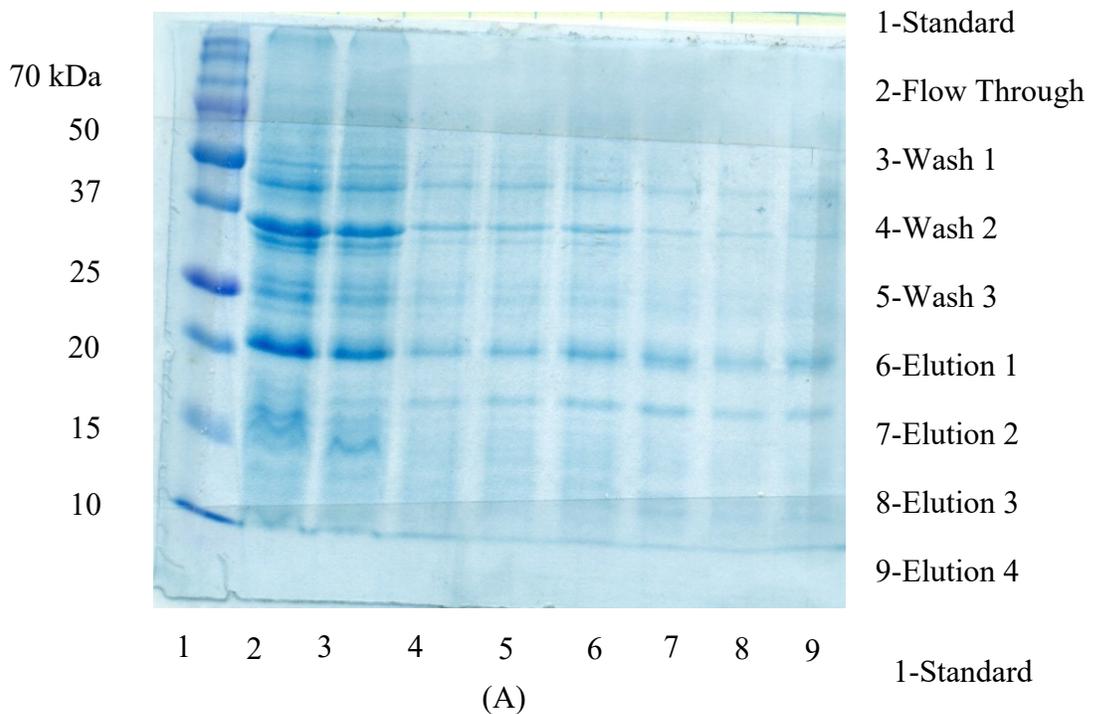
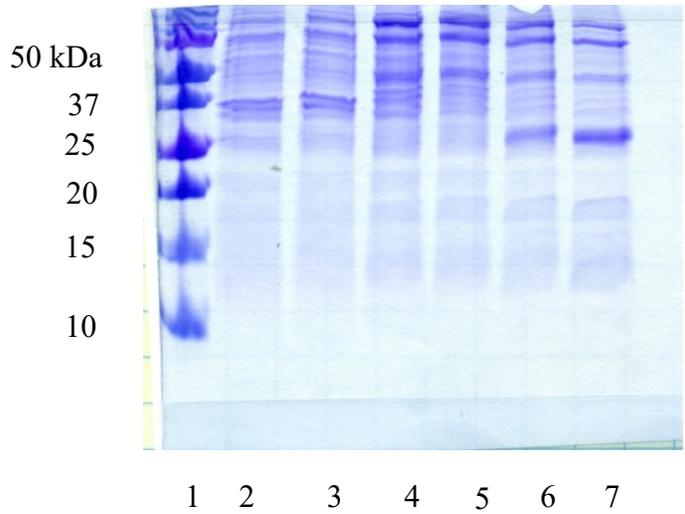
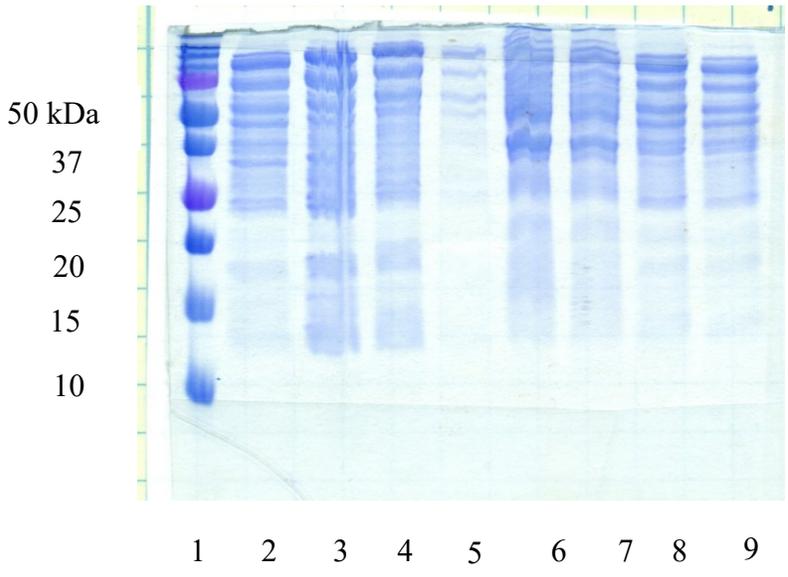


Figure 3-6 SDS-PAGE of purified fractions for SHA2-TM<sub>21</sub>. The cells were first lysed in PBS followed by centrifugation to separate the insoluble material enriched in RP from soluble fraction enriched in bacterial proteins. The procedure was repeated three times and the pellet was then solubilized with the buffer containing either (A) Urea; or (B) 0.5% (w/v) sarkosyl (SRC) with subsequent Co<sup>2+</sup> affinity purification.



- 1-Std
- 2-Flow through
- 3-Wash 1
- 4-Wash 2
- 5-Wash 3
- 6-Elution 1
- 7-Elution 2

(A)

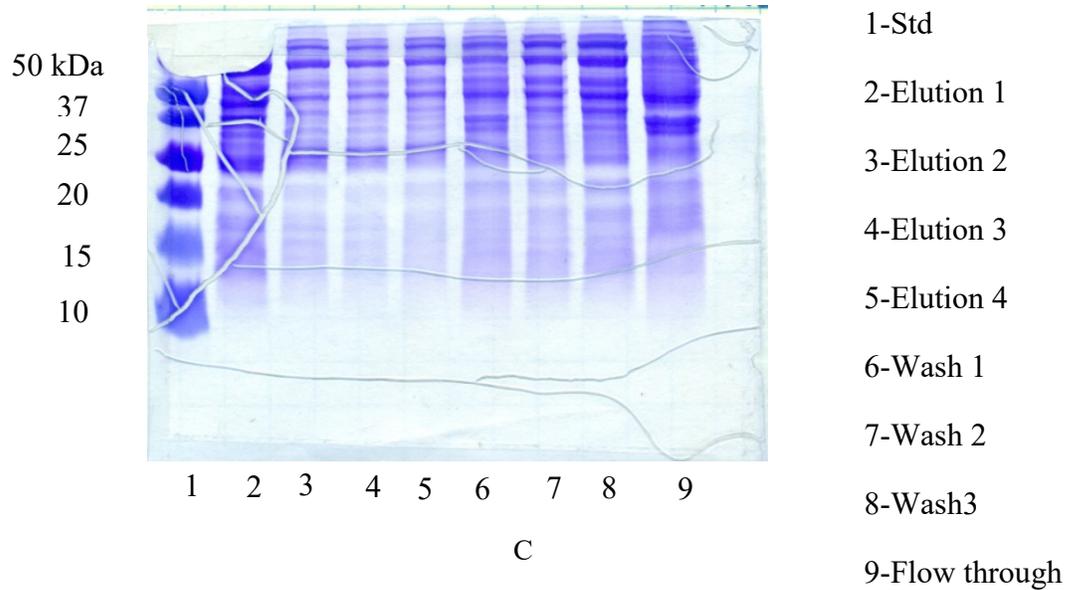


- 1-Std
- 2-Flow through
- 3-Wash 1
- 4-Wash 2
- 5-Wash 3
- 6-Elution 1
- 7-Elution 2
- 8-Elution 3
- 9-Elution 4

(B)

Figure 3-7 SDS-PAGE of affinity purified HA2 . The cells were first lysed in PBS

Figure 3-7 (cont'd)



followed by Centrifugation to separate the insoluble material enriched in RP from soluble fraction enriched in bacterial proteins. The procedure was repeated three times and the pellet was then solubilized with the buffer containing either (A) 0.17% (B) 0.1% DDM (C) 0.1% DPC with subsequent  $\text{Co}^{2+}$  affinity purification.

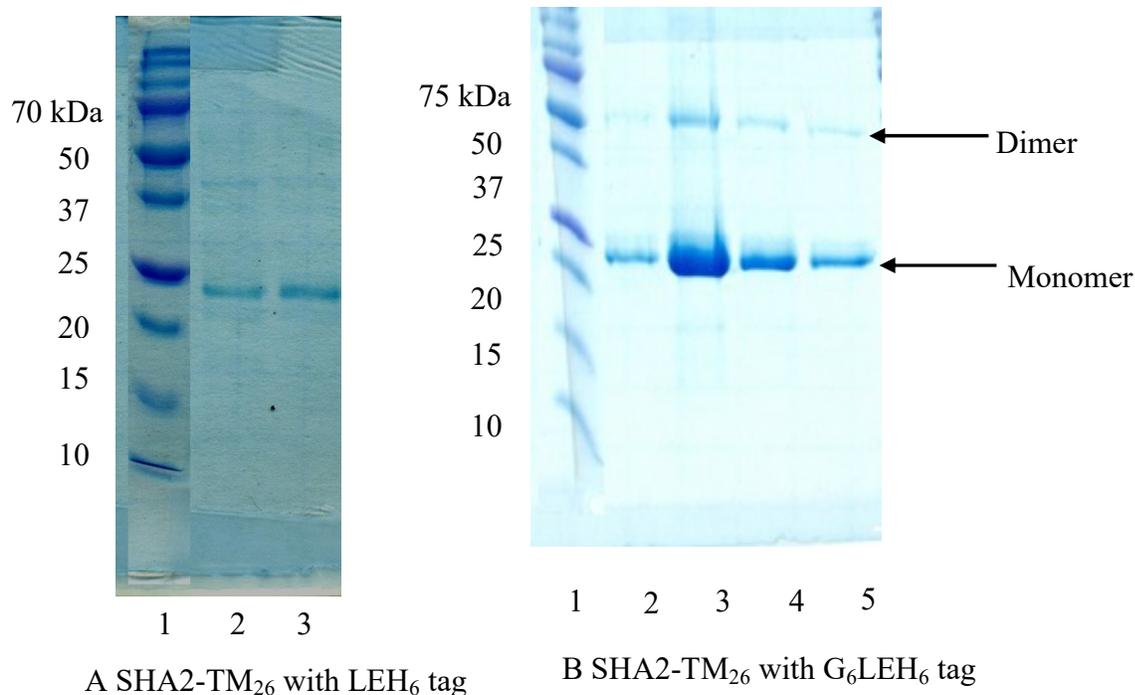


Figure 3-8 SDS-PAGE of purification eluents for SHA2-TM<sub>26</sub> with either a (A) LEH<sub>6</sub> or (B) G<sub>6</sub>LEH<sub>6</sub> terminal tag. Sequential 0.5 mL elutions are shown in (A) lanes 2 and 3 and (B) lanes 2-5. Intermediate lanes in gel A were cropped for clarity.

### 3-3-2 RP confirmation by mass spectrometry

The SDS-PAGE band thought to be RP was digested with trypsin and the fragment peptides were sequenced by tandem mass spectrometry (Figure 3-9 (A)). For all RP's, there is good matching between the peptides and continuous regions of the RP sequence. It is particularly significant that full-length HA2 has 81% sequence coverage including the FP and endodomain regions. Additional confirmation was obtained from LC-MS of the RP eluents with each experimental mass within 3 Da (0.01%) of the expected mass (Figure 3-9 (B)).

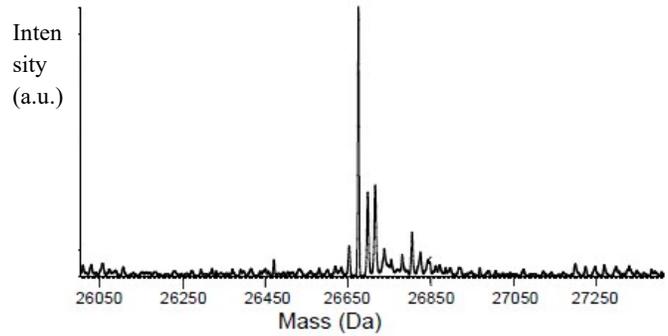
### A Peptide mapping

**HA2**

```

1  GLFGAIAGFIENGWEGMIDGWYGFRHQNSE
31  GTGQAADLKSTQAAIDQINGKLNRVIEKTN
61  EKFHQIEKEFSEVEGRIQDLEKYVEDTKID
91  LWSYNAELLVALENQHTIDLTDSEMNKLFE
121 KTRRQLRENAEEMGNSFKIYHKADNAAIE
151 SIRNGTYDHDVYRDEALNNRFQIKGVELKS
181 GYKDWILWISFAISAFLLAVVLLGFIMWAA
211 QRGNIRANIAIGGGGGGLEHHHHHH
  
```

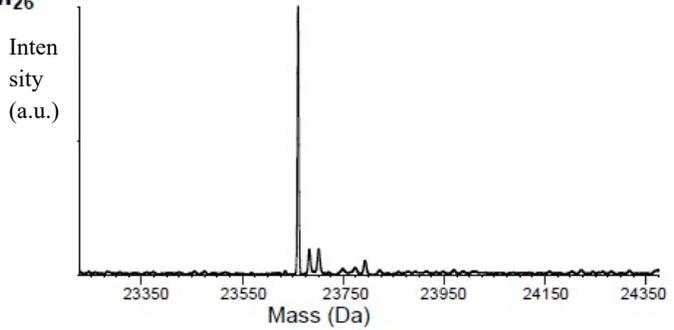
### B Mass spectrum



**SHA2-TM<sub>26</sub>**

```

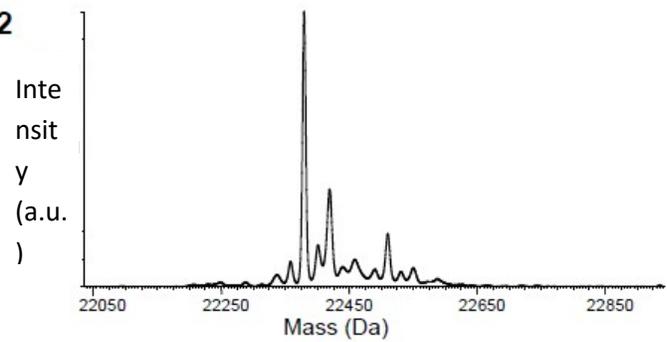
          20 CWYGFRHQNSE
31  GTGQAADLKSTQAAIDQINGKLNRVIEKTN
61  EKFHQIEKEFSEVEGRIQDLEKYVEDTKID
91  LWSYNAELLVALENQHTIDLTDSEMNKLFE
121 KTRRQLRENAEEMGNSFKIYHKADNAAIE
151 SIRNGTYDHDVYRDEALNNRFQIKGVELKS
181 GYKDWILWISFAISAFLLAVVLLGFIMWAA
211 QGGGGGGLEHHHHHH
  
```



**FHA2**

```

1  GLFGAIAGFIENGWEGMIDGWYGFRHQNSE
31  GTGQAADLKSTQAAIDQINGKLNRVIEKTN
61  EKFHQIEKEFSEVEGRIQDLEKYVEDTKID
91  LWSYNAELLVALENQHTIDLTDSEMNKLFE
121 KTRRQLRENAEEMGNSFKIYHKADNAAIE
151 SIRNGTYDHDVYRDEALNNRFQIKGVELKS
181 GYKDWLEHHHHHH
  
```



**SHA2**

```

          20 CWYGFRHQNSE
31  GTGQAADLKSTQAAIDQINGKLNRVIEKTN
61  EKFHQIEKEFSEVEGRIQDLEKYVEDTKID
91  LWSYNAELLVALENQHTIDLTDSEMNKLFE
121 KTRRQLRENAEEMGNSFKIYHKADNAAIE
151 SIRNGTYDHDVYRDEALNNRFQIKGVELKS
181 GYKDWLEHHHHHH
  
```

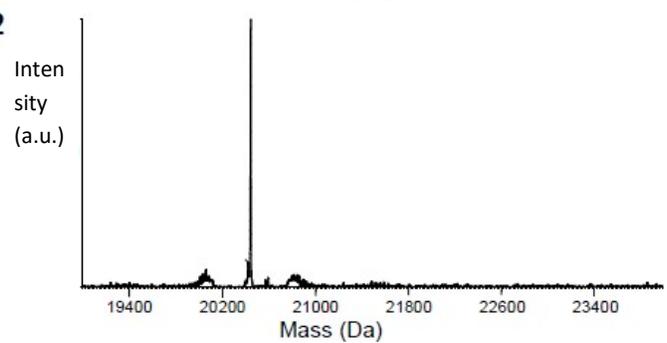
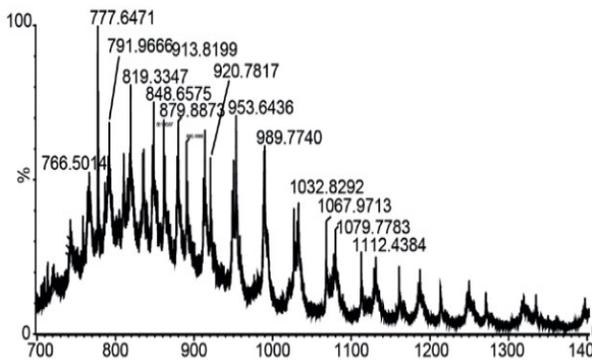


Figure 3-9 (A) Peptide mapping via trypsin digestion and mass spectrometry; and (B)

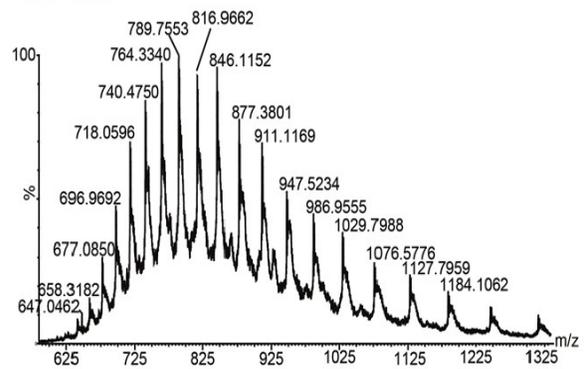
Figure 3-9 (cont'd) analysis of liquid chromatography-electrospray ionization mass spectra. The mapping in panel A was done for the most intense monomer band of the SDS-PAGE of the purification eluent (Figure 3-3). The underlined regions of a sequence correspond to detected peptides and the percent sequence coverages are: HA2, 81%, SHA2-TM<sub>26</sub>, 55%; FHA2, 77%; and SHA2, 65%. The color coding of the sequences matches Figure 3-1. The experimental peak masses in panel B and expected masses are: HA2, 26678 and 26676 Da; SHA2-TM<sub>26</sub>, 23663 and 23663 Da; FHA2, 22379 and 22378 Da; and SHA2, 20444 and 20444 Da. The ESI mass spectra are presented in figure 3-10.

HA2



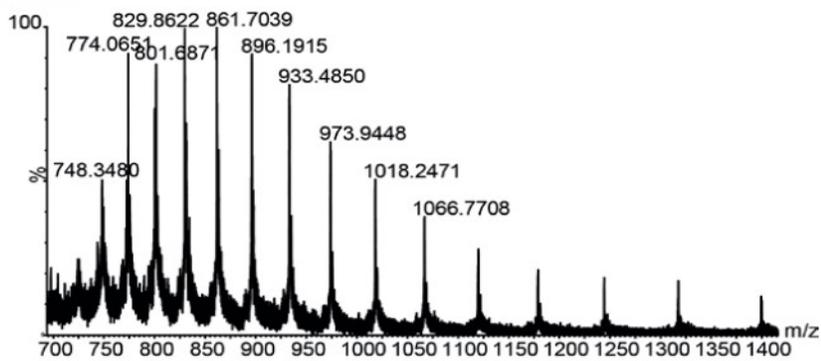
(A)

SHA2-TM<sub>26</sub>



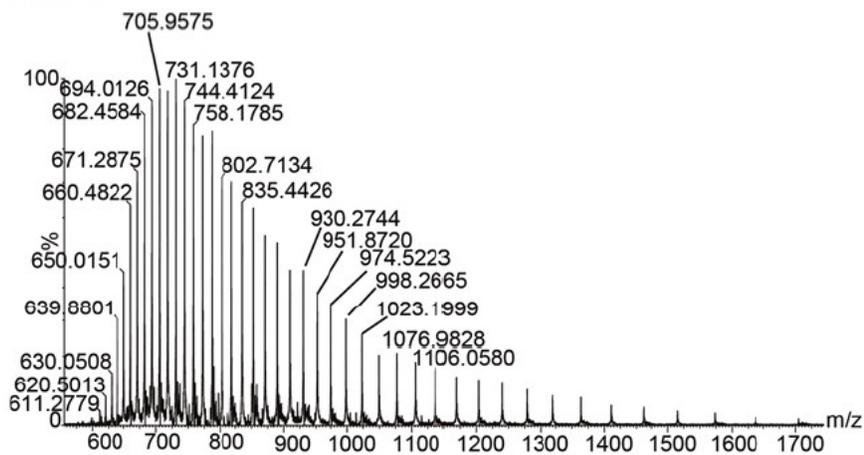
(B)

FHA2



(C)

SHA2



(D)

Figure 3-10 Mass Spectra of (A) HA2 (B) SHA2-TM<sub>26</sub> (C) FHA2 (D) SHA2 after liquid chromatography and electrospray ionization.

### ***3-3-4 Hyperthermostable $\alpha$ helical structure***

Figure 3-11 displays CD spectra of the HA2 constructs in 0.17% DM detergent at pH 7.4. The spectra of the four constructs are similar and have the profile characteristic of proteins with high  $\alpha$ -helical content including minima at 208 and 222 nm. The spectra have similar appearance in 0.30% DM or at pH 9.0 (Figure 3-12 (A) and (B)). The proteins visibly aggregate at either pH 5.0 and 3.0 so spectra were not obtained at these lower pH values. A ~65%  $\alpha$ -helical content is estimated from the experimental  $|\theta_{222}| \approx 2 \times 10^4$  deg cm<sup>2</sup>/dmole-residue and agrees semi-quantitatively with ~60 %  $\alpha$ -helical content calculated for HA2 based on the high-resolution structures of FP and SE fragments and the predicted  $\alpha$ -helical structure in the ~25-residue TM domain [23, 25]

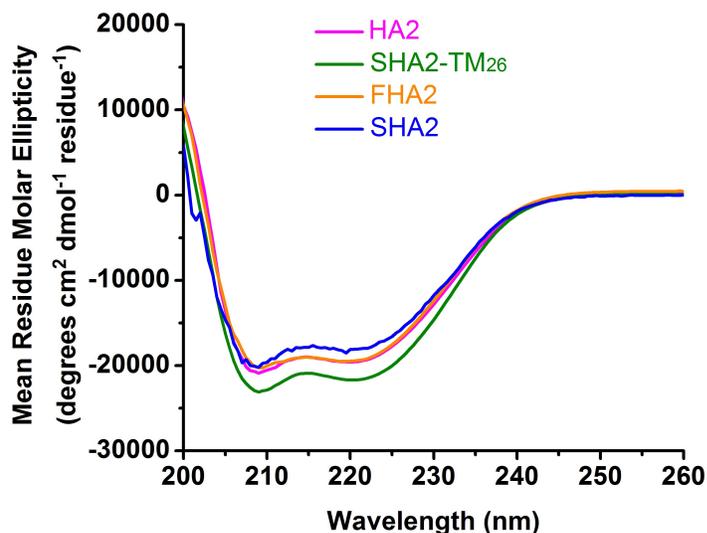


Figure 3-11 Circular dichroism spectra of samples containing  $\sim 20 \mu\text{M}$  protein in buffer with 10 mM Tris-HCl, 0.17% DM detergent, and 1 mM DTT at pH 7.4. The spectra were obtained at ambient temperature.

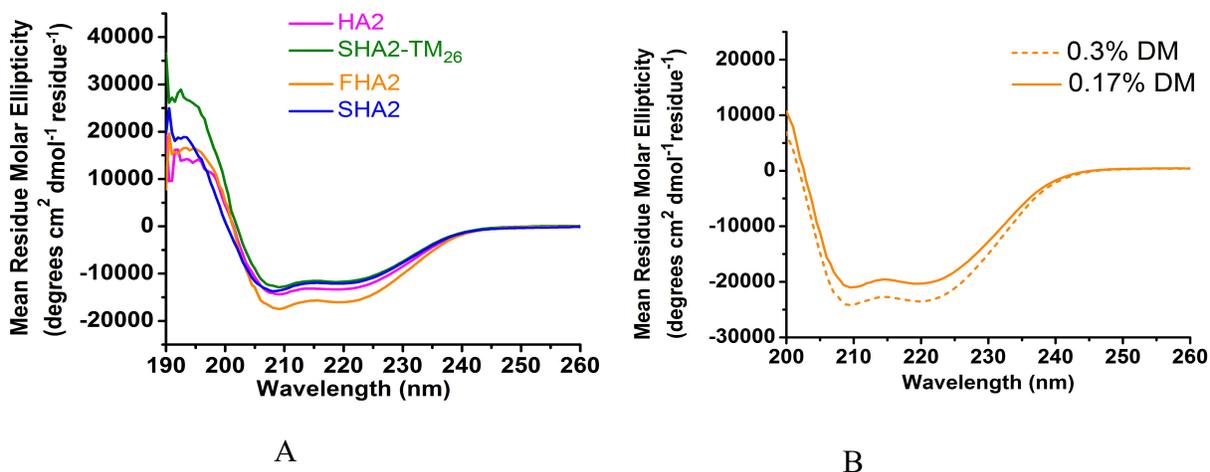


Figure 3-12 Circular dichroism spectra of samples containing  $\sim 20 \mu\text{M}$  protein in (A) 0.17% DM at pH 9.0 and (B) 0.17% and 0.30% DM at pH 7.4.

Spectra were obtained for a series of temperatures between 20 and 85 °C (Figure 3-14) and the  $|\theta_{222}|$  were plotted vs temperature (Figure 3-13). SHA2 exhibits a small linear decrease in  $|\theta_{222}|$

between 20 and 55 °C, greater linear decrease between 55 and 75 °C, and even larger decrease between 75 and 85 °C. This multi-step behavior may be due to the aggregation of SHA2 in DM and greater description is provided in the Discussion section. FHA2, SHA2-TM<sub>26</sub>, and HA2 behave more simply and exhibit small linear decreases in  $|\theta_{222}|$  between 20 °C and ~75, 80, and 85 °C, respectively. For FHA2 and SHA2-TM<sub>26</sub>, there are more significant decreases approaching 85 °C which indicate onset of unfolding. For FHA2, the estimated  $T_m \approx 85$  °C based on  $|\theta_{222}(85$  °C)/  $|\theta_{222}(20$  °C)  $\approx 1/2$  and for SHA2-TM<sub>26</sub>, the  $T_m > 85$  °C. There is no indication of unfolding of full-length HA2 for temperatures  $\leq 85$  °C, so HA2 appears to adopt a hyperthermostable structure which is likely the final HA2 state in HA-mediated fusion. Overall, the CD-detected melting in DM support a thermostable SE in HA2 and the ordering of stability FHA2 < SHA2-TM<sub>26</sub> < HA2 supports additional stabilization associated with the FP and TM.

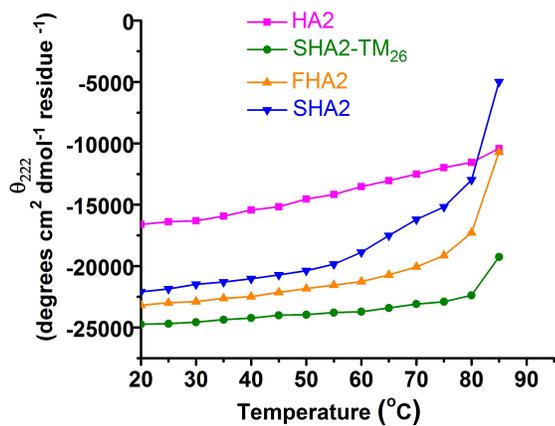


Figure 3-13 Thermal melts of samples containing ~20  $\mu\text{M}$  protein in buffer with 10 mM Tris-HCl, 0.17% DM detergent, and 1 mM DTT at pH 7.4. Temperature is plotted Vs the mean-residue molar ellipticity at 222 nm.

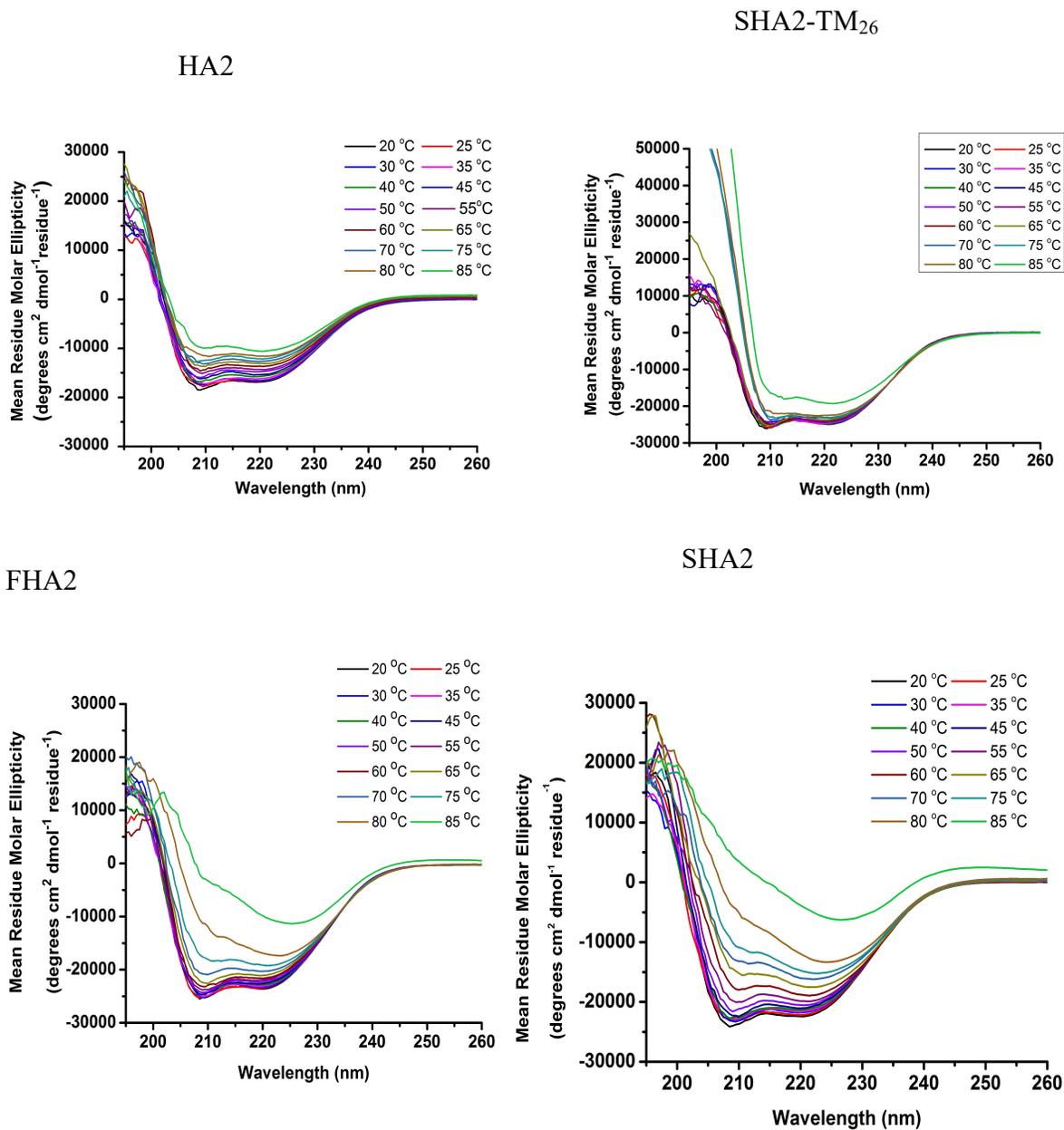


Figure 3-14 Circular dichroism spectra at pH 7.4 in 0.17% DM detergent at temperature

20 °C-85 °C

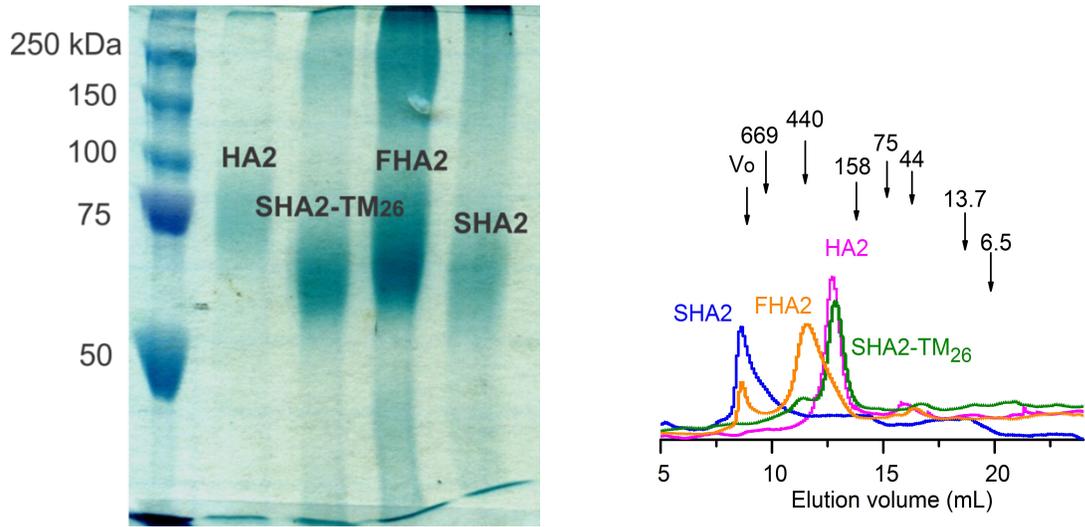
### ***3-3-5 Trimers and higher-order oligomers at pH 7.4***

For [protein] = 0.8 mg/mL at 4 °C, all HA2 constructs remain soluble in pH 7.4 buffer that contains either 0.17% DM or 0.10% SRC detergent, e.g. after one day, there were no visible precipitates after centrifugation of these solutions. The protein oligomeric states were investigated with these two detergent conditions. The cross-linking experiments used [protein] = 0.8 mg/mL and with either DM or SRC, the protein: detergent mole ratio is ~1:110. Although the loading [protein] = 0.8 mg/mL in the SEC, the running concentration is about ten-fold lower, so that the protein: detergent mole ratio is ~1:1100. The 0.17% DM concentration is about 2×CMC and the aggregation number is ~70 so that protein: micelle ratio is ~1:1.6 in cross-linking and ~1:16 in SEC. Although the 0.10% SRC concentration is below its CMC, the aggregation number is ~2 so the SRC “micelle” is just a dimer and there isn’t significant difference in detergent state below vs above the CMC.

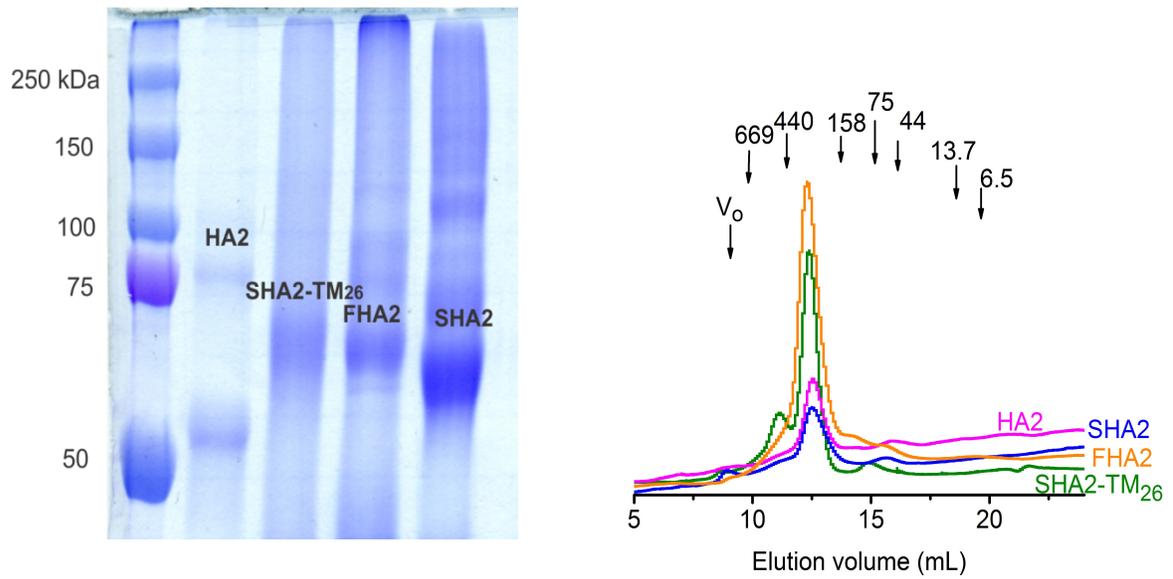
Figure 3-15 displays SDS-PAGE of HA2 constructs after one hour of chemical cross-linking. For protein in either detergent, there is typically a band in the 60-80 kDa range that likely corresponds to a trimer. For SHA2 and FHA2 in DM, there are also bands at higher MW’s that correspond to larger oligomers and for SHA2, a significant fraction of these oligomers are so massive that they don’t migrate in SDS-PAGE. There is only a small fraction of monomer protein (Figure 3-17). The cross-linking is done with folded proteins with significant bound detergent mass, but this mass is not apparent in the SDS-PAGE because of SDS-induced unfolding of proteins, including standards, and SDS binding to the unfolded proteins and replacement of the bound DM or SRC in the HA2 proteins. The MW determination by SDS-PAGE via comparison with the protein standards therefore results in protein-only rather than protein + SDS masses.

SEC was performed with folded non-cross-linked HA2 proteins which migrate with bound detergent whereas the soluble protein standards do not migrate with detergent. The MW's determined by SEC are therefore for the HA2+detergent complexes and include significant contributions from both protein and detergent masses. The SEC of all constructs in SRC shows a dominant oligomeric species with  $MW_{\text{Prot+Det}} \approx 200$  kDa (Figure 3-15 (B)). This is also the dominant species for HA2 and SHA2-TM<sub>26</sub> in DM whereas the dominant species for FHA2 and SHA2 are larger with respective  $MW_{\text{Prot+Det}}$  of  $\sim 400$  kDa and  $\geq 2$  MDa with most SHA2 in the column void volume. For all constructs in DM, there is also a small population with  $MW_{\text{Prot+Det}} \approx 50$  kDa. Figure 3-16 displays the SEC of the MW standards.

### Protein in DM detergent



### Protein in SRC detergent

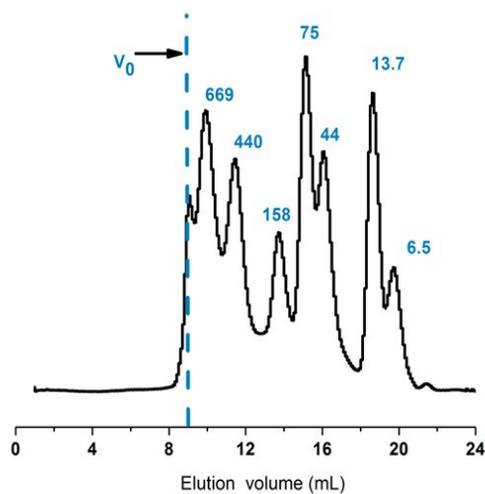


A SDS-PAGE after cross-linking

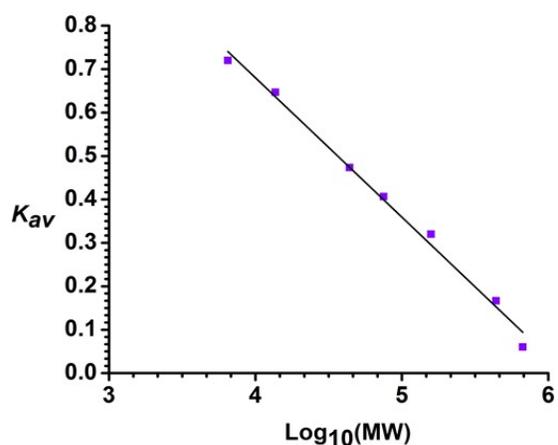
B Size-exclusion chromatography

Figure 3-15 (A) SDS-PAGE after chemical cross-linking and (B) SEC of the proteins in the

Figure 3-15 (cont'd) absence of cross-linking. Solutions contained either (top) 0.17% DM or (bottom) 0.10% SRC detergent at pH 7.4. Cross-linking was done with pure proteins (Figure 3-3) using bis (sulfosuccinimidyl) suberate and the displayed SDS-PAGE are 8% gels with band diffuseness at low MW's. The trimer region is marked. SEC was done with  $A_{280}$  detection and  $\sim 0.8$  mg protein/mL loading. The vertical arrows in the plots mark the elution times of the MW standards with  $V_o \equiv$  column void volume. The SEC of the MW standards is provided in Figure 3-16. A fraction of protein molecules are typically in oligomeric states larger than a trimer as evidenced by diffuse high MW intensity in SDS-PAGE and by  $A_{280}$  intensity at shorter SEC elution times.



(A) SEC of MW standards



(B) SEC analysis

Figure 3-16 SEC (A) plot and (B) analysis of MW standards. The SEC was performed with a Superdex 200 column with  $A_{280}$  detection. For each standard,  $K_{av}$  was calculated as  $[(V_e - V_o)/(V_c - V_o)]$  using  $V_e \equiv$  elution volume of the standard,  $V_o \equiv$  column void volume, and  $V_c \equiv$  column volume. The standards include the thyroglobulin (669 kDa), ferritin (440 kDa), aldolase (158 kDa), conalbumin (75 kDa), Ovalbumin (44 kDa), carbonic anhydrase (29 kDa), ribonuclease A (13.7 kDa), and aprotinin (6.5 kDa).  $V_o$  corresponded to blue dextran with  $\text{MW} = 2 \text{ MDa}$ .

The ~200 kDa species and ~50 kDa species in SEC are respectively assigned to protein trimer and monomer. These assignments are based in part on correlation to the cross-linking data showing major trimer and minor monomer species. The larger MW species in the SEC of FHA2 and SHA2 in DM are also consistent with the cross-linking data. The Discussion section describes additional support for these assignments by correlations with earlier studies of HA2 and HIV gp41 constructs. All constructs form visible aggregates at pH 5.0 in 0.17% DM. In addition, SHA2 forms aggregates at pH 7.4 in the absence of detergent.

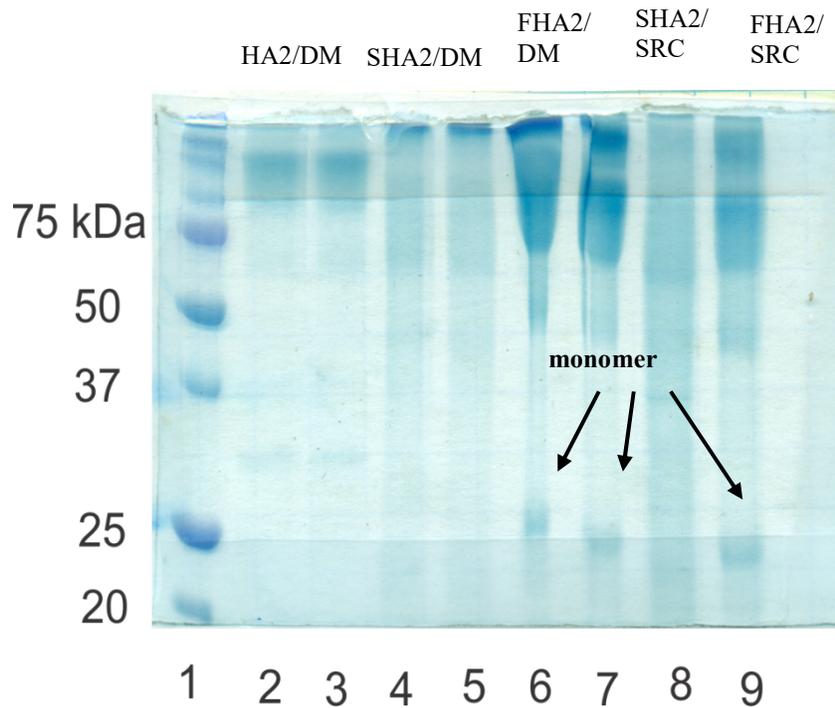


Figure 3-17 SDS PAGE of protein after one hour of cross-linking. The faint marked bands are assigned as protein monomer. Cross-linked oligomers are represented by the diffuse bands with MW's typically >60kDa. The bands of oligomers of different MW's are not resolved. This is a 15% gel.

### ***3-3-6 Vesicle fusion.***

The HA2 fusion activities were probed by monitoring mixing of lipids between vesicles after addition of an aliquot of protein stock. The pH of the vesicle solution was either 7.4 or 5.0 with the latter pH closer to the 5-6 range of influenza virus fusion with endosomes. Figure 3-18 displays fusion at pH 5.0 of vesicles containing 20 mole% negatively charged lipid. All constructs induce significant and comparable rates and extents of vesicle fusion with increased extents for constructs that include the FP, e.g. HA2 vs SHA2-TM<sub>26</sub> and FHA2 vs SHA2. There is typical  $\pm 1\%$  variation in fusion extents among assay replicates so the extents for HA2 and FHA2 are comparable and the extents for SHA2-TM<sub>26</sub> and SHA2 are comparable. Thus, there is not a significant effect of inclusion of the TM- or TM- + endodomains and this is also evidenced by comparable fusion for SHA2-TM constructs with different length segments of the TM domain (Figure 3-19 (A)).

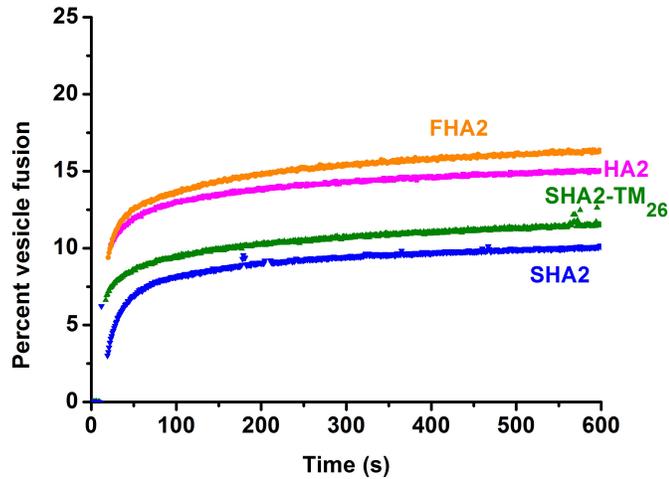


Figure 3-18 Protein-induced fusion of negatively-charged vesicles at pH 5.0. The protein:lipid mole ratio = 1:280. The vesicle composition is POPC:POPG (4:1). For each trace, protein is added a few seconds before the first displayed point. There is typically  $< \pm 1\%$  variation in final extent of vesicle fusion among replicate assays and no fusion is observed with addition of detergent solution without protein .

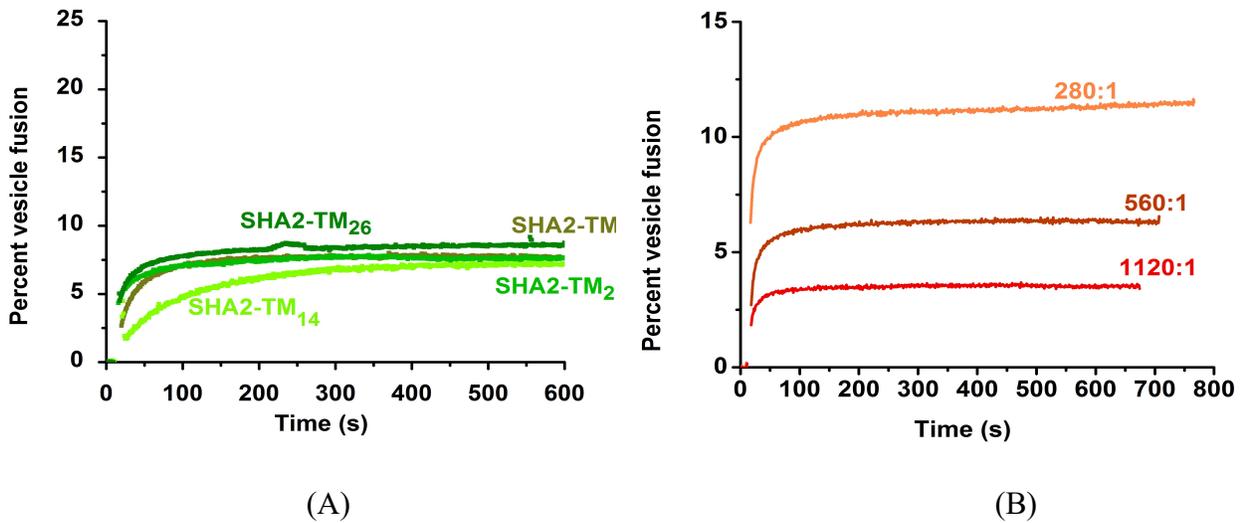


Figure 3-19 Protein-induced fusion of negatively-charged vesicles at pH 5.0 (A) different SHA2-TM constructs (B) dose response for FHA2. The protein: lipid mole ratio for figure (A) = 1:280. The vesicle composition is POPC:POPG (4:1).

The typical fusion rate is estimated to be  $\sim 200 \text{ s}^{-1}$  based on typical achievement of at least half the final fusion extent during the  $\sim 5 \text{ s}$  dead-time of the assay. Dose dependence is observed with significant fusion by FHA2 for protein: lipid = 1:1120 (Figure 3-19 (B)). This corresponds to  $\sim 15$  trimers per 100 nm diameter vesicle (assuming quantitative binding) which is much smaller than the typical number of HA trimers per virion ( $\sim 400$ ).

There is little fusion of negatively-charged vesicles at pH 7.4 (Figure 3-20 (A)). One reason for this pH-dependence is attractive vs repulsive protein/vesicle electrostatic energies at pH 5.0 and 7.4, respectively, corresponding to the protein charges of  $\sim +8$  and  $\sim -9$ . This electrostatic contribution is also evident in the reversal of the pH-dependence of vesicle fusion for positively charged vesicles (Figure 3-20 (B)). The fusion extent of positively-charged vesicles at pH 7.4 appears to be significantly larger than the fusion of negatively-charged vesicles at pH 5.0. For HA2 in 0.17% DM detergent, the major biophysical difference between pH 7.4 vs 5.0 is predominant trimer vs larger (likely multiple trimer) species. There may be a similar pH-dependent difference in oligomeric state in membrane that correlates to the different fusion extents.

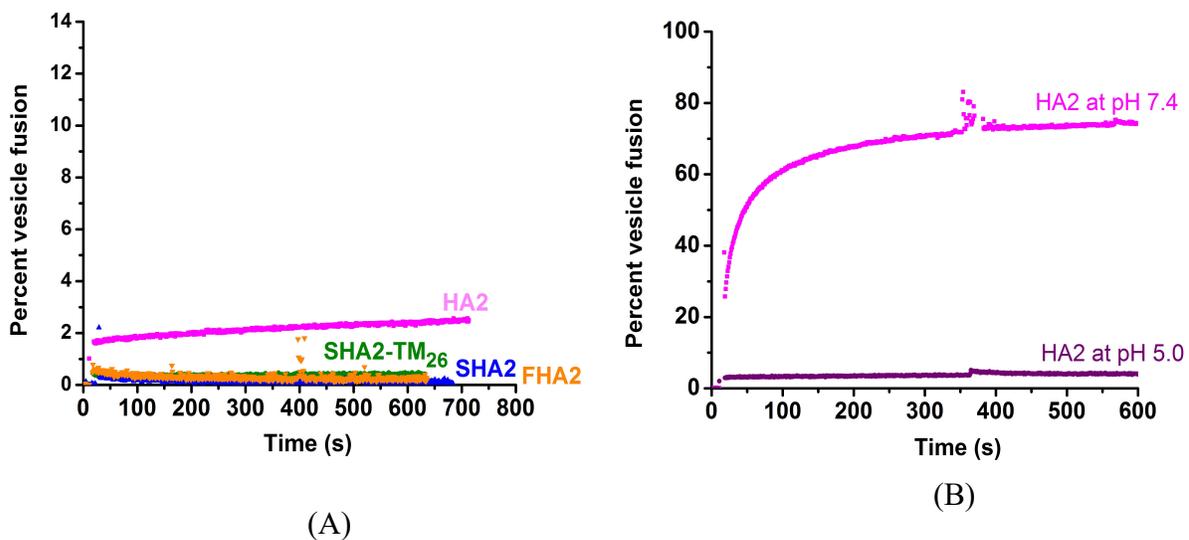


Figure 3-20 Protein-induced fusion of (A) negatively-charged vesicles at pH 7.4 and (B) positively charged vesicles at pH 5.0 and pH 7.4. (C) dose response for SHA2 at pH 5.0. The protein:lipid mole ratio = 1:280 in panel(A) and (B). The vesicle composition is (A) POPC : POPG (4:1) and (B) POPC : DOTAP (4:1). For each trace, protein is added a few seconds before the first displayed point. There is typically  $< \pm 1\%$  variation in final extent of vesicle fusion among replicate assays and no fusion is observed with addition of detergent solution

HA2-induced fusion is also compared for neutral PC-only and negatively-charged PC:PG (4:1) vesicles (Figure 3-21). At 50 s after HA2 addition, there is at least twice the fusion of negatively-charged vesicles relative to neutral vesicles at pH 5.0 with little fusion of either vesicle-type at pH 7.4 [21]. These observations are consistent with HA2 binding to vesicles in descending order PC: PG (pH 5.0) > PC (pH 5.0) > PC (pH 7.4) > PC: PG (pH 7.4) with respective estimated [protein] bound/protein free ratios of ~80, 3, 0.5, and 0. The first three values are calculated using: (1) the experimental binding constants to PC and PC: PG vesicles for the positively-charged HA3fp20 fusion peptide; and (2) [outer-leaflet lipid]  $\approx 8 \times 10^{-5} M$  in our vesicle fusion assays [33]. The HA3fp20 binding constants are used because to our knowledge, there aren't yet constants for larger HA2 constructs. Negligible HA2 binding to PC:PG vesicles at pH 7.4 is based on repulsive HA2/vesicle electrostatic energy.

Solution transparency was retained after protein addition for PC:PG vesicles at pH 5.0 and 7.4 and for PC vesicles at pH 7.4. In some contrast, the PC vesicle solution at pH 5.0 became cloudy after addition of protein. We previously observed that HA2 constructs aggregate at pH 5.0 but not 7.4 and we attribute the cloudiness to aggregation of unbound protein.

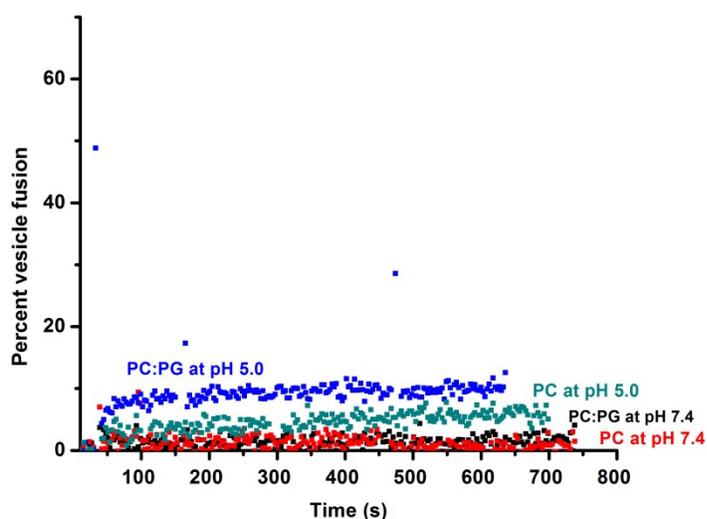


Figure 3-21 HA2-induced fusion for POPC-only and POPC:POPG (4:1) vesicles with protein:lipid = 1:140. Replicate assays consistently showed at least twice the fusion for PC:PG vesicles relative to PC-only vesicles at 50 s after protein addition. Because of instrument breakage, this vesicle fusion assays were done with a different fluorimeter than the rest of the assays of this study. Because this fluorimeter had lower sensitivity, we needed to increase the integration and sampling times to 3 s.

### 3-4-1 Significant findings

The present paper describes production of mg quantities of purified full-length HA2 as well as shorter HA2 constructs via expression in *E. coli* without a large solubility tag. Folding of the proteins in detergent at pH 7.4 was evidenced by CD spectra consistent with helical structure and by predominant trimer species evidenced by cross-linking and SEC. The proteins are highly-thermostable in DM detergent with  $T_m \approx 85$  °C for FHA2 and  $T_m > 90$  °C for HA2 with major stabilization provided by the SE and additional stabilization from the FP and TM domains. The

proteins efficiently fused negatively-charged vesicles at pH 5.0 but not pH 7.4 which corresponding positive and negative protein charges, respectively. The role of protein/vesicle electrostatic energy in vesicle fusion was further evidenced by much greater fusion of positively-charged vesicles at pH 7.4 than pH 5.0. Comparison between HA2 constructs showed a moderate enhancement of fusion with inclusion of the FP and little effect by inclusion of the TM. Efficient vesicle fusion is evidenced by its requirement of fewer than 15 HA2 trimers per vesicle which is lower by at least a factor of 10 from the number of HA2 trimers in the viral envelope.

#### ***3-4-2 Full-length HA2 purified from E. coli***

To our knowledge, this is the first report of mg purified quantities from *E. coli* of full-length HA2 without a large *N*-terminal solubility tag. Our accomplishment compares favorably with an earlier report (also reproduced in our laboratory) of a maltose-binding protein-HA2 construct that could be purified in mg quantities from *E. coli* but for which little HA2 was subsequently recovered after attempted cleavage of the maltose-binding protein [30]. Key points of our approach included: (1) removal of soluble *E. coli* proteins from the cell pellet prior to solubilization of the remaining proteins with SRC; and (2) a G6LEH6 rather than a LEH6 tag to increase exposure of the H6 region during affinity purification. The purified FHA2 yield of the present study is ~10 mg/L culture and is compared with >100 mg/L expression of FHA2. There is probably lower expression of full-length HA2 which includes the TM and corresponding lower purified yield of ~2 mg/L (Figure 3-3 (A)). Although these yields are good for a non-bacterial membrane protein expressed in *E. coli*, there are >90% losses during solubilization and purification. The purification losses have previously been noted for FHA2 using SDS-PAGE with lanes for the soluble cell lysate and for washes and elutions from the Co<sup>2+</sup>-affinity column

with bound protein [29]. The lysate and wash lanes are similar with a band corresponding to FHA2 as well as bands with significant intensity corresponding to higher MW proteins. This supports some loss of FHA2 during the column washes. The eluent lane shows a dominant FHA2 band which is ~3 times more intense than the FHA2 band in the lysate. However, the eluent volume is ~10 times smaller than the lysate volume which also supports an overall loss of FHA2 during purification.

### ***3-4-3 Dominant thermostable trimer supported by earlier biophysical studies***

To our knowledge, the present study is the first to produce and study the full-length HA2, SHA2  $\equiv$  SE, and SHA2-TM26  $\equiv$  SE+TM proteins. FHA2  $\equiv$  FP+SE has been previously studied as has the construct “F185”  $\equiv$  FLAG+FHA2 where FLAG  $\equiv$  DYKDDDDK [34]. Because of the 8-residue non-native hydrophilic FLAG tag, F185 is soluble at pH 7.0 in the absence of detergent and cross-linking and SEC shows that F185 forms a dominant trimer species. The major SEC peak of F185 reflects migration of protein-only and the mass agrees quantitatively with the expected value of a protein trimer. The CD spectrum of F185 in the absence of detergent is similar to our CD spectrum of FHA2 in detergent in Figure 3-10 and the  $|\theta_{222}|$  values correspond to ~65%  $\alpha$ -helical structure which is close to the ~60% value calculated using the high-resolution structure of the FP and the structure of a large region of the SE. In addition, the  $|\theta_{222}|$  vs temperature for both constructs have sigmoidal shape with  $T_m \approx 85$  °C. These good correlations between F185 and FHA2 support dominant trimers of our HA2 constructs with thermostable SE hairpin structure. The major peaks in our SEC’s are typically at ~200 kDa which is similar to the mass previously observed in the SEC of FHA2 in Brij-35 non-ionic detergent [34]. We assigned this peak to a protein trimer with ~80 kDa protein and ~120 kDa detergent contributions. This assignment is consistent with an earlier study on a MBP+HA2 construct that supported a

dominant trimer species with hairpin HA2 SE structure based on analytical ultracentrifugation, electron microscopy, antibody-binding, and cross-linking data [30]. The MBP+HA2 had been solubilized at neutral pH in non-ionic detergents which included 0.17% DM which is used in our study. The major SEC peak for MBP+HA2 was at ~350 kDa with respective protein trimer and detergent mass contributions of ~200 and ~150 kDa. MBP is very soluble so it is reasonable that most of the detergent binds to HA2. It is therefore reasonable that similar detergent masses bind to trimeric MBP+HA2 and to trimeric HA2.

There is also semi-quantitative agreement between our SEC and the published SEC in non-ionic detergent of an ectodomain + TM construct of the HIV gp41 membrane fusion protein [35]. The HIV construct is functionally and structurally homologous to HA2. The gp41 SEC had a peak at ~180 kDa that was assigned to a protein trimer with respective protein and detergent mass contributions of ~60 and ~120 kDa.

#### ***3-4-4 SHA2 aggregation in DM***

The FHA2 CD data in DM detergent at pH 7.4 is compared with a previous study in which the FHA2 was solubilized in 0.14 M NaF at pH 7.4 [21]. At ambient temperature, the CD-derived  $|\theta_{222}|$  in NaF is about half that in DM. The  $|\theta_{222}|$  vs temperature are different in NaF and DM with the NaF data resembling those of SHA2 in DM (Figure 3-13). There was a large decrease in  $|\theta_{222}|$  between 55 and 65 °C, followed by leveling, and then a second large decrease between 75 and 85 °C. This behavior differs from the single sigmoidal curve for FHA2 in DM with  $T_m \approx 85$  °C. For our SEC in DM, SHA2 forms large (>2 MDa) aggregates which is different from the other three constructs which are not aggregated (Figure 3-15 (B)). There was no SEC characterization of

FHA2 in the earlier study but we propose that this protein was also aggregated which is reflected in the  $|\theta_{222}|$  vs temperature. For SHA2 in DM and FHA2 in NaF, we propose that the 55-65 °C component corresponds to dissociation of the protein aggregates and the 75-85 °C component corresponds to unfolding of the SE. The latter assignment is consistent with the  $T_m$  of FHA2 in DM for which large aggregates are not formed.

### ***3-4-5 Viral fusion overview***

Influenza fusion occurs within endosomes which themselves are undergoing morphological and chemical changes in time [3]. Following initial endocytosis, the early endosome forms in ~5 minutes and is characterized by: (1) enlargement due to fusion with other endocytic vesicles; (2) pH reduced to ~6.2; and (3) migration from the plasma membrane towards the nucleus [2]. The late endosome forms in the next ~3 minutes and is characterized by: (1) creation of vesicles within the endosomal lumen; and (2) further lowering of pH to ~5.5. The late endosome may then fuse with lysosomes which contains hydrolase enzymes that degrade proteins and whose pH < 5. The membrane compositions of both the endosome and interior vesicles are different from one another and change during these maturation steps in a manner correlated to use of the endosomal pathway to transport cholesterol in and out of the cell [36, 37].

The fusion trigger pH of 5-6 for influenza supports fusion within the late endosome which is formed after the early endosome and prior to the endosome-lysosome hybrid. To our knowledge, it is not clear whether the virus fuses first with the outer membrane of the late endosome or with the membrane of one of the vesicles within the endosomal lumen. For the former circumstance, the viral nucleocapsid is released directly into the cytosol close to the

nucleus whereas in the latter circumstance, nucleocapsid release into the cytosol requires an additional event such as back-fusion with the endosomal membrane. The influenza virus can fuse with membranes with a variety of compositions so fusion is probably not influenced by composition differences between outer vs interior endosomal membranes or by differences at different stages of endosome maturation [7].

To our knowledge, there has been little detailed study of virus/endosome fusion. There have been some studies of virus/vesicle fusion but most detailed studies have examined fusion between cells expressing HA (HA-cells) and cells containing sialic acid but not HA (often Red Blood Cells – RBC's) [6, 8, 9]. One dye (typically small-molecule) is incorporated in the RBC membrane and a different dye incorporated into the RBC cytoplasm. Fusion is triggered by lowered pH. Subsequent intercellular lipid mixing is quantitated by the percentage of HA-cells containing membrane dye while intercellular contents mixing is quantitated by the percentage cells containing cytoplasmic dye. Lipid mixing without contents mixing is often considered evidence for arrest at a hemifusion intermediate state. In cases in which neither lipid nor contents mixing are observed with lowered pH, a small molecule like chlorpromazine is sometimes added and becomes membrane-incorporated. Subsequent lipid and contents mixing has been interpreted to mean that HA alone had induced “hemifusion with restricted lipid movement” where the restriction prevented movement of the RBC membrane dye into the HA-cell membrane [9].

Although cell/cell fusion shares common features with virus/endosome fusion, one topological difference is initial physical separation of the two cells vs initial enclosure of the virus within the endosome. In addition, unlike cell/cell fusion, there may be significant content leakage during virus/endosome fusion. Leakage has been observed in virus/vesicle fusion and

leakage in intracellular virus/endosome fusion probably doesn't hurt cell viability so there isn't evolutionary selection against it [7].

### ***3-4-6 Common fusion features***

The present study describes HA2-catalyzed vesicle fusion using stock solutions with predominant folded hairpin trimer species. Such protein characterization has typically not been done in vesicle fusion studies but is very useful because it provides information about the likely protein structure during initial interaction with the membrane (Figure 3-22). This may also be the fusogenic structure. Unlike vesicle/vesicle fusion for which HA2 initially has hairpin structure, virus/endosome and HA-cell/RBC fusion initially have HA2 in non-hairpin structure in complex with HA1. Upon pH reduction, HA1 dissociates and HA2 folds into the final hairpin state. It has often been assumed that some of the free energy released upon folding ( $\Delta G_{fold}$ ) provides activation energy for fusion ( $\Delta G^{\ddagger}_{fusion}$ ) [6, 38]. For vesicle/vesicle fusion, HA2 always has hairpin structure so  $\Delta G_{fold} \approx 0$  and  $\Delta G^{\ddagger}_{fusion}$  would be influenced by membrane interactions of the hairpin SE and the FP.

### Common fusion intermediate

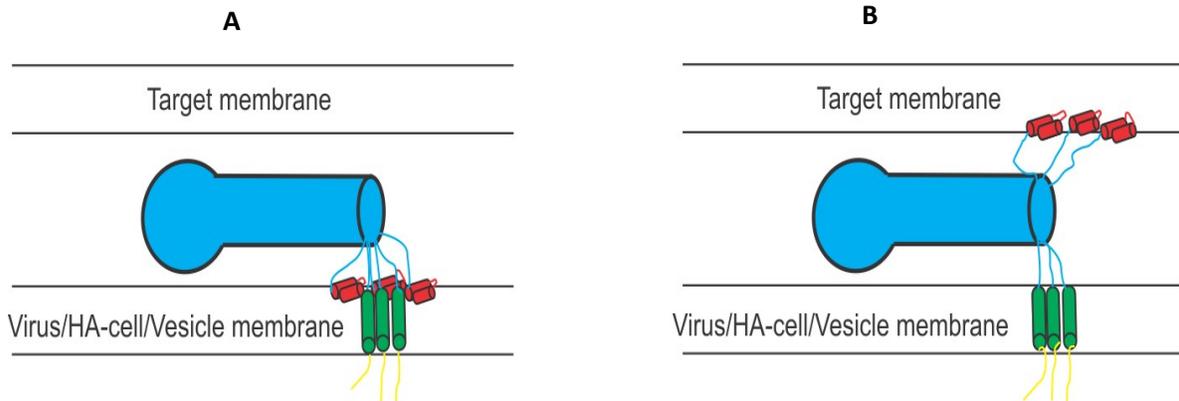


Figure 3-22 Model of trimeric hairpin HA2 prior to fusion with the FP and TM regions in (A) the same membrane or (B) different membranes. The two membranes would correspond to: (1) two vesicles; (2) HA-cell and RBC; or (3) virus and endosome. For cases (2) and (3), the TM is respectively in the HA-cell and the viral membrane. The color coding is the same as Figure 3-1 with FP (red), SE (blue), TM (green), and Endo (yellow). A single HA2 trimer is shown with structural elements: FP, helix-(tight-turn)-helix; SE, trimer-of-hairpins; and TM, helix. For either fusion type at  $\text{pH} \approx 5$ , the SE binds to both membranes in part because of attractive electrostatic energy between the positively-charged protein and negatively-charged membrane. The resultant membrane apposition and membrane perturbation aid catalysis of the subsequent membrane fusion.

There is evidence that the HA2 hairpin is a fusogenic structure in cell/RBC fusion. In particular, FHA2 and truncated FHA2trunc (residues 1-127) catalyze cell/RBC fusion including the steps of intercellular lipid mixing and small-molecule contents mixing [21, 39]. Both constructs are presumed to be in the final hairpin state. Large pores are not formed but this may be due to no

TM in either FHA2 construct. The TM is required for formation of large pores in HA-cell/ RBC fusion[6] There are many correlations between HA-cell/RBC fusion and FHA2- catalyzed cell/RBC fusion including: (1) low pH requirement; (2) reduction in fusion with the G1E and V173E mutations which are respectively in the FP and SE domains; (3) reduction in fusion with addition of lysophosphatidylcholine detergent with positive curvature; (4) reduction in fusion at 4 °C and fusion recovery upon return to 37 °C; and (5) increased fusion under some circumstances with addition of chlorpromazine dye [9] All these correlations are consistent with similar mechanisms of HA-cell/RBC fusion and FHA2-catalyzed cell/RBC fusion. Vesicle fusion induced by HA2 constructs also show pH dependence as well as reduction in fusion with the G1E and other mutations [19, 21]. It is likely there are at least qualitative correlations between the mechanisms of HA-cell/RBC fusion and FHA2-catalyzed cell/RBC and vesicle fusion. Overall, these studies support the fusogenic nature of the trimer-of-hairpins structure including the significances of the SE and FP regions.

The Figure 3-22 pictures of trimeric hairpin HA2 prior to fusion show two topologies in which the FP and TM domains are in (A) the same or (B) different membranes. The pictures may be relevant catalytic intermediates for: (1) vesicle/vesicle, (2) HA-cell/RBC; and (3) virus/endosome fusion with the TM in the (2) HA-cell or (3) viral membrane. To our knowledge, there aren't yet data to distinguish between these topologies for any of these three cases. It is also possible that there is a distribution of FP's between the two membranes. For any FP/TM topology, virus/endosome fusion may be efficient because of the high virus/endosome collision rate in the restricted volume of the endosome.

Virus/vesicle and HA-cell/RBC fusion are greatly reduced when preceded by ~5 minute

low-pH pre-incubation without target membrane [21] This reduction is often ascribed to insertion of the FP in the same membrane as the TM and it is therefore inferred that successful fusion requires FP insertion in the different (target) membrane. However, Figure 3-18 and 3-19(B) in the present paper shows little vesicle fusion for times  $\geq 10$  s following protein addition. The absence of cell-cell fusion with pre-incubation for 5 minutes may therefore be due to intrinsic loss of protein activity rather than initial insertion of the FP in the same membrane as the TM.

For HA2-catalyzed vesicle fusion, the inter-vesicle collision rate is  $\sim 10$  s<sup>-1</sup> and the fusion rate is  $\sim 0.2$  s<sup>-1</sup> so there are  $\sim 50$  collisions prior to fusion. This may result in FP's and TM's in different vesicles prior to fusion. Alternatively, they could plausibly insert in the same vesicle during the initial HA2/vesicle binding.

### ***3-4-7 HIV vs influenza fusion***

HIV fusion with the plasma membrane and influenza virus fusion with the endosome membrane are respectively catalyzed by the gp41 and HA2 proteins which are non-homologous in sequence but have similar final hairpin structures of the SE. Successful infection by HIV requires nucleocapsid release in the interior of the cell body with which the virus fuses, whereas infection by influenza requires release to the exterior of the body (endosome). Monomer species exist for both gp41 and HA2 proteins and may be functionally important because folding of the individual protein monomers to the final hairpin state is topologically more straightforward than folding of trimers [35, 40-43]. The monomer is dominant for gp41 at pH 3-4 whereas the trimer is dominant for HA2 at pH 7.4 (Figure 3-15). Hexamers containing the folded gp41 ectodomain are also observed and correlate with observation of HA2 oligomers larger than the trimer under some conditions [40, 44]. Both gp41 and HA2 typically aggregate at their respective fusion pH's

of  $\sim 7$  and  $\sim 5$  which may correlate to previous detection of multiple trimers for both proteins at the fusion site [45-48]. For HA2, the SE hairpin trimer has  $T_m \approx 85^\circ\text{C}$  whereas the corresponding monomer hairpin SE of HIV gp41 has  $T_m \approx 110^\circ\text{C}$  [49]. In addition, the HA2  $T_m$  is higher with inclusion of the FP and/or TM. To our knowledge, there isn't evidence yet for this effect in monomer gp41.

Like HA2, vesicle fusion induced by gp41 is also enhanced by attractive protein/vesicle electrostatic energy [40, 49, 50]. Interestingly, fusion extent for gp41 is inversely-dependent on the magnitude of vesicle charge which probably reflects the effect of inter-vesicle electrostatic repulsion. For both proteins, the hairpin SE makes an important contribution to fusion with moderate enhancement with inclusion of the FP in the construct.

### **3-5 Conclusions**

Full-length influenza virus hemagglutinin II membrane fusion protein as well as shorter constructs have been expressed and purified in mg quantities. The constructs typically adopt a folded trimer-of-hairpins structure in detergent and the full-length protein is hyperthermostable. The proteins induce vesicle fusion with significant contributions from the soluble ectodomain and fusion peptide and little contribution from the TM domain. The present as well as previous studies support a role for the final hairpin structure in catalysis of fusion between the viral and endosomal membranes.

## **REFERENCES**

## REFERENCES

1. White, J.M., et al., *Structure and mechanism of viral membrane fusion proteins: multiple variations on common theme*. Critical Reviews in Biochemistry and Molecular Biology 2008. **43** p. 189 – 219.
2. Lakadamyali, M., et al., *Visualizing infection of individual influenza viruses*. Proceedings of the National Academy of Sciences of the United States of America, 2003. **100**(16): p. 9280-9285.
3. Gruenberg, J., *Lipids in endocytic membrane transport and sorting*. Current Opinion in Cell Biology, 2003. **15**(4): p. 382-388.
4. Gruenberg, J., *Viruses and endosome membrane dynamics*. Current Opinion in Cell Biology, 2009. **21**(4): p. 582-588.
5. Le Blanc, I., et al., *Endosome-to-cytosol transport of viral nucleocapsids*. Nature Cell Biology, 2005. **7**(7): p. 653-U25.
6. Kemble, G.W., T. Danieli, and T.M. White, *Lipid anchored influenza hemagglutinin promotes hemi fusion not complete fusion*. Cell 1994. **76**: p. 383-391.
7. Tong, S.G., D. Alford, and J. Bentz, *Influenza virus liposome lipid mixing is leaky and largely insensitive to the material properties of the target membrane*. Biochemistry, 1996. **35**(15): p. 4956-4965.
8. Park, H.E., J.A. Gruenke, and J.M. White, *Leash in the groove mechanism of membrane fusion*. Nature Structural Biology, 2003. **10**(12): p. 1048-1053.
9. Borrego-Diaz, E., et al., *Completion of trimeric hairpin formation of influenza virus hemagglutinin promotes fusion pore opening and enlargement*. Virology, 2003. **316**(2): p. 234-244.
10. Durrer, P., et al., *H<sup>+</sup>-induced membrane insertion of influenza virus hemagglutinin involves the HA2 amino-terminal fusion peptide but not the coiled coil region*. Journal of Biological Chemistry, 1996. **271**(23): p. 13417-13421.
11. Grewe, C., A. Beck, and H.R. Gelderblom, *HIV-Early virus-cell interactions*. Journal of Acquired Immune Deficiency Syndromes and Human Retrovirology, 1990. **3**(10): p. 965-974.
12. Nobusawa, E., et al., *Comparison of complete amino acid sequences and receptor binding properties among 13 serotypes of hemagglutinins of influenza A-viruses*. Virology, 1991. **182**(2): p. 475-485.

13. Armstrong, R.T., A.S. Kushnir, and J.M. White, *The transmembrane domain of influenza hemagglutinin exhibits a stringent length requirement to support the hemifusion to fusion transition*. *J. Cell Biology* 2000. **151**: p. 425-437.
14. Durell, S.R., et al., *What studies of fusion peptides tell us about viral envelope glycoprotein-mediated membrane fusion*. *Molecular Membrane Biology*, 1997. **14**(3): p. 97-112.
15. Qiao, H., et al., *A specific point mutant at position 1 of the influenza hemagglutinin fusion peptide displays a hemifusion phenotype*. *Molecular Biology of the Cell*, 1999. **10**(8): p. 2759-2769.
16. Apellaniz, B., et al., *The three lives of viral fusion peptides*. *Chemistry and Physics of Lipids*, 2014. **181**: p. 40-55.
17. Wharton, S.A., et al., *Membrane- fusion by peptide analogs of influenza virus hemagglutinin*. *Journal of General Virology*, 1988. **69**: p. 1847-1857.
18. Epand, R.F., et al., *The ectodomain of HA2 of influenza virus promotes rapid pH dependent membrane fusion*. *Journal of Molecular Biology*, 1999. **286**(2): p. 489-503.
19. LeDuc, D.L., et al., *Factors determining vesicular lipid mixing induced by shortened constructs of influenza hemagglutinin*. *Biochemistry*, 2000. **39**(10): p. 2733-2739.
20. Curtis-Fisk, J., et al., *Solid-State NMR Structural Measurements on the Membrane-Associated Influenza Fusion Protein Ectodomain*. *JACS*, 2007. **129** p. 11320-11321.
21. Kim, C.S., et al., *The Final Conformation of the Complete Ectodomain of the HA2 Subunit of Influenza Hemagglutinin Can by Itself Drive Low pH-dependent Fusion*. *Journal of Biological Chemistry*, 2011. **286**(15): p. 13226-13234.
22. Wilson, I.A., J.J. Skehel, and D.C. Wiley, *Structure of the haemagglutinin membrane glycoprotein of influenza virus at 3 Å resolution*. *Nature* 1981. **289** p. 366- 373.
23. Chen, J., J.J. Skehel, and D.C. Wiley, *N- and C-terminal residues combine in the fusion-pH influenza hemagglutinin HA(2) subunit to form an N cap that terminates the triple-stranded coiled coil*. *Proceedings of the National Academy of Sciences of the United States of America*, 1999. **96**(16): p. 8967-8972.
24. Han, X., et al., *Membrane structure and fusion triggering conformational change of the fusion domain from influenza haemagglutinin*. *Nature Structural Biology* 2001. **8** p. 715-720.
25. Lorieau, J.L., J.M. Louis, and A. Bax, *The complete influenza hemagglutinin fusion domain adopts a tight helical hairpin arrangement at the lipid:water interface*.

- Proceedings of the National Academy of Sciences of the United States of America, 2010. **107**(25): p. 11341-11346.
26. Du, T.P., L. Jiang, and M.L. Liu, *NMR structures of fusion peptide from influenza hemagglutinin H3 subtype and its mutants*. Journal of Peptide Science, 2014. **20**(4): p. 292-297.
  27. Sun, Y. and D.P. Weliky, *<sup>13</sup>C-<sup>13</sup>C correlation spectroscopy of membrane associated influenza virus fusion peptide strongly supports a helix-turn-helix motif and two turn conformations*. JACS Communication, 2009. **131** p. 13228-13229.
  28. Ghosh, U., L. Xie, and D.P. Weliky, *Detection of closed influenza virus hemagglutinin fusion peptide structures in membranes by backbone (CO)-C-13-N-15 rotational-echo double-resonance solid-state NMR*. Journal of Biomolecular Nmr, 2013. **55**(2): p. 139-146.
  29. 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.
  30. 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.
  31. Vogel, E.P. and D.P. Weliky, *Quantitation of Recombinant Protein in Whole Cells and Cell Extracts via Solid-State NMR Spectroscopy*. Biochemistry, 2013. **52**(25): p. 4285-4287.
  32. 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.
  33. Han, X. and L.K. Tamm, *A host-guest system to study structure-function relationships of membrane fusion peptides*. Proceedings of the National Academy of Sciences of the United States of America, 2000. **97**(24): p. 13097-13102.
  34. Chen, J., J.J. Skehel, and D.C. Wiley, *A polar octapeptide fused to the N-terminal fusion peptide solubilizes the influenza virus HA(2) subunit ectodomain*. Biochemistry, 1998. **37**(39): p. 13643-13649.
  35. Lakomek, N.A., et al., *Internal Dynamics of the Homotrimeric HIV-1 Viral Coat Protein gp41 on Multiple Time Scales*. Angewandte Chemie-International Edition, 2013. **52**(14): p. 3911-3915.

36. Kobayashi, T., et al., *Separation and characterization of late endosomal membrane domains*. Journal of Biological Chemistry, 2002. **277**(35): p. 32157-32164.
37. Mobius, W., et al., *Recycling compartments and the internal vesicles of multivesicular bodies harbor most of the cholesterol found in the endocytic pathway*. Traffic, 2003. **4**(4): p. 222-231.
38. Carr, C.M. and P.S. Kim, *A spring loaded mechanism for the conformational change of influenza hemagglutinin*. Cell, 1993. **73**(4): p. 823-832.
39. Leikina, E., et al., *The 1-127 HA2 construct of influenza virus hemagglutinin induces cell-cell hemifusion*. Biochemistry, 2001. **40**(28): p. 8378-8386.
40. 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.
41. Roche, J., et al., *Dissociation of the trimeric gp41 ectodomain at the lipid-water interface suggests an active role in HIV-1 Env-mediated membrane fusion*. Proceedings of the National Academy of Sciences of the United States of America, 2014. **111**(9): p. 3425-3430.
42. Caffrey, M., et al., *Monomer-trimer equilibrium of the ectodomain of SIV gp41: Insight into the mechanism of peptide inhibition of HIV infection*. Protein Science, 1999. **8**(9): p. 1904-1907.
43. Dai, Z., et al., *Conditional Trimerization and Lytic Activity of HIV-1 gp41 Variants Containing the Membrane-Associated Segments*. Biochemistry, 2015. **54**(8): p. 1589-1599.
44. Gao, G.F., et al., *Designing a Soluble Near Full-length HIV-1 gp41 Trimer*. Journal of Biological Chemistry, 2013. **288**(1): p. 234-246.
45. Freed, E.O., et al., *A mutation in the human immunodeficiency-virus type-1 transmembrane glycoprotein-gp41 dominantly interferes with fusion and infectivity*. Proceedings of the National Academy of Sciences of the United States of America, 1992. **89**(1): p. 70-74.
46. Bentz, J., *Minimal aggregate size and minimal fusion unit for the first fusion pore of influenza hemagglutinin-mediated membrane fusion*. Biophysical Journal, 2000. **78**(1): p. 227-245.
47. Sougrat, R., et al., *Electron tomography of the contact between T cells and SIV/HIV-1: Implications for viral entry*. Plos Pathogens, 2007. **3**(5): p. 571-581.

48. Magnus, C., et al., *Estimating the Stoichiometry of Human Immunodeficiency Virus Entry*. Journal of Virology, 2009. **83**(3): p. 1523-1531.
49. Sackett, K., et al., *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, 2010. **397**(1): p. 301-315.
50. 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 conformation*. Biochimica Et Biophysica Acta-Biomembranes, 2015. **1848**(1): p. 289-298.
51. Ghosh, U., Xie, L., Jia, L., Liang, S., 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. Journal of the American Chemical Society **137**, 7548-7551.

**Chapter 4 Investigation of pH-dependent vesicle fusion induced by  
the ectodomain constructs of human immunodeficiency virus  
membrane fusion protein gp41**

## 4-1 Introduction

Acquired Immune Deficiency Syndrome (AIDS) is caused by the Human Immunodeficiency Virus (HIV). World Health Organization reported that, there are about 1.6 million deaths in each year around worldwide and there are about 1.1 million people living with HIV in USA. Human immunodeficiency virus (HIV) is a class I enveloped virus surrounded by a membrane obtained from an infected host cell during viral budding [4]. Uninfected macrophage and T helper cells are targeted by HIV with subsequent joining (fusion) of the HIV and cell membranes and deposition of the viral capsid inside the cell. Infection is mediated by gp160 which contains non-covalently-associated gp41 and gp120 subunit proteins. Virion-associated gp160 is likely trimeric [7]. The gp41 protein contains ~180-residue extraviral and ~150-residue intraviral domains (figure 4-1). Gp120 is ~500 residues and associated with the gp41 ectodomain. The gp41 ectodomain plays a critical role in fusion and includes a ~20-residue fusion peptide (FP), *N*-heptad repeat (NHR), loop, *C*-heptad repeat (CHR), and membrane-proximal external region (MPER). Fusion models typically show partial insertion of the FP in the cell membrane and partial insertion of the MPER in the viral membrane. High resolution structures of the gp41 ectodomain lacking the FP and MPER show a helical hairpin with antiparallel NHR and CHR helices [8-10]. In addition, the structures (at >2 mM protein concentration and pH 3) show a molecular trimer containing an interior bundle of three parallel helical NHRs and three antiparallel helical CHRs closely packed in the exterior grooves of the NHR bundle. Because of its thermostability, the hairpin structure has been considered to be the final state of the gp41 ectodomain during fusion.

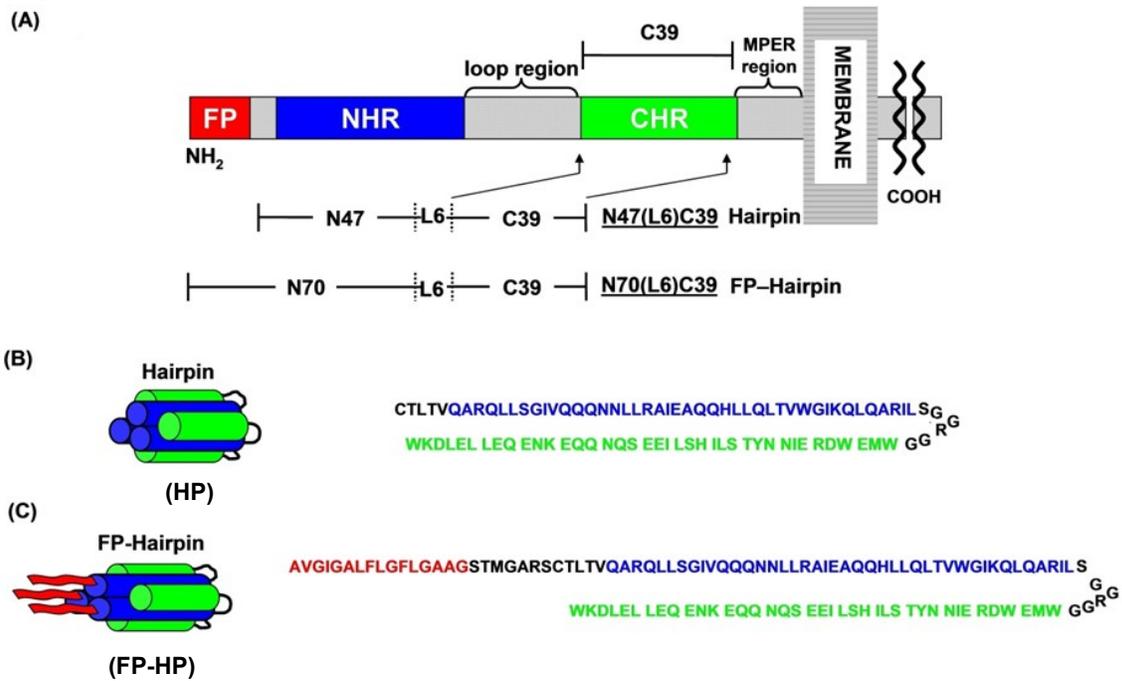


Figure 4-1 (A) The diagram showing the ectodomain of the HIV gp41 protein including fusion peptide (FP), *N*-heptad repeat (NHR), loop, *C*-heptad repeat (CHR), and membrane-proximal external region (MPER). (B,C) Amino acid sequences of the hairpin (HP) and FP-hairpin (FP-HP) proteins with color-matching with panel A. The sequences of HP and FP-HP are from the HXB2 laboratory strain of HIV-1. According to the gp160 numbering HP includes NHR [(535-581(M535C))-[SGGRGG]-CHR-[628-666]. FP-HP includes [512-581 (M535C)]-[SGGRGG]-[628-666]. The first bracket for HP includes much of the NHR for HP and the first bracket for FP-HP includes the full FP and much of the NHR. For both constructs, the third bracket includes much of the CHR. The intervening nonnative loop (SGGRGG) is shorter than ~20 residue native loop but still allows formation of stable hairpin with antiparallel NHR and CHR helices. [1-3] .

At lower protein concentrations, the ectodomain remains highly helical but is monomeric [11, 12]. The ectodomain can also be hexameric or more highly aggregated and the oligomerization states depend on pH, protein sequence, and solution additives [12, 13]. Protein monomers are also predominant at low pH for gp41 constructs that include the TM domain although it is not known whether or not the TM domains of the trimer dissociate during HIV/cell fusion [14].

Experiments and models support lipid mixing between the outer leaflets of the HIV and cell membranes as the first membrane topological change during fusion [4]. The resultant membrane state is “hemifusion”, i.e. a shared bilayer interface between HIV and the cell but no mixing of contents between the two bodies. The shared bilayer then breaks to form a “fusion pore” which then expands to create a single contiguous bilayer which envelopes the HIV capsid and cell. Although the relative timings of ectodomain and membrane topological changes are still under investigation, some experiments have been interpreted to support formation of a final gp41 state after lipid mixing and fusion pore creation and before pore expansion [15]. The effects of gp41 mutations on gp160-mediated fusion and HIV infection have also been investigated. One significant result was trans-dominant reduction in fusion and infection with the V2E mutation in the FP region of gp41 [16]. This result supports a requirement of multiple gp41 trimers for fusion [17].

The present chapter focuses on the “Hairpin” (HP) and “FP-Hairpin” (FP-HP) large ectodomain constructs that respectively include much of the NHR+CHR and FP+NHR+CHR regions [1]. Figure 4-1 displays their sequences which are derived from the HXB2 laboratory strain of HIV. Both proteins have a stable NHR/CHR helical fold and induce vesicle fusion under some conditions [2, 18]. The final state is typically a multi-vesicle aggregate similar to the

multicellular “syncytia” aggregates that form in cell culture with HIV infection [19]. Intervesicle contents mixing has not typically been observed in protein induced vesicle fusion because of the rapid leakage from the vesicles.

Studies-to-date on large ectodomain constructs like HP and FP-HP have both provided ideas and raised questions relevant to fusion. Some results support requirements of low pH (3-4) and anionic lipids for rapid and extensive vesicle fusion [3, 18, 20]. To date, different studies used vesicles with either phosphatidylserine (PS) or phosphatidylglycerol (PG) anionic lipids but there has not been direct comparison between the two lipid types. For vesicles with PG, there was little fusion at pH 7. The ectodomain is positively charged at pH 4 and negatively charged at pH 7 so the combination of low pH and anionic lipids may reflect an underlying requirement for fusion of attractive protein/vesicle electrostatic energy. This energy could increase protein binding and/or contribute to the activation energy of fusion, and/or locally heat the membrane. At pH 4, the electrostatic energy will be different for PG and PS lipids because they carry full and partial negative charges, respectively. Protein binding to vesicles at pH 7 and consequent fusion may also be impacted by aggregation of ectodomain protein in the pH 7 solution [12]. By contrast, the protein is soluble at  $\text{pH} \leq 4$  and is predominantly monomeric for  $[\text{protein}] < 100 \mu\text{M}$ . Because of gp41 ectodomain aggregation at neutral pH, many biophysical studies including high-resolution structures have used stock protein solutions with  $\text{pH} \leq 4$ . Example studies and results for HP protein include circular dichroism which showed >95% helicity at ambient temperature and calorimetry which showed a melting temperature of  $110^\circ \text{C}$  and melting enthalpy of 65 kcal/mole [1, 2]. Similar results were obtained for FP-HP and other large ectodomain constructs [3].

Solid-state NMR (SSNMR) spectroscopy has also been applied to probe the structure of FP-HP in its membrane-associated form [2]. Protein stock at low pH was added dropwise to

vesicle suspensions maintained at neutral pH, followed by centrifugation and harvesting the pellet for SSNMR. The spectra were consistent with a highly helical Hairpin domain and a ~1:1 mixture of molecules with either helical or  $\beta$  sheet FPs. The  $\beta$  sheet was formed by multiple proteins whose FPs were arranged in antiparallel rather than parallel geometry [21]. There was also a distribution of antiparallel  $\beta$  sheet registries. If the highly helical Hairpin region is a molecular trimer, the antiparallel FP  $\beta$  sheet could reflect interleaved FPs from two different trimers. This result correlates with predominant hexamers for gp41 ectodomain constructs under some conditions [12, 13]. Because of protein aggregation in neutral pH aqueous solution, the SSNMR samples could also have contained a fraction of aggregated protein that was not membrane-associated. The relative fractions of membrane-associated and non-associated FP-HP with  $\beta$  or  $\alpha$  FPs were not known.

The present work addresses some questions raised by these earlier studies. Significant results include negligible protein-induced fusion of neutral vesicles at either low or neutral pH and negligible fusion of anionic vesicles at neutral pH. Rapid and extensive fusion was observed for anionic vesicles at low pH. Fusion rates and extents were analyzed for vesicles with different fractions of PG and PS lipids and two kinetic processes were detected. The rate of the faster process was strongly positively correlated with vesicle anionic charge while the extent of the slow process was inversely correlated with this charge. The slow process may be more relevant for HIV/cell fusion at neutral pH.

Fully-membrane-associated FP-HP samples were prepared by coauthors Sackett, K, Nethercott, M.J. and Weliky D.P. using anionic vesicles and low pH where the protein did not aggregate. The SSNMR spectra of these samples showed that the  $\beta$  FP structure was predominant. The  $\beta$  structure was also predominant when the sample pH was increased to 7.

Protein to-lipid head group contact in both low and neutral pH samples was probed by protein  $^{13}\text{C}$  to lipid  $^{31}\text{P}$  SSNMR. There was greater contact for the NHR+CHR region at low pH relative to neutral pH which is consistent with protein/membrane electrostatics. There was close membrane contact of the FP at both pH's which is consistent with FP insertion via the hydrophobic effect [24].

## 4-2 Materials and Methods

### 4-2-1 Materials

Lipids were purchased from Avanti Polar Lipids. Most other reagents were purchased from Sigma-Aldrich. HP contains *N*-heptad repeat (NHR) + short non-native loop + *C*-heptad repeat (CHR) and FP-HP contains FP + HP (Figure 4-1). Protein with non-native loop has similar physical properties to protein with the ~20-residue native loop including very high fractional helicity and  $T_m \approx 110^\circ\text{C}$  [1-3]. The gp160-numbered amino-acid sequences from the HXB2 laboratory strain of HIV-1 are: FP (512-534); NHR (535(M535C)-581); and CHR (628-666).

HP and FP-HP proteins were produced and purified as previously described [1, 18, 21]. Briefly, FP was synthesized by *t*-Boc chemical synthesis and the subsequent cleavage with hydrogen fluoride was done by Midwest Biotech. HP was synthesized by bacterial expression and FP-HP was produced using native chemical ligation of purified FP and HP. Purifications were done by reversed-phase HPLC and final purities were typically >95% as assessed by mass spectrometry. Typical purified yields were >100 mg FP per *t*-Boc synthesis, ~50 mg HP per L bacterial culture, and ~5 mg FP-HP from a ligation reaction with ~10 mg FP and ~50 mg HP.

#### ***4-2-2 Protein-induced vesicle fusion***

Vesicles were prepared with lipid and cholesterol (Chol) in 2:1 mole ratio which was intermediate between the ratios for the membrane of host cells and the membrane of HIV [22]. The membranes always contained electrically neutral phosphatidylcholine (PC) lipids because this is the most common head group of membranes of cells infected by HIV. The membranes typically contained an anionic lipid (AL) with either phosphatidylglycerol (PG) or phosphatidylserine (PS) head group and the AL: PC mole ratio varied between 0 and 1. Lipids typically contained 1-palmitoyl-2-oleoyl acyl chains.

A corollary “labeled” vesicle sample was prepared for each “unlabeled sample” described above and contained additional 0.02 mole fractions of fluorescent lipid [*N*-(7-nitro-2,1,3-benzoxadiazol-4-yl)(ammonium salt) dipalmitoylphosphatidylethanolamine] and quenching lipid [*N*-(lissamine rhodamine B sulfonyl)(ammonium salt)dipalmitoylphosphatidylethanolamine]. For each sample, lipid and Chol were dissolved in chloroform:methanol (9:1 v/v) and excess solvent was removed by rotary evaporation. The film was suspended in 25 mM citrate buffer, homogenized by freeze-thaw cycles, and extruded through a filter with 100 nm diameter pores. Unlabeled and labeled vesicle solutions were mixed in a 9:1 ratio and the pH was adjusted using 1 M HCl or 1 M NaOH. Protein-induced vesicle fusion was detected via lipid mixing between labeled and unlabeled vesicles with consequent higher fluorescence because of longer average fluorophore-quencher distance. Fluorescence was detected: (1) for the initial labeled+unlabeled vesicle solution; (2) with time-resolution after addition of an aliquot of protein solution; and (3) after addition of an aliquot of Triton X-100 detergent that solubilized the vesicles and maximized

fluorescence. The respective fluorescence values were  $F_0$ ,  $F(t)$ , and  $F_{max}$  and the percent vesicle fusion was calculated:

$$M(t) = [(F(t) - F_0)/(F_{max} - F_0)] \times 100 \quad \text{----- Equation 4-1}$$

Typical experimental conditions included: (1) PTI QW4 fluorimeter with 467 nm excitation and 530 nm detection wavelengths; (2) vesicles in 1400  $\mu\text{L}$  volume and [total lipid] = 150  $\mu\text{M}$  and [Chol] = 75  $\mu\text{M}$ ; (3) constant stirring; (4) protein:lipid (PC+AL) = 1:700 achieved with a 7.5  $\mu\text{L}$  aliquot of 40  $\mu\text{M}$  protein in 10 mM formate buffer at pH 3.2 with 0.2  $\mu\text{M}$  TCEP reducing agent; (5) syringe injection of protein solution to reduce assay dead time to  $\sim 1$  s; (6) 1 s time increments for  $F(t)$ ; and (7) 0.08% w/v Triton X-100 achieved with a 12  $\mu\text{L}$  aliquot of 10% Triton X-100. The protein:lipid = 1:700 ratio was chosen in part because there was light scattering from the fused vesicles for higher ratios.

## 4-3 Results

### *4-3-1 Vesicle fusion requires attractive electrostatic energy*

For either HP or FP-HP, the protein charge is  $\sim +9$  at pH 3.5,  $\sim +7$  at pH 4.0 and  $\sim -2$  at pH 7.0. For protein: lipid mole ratio = 1:700, fusion was only observed at low pH with anionic vesicles (Figure 4-2). There was negligible fusion of anionic vesicles at neutral pH and also little fusion with neutral vesicles at either low or neutral pH. These data support a fusion requirement of attractive protein/vesicle electrostatic energy which is likely needed for significant binding.

For a given protein and vesicle composition, there was typically greater fusion at pH 3.5 relative to pH 4.0.

*Fast and slow fusion.*

The time courses of fusion of anionic vesicles at low pH,  $M(t)$ , were well-fitted with either the sum of a fast and slow exponential buildup:

$$M(t) = M_f(1 - e^{-k_f t}) + M_s(1 - e^{-k_s t}) \text{----- Equation 4-2a}$$

or a single slow exponential buildup:

$$M(t) = M_s(1 - e^{-k_s t}) \text{----- Equation 4-2b}$$

The dark lines in figure 4-2 are the best-fits of the experimental data. Use of one vs two buildups for fitting was determined by visual comparison of the best-fit and the experimental data. In practice, the fast buildup was used for  $|\text{charge/lipid}| \geq 0.12$ . The  $k_f$  and  $k_s$  are the fusion rate constants with  $k_f$  in the  $30 - 200 \text{ ms}^{-1}$  range and  $k_s$  in the  $1 - 9 \text{ ms}^{-1}$  range. The  $M_f$  and  $M_s$  are the long-time fusion extents of the fast and slow processes, respectively. Table 4-1 lists the fitted rates and extents and figure 4- 3 displays bar plots of the pH 3.5 values. For each set of assay conditions, data were acquired in triplicate and each data set was fitted independently. Each reported parameter uncertainty is the standard deviation among these replicate values. The uncertainty from a single replicate fitting is much smaller, typically  $<1\%$  of the best-fit parameter value.

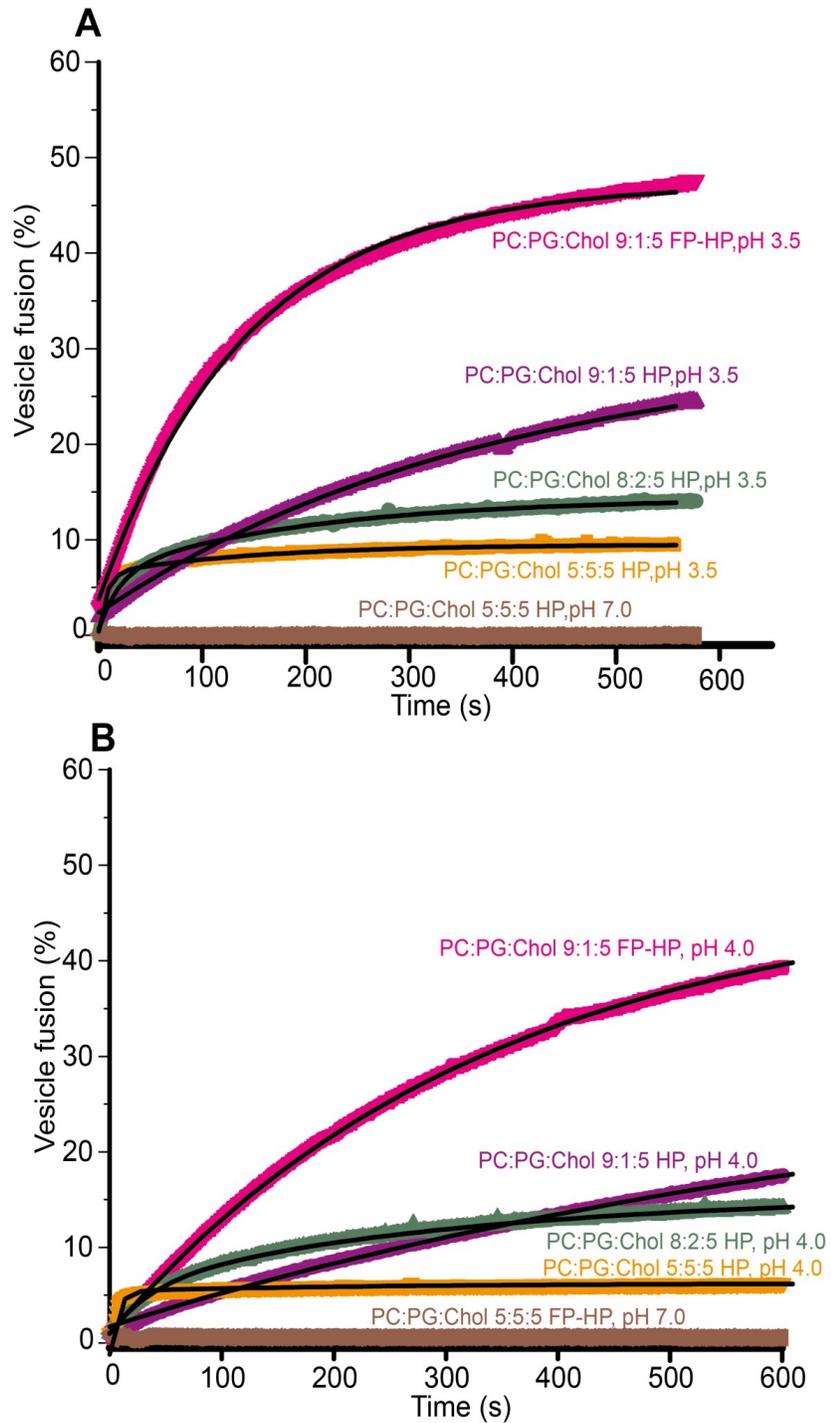


Figure 4-2 Representative time-courses of protein-induced vesicle fusion for protein:total lipid mole ratio = 1:700 at pH (A) 3.5 and (B) 4.0. Chol is not considered in the total lipid quantity. Best-fits are displayed in black.

#### ***4-3-2 Fast rate is strongly positively correlated with lipid charge***

The rate of fast fusion ( $k_f$ ) is highly positively correlated with the average anionic lipid charge, e.g.  $k_f \approx 200 \text{ ms}^{-1}$  for charge/lipid  $\approx -0.5$  and  $k_f \approx 30 \text{ ms}^{-1}$  for charge/lipid  $\approx -0.2$ . The fast buildup is not observed for  $|\text{charge/lipid}| \leq 0.1$  which suggests that the slow process becomes dominant under this condition. The correlation of  $k_f$  with lipid charge is further supported by similar values of  $k_f$  with vesicles with different fractions of PG or PS but similar charge/lipid, e.g. similar  $k_f$  for PC:PG:Chol = 8:2:5 and PC:PS:Chol = 5:5:5. The  $k_f$  value may be more specifically correlated to the magnitude of the attractive protein/vesicle electrostatic energy because for the same protein and vesicle composition, the  $k_f$  at pH 3.5 (protein charge  $\approx +9$ ) is typically  $\sim 30\%$  higher than the  $k_f$  at pH 4.0 (protein charge  $\approx +7$ ). The figure 4-4 (A) plot of  $k_f$  vs  $|\text{charge/protein} \times \text{charge/lipid}|$  for HP at pH 3.5 and 4.0 illustrates the positive correlation between  $k_f$  and electrostatic energy. The typical long-time extent of fast buildup,  $M_f$  is  $\sim 5\%$  with little dependence on charge/lipid or pH, i.e. variation in protein/vesicle electrostatic energy.

#### ***4-3-3 Slow extent is inversely dependent on lipid charge***

The slow process rate constant  $k_s$  is typically in the  $1 - 9 \text{ ms}^{-1}$  range. As with  $k_f$ , for a given protein and pH, there is typically a positive correlation between  $k_s$  and  $|\text{charge/lipid}|$  and an inverse correlation with pH. The dependences likely reflect a contribution from attractive protein/vesicle electrostatic energy although the  $k_s$  dependence has much smaller magnitude than the corresponding  $k_f$  dependence. There is also strong inverse correlation of the extent  $M_s$  with  $|\text{charge/lipid}|$ , e.g. for FP-HP induced fusion of PC:PG:Chol vesicles at pH 3.5, the decrease of charge/lipid from -0.5 to -0.1 correlates with an increase in  $M_s$  from  $\sim 5\%$  to  $\sim 50\%$ . The inverse

dependence of  $M_s$  on  $|\text{charge/lipid}|$  is likely a consequence of the decrease in repulsive intervesicle electrostatic energy with decreasing charge. For example, the figure 4-4 (B) plot of  $M_s$  vs  $(\text{charge/lipid})^{-2}$  has positive slope for  $|\text{charge/lipid}| > 0.1$ . The  $M_s$  is approximately constant for  $|\text{charge/lipid}| < 0.1$ .

#### 4-3-4 Higher fusion for FP-HP

For the same vesicle composition and pH, the  $k_f$  or  $k_s$  for FP-HP is typically 1.5 – 2 times larger than the corresponding  $k_f$  or  $k_s$  for HP. The  $M_f$  or  $M_s$  are similarly larger too.

Table 4-1: Best-fit parameters of vesicle fusion.<sup>a,b</sup>

<b>Hairpin pH 3.5</b>					
Vesicle composition	Charge/lipid <sup>c</sup>	$k_f(\text{ms})^{-1}$	$M_f(\%)$	$k_s(\text{ms})^{-1}$	$M_s(\%)$
PC:PG:Chol (5:5:5)	- 0.485	178(26)	7.1(9)	8(1)	3.4(4)
PC:PG:Chol (8:2:5)	- 0.194	42(5)	6.8(5)	4.2(5)	7.8(1)
PC:PG:Chol (9:1:5)	- 0.097	-	-	2.4(2)	29.1(9)
PC:PS:Chol (5:5:5)	- 0.120	34(4)	3.6(7)	4.5(4)	12(4)
PC:PS:Chol (8:2:5)	- 0.048	-	-	1.4(1)	37(3)
PC:PS:Chol (9:1:5)	- 0.024	-	-	1.4(2)	35(4)

<b>FP-Hairpin pH 3.5</b>					
Vesicle composition	Charge/lipid	$k_f(\text{ms})^{-1}$	$M_f(\%)$	$k_s(\text{ms})^{-1}$	$M_s(\%)$
PC:PG:Chol (5:5:5)	- 0.485	216(61)	9.6(5)	9(3)	4.2(8)
PC:PG:Chol (8:2:5)	- 0.194	51(12)	10(2)	5.2(7)	10(1)
PC:PG:Chol (9:1:5)	- 0.097	-	-	8(1)	46(5)
PC:PS:Chol (5:5:5)	- 0.120	58(11)	15(3)	9(2)	14(3)
PC:PS:Chol (8:2:5)	- 0.048	-	-	3.7(1)	50(1)
PC:PS:Chol (9:1:5)	- 0.024	-	-	1.9(1)	50(2)

Table 4-1 (cont'd)

**Hairpin pH 4.0**

Vesicle composition	Charge/lipid	$k_f(\text{ms})^{-1}$	$M_f$ (%)	$k_s(\text{ms})^{-1}$	$M_s$ (%)
PC:PG:Chol (5:5:5)	- 0.495	135(17)	5.3(2)	4(2)	1.3(4)
PC:PG:Chol (8:2:5)	- 0.198	27(1)	5(1)	2.7(1)	9(2)
PC:PG:Chol (9:1:5)	- 0.099	-	-	1.3(3)	20(2)
PC:PS:Chol (5:5:5)	- 0.250	32(3)	3.6(9)	4.0(5)	7.7(7)
PC:PS:Chol (8:2:5)	- 0.100	-	-	1.4(1)	23(6)
PC:PS:Chol (9:1:5)	- 0.050	-	-	1.0(3)	11(2)

**FP-Hairpin pH 4.0**

Vesicle composition	Charge/lipid	$k_f(\text{ms})^{-1}$	$M_f$ (%)	$k_s(\text{ms})^{-1}$	$M_s$ (%)
PC:PG:Chol (5:5:5)	- 0.495	201(60)	6.0(5)	15(2)	2.4(6)
PC:PG:Chol (8:2:5)	- 0.198	27(3)	12(1)	4.3(8)	10.8(5)
PC:PG:Chol (9:1:5)	- 0.099	-	-	3.0(1)	43(3)
PC:PS:Chol (5:5:5)	- 0.250	25(4)	6(1)	3(1)	3(1)
PC:PS:Chol (8:2:5)	- 0.100	-	-	3.7(2)	32(3)
PC:PS:Chol (9:1:5)	- 0.050	-	-	1.4(3)	29(3)

<sup>a</sup> The assay for each set of conditions was repeated in triplicate with separate fitting of each trial. Each reported parameter value and uncertainty is respectively the average and standard deviation of the three replicate fittings. For a single data set, the fitting uncertainty was typically <1% of the fitted parameter value.

<sup>b</sup> Protein: total lipid mole ratio = 1:700. Chol is not included in the total lipid quantity.

<sup>c</sup> Charge/ lipid was calculated using a model for which: (1) PC is uncharged; and (2) and PG and PS are each characterized by an equilibrium  $\text{HA} \leftrightarrow \text{H}^+ + \text{A}^-$  with  $\text{pK}_a = 2$  for PG and 4 for PS [6].

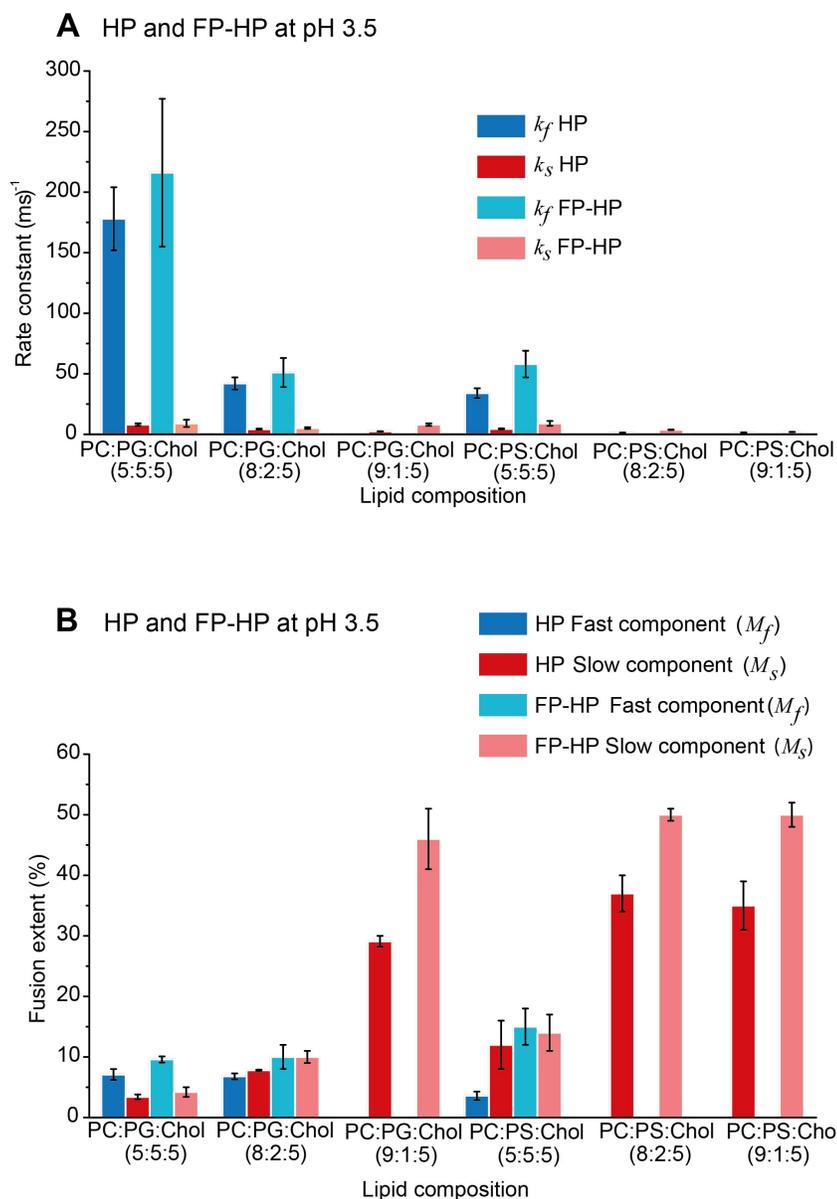


Figure 4-3 (A) Buildup rate and (B) final extent best-fit parameters at pH 3.5. Each displayed parameter value and uncertainty is respectively the average and standard deviation among three replicates with separate fitting of each trial. The fitting uncertainty of a parameter for an individual trial is typically <1% of the best-fit value. For PC:PG:Chol = 5:5:5 and 8:2:5 and PC:PS:Chol = 5:5:5, the fitting model is a sum of a fast and a slow exponential-buildup while for other compositions, the model is a single slow buildup. Choice of fitting model was based on visual inspection of goodness-of-fit.

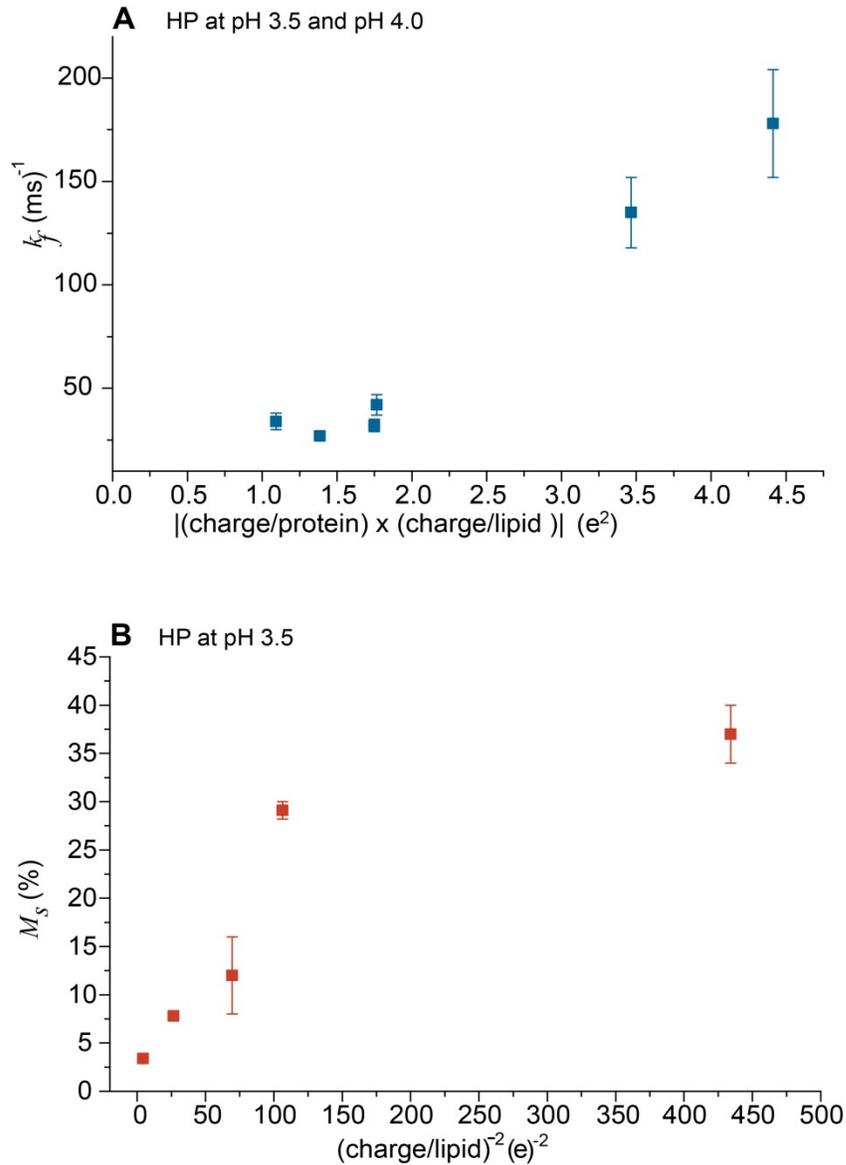


Figure 4-4 Plots of (A)  $k_f$  vs  $|(\text{charge/protein}) \times (\text{charge/lipid})|$  and (B)  $M_s$  vs  $(\text{charge/lipid})^{-2}$ . The positive correlation of plot A supports the fast process resulting from release of electrostatic energy from protein/vesicle binding. Plot B supports an inverse correlation between the extent of slow fusion and inter-vesicle repulsion. Plot A displays data for HP at pH 3.5 and 4.0 and plot B displays data for HP at pH 3.5.

## 4-5 Discussion

### 4-5-1 Fast and slow fusion processes

One important result of the present study was detection of fast  $\sim 100 \text{ ms}^{-1}$  and slow  $\sim 5 \text{ ms}^{-1}$  components of vesicle fusion. The rate constant of the fast component  $k_f$  is strongly positive correlated with  $|\text{charge/lipid}|$  and therefore the magnitude of attractive protein/vesicle electrostatic energy (figure 4-4(A)). The fusion extent  $M_f$  is approximately independent of this energy. The magnitude of this energy is 60 kJ/(mole protein) as calculated for: (1) vesicles with 20% negatively charged lipid; (2) monomer protein with +7 charge; (3) bound protein separated by 20 Å from 50 lipid phosphate groups; and (4) aqueous media with dielectric constant of 80. Fast fusion is attributed to deposition of this energy in the membrane and the consequent local membrane perturbation. After dissipation of this energy, fusion may be arrested by intervesicle repulsion whose magnitude also increases with  $|\text{charge/lipid}|$ .

The  $k_f$ 's are  $\sim 1\%$  of the rates of protein/vesicle ( $k_{pv}$ ) and vesicle/vesicle ( $k_{vv}$ ) collisions [23]. Protein/vesicle electrostatics are attractive and orientation-independent, so a protein likely binds to the vesicle after a single collision. For protein: lipid = 1:700, each 100 nm diameter vesicle has  $\sim 80$  bound proteins that cover  $\sim 10\%$  of the vesicle surface. The ratio for  $k_f/k_{vv}$  of  $\sim 0.01$  may reflect the  $(0.1)^2$  fraction of inter-vesicle collisions with contact between protein regions on both vesicles. Other collision geometries may not lead to fusion because of repulsion between the negatively-charged vesicle surfaces.

Relative to the  $k_f$ 's, the  $k_s$ 's show weaker dependence on protein and vesicle charges. The positive  $k_s$  dependence likely reflects a contribution from attractive protein/vesicle electrostatic

energy. The  $M_s$ 's have strong inverse correlation on  $|\text{charge/lipid}|$  in the 0.1 to 0.5 range, with no further increases for  $|\text{charge/lipid}|$  lower than 0.1. We assign the slow process to fusion after dissipation of the attractive protein/vesicle electrostatic energy with higher extents at smaller  $|\text{charge/lipid}|$  attributed to reduced intervesicle repulsion (figure 4-4(B)). The slow process occurs on the  $\sim 200$  s timescale and reflects rare fusion events that occur after hundreds of vesicle/vesicle collisions. The probability of a fusogenic collision likely increases with decreased inter-vesicle repulsive energy. The approximately constant  $M_s$  for  $|\text{charge/lipid}| \leq 0.1$  may reflect competing effects of smaller intervesicle repulsion and smaller protein/vesicle attraction that results in fewer bound proteins. This latter effect is consistent with negligible fusion of electrically neutral PC:Chol vesicles at either low or neutral pH.

Figure 4-5 presents a qualitative model of gp41 ectodomain-induced vesicle fusion at low pH. The fast process is attributed to bound protein monomers that reflect the predominant protein state in aqueous solution. The slow process is attributed to membrane-bound protein oligomers which are the likely equilibrium structural state and which may form through protein diffusion on the membrane surface on the  $\sim 200$  s time scale of the slow process. For some conditions, the displayed hexamer  $\equiv$  two trimer state is the predominant oligomerization state for large ectodomain constructs such as HP [12, 13].

Protein  $^{13}\text{C}$ -to-lipid  $^{31}\text{P}$  REDOR SSNMR spectra obtained by the co-authors Sackett. K, Nethercott. M. J, and Weliky. D.P showed that the helical NHR and CHR regions of FP-HP are closer to the membrane surface at low pH than at neutral pH which correlates with the respective attractive and repulsive protein/membrane electrostatic energy [24].

#### ***4-5-2 Slow vesicle fusion as a model for HIV hemifusion***

The lipid mixing assay of ectodomain-induced vesicle fusion correlates with the initial hemifusion (joining) of viral and cell membranes. At low pH, the protein is monomeric in solution and binding to anionic membranes is favored by attractive electrostatics. At neutral pH, the protein aggregates in solution and binding is inhibited by repulsive electrostatics. HIV/cell fusion probably occurs at neutral pH with anionic membranes but is different from vesicle fusion at neutral pH because viral protein aggregation is disfavored and viral protein binding to the cell membrane is favored. These properties are the result of: (1) viral gp41 being an integral membrane protein; and (2) spatial confinement of the viral gp41 ectodomain to the region between viral and host cell membranes because of initial binding between the extra viral gp120 protein and cell membrane proteins. Viral fusion is therefore better modeled by ectodomain-induced vesicle fusion at low pH because there is minimal protein aggregation in solution and high membrane-binding.

In addition, virus/cell hemifusion is probably better modeled by the slow than the fast component of vesicle fusion. The fast process is highly positively correlated with the attractive protein/membrane electrostatic energy and this energy is likely small or even repulsive for viral fusion at neutral pH. In addition, the  $|\text{charge}/\text{lipid}|$  is  $\sim 0.15$  for HIV and  $\sim 0.07$  for the host cell membrane [22]. For this regime, there is maximal slow and minimal fast vesicle fusion. To our knowledge, the viral hemifusion rate has not been measured but the typical overall timescale for viral fusion is  $\sim 30$  min [25]. Relative to the fast timescale of  $\sim 0.1$  min, the slow timescale of  $\sim 3$  min also seems more reasonable for viral fusion. Correlation of slow vesicle fusion with viral fusion suggests that the active state of the viral gp41 ectodomain may be oligomeric with membrane-inserted FP's in an intermolecular antiparallel  $\beta$  sheet (figure 4-5). This is consistent

with earlier data supporting the significance of FP oligomers in viral fusion and infection [16, 17, 23].

Although the fusion of anionic vesicles requires low pH for attractive protein/vesicle electrostatic energy and quantitative binding, viral fusion likely occurs near neutral pH. This holds for either direct fusion between the viral and plasma membranes or for endocytosis with subsequent fusion between the viral and endosomal membranes [26, 27]. There is no evidence yet for acidification of HIV-containing endosomes.

The whole gp41 protein including TM and endodomain is likely minimally trimeric during all steps of viral fusion. The initial gp160 complex contains three gp41 and three gp120 subunits with the three NHR helices forming a bundle at the center of the complex [28]. After binding of the three gp120's to host cell receptors, the gp120's move away from the gp41's. The NHR bundle may dissociate with subsequent formation of three NHR/CHR monomer ectodomain hairpins. These monomer hairpins are hyperthermostable with  $T_m \approx 110^\circ\text{C}$  [2, 3]. Evidence for the ectodomain monomer as an intermediate in viral fusion includes fusion inhibitors that could bind to the ectodomain hairpin monomer but not to the final trimer of hairpins (six-helix bundle) state. The inhibitors would reduce the oligomer populations of trimer and hexamer and consequently the intermolecular antiparallel FP  $\beta$  sheet [15, 25]. This interpretation of the peptide inhibition supports the role of this sheet for hemifusion and viral infection.

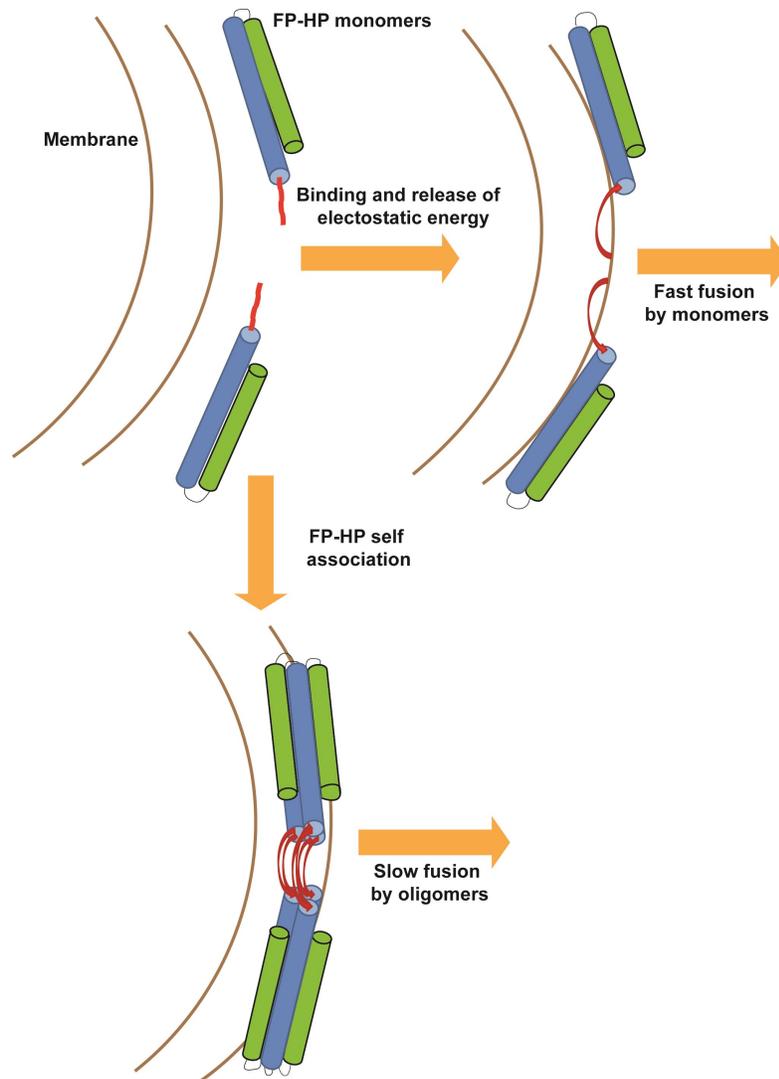


Figure 4-5 Qualitative model of FP-HP-induced fusion of anionic vesicles with FP-HP color coding matched to Fig. 4-1. Monomers from solution bind to the vesicles with release of electrostatic energy that results in fast fusion. The bound monomers also diffuse on the vesicle surface and self-associate as small oligomers with formation of an intermolecular antiparallel FP  $\beta$  sheet which is associated with slow fusion [5].

## REFERENCES

## REFERENCES

1. Sackett, K., et al., *Hairpin Folding of HIV gp41 Abrogates Lipid Mixing Function at Physiologic pH and Inhibits Lipid Mixing by Exposed gp41 Constructs*. *Biochemistry*, 2009. **48**(12): p. 2714-2722.
2. Sackett, K., et al., *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*, 2010. **397**(1): p. 301-315.
3. Lev, N., et al., *Conformational Stability and Membrane Interaction of the Full-Length Ectodomain of HIV-1 gp41: Implication for Mode of Action*. *Biochemistry*, 2009. **48**(14): p. 3166-3175.
4. White, J.M., et al., *Structure and mechanism of viral membrane fusion proteins: multiple variations on common theme*. *Critical Reviews in Biochemistry and Molecular Biology* 2008. **43** p. 189 – 219.
5. Ott, M., Y. Shai, and G. Haran, *Single-Particle Tracking Reveals Switching of the HIV Fusion Peptide between Two Diffusive Modes in Membranes*. *Journal of Physical Chemistry B*, 2013. **117**(42): p. 13308-13321.
6. Tsui, F.C., D.M. Ojcius, and W.L. Hubbell, *The intrinsic pKa Values for phosphatidylserine and phosphatidylethanolamine in phosphatidylcholine host bilayers*. *Biophysical Journal*, 1986. **49**(2): p. 459-468.
7. Liu, J., et al., *Molecular architecture of native HIV-1 gp120 trimers*. *Nature*, 2008. **455**(7209): p. 109-U76.
8. Caffrey, M., et al., *Three-dimensional solution structure of the 44 kDa ectodomain of SIV gp41*. *Embo Journal*, 1998. **17**(16): p. 4572-4584.
9. Yang, Z.N., et al., *The crystal structure of the SIV gp41 ectodomain at 1.47 angstrom resolution*. *Journal of Structural Biology*, 1999. **126**(2): p. 131-144.
10. Buzon, V., et al., *Crystal Structure of HIV-1 gp41 Including Both Fusion Peptide and Membrane Proximal External Regions*. *Plos Pathogens*, 2010. **6**(5).
11. Caffrey, M., et al., *Monomer-trimer equilibrium of the ectodomain of SIV gp41: Insight into the mechanism of peptide inhibition of HIV infection*. *Protein Science*, 1999. **8**(9): p. 1904-1907.
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. Gao, G.F., et al., *Designing a Soluble Near Full-length HIV-1 gp41 Trimer*. *Journal of Biological Chemistry*, 2013. **288**(1): p. 234-246.
  14. Lakomek, N.A., et al., *Internal Dynamics of the Homotrimeric HIV-1 Viral Coat Protein gp41 on Multiple Time Scales*. *Angewandte Chemie-International Edition*, 2013. **52**(14): p. 3911-3915.
  15. Markosyan, R.M., F.S. Cohen, and G.B. Melikyan, *HIV-1 envelope proteins complete their folding into six-helix bundles immediately after fusion pore formation*. *Molecular biology of the cell*, 2003. **14**(3): p. 926-938.
  16. Freed, E.O., et al., *A mutation in the Human Immunodeficiency-Virus Type-1 Transmembrane glycoprotein-gp41 dominantly interferes with fusion and infectivity*. *Proceedings of the National Academy of Sciences of the United States of America*, 1992. **89**(1): p. 70-74.
  17. Magnus, C., et al., *Estimating the Stoichiometry of Human Immunodeficiency Virus Entry*. *Journal of Virology*, 2009. **83**(3): p. 1523-1531.
  18. 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.
  19. Pereira, F.B., et al., *Permeabilization and fusion of uncharged lipid vesicles induced by the HIV-1 fusion peptide adopting an extended conformation: Dose and sequence effects*. *Biophysical Journal*, 1997. **73**(4): p. 1977-1986.
  20. Cheng, S.F., et al., *The fusion peptide domain is the primary membrane-inserted region and enhances membrane interaction of the ectodomain of HIV-1 gp41*. *Molecular Membrane Biology*, 2010. **27**(1): p. 31-44.
  21. Sackett, K., et al., *Solid-State NMR Spectroscopy of the HIV gp41 Membrane Fusion Protein Supports Intermolecular Antiparallel 13 Sheet Fusion Peptide Structure in the Final Six-Helix Bundle State*. *Journal of Molecular Biology*, 2014. **426**(5): p. 1077-1094.
  22. Brugger, B., et al., *The HIV lipidome: A raft with an unusual composition*. *Proceedings of the National Academy of Sciences of the United States of America*, 2006. **103**(8): p. 2641-2646.

23. Yang, R., et al., *A trimeric HIV-1 fusion peptide construct which does not self-associate in aqueous solution and which has 15-fold higher membrane fusion rate*. Journal of the American Chemical Society, 2004. **126**(45): p. 14722-14723.
24. 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 conformation*. Biochimica Et Biophysica Acta-Biomembranes, 2015. **1848**(1): p. 289-298.
25. Gallo, S.A., et al., *The HIV Env-mediated fusion reaction*. Biochimica Et Biophysica Acta-Biomembranes, 2003. **1614**(1): p. 36-50.
26. Grewe, C., A. Beck, and H.R. Gelderblom, *HIV Early virus-cell interactions*. Journal of Acquired Immune Deficiency Syndromes and Human Retrovirology, 1990. **3**(10): p. 965-974.
27. Melikyan, G.B., *HIV entry: a game of hide-and-fuse?* Current Opinion in Virology, 2014. **4**: p. 1-7.
28. Julien, J.P., et al., *Crystal Structure of a Soluble Cleaved HIV-1 Envelope Trimer*. Science, 2013. **342**(6165): p. 1477-1483.

**Chapter 5 Characterization of full length HA2 using hydrogen deuterium exchange mass spectrometry (HDX)**

## 5-1 Introduction

The research presented in this dissertation was the first successful attempt to purify full length HA2 protein without the presence of a large solubilization tag. Full length HA2 is a helical, trimer in decylmaltoside (DM) detergent at pH 7.4 with  $T_m > 90^\circ \text{C}$ . The full length HA2 is likely in a trimer of hairpin structure (described briefly in the discussion part of chapter 3). The crystal structures and functional studies performed by other groups were based on the truncated versions of HA2 [1-3]. Based on the proposed fusion model as described in chapter 1, it is hypothesized that a trimer of hairpin structure is formed [5]. In the final stage of trimeric hairpin structure, it is hypothesized that a complex is formed between FP and TM as depicted in the chapter 1 figure 1-4 (f). However, there is no experimental evidence to support the formation of this complex in between FP and TM. We designed a hydrogen – deuterium exchange experiment to investigate this hypothesis. The hypothesis we proposed was that, if FP and TM formed a complex in the final trimer of hairpin structure of full length HA2 protein, the exchange rate of deuterons present in the FP region of HA2 can be different from the exchange rate of deuterons present in the FP region of FHA2 which lacks TM. Additionally, this experiment enables the determination of solvent exposed and unexposed regions of a protein. This information will allow conclusions to be made about the solvent accessible nature of the FP (ex: if FP is located in a hydrophobic core, it will display less HD exchange whereas if it is located at a solvent accessible place, high level of HD exchange can be expected).

## 5-2 Experimental design

The basic procedure used in HD exchange experiment is depicted as flowchart in figure 5-1. The specific experimental procedures used in this study are briefly described under section 5-5. Liquid chromatography-mass spectrometry (LC-MS) based HD exchange experiments requires the initial identification of the peptides generated from a pepsin digestion of a given protein so that exchange can be assigned to specific parts of the protein to determine the HD exchange rate under different experimental conditions. Therefore, before proceeding with actual HD exchange experiment, identification of the peptic peptides was performed. During the HD exchange experiment the deuterium exchange of these identified peptides was monitored. For the purpose of identification of peptides resulting from FP region of the protein being studied, first, pepsin digestion followed by LC-MS characterization was performed only for synthetically prepared FP. Since synthetically prepared FP (figure 5-3) has only 27 residues (along with the solubilization tag) chromatogram resulting from FP had fewer chromatographic peaks compared to the chromatograms resulting from proteins. This is because pepsin digestion of FP results in fewer peptides while full length HA2 (which has ~220 residues) digestion using pepsin results in more peptides. Accurate mass of the resulting peptides from the pepsin digestion was matched with the theoretical peptide mass predicted by the peptide cutter software. Based on accurate mass match (mass accuracy  $\pm 10$  ppm) between theoretical mass and observed mass, chromatographic peaks were assigned to predicted peptides (Calculation of mass accuracy is given below under the section 5-5-1).

Pepsin was used for the digestion, since it is capable of rapid digestion of the protein and is active in low pH levels (less than pH 3.0). This is necessary since the HD exchange of the protein is quenched by lowering the pH (to below pH 3.0) to minimize the back exchange of the

deuterons with protons during LC-MS analysis. Therefore an enzyme active at low pH (such as pepsin) is necessary to cleave the protein into the peptides that are detected. Pepsin is an enzyme which cleaves peptide bonds between hydrophobic amino acids. Pepsin shows its maximum activity below pH 3.0 and it becomes inactive at pH values above 6.5. Before doing actual HD exchange experiment peptides resulting from FP region were identified. Next, peptides resulting from HA2 and FHA2 were analyzed to verify the presence of the peptides identified from FP since both HA2 and FHA2 also has FP in them. The longest peptide representing the FP was selected as the signature peptide for FP and this was monitored during HD exchange experiment. HD exchange was performed at different time points for both FHA2 and HA2, and the number of deuterons exchanged in the signature peptide representing the FP was calculated.

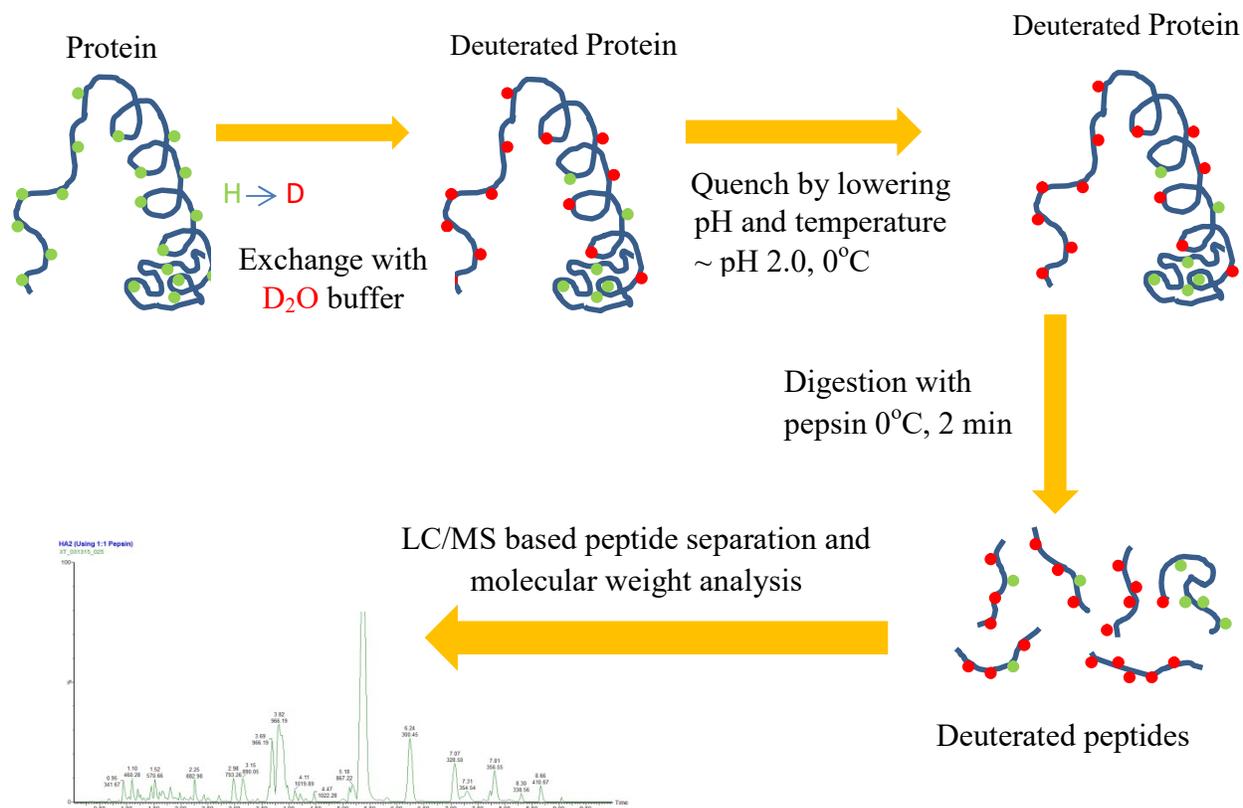


Figure 5-1 Diagram showing the basic procedure used for the HD exchange experiment [4].

### 5-3 Reagent and Chemicals

$D_2O$ , acetonitrile (ACN) and formic acid were purchased from Sigma. Enzymate BEH Pepsin column (2.1 mm $\times$ 30 mm), XBridge BEH , (2.1  $\times$  50 mm) 2.5  $\mu m$  particle size, C18 column, XBridge BEH Vanguard pre-column (trap column) ( 2.1 $\times$ 5mm) 2.5  $\mu m$  particle size were purchased from Waters Inc.

## **5-4 LC-MS analysis of peptides generated from the pepsin digestion of FP, HA2 and FHA2 protein**

Initial pepsin digestions were performed in solution (off column) to identify the signature peptides for FP from synthetic FP, HA2 and FHA2. Once these were identified, the on-column digestion procedure was optimized.

### ***5-4-1 Off-column digestion of fusion peptide and the proteins followed by LC-MS analysis***

Synthetic Fusion peptide was dissolved in 50% formic acid solution prepared in water at a concentration of 0.1 mg/mL. HA2 and FHA2 were purified as described in chapter 3. Both samples were dialyzed into buffer containing 0.17% DM, 10mM Tris, at pH 7.4. After two days of dialysis, protein samples were concentrated up to 1 mg/mL.

1 mg/mL solution of pepsin was prepared in 1% formic acid solution. 90  $\mu$ L sample (fusion peptide solution or the proteins solution) was mixed with 10  $\mu$ L of the pepsin solution. The mixtures were incubated at room temperature for 2 min and analyzed using LC-MS. Chromatographic gradient used is shown in table 5-1.

### ***5-4-2 On-column digestion of the proteins followed by LC-MS analysis***

This was performed on a Waters Enzymate Pepsin column. The column matrix has Pepsin immobilized on the surface. Protein that needs to be digested is first dissolved in 1% formic acid solution at 1.0 mg/mL and 50  $\mu$ L of this solution was injected to the external

injection port (figure 5-2). This is accomplished by injecting the sample using a 100  $\mu$ L syringe into the external injection port when it is set to “load” position. Then the port is set to “inject” position and the external pump flow rate is set to 0.1 ml/min and is started. This flows the 0.1% formic acid in the solvent reservoir connected to the external pump through the injection port and the enzymate column. This results in transfer of the sample from the “sample loop” into the Enzymate column (Figure 5-2). After 1 min, the external pump liquid flow was stopped to “park” the protein molecules on the Enzymate column for 2 min. After 2 min, external pump flow was restarted at 0.3 ml/min to elute the peptides formed from digestion on the Enzymate column. These peptides were captured onto a C18 trap column located at the injection port connected to the LC system coupled to the mass spectrometer. The trap column was used to trap the peptides resulting from the peptide digestion so that they can be introduced as a single sample plug onto the C18 analytical column. Then, the LC-MS analysis is initiated from Masslynx software, and this starts the chromatographic gradient. As the chromatographic gradient is initiated, it washes off the peptide onto the analytical C18 column by pumping UPLC mobile phase (0.1% formic acid in water and acetonitrile) at 0.3 mL/min. The peptides were separated on the analytical column using gradient elution given in table 5-1 and were detected by the mass spectrometer. Based on the accurate mass, the peptides were assigned to predicted pepsin digestion products from Peptide Cutter software. Waters Xevo G2S QToF Mass Spectrometer coupled to a Waters Acuity UPLC system controlled using Waters Masslynx 4.1 software was used for data acquisition and analysis.

Table 5-1: Solvent gradient used for the separation of pepsin digestion products of proteins on the C18 analytical column.

<b>Time (min)</b>	<b>%A (Water with 0.1% Formic Acid)</b>	<b>%B (Acetonitrile)</b>	<b>Flow rate (ml/min)</b>
0.01	99	1.0	0.3
1.00	99	1.0	0.3
3.00	70	30	0.3
8.00	1.0	99	0.3
9.00	1.0	99	0.3
9.01	99	1.0	0.3
10.00	99	1.0	0.3

#### ***5-4-3 Hydrogen – Deuterium exchange experiment***

10  $\mu$ L of 1mg/ml of HA2 and FHA2 was mixed with 90  $\mu$ L of D<sub>2</sub>O containing 5 mM HEPES and 10 mM MES buffer at pH 7.4 (Addition of D<sub>2</sub>O will dilute the detergent concentration of the medium in which the protein is solubilized. Therefore, it is advisable to use buffer containing detergent of 0.17% DM in D<sub>2</sub>O for future experiments). HD exchange incubation was performed for different time points (0.5 min, 1.5 min, 3.0 min, 10 min and 24 hour). Next, the exchange was quenched by lowering the pH of the reaction to less than 3 by addition of lab stock formic acid (10  $\mu$ L) and immediately placed on ice. Then, the 50  $\mu$ L of sample was injected into the external sample loop. Then the procedure discussed under section 5-4-2 was performed. All connection tubing, Enzymate pepsin column, C18 analytical column, external injection port, trap column and all solvents used were prechilled to 4 <sup>0</sup>C and were maintained at that temperature by keeping on ice.

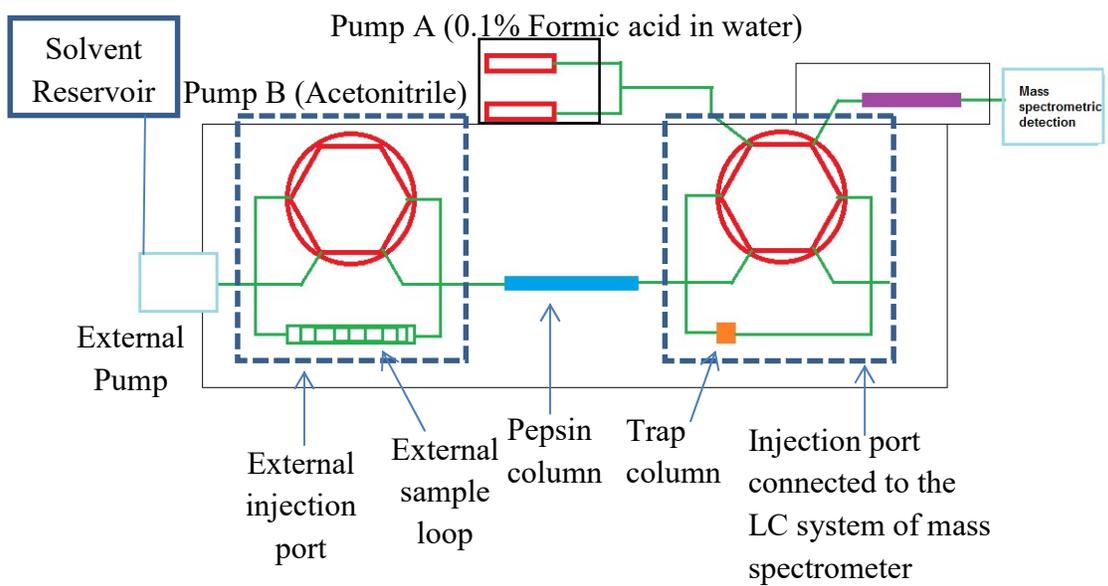


Figure 5-2 H-D exchange apparatus used for the experiments [6].

## 5-5 Results

G<sub>1</sub>LFGAIAGFIENGWEGMIDGGGK<sub>4</sub>KKK<sub>2</sub>G<sub>1</sub><sub>27</sub>

Figure 5-3 Amino acid sequence (27 residues) of the fusion peptide used for the preliminary identification of peptides resulted from FP region of the proteins. Nonnative G<sub>2</sub>K<sub>4</sub>G<sub>1</sub> tag present in the C-terminus was added for the purpose of solubilization and purification with HPLC.

Using peptide cutter software, masses of the peptides resulting from the pepsin digestion of FP were predicted. Using the amino acid sequence of the predicted peptides the theoretical accurate mass is calculated. Based on accurate mass match between the spectra of each chromatographic peak observed and predicted mass, observed spectra and chromatographic peaks were assigned to each predicted peptide.

Table 5-2: Table showing the sequence of the predicted pepsin digestion products for FP and their accurate masses. Yellow color highlighted columns showing the peptides that were identified by the experiment.

Peptide sequence	Predicted mass of +1 charge state
IAGFIENGWEGMIDGGGKKKKG	2292.1804
GFIENGWEGMIDGGGKKKKG	2108.0593
IENGWEGMIDGGGKKKKG	1903.9694
NGWEGMIDGGGKKKKG	1661.8427
EGMIDGGGKKKKG	1304.699
FGAIAGFIENGW	1281.6262
GAIAGFIENGWE	1263.6004
GMIDGGGKKKKG	1175.6565
IAGFIENGWE	1135.5418
GAIAGFIENGW	1134.5578
GLFGAIAGFIE	1094.588
IAGFIENGW	1006.4992
GFIENGWE	951.4206
FGAIAGFIE	924.4825
<b>GLFGAIAGF</b>	<b>852.4614</b>
GFIENGW	822.378
GAIAGFIE	777.4141
IENGWE	747.3308
FGAIAGF	682.3558
IAGFIE	649.3555
GLFGAIA	648.3715
IENGW	618.2882
<b>GAIAGF</b>	<b>535.2874</b>
NGWE	505.2041
FGAIA	478.266
GFIE	465.2343
GLFGA	464.2503
IAGF	407.2289
NGW	376.1615
<b>GLF</b>	<b>336.1918</b>
GAIA	331.1976
FGA	294.1448
GA	147.0764

Given below represents the chromatogram (figure 5-4 (A)) obtained by separating the pepsin digestion products of FP on the C18 analytical column. Mass spectra corresponding to peaks a, b and c without HDX are shown in figure 5-4 (B). The unassigned peaks did not show matches with the theoretical mass of peptides predicted by peptide cutter software. Therefore they were not assigned.

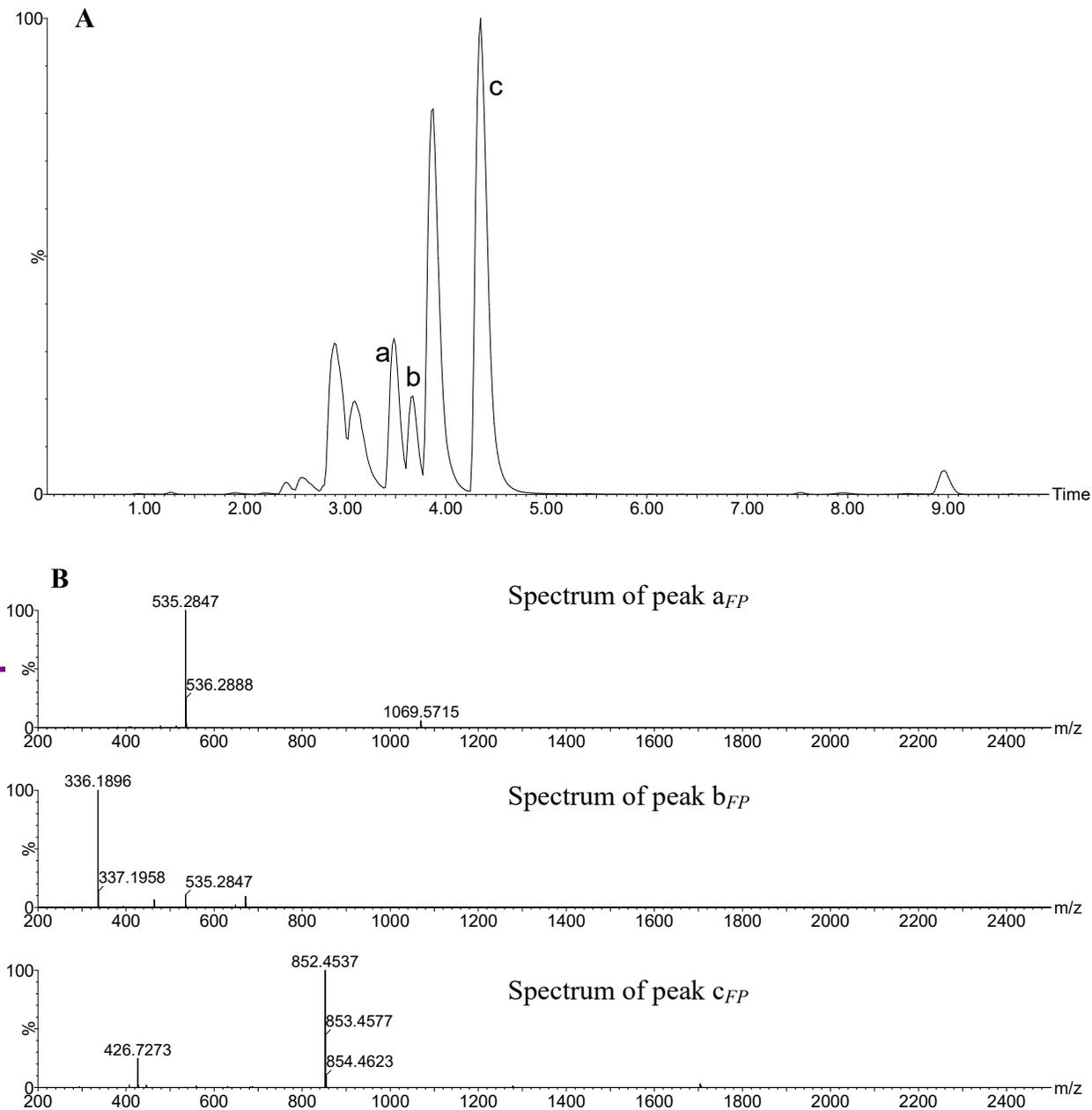


Figure 5-4 (A) Chromatogram obtained by separating the pepsin digestion peptides of FP on the C18 analytical column (B) Mass spectra corresponding to peaks a, b and c.

### 5-5-1 Calculating Mass accuracy in ppm

Mass accuracy is calculated in ppm as follows.

$$\text{Mass accuracy} = \frac{\text{Observed mass} - \text{Theoretical mass}}{\text{Theoretical mass}} \times 10^6 \text{ ppm}$$

. Ex: Peak “a” was assigned to the sequence GAIAGF (calculated  $m/z$  for +1 charge state is 535.2874 and observed  $m/z$  is 535.2847. Difference is -5 ppm). Peak “b” was assigned to the sequence GLF (calculated  $m/z$  for +1 charge state is 336.1918 and observed  $m/z$  is 336.1896. Difference is -6 ppm). Peak “c” was assigned to GLFGAIAGF sequence based on mass accuracy. (Calculated  $m/z$  for +1 charge state is 852.4614 and observed  $m/z$  is 852.4537. Difference is -9 ppm). Of these three sequences, GLFGAIAGF was selected as a “signature” peptide representative of the fusion peptide. And it was decided to investigate the occurrence of this sequence in FHA2 and HA2 pepsin digestion products.

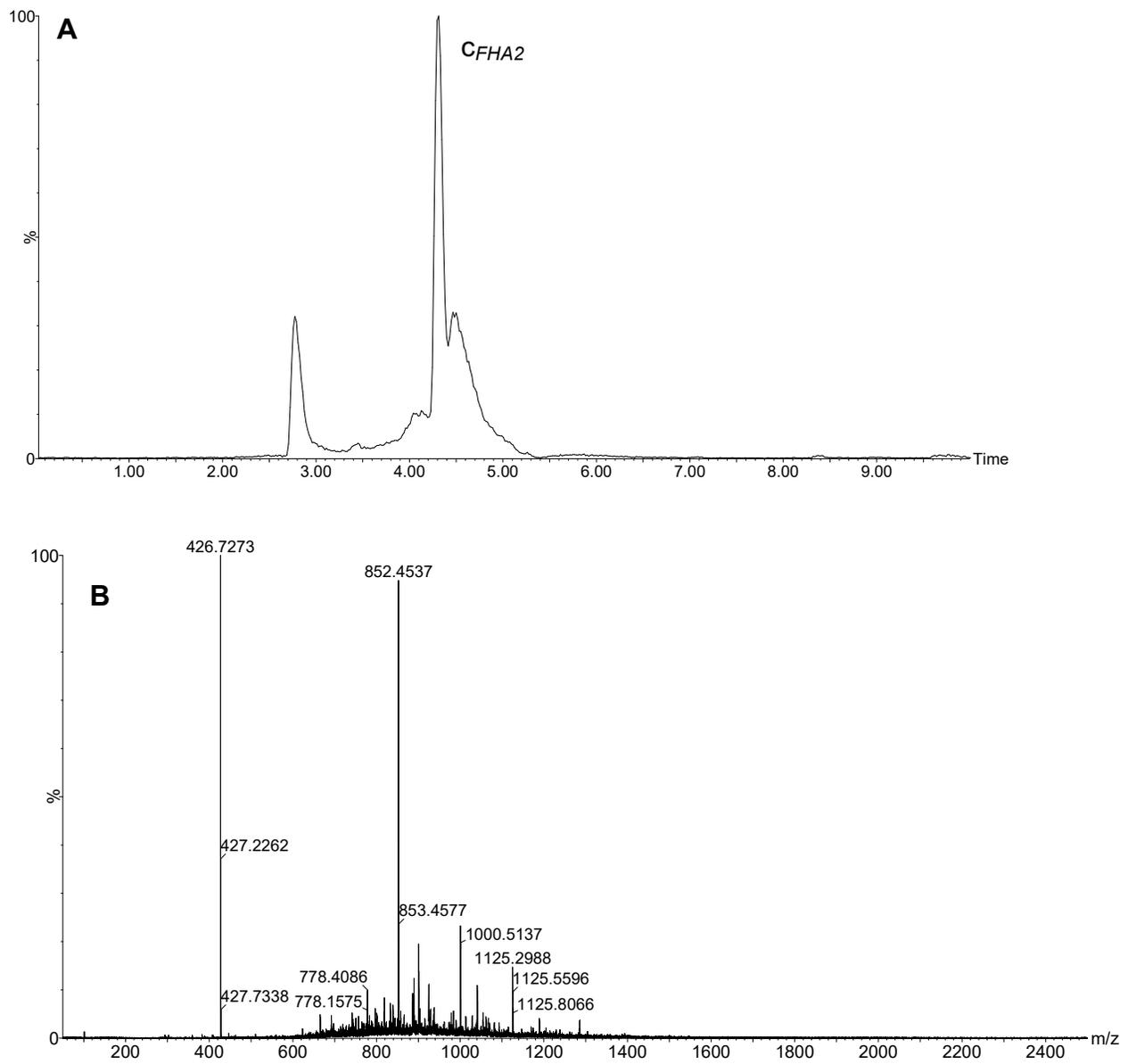


Figure 5-5 (A) Extracted ion chromatogram for  $m/z$  852.46 from the chromatographic profile of pepsin digestion products of FHA2. (B) Spectrum corresponding to the chromatographic peak “ $c_{FHA2}$ ” observed in A.  $m/z$  852.4637 represents  $[m+H]^+$  the signature peptide selected for FP in FHA2.

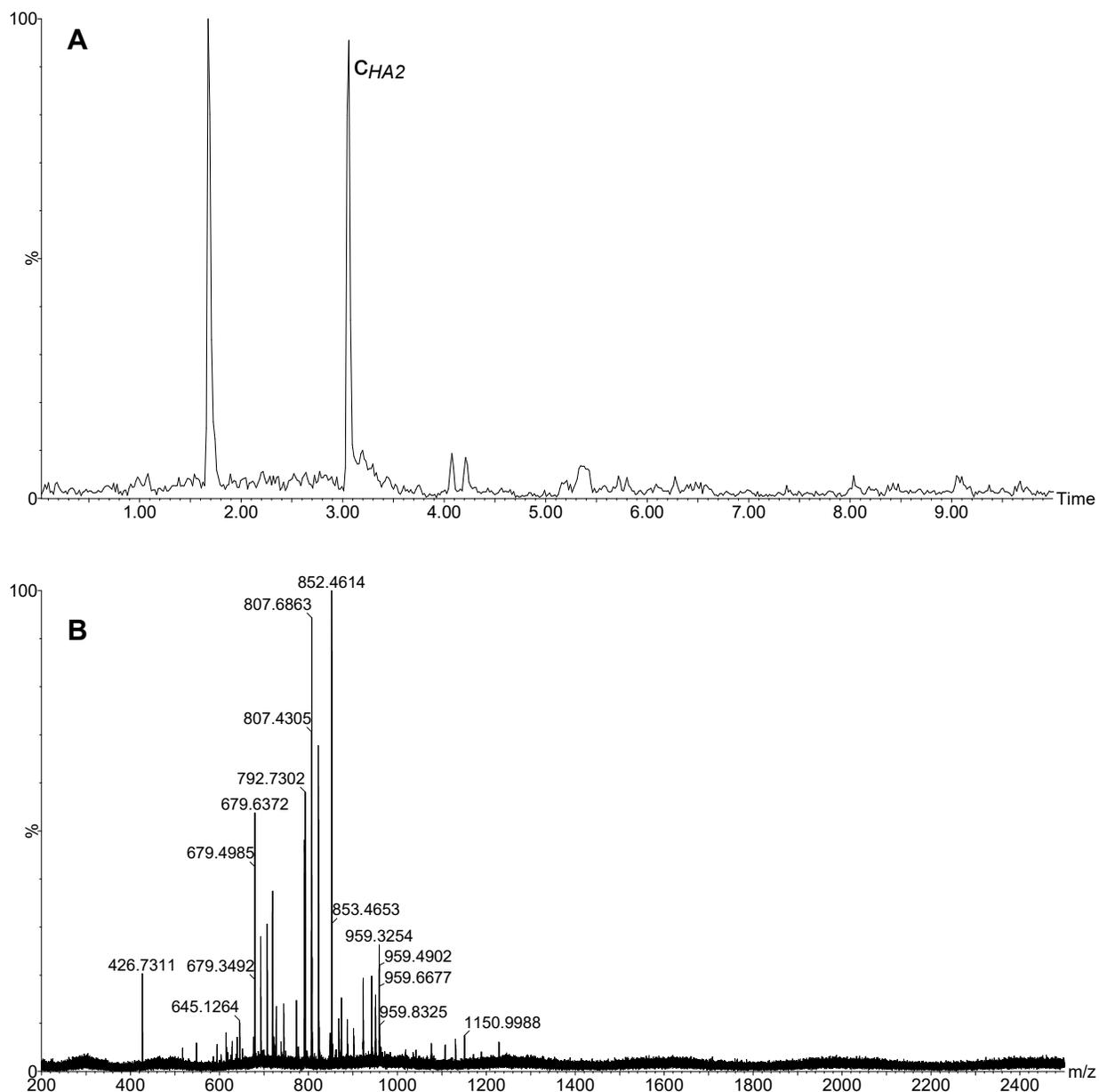


Figure 5-6 (A) Extracted ion chromatogram for  $m/z$  852.46 from the LC/MS analysis of pepsin digestion products of HA2. (B) Spectrum corresponding to the chromatographic peak “ $C_{HA2}$ ” observed in A. The peak at  $m/z$  852.4614 observed here represents the  $[M+H]^+$  of the signature peptide present in HA2. Since the signature peptide for FP was observed in FHA2 and HA2, the next step was to perform the deuteration and characterize the HD exchange rates of the different proteins under different time points.

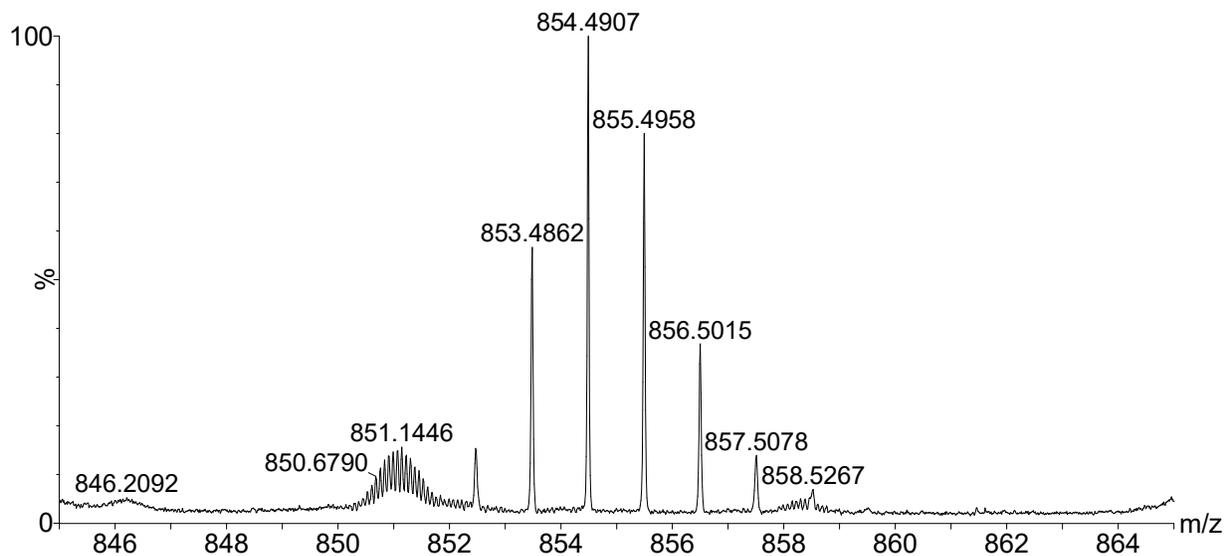


Figure 5-7: Deuterium incorporation into the signature peptide for FP in FHA2 after 0.5 min of exchange

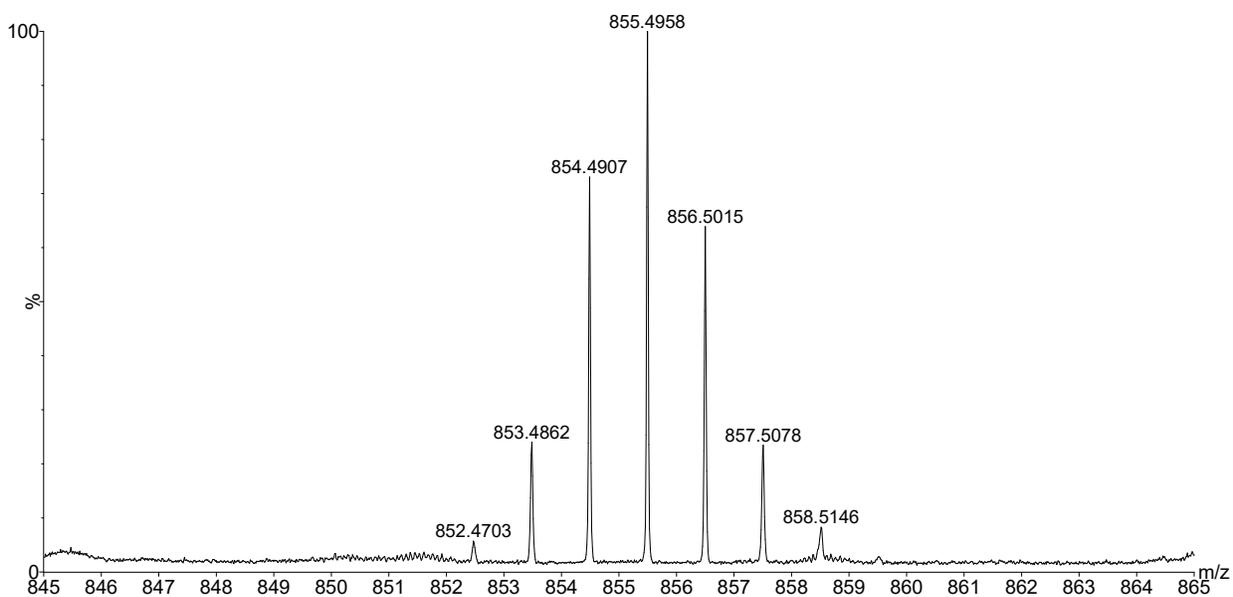


Figure 5-8: Deuterium incorporation into the signature peptide for FP in FHA2 after 1.5 min of exchange

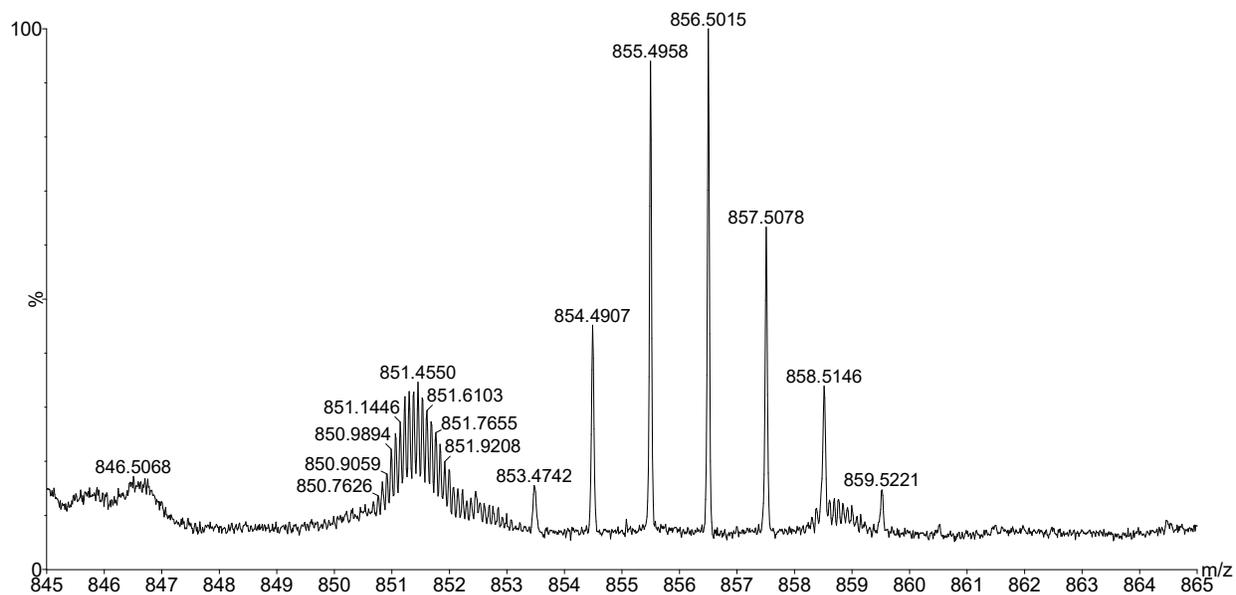


Figure 5-9: Deuterium incorporation into the signature peptide for FP in FHA2 after 3.0 min of exchange

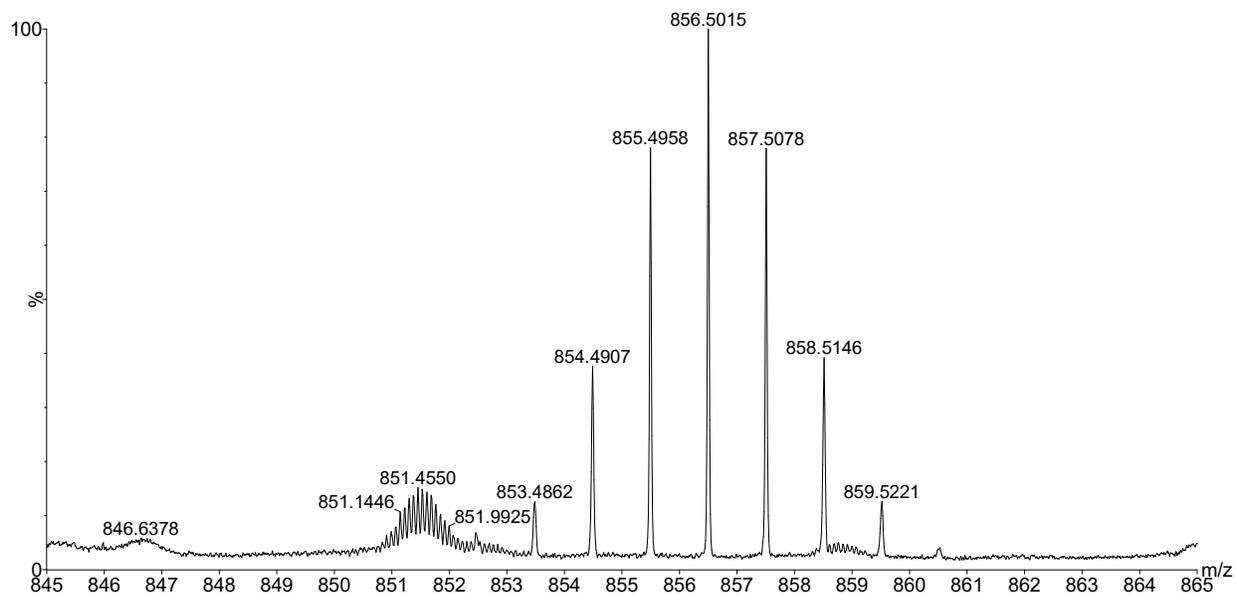


Figure 5-10 Deuterium incorporation into the signature peptide for FP in FHA2 after 10.0 min of exchange.

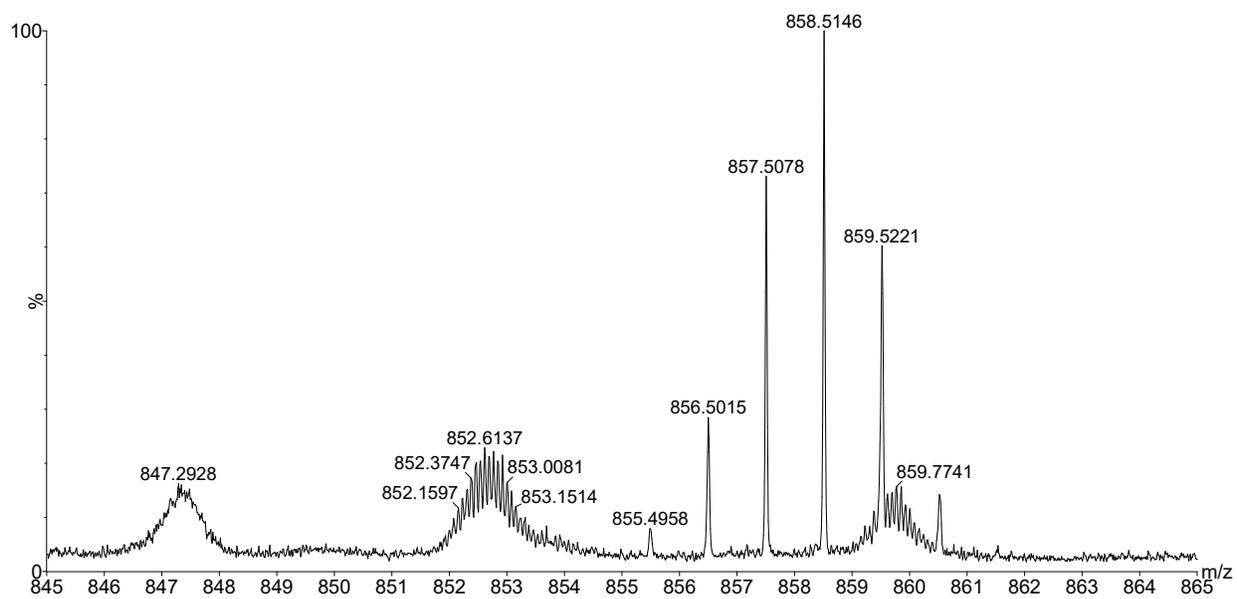


Figure 5-11: Deuterium incorporation into the signature peptide of FP in FHA2 after 24 hour of exchange.

## Results of HD exchange for the signature peptide in HA2

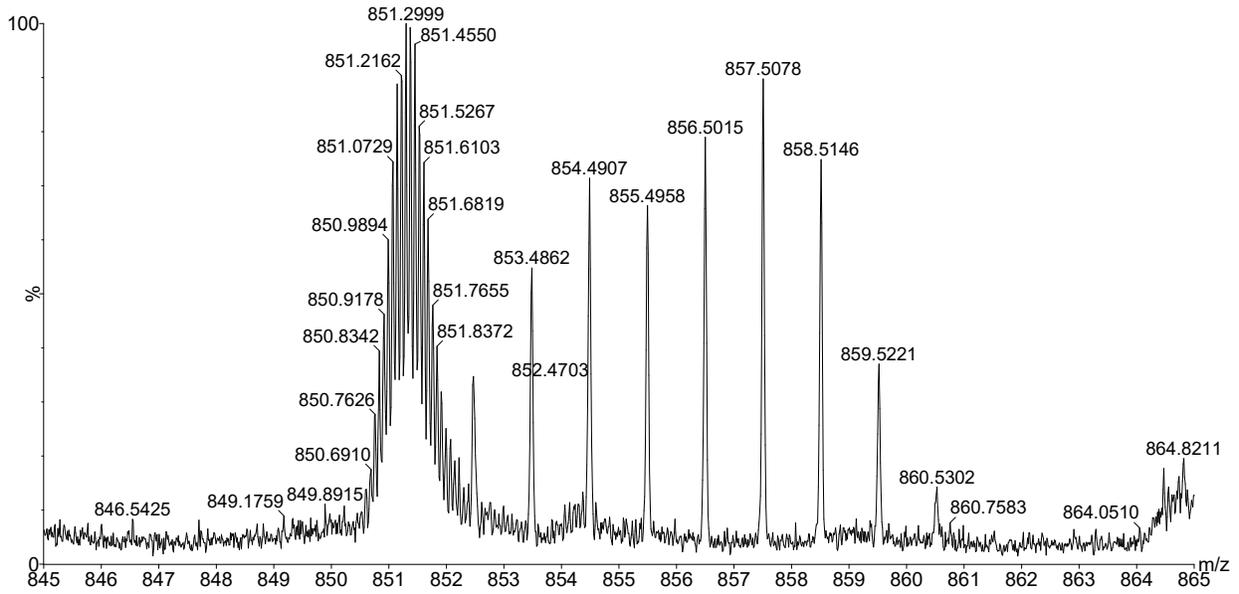


Figure 5-12: Deuterium incorporation into the signature peptide for FP in HA2 after 1.5 min of exchange.

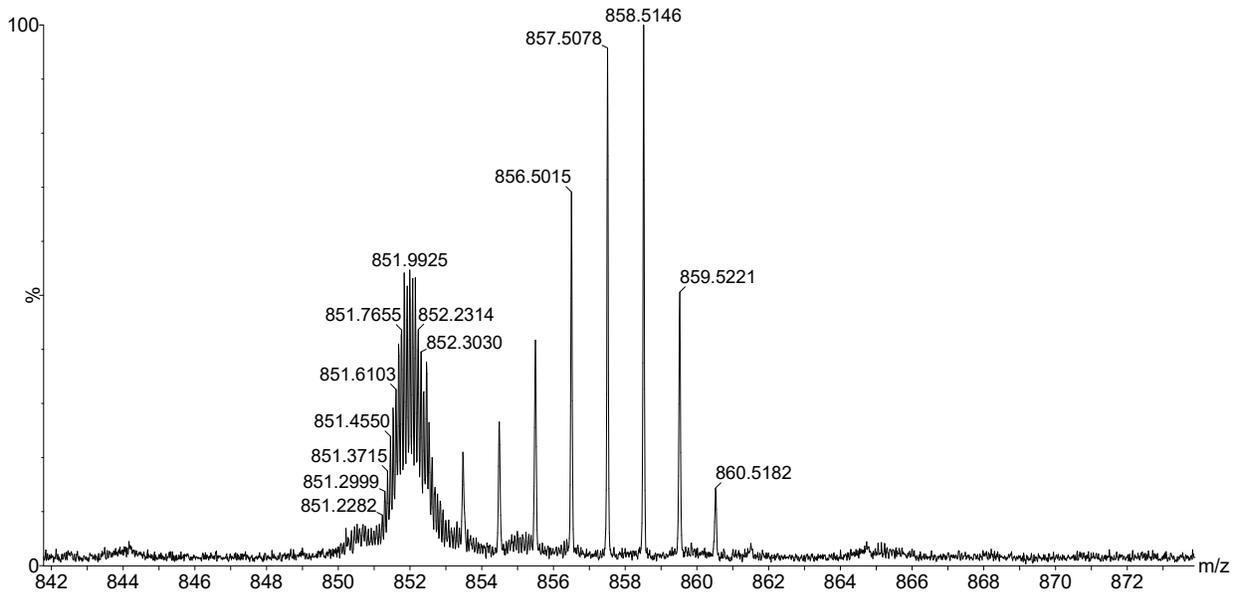


Figure 5-13: Deuterium incorporation into the signature peptide for FP in HA2 after 10 min of exchange.

Summary of the results of experiments completed is shown in table 5-3. Difference between the  $m/z$  of the most intense peak observed and the mass of the unexchanged peptide ( $m/z$  852.4614) represents the number of protons exchanged.

Table 5-3: Preliminary results of HD exchange experiment (ND – Not Determined). Due to instrument unavailability, 0.5 min, 3 min and 24 hour time points for HA2 could not be performed.

Time	FHA2 $m/z$ of most intense peak	# of protons exchanged in the most intense peak (FHA2)	HA2 $m/z$ of most intense peak	# of protons exchanged in the most intense peak (HA2)
0.5	854.4907	2	ND	ND
1.5	855.4958	3	857.5078	5
3.0	856.5015	4	ND	ND
10.0	856.5015	4	858.5146	6
24 hour	858.5146	6	ND	ND

## 5-7 Discussion

During the HD exchange experiment, maintaining the enzyme column, solvents, sample loop, C18 analytical column and connection tubing under ice cold conditions is important to minimize the back exchange of the protons.

Further experimentation is required to confirm the differences of exchange rates for the signature peptide for FP in the two proteins (FHA2 and HA2). Each time point has to be repeated at least 3 times to perform statistical analysis. However, these preliminary results indicate that the GLFGAIAGF sequence in FP in both FHA2 and HA2 are located in a solvent accessible

location in their trimeric form rather than at a hydrophobic core. Furthermore, these preliminary results also lead to the conclusion that the exchangeable protons in FP of HA2 exchange rapidly compared to those in FHA2 (since within 10 minutes 6 protons of HA2 exchanged, while FHA2 took 24 hours for exchange of 6 protons). This preliminary result supports that the trimeric hairpin structure formed does not possess a complex between TM and FP in HA2. This is because if a complex is formed between the highly hydrophobic TM and FP in HA2, it can be expected that the FP part of HA2 will show less propensity to exchange protons within a given time frame.

## **REFERENCES**

## REFERENCES

1. Epanand, R.F., et al., *The ectodomain of HA2 of influenza virus promotes rapid pH dependent membrane fusion*. Journal of Molecular Biology, 1999. **286**(2): p. 489-503.
2. Chen, J., J.J. Skehel, and D.C. Wiley, *N- and C-terminal residues combine in the fusion-pH influenza hemagglutinin HA2 subunit to form an N cap that terminates the triple-stranded coiled coil*. Proceedings of the National Academy of Sciences of the United States of America, 1999. **96**(16): p. 8967-8972.
3. Bullough, P.A., et al., *Structure of influenza hemagglutinin at the pH of membrane fusion*. Nature, 1994. **371** p. 37-43.
4. Zhang, Z.Q. and D.L. Smith, *Determination of amide hydrogen exchange by mass spectrometry –A- New tool for protein-structure elucidation*. Protein Science, 1993. **2**(4): p. 522-531.
5. White, J.M., et al., *Structure and mechanism of viral membrane fusion proteins: multiple variations on common theme*. Critical Reviews in Biochemistry and Molecular Biology 2008. **43** p. 189 – 219.
6. Wang, L.T., H. Pan, and D.L. Smith, *Hydrogen exchange-mass spectrometry - Optimization of digestion conditions*. Molecular & Cellular Proteomics, 2002. **1**(2): p. 132-138.

**Chapter 6 Sample preparation and preliminary studies for the  
crystallization of influenza full length HA2 and mini projects  
completed**

## 6-1 Introduction

The crystal structure of HA2 ectodomain (residue 1-175) along with HA1 (residue 4-328) was determined and showed the trimeric nature of HA2 (figure 1-3). This crystal structure does not provide information regarding TM and endodomain [1]. Another crystal structure consisting of soluble ectodomain (residues 23-185) of HA2 also supports the trimeric nature. In this structure each molecule of HA2 forms a hairpin structure where N and C terminal residues are present in the same end of the hairpin molecule figure (1-5 (B)) [3]. Since the research presented in this dissertation described purification of full length HA2 along with the identification of detergent conditions where it forms trimeric form, next step was to attempt to crystallize this protein. Such a crystal structure of the full length HA2 protein will provide information that was missing in the previously established crystal structures such as location of the FP, structural details of the FP, whether FP and TM forms a complex and structural details of the TM and endodomain.

In general, the protein samples used for the crystallization trials should be pure, homogeneous (single oligomeric state) and stable up to three days. Protein should be concentrated up to ~10 mg/ml and SEC profile (gel filtration) should show a monodisperse single oligomeric state. The stability of the protein is checked by keeping concentrated protein in room temperature and running a SEC to find if its oligomeric state changes with time upon incubation at room temperature. If it shows the same oligomeric state that it has shown before incubation even after a few days it is an indication that the protein is stable. Screening of different detergents for the solubilization and purification is helpful to find which detergent is amenable for the purification and solubilization of the protein, and is able to give physiologically relevant oligomeric state.

### 6-1-1 Stability test for the HA2

HA2 was purified from inclusion bodies using the method explained in chapter 3. HA2 samples were concentrated up to ~8 mg/ml using concentration tubes with 100 kDa molecular weight cut off (MWCO). This type of concentration tubes were used since the molecular weight of detergent + micelle is around 200 kDa based on the information obtained from the preliminary gel filtration results. Gel filtration was performed after the concentration step using mobile phase consisting of 0.17% DM, Tris 10mM and 150mM NaCl at pH 7.4. To check the stability, concentrated protein sample was left in the room temperature for 3 days. Gel filtration experiment was performed on the 3<sup>rd</sup> day (figure 6-1).

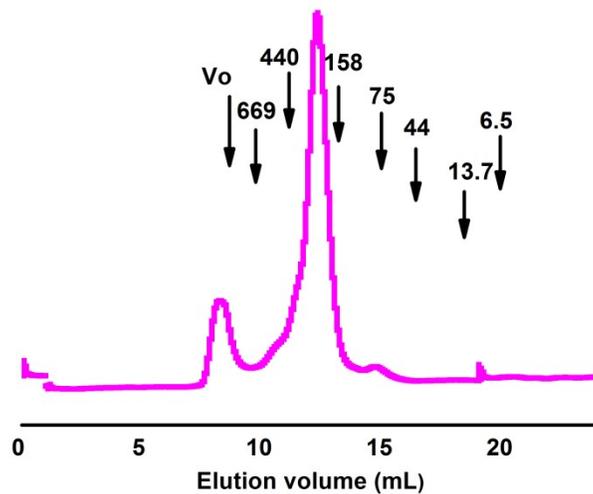


Figure 6-1 SEC  $A_{280}$  profile of concentrated (~ 8 mg/mL) full length HA2 protein after keeping 3 days in room temperature.

### ***6-1-2 Setting up the crystallographic trays***

For setting up the crystallography, HA2 was purified using 75 g of wet cells. The peak eluting at 12.5 ml retention volume corresponds to trimeric state of the protein and it was collected followed by re-concentrating to about 8mg/ml. Starting with about 75 g of wet cells, resulted ~80uL of the trimeric protein with concentration of 8mg/ml after performing the SEC. Then crystallization plates were set up in Dr. Jian Hu's lab at the Department of Biochemistry and Molecular Biology at Michigan State University. We have screened ~ 400 conditions using protein crystallization screening kits - Molecular dimensions Inc. memgold MD1-41, Rigaku crystallography screens wizard classic 1& 2, wizard classic 3&4 and Hampton research Inc. HR2-130. These trays were examined under a microscope in 1 week intervals since set up. After about 3 weeks of setting up the crystallographic trays, only in one condition rice weed shaped transparent crystals were observed. Crystals were observed in the Rigaku crystallography screen wizard 1 and 2 at C11 well. The conditions corresponding to this is 20% (v/v) 1, 4 butanediol, 100 mM sodium acetate, acetic acid at pH 4.5. Optimization of the conditions was performed as shown in the diagram below (table 6-2). In these optimization trials again the crystals were observed. The conditions where I saw the crystals are highlighted in the table below. Crystals were harvested while looking through the microscope and were sent to Argon National Lab. However none of the crystals resulted diffraction that enabled a crystal structure to be established.

Butanediol (v/v) \ pH	15%	17%	20%	23%	25%
pH 4.0					
pH 4.3					
pH 4.5		Crystals			
pH 4.7			Crystals		
pH 5.0		Crystals			Crystals

Table 6-1 Crystallography set up used for the optimization of full length HA2.

Highlighted slots show the spots that showed the crystals.

### ***6-1-3 Purification of cell palette with other detergents***

Since initial trials of crystallography failed, resulting cell palette after centrifugation was subjected to lysis using different lysis buffers containing different detergents (0.17% DM, 0.1% DDM, 0.1% DPC) as described in chapter 3. The reason for evaluating different lysis buffers was to identify the detergent that would enable crystals to form since the exact detergent that will result in the crystallization of full length HA2 is unknown. But none of the detergents (except sarkosyl) that I have tried were able to solubilize and purify the full length HA2 from the cell palette. Those detergents resulted poor yield and the purity (Figure 3-7). (Described briefly in chapter 3).

#### **6-1-4 Extracting HA2 from membrane fraction**

In addition to the inclusion body fraction, since HA2 is a membrane protein, there is a chance that some of the protein expressed by the *E.coli* can go in to the cell membrane fraction as well. In this case, the protein is in a more biologically relevant environment; probably the protein retains its correct native oligomeric state. Even though we got few crystals from the initial attempt, those crystals did not result diffraction that would allow a crystal structure to be established. Therefore in addition to purifying the protein from cell palette, we thought to extract HA2 from membrane fraction also.

Five grams of cells were lysed in PBS buffer (13.7 mM NaCl, 2.7 mM KCl, 10 mM Na<sub>2</sub>HPO<sub>4</sub>, 2 mM KH<sub>2</sub>PO<sub>4</sub>) with 4 rounds of 1 minute sonication cycles. There was a 1 minute interval between each sonication cycle. Each sonication cycle consisted of 80% amplitude with 0.8 s on and 0.2 s off. Cell lysate was centrifuged at 48000g for 20 min at 4°C. Next, the supernatant resulting after centrifugation was collected. Supernatant was subjected to ultra-centrifugation at 244720 g for 1 hour at 4 °C. Brown color transparent palette (membrane fraction) resulting from ultra-centrifugation was lysed in buffers A containing (50 mM sodium phosphate, 0.5% N-lauroylsarcosine (SRC) detergent, 300 mM NaCl, 10 mM imidazole at pH 8.0) using the same sonication setting. After lysis, centrifugation was performed again and supernatant was mixed with 1 mL of cobalt His select affinity resin. Supernatant + affinity resin was agitated using labquake shaker in room temperature for 1 hour. The mixture was reloaded to fritted column and washed 3 times with 0.75 mL fraction of wash buffer (same lysis buffer (A)). The retained protein was eluted 4 times with 0.5 mL fractions of elution buffer (A) containing 50 mM phosphate, 50 mM sodium phosphate, 0.5% N-lauroylsarcosine (SRC) detergent, 300 mM

NaCl, 250 mM imidazole at pH 8.0. Membrane fraction solubilization and purification was also performed using 0.17% DM, 0.04% LDAO, and 0.1% DDM.

### ***6-1-5 CD with protein purified from membrane fraction***

CD spectra were obtained after dialysis of the elutions from the affinity column directly in to the dialysis buffer containing 0.17% DM, 10 mM Tris at pH 7.4. Refolding was not required for HA2 extracted from membrane fraction. CD spectra were obtained using a CD instrument (Chirascan, Applied Photophysics, Surrey, United Kingdom) with a cuvette of 1 mm path length, a 260 nm-190 nm spectral window, wavelength points separated by 0.5 nm and 1.5 s signal average per point. For each sample, 3 scans were averaged. The concentrations of protein samples used for CD analysis were ~ 20  $\mu$ M. Reference spectra were obtained for all samples containing only the buffer components and the final spectrum is the difference of the protein sample spectrum after subtraction of the buffer spectrum.

## **6-2 Results and Discussion**

Protein extracted from membrane fraction was run in a SDS PAGE to check the yield of protein and its purity. According to the results, 0.5% sarkosyl and 0.17% DM (figure 6-3, 6-4, 6-5, 6-6) were able to extract the protein from membrane fraction. But the yield of the protein is about 5 times low when we do the solubilization and purification using 0.17% DM compared to that obtained using sarkosyl. The CD spectrum showed the  $\alpha$ -helical nature with the minima at 208 nm and 222 nm (figure 6-7) for the protein purified using membrane fraction. The protein obtained from the membrane fraction did not require the refolding before the CD experiment to

obtain spectra corresponding to a correctly folded protein. This indicates that protein obtained from membrane fraction is already present in the correctly folded form. However, the protein obtained from inclusion bodies did not result  $\alpha$ -helical curve before undergoing refolding process. This indicates that protein from inclusion bodies requires refolding prior to performing CD experiment. Before setting up crystallography trays with the membrane extracted HA2, gel filtration has to be performed to assess if the protein isolated from membrane fraction is in its trimeric oligomeric state. Then stability of the protein has to be checked after keeping concentrated protein sample (~8 mg/ml) for 3 days in room temperature, followed by the gel filtration on 3<sup>rd</sup> day. If protein displays the trimeric oligomerization even after 3 days then crystallography experiment can be set up.

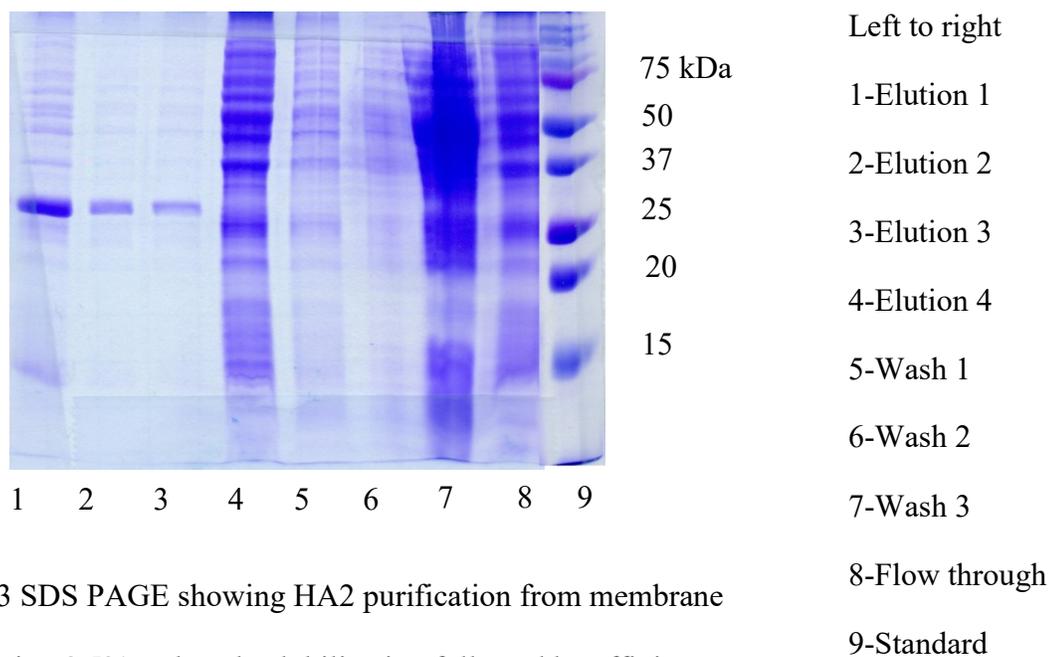


Figure 6-3 SDS PAGE showing HA2 purification from membrane fraction using 0.5% sarkosyl solubilization followed by affinity purification with 0.5% sarkosyl

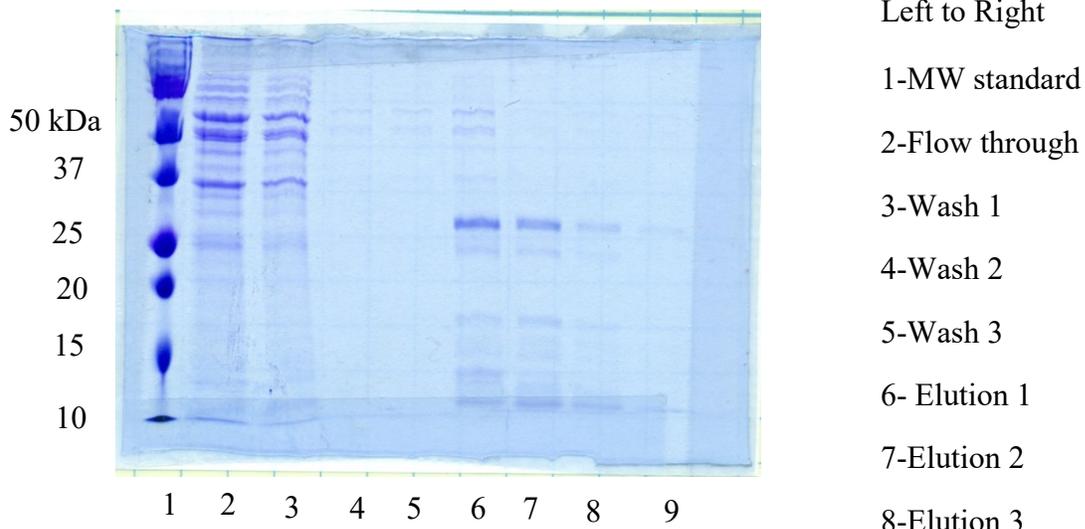


Figure 6-4 SDS PAGE showing HA2 purification from membrane fraction using 0.17% DM solubilization followed by affinity purification with 0.17% DM

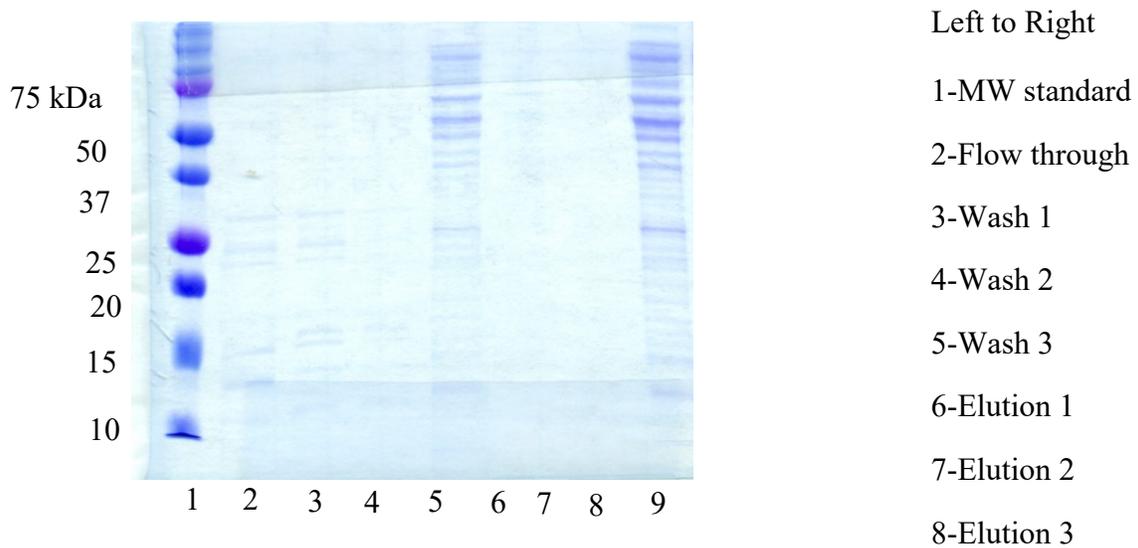


Figure 6-5 SDS PAGE showing HA2 purification from membrane fraction using 0.1% DDM solubilization followed by affinity purification with 0.1% DDM

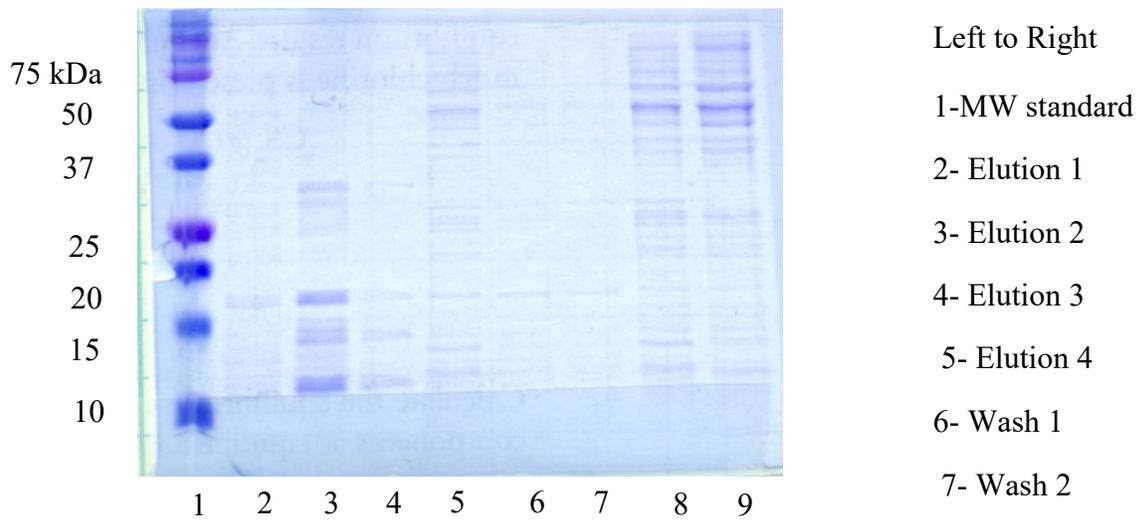


Figure 6-6 SDS PAGE showing HA2 purification from membrane fraction using 0.04% LDAO solubilization followed by affinity purification with 0.04% LDAO.

### 6-3 Future work

Lipidic cubic phase (LCP) and bicell crystallography are new techniques that used for the membrane protein crystallography. In bicell crystallization, purified membrane protein is incorporated in to mixture of lipid and detergent. In LCP purified membrane protein is incorporated to a lipid [4-7]. Membrane proteins have hydrophobic regions embedded in the lipid bilayer and both these methods provide more native like lipidic environment to the membrane proteins. Structural biologists were able to get high resolution structures of membrane protein using these two methods. Future graduate students in the Weliky lab can set up HA2 crystallography trials using these two methods.

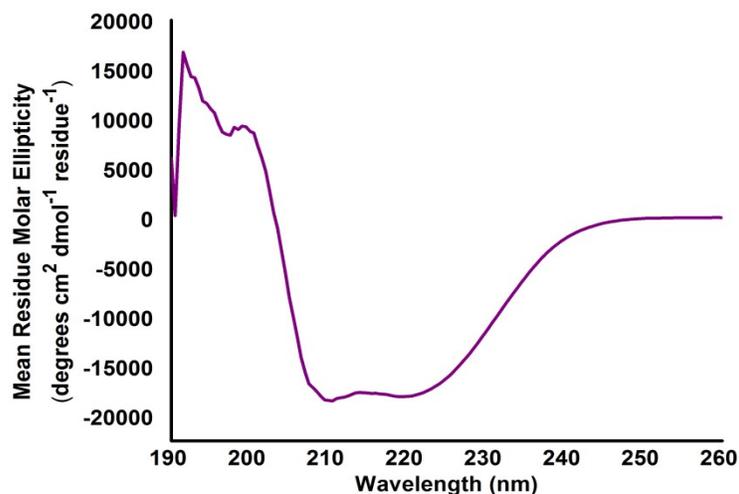


Figure 6-7 Circular dichroism of full length HA2 (~20 $\mu$ M) extracted from membrane fraction using 0.5% sarkosyl followed by dialysis in to buffer containing 10mM Tris-HCl, 0.17% DM at pH 7.4.

## 6-4 Mini projects completed

### 6-4-1 Subcloning of FPHM, HMTM and FPHMTM

As was the case for hemagglutinin constructs, it is interesting to study gp 41 construct containing different domains. Successful approach of the cloning that I have used for the hemagglutinin constructs lead us to prepare gp41 constructs. Genes of FPHM (FP+HM), HMTM (HM+TM), FPHMTM (FP+HM+TM) was subcloned in to pet24a+ vector. FPHM, HMTM, FPHMTM genes were ordered form Genscript. Genes were originally present in the pUC57-Kan vector and genes were inserted to pet24a+ (protein expression vector) using the cloning protocol discussed in chapter 2. Nde1 and Xho1 restriction sites were used for the subcloning process. G6 tag which is nonnative to the protein sequence was added before the tag at C-terminus.

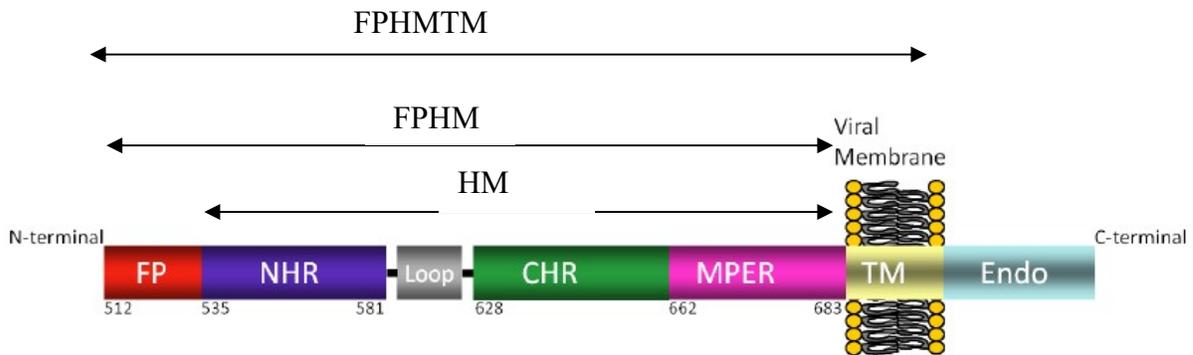


Figure 6-8 Schematic diagram of the HIV-gp41 where FP; fusion peptide, NHR; *N*-heptad repeat, CHR; *C*-heptad repeat, MPER; MPER, TM; Transmembrane domain.

Amino acid sequences given below color coded according to the different domains [2].

FPHM (FP+HM) protein sequence

AVGIGALFLGFLGAAGSTMGARSMTLTVQARQLLSGIVQQQN  
NLLRAIEAQQHLLQLTVWGIKQLQARILSGGRGGWMEWDREI  
NNYTSLIHSLIEESQNQQEKNEQELLELDKWASLWNWFNITNW  
LWYIKGGGGGG

FPHM DNA sequence

GCCGTGGGTATCGGTGCTCTGTTCTGGGTTTCCTGGGTGCTGCTGGTTCGACGATG  
GGTGCCCGCTCAATGACGCTGACGGTCCAAGCACGTCAGCTGCTGAGCGGCATTGT  
GCAGCAACAGAACAAATCTGCTGCGCGGATCGAAGCCCAACAGCATCTGCTGCAGC  
TGACCGTTTGGGGTATTAACAACACTGCAGGCTCGTATCCTGAGCGGCGGTGCGGGCG  
GTTGGATGGAATGGGATCGTGAAATTAACAATTATACGAGCCTGATTCACTCTCTGA  
TCGAAGAAAGTCAAACCAACAGGAGAAAAACGAACAGGAACTGCTGGAAGTGA  
CAAATGGGCCTCCCTGTGGAAGTGGTTAACATTACGAACTGGCTGTGGTACATCAA  
AGGCGGCGGTGGCGGTGGT

HMTM (HM+TM) protein sequence

CTLTVQARQLLSGIVQQQNLLRAIEAQQHLLQLTVWGIKQLQ  
ARILSGGRGGWMEWDREINNYTSLIHSLIEESQNQQEKNEQEL  
LELDKWASLWNWFNITNWLWYIKLFIMIVGGLVGLRIVFAVLS  
IVGGGGGG

HMTM sequence

TGTACCCTGACGGTCCAAGCTCGCCAAGCTGCTGAGTGGTATCGTTCAACAACAAAAC  
AATCTGCTGCGTGCTATCGAAGCCCAACAACATCTGCTGCAGCTGACCGTGTGGGGC  
ATTAAACAGCTGCAGGCCCGTATCCTGAGCGGCGGTTCGCGGGCGGTTGGATGGAATG  
GGATCGTGAAATTAACAATTATACGAGCCTGATTCACTCTCTGATCGAAGAAAGTCA  
GAATCAGCAAGAGAAAAACGAACAAGAACTGCTGGAACTGGACAAATGGGCCTCC  
CTGTGGAAGTGGTTTAATATTACCAACTGGCTGTGGTACATCAAAGTTCATTATG  
ATCGTTGGCGGTCTGGTTCGGTCTGCGTATTGTGTTTGCTGTCCTGAGTATTGTCGGTG  
GCGGTGGCGGCGGT

FPHMTM (FP+HM+TM) protein sequence

A V G I G A L F L G F L G A A G S T M G A R S M T L T V Q A R Q L L S G I V Q Q Q N  
N L L R A I E A Q Q H L L Q L T V W G I K Q L Q A R I L S G G R G G W M E W D R E I  
N N Y T S L I H S L I E E S Q N Q Q E K N E Q E L L E L D K W A S L W N W F N I T N W  
L W Y I K L F I M I V G G L V G L R I V F A V L S I V G G G G G G

FPHMTM DNA sequence

GCGGTTGGTATCGGTGCTCTGTTCCCTGGGTTTCCTGGGTGCGGCTGGCTCCACGATG  
GGTGCTCGCTCAATGACGCTGACGGTCAAGCACGTCAGCTGCTGAGCGGCATTGTG  
CAGCAACAGAACAATCTGCTGCGCGCGATCGAAGCCCAACAGCATCTGCTGCAGCT  
GACCGTTTGGGGTATTAACAAGTGCAGGCTCGTATCCTGAGCGGCGGTTCGCGGCG  
GTTGGATGGAATGGGATCGTGAAATTAACAATTATACGAGCCTGATTCACTCTCTGA  
TCGAAGAAAGTCAAATCAACAGGAGAAAAACGAACAGGAACTGCTGGAACTGGA  
CAAATGGGCCTCCCTGTGGAAGTGGTTTAATATTACCAACTGGCTGTGGTACATCAA  
ACTGTTTCATTATGATCGTTCGGCGGTCTGGTTGGTCTGCGTATCGTTTTTTCGGTCCCTG  
AGCATCGTTGGTGGTGGCGGTGGCGGT

Although the purification protocol for the HM protein has been developed in the lab [2], usage of guanidinium chloride for the purification is cumbersome since it has to be dialyzed in to 8 M urea solution before running a SDS PAGE. SDS and guanidine hydrochloride form a precipitate within the gel, making protein separation impossible. And dialysis has to be performed for a minimum of 2 days before running a SDS PAGE. To minimize this 2 days delaying time, alternative buffer system was checked for the purification of HM.

#### ***6-4-2 Purification of HM and FPHM***

HM cells were grown using the method described in literature [2]. The same method was used for growing cells for FPHM too. About 5g of wet cells were lysed with 40ml of PBS buffer (13.7 mM NaCl, 2.7 mM KCl, 10 mM Na<sub>2</sub>HPO<sub>4</sub>, 2 mM KH<sub>2</sub>PO<sub>4</sub>) with 4 rounds of 1 minute sonication cycles. There was a 1 minute interval between each sonication cycle. Each sonication cycle consisted of 80% amplitude with 0.8 s on and 0.2 s off. The cell pallet was separated from the soluble cell lysate by centrifugation (48000 g, 20 min, 4 °C). Resultant cell pellet was dissolved in PBS followed by sonication and centrifugation with same settings. Three rounds of sonication followed by centrifugation was performed to obtain pure cell palette devoid of soluble proteins. The resultant cell pellet III was dissolved in lysis buffer containing 8 M urea + 0.8 % sarkosyl (w/v N-lauroylsarcosine), 13.7 mM NaH<sub>2</sub>PO<sub>4</sub>.H<sub>2</sub>O, 0.3 M NaCl, and 10 mM imidazole at pH 8.0 and sonicated with the same settings. After the cell lysis, cell suspension was kept for one hour in the stirrer. Then centrifugation was performed using the same setting as before. No obvious cell pallet was seen after the centrifugation indicating that most of the recombinant protein was dissolved in 8 M urea+ 0.8% sarkosyl containing buffer. Resultant cell lysate was mixed with 1 mL of cobalt His select resin and agitated in labquake shaker in room temperature for 1 hour. The mixture was reloaded to fritted column and washed 3 times with 0.75mL

fractions of wash buffer (same as lysis buffer) followed by elution of the retained protein using 4 times with 0.5 mL fractions of elution buffer containing 8 M urea, 0.8% sarkosyl (w/v), 13.7 mM NaH<sub>2</sub>PO<sub>4</sub>·H<sub>2</sub>O, 0.3 M NaCl, 250 mM Imidazole, at pH 8.0).

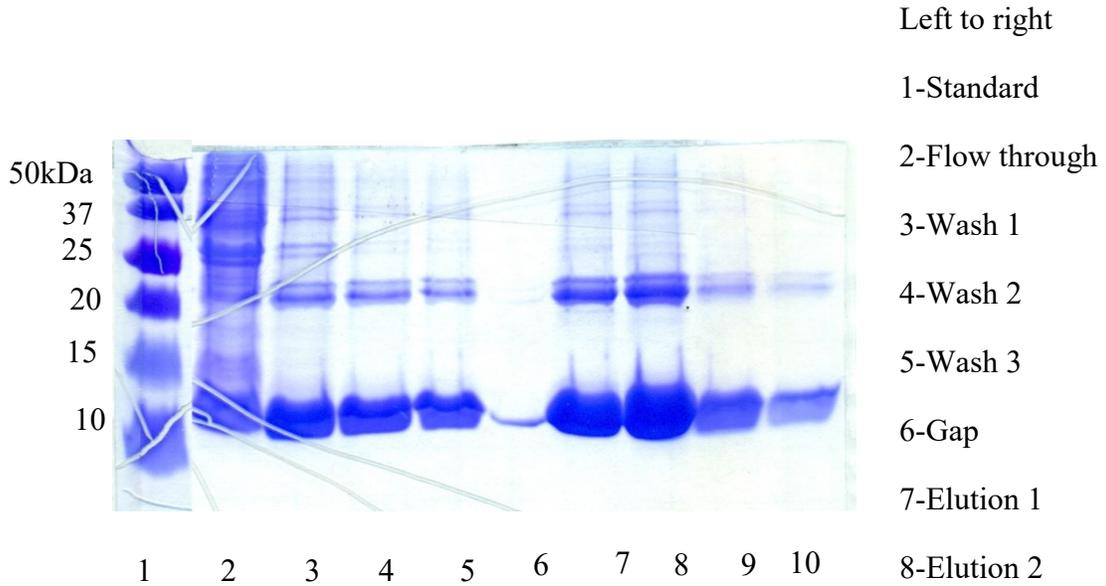


Figure 6-9 SDS PAGE showing HM purification from cell palette using 0.8% sarkosyl+ 8 M urea solubilization followed by affinity purification with 0.8% sarkosyl+8 M urea.

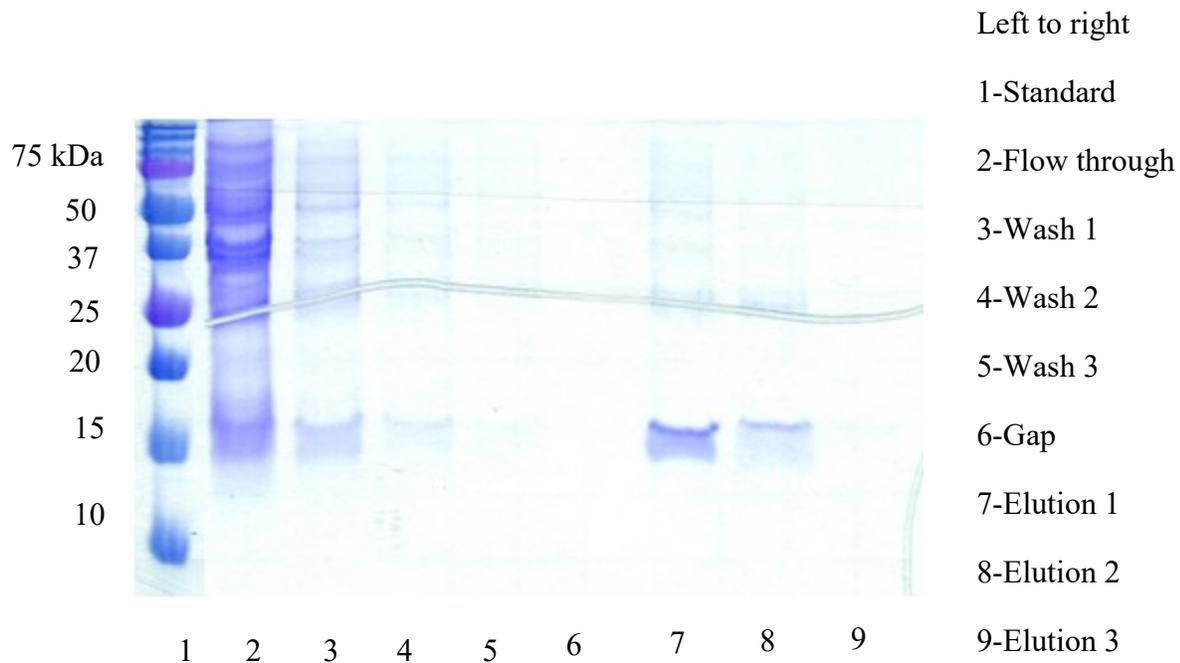


Figure 6-10 SDS PAGE showing FPHM purification from cell palette using 0.8% sarkosyl+urea solubilization followed by affinity purification with 0.8% sarkosyl+8 M urea.

## 6-5 Results and Discussion

The yield of the HM with this purification is 10 mg/ml and with FPHM it was ~ 0.3 mg/ml.

## 6-6 Future Work

For the HMTM and FPHMTM purification protocols have to be developed. For FPHM, HMTM and FPHMTM accurate mass experiments and proteomics analysis (as described under chapter 3) has to be performed to verify the presence of correct protein. After optimizing the

purification conditions for FPHM, HMTM, FPHMTM characterization of the protein can be performed using CD spectroscopy, SEC and lipid mixing assays.

HD exchange experiments also can be performed for FPHMTM and FPHM and will be able to disclose if FP region is exposed to the solvent or if it is located inside a hydrophobic core the same way explained in the chapter 5. As explained in chapter 5, it is hypothesized that in the final stage of fusion, gp41 hairpin molecule shows the formation of a complex between FP and TM (chapter 1 figure 1-4 (f)). However there is no experimental evidence to support this hypothesis for gp41 constructs also. We can design hydrogen-deuterium experiment to investigate this hypothesis as the way explained in the chapter 5.

## REFERENCES

## REFERENCES

1. Wilson, I.A., J.J. Skehel, and D.C. Wiley, *Structure of the haemagglutinin membrane glycoprotein of influenza virus at 3 Å resolution*. Nature 1981. **289** p. 366- 373.
2. 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.
3. Chen, J., J.J. Skehel, and D.C. Wiley, *N- and C-terminal residues combine in the fusion-pH influenza hemagglutinin HA(2) subunit to form an N cap that terminates the triple-stranded coiled coil*. Proceedings of the National Academy of Sciences of the United States of America, 1999. **96**(16): p. 8967-8972.
4. Ujwal, R. and J.U. Bowie, *Crystallizing membrane proteins using lipidic bicelles*. Methods, 2011. **55**(4): p. 337-341.
5. Ujwal, R. and J. Abramson, *High-throughput Crystallization of Membrane Proteins Using the Lipidic Bicelle Method*. Jove-Journal of Visualized Experiments, 2012(59).
6. Yin, X., et al., *GPCR Crystallization Using Lipidic Cubic Phase Technique*. Current Pharmaceutical Biotechnology, 2014. **15**(10): p. 971-979.
7. Cherezov, V., *Lipidic cubic phase technologies for membrane protein structural studies*. Current Opinion in Structural Biology, 2011. **21**(4): p. 559-566.