

STRUCTURE AND FUNCTION OF INFLUENZA VIRUS MEMBRANE FUSION PROTEIN

By

Ahinsa Ranaweera Kankanamge

A DISSERTATION

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

Chemistry - Doctor of Philosophy

2019

## ABSTRACT

### STRUCTURE AND FUNCTION OF INFLUENZA VIRUS MEMBRANE FUSION PROTEIN

By

Ahinsa Ranaweera Kankanamge

The entry of influenza virus into a host cell is mediated by the viral protein hemagglutinin (HA), which forms an initial complex that contains three HA1 subunits in association with a bundle containing the ectodomains of three HA2 subunits. Each ectodomain has a N-terminal ~25-residue fusion peptide (FP) and ~160-residue soluble ectodomain (SE). HA2 also has a ~25-residue transmembrane domain (TM) and a ~10-residue endodomain. Entry begins with HA1 binding to target cell sialic acids, followed by endocytosis of the virus and endosome maturation with pH reduction to <6. The HA1 subunits move away from the HA2 bundle, which then undergoes a large structural rearrangement to a final hyperthermostable state with a trimer-of-SE hairpins. The pH reduction is also correlated with HA2-mediated joining/fusion of the virus and endosome membranes, and consequent deposition of the viral nucleocapsid in the cytoplasm. The FP and TM are the only HA2 segments that are deeply inserted in the fused membrane. Residues 38-105 from the SE hairpins form an interior parallel helical bundle, while 154-176 are strands within the grooves of the bundle exterior. The FP forms a separate hairpin structure with antiparallel and close packing of two FP helices. My research focuses on better understanding the HA2 structure and fusion mechanism, which may aid the development of antiviral drugs and vaccine.

One main project in this dissertation compares WT-HA2 with G1E(FP) and I173E(SE strand) mutants. WT-HA2 induces vesicle fusion at pH 5.0, whereas fusion is greatly reduced for both mutants. Circular dichroism for HA2 and FHA2=FP+SE constructs show dramatic losses in stability for the mutants, including a melting temperature reduced by 40 °C for I173E-FHA2. This



evidences destabilization of SE hairpins via dissociation of strands from the helical bundle, which is also supported by larger monomer fractions for mutant vs. WT proteins. The G1E mutant may have disrupted FP hairpins, with consequent non-native FP binding to dissociated SE strands. It is commonly proposed that free energy released by the HA2 structural rearrangement catalyzes HA-mediated fusion. The present study supports an alternate mechanistic model in which fusion is preceded by FP insertion in the target membrane and formation of the final SE hairpin. Lower fusion by the mutants is due to loss of hairpin stability and consequent reduced membrane apposition of the virus and target membranes.

Another project describes the hydrogen-deuterium exchange-mass spectrometry (HDX-MS) on HA2. Fusion models often depict a FP/TM complex in the final HA2 state with membrane traversal by both domains, and a role for this complex in membrane pore expansion. The HDX-MS data are inconsistent with this complex, and show respective high FP and low TM aqueous exposures. The data support independent FP and TM with respective membrane interfacial and traversal locations. The data also support low aqueous exposure of the 22-38 segment, consistent with a structured and protected region and with retention of the 23-35 antiparallel  $\beta$  sheet observed in the HA2/HA1 complex. The  $\beta$  sheet is a semi-rigid connector between FP and SE that helps maintain close membrane apposition during fusion. The 22-69 and 150-191 regions exhibit much greater HDX for the I173E mutant vs. WT, which correlates with dissociation of C-terminal strands from interior N-helices of the trimer-of-hairpins. Similar trends are observed for G1E mutant vs. WT, as well as smaller HDX for G1E FP, which correlates with unfolding of the G1E FP helical hairpin and FP/SE binding. Both mutants exhibit highly-impaired fusion that may be due to increased distance between the initial apposed membranes and for G1E, FP binding to SE rather than target membrane.

Dedicated to mom, dad and hubby.  
Thank you for always believing in me.

## ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to my advisor Prof. David P. Weliky for giving me the opportunity to work in this wonderful research group, for his patience and immense knowledge. I highly appreciate his guidance and encouragement in overcoming numerous obstacles I have been facing through my research. Besides my advisor, I also want to thank the rest of my dissertation committee, Dr. Heedeok Hong, Prof. A. Daniel Jones, and Prof. James H. Geiger, for their comments and hard questions during comprehensive exams, which motivate me to explore my research from different perspectives.

I highly appreciate the support from past and present Weliky group members, with whom I shared happiness and sorrow. I would like to thank my loving sister Dr. Punsisi Ratnayake who guided me to join the Weliky group, and for her immense support throughout my graduate life. She has shown me, by her example, what a good scientist and person should be. I am grateful for her assistance with the basic biochemical and molecular biology techniques during my first semester in the lab. I would also like to thank Dr. Ujjayini Ghosh, Dr. Shuang Liang and Dr. Lihui Jia for useful discussion regarding research projects. I am thankful for the present group members Rob Wolfe and Yijin Zhang for being friendly and helpful, and also for the undergraduate students in our lab Robin, Lindsay, Grace and Abdullah for their immense support.

I am grateful to all the collaborators with whom I have had the pleasure to work during my PhD projects: Dr. Leonid Chernomordik and Dr. Leikina, Eugenia at NIH for their tremendous support for cell fusion studies; Dr. Shelagh Ferguson-Miller and Dr. Jian Liu at MSU Biochemistry department; and Dr. Kai Yu, Dr. Pan Pan and Dr. Michelle White at Creative Biostructure for their support for protein crystallography. I would also like to thank Dr. Tony Schillmiller and Sundari Chodavarapu for their assistance and useful discussions regarding the mass spectroscopy studies,

and Dr. Thomas Carter in MSU IT support for assistance with figures. I would like to extend my thank for Dr. Lisa Lapidus, and Dr. Babak Borhan for allowing to use the CD instrument and fluorimeter, and all the members of the Dr. Geiger and Dr. Hong research groups for being supportive throughout my graduate life.

My husband Mekulu's constant encouragement, support and love always help me to overcome the obstacles in my life. I would like to thank my family, especially my mom, dad and sisters, whose love and guidance are with me in whatever I pursue. I would also like to mention my friends at MSU: Evert, Benjamin, Gayanthi, Shaima and Mihiravi with whom I share happiness and sorrow. I am blessed to have you all in my life.

## TABLE OF CONTENTS

LIST OF TABLES.....	xi
LIST OF FIGURES .....	xii
KEY TO ABBREVIATIONS.....	xx
Chapter 1 Introduction to influenza virus fusion protein hemagglutinin II and viral-membrane fusion.....	1
1.1 Introduction.....	2
1.2 Influenza (flu) .....	3
1.3 Influenza virus .....	4
1.3.1 Influenza type A virus structure.....	4
1.3.2 Influenza type A viral life cycle.....	7
1.4 The influenza virus fusion protein .....	10
1.4.1 Structural studies of hemagglutinin .....	11
1.4.2 Low pH induced conformational changes of HA2 .....	14
1.5 Membrane fusion .....	16
1.5.1 Fusion of protein-free lipid bilayers .....	16
1.5.2 Protein-mediated fusion and proposed mechanisms of HA-mediated fusion.....	20
1.5.2.1 Model 1 .....	22
1.5.2.2 Model 2 .....	27
1.6 The membrane interacting domains of HA2.....	29
1.6.1 Influenza fusion peptide (FP).....	29
1.6.2 Influenza hemagglutinin transmembrane domain (TM) .....	32
1.7 Towards a universal flu vaccine .....	33
REFERENCES .....	37
Chapter 2 Methodology .....	47
2.1 Introduction.....	48
2.2 Polymerase chain reaction (PCR) .....	48
2.3 Protein Expression .....	52
2.4 Solubilization and purification of the expressed protein .....	54
2.4.1 Protein solubilization .....	54
2.4.2 Affinity chromatography .....	56
2.5 Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) .....	57
2.6 Cross-linking experiment.....	58
2.7 Size Exclusion Chromatography (SEC).....	59
2.8 Circular Dichroism (CD) spectroscopy .....	62
2.8.1 Protein secondary structure analysis.....	63
2.8.2 Protein thermostability analysis .....	64
2.9 Lipid mixing assay .....	66
2.10 Liquid Chromatography-Mass Spectrometry (LC-MS).....	68
2.11 Hydrogen-Deuterium Exchange Mass Spectrometry .....	69

REFERENCES .....	74
Chapter 3 Structure-Function Relationship of HA2 by Site-Specific Mutagenesis .....	77
3.1 Introduction.....	78
3.2 Materials and Methods.....	82
3.2.1 Materials .....	82
3.2.2 Protein constructs, expression, and purification .....	82
3.2.3 CD, SEC, Cross-linking, and vesicle fusion.....	87
3.3 Results.....	89
3.3.1 Protein preparation and folding .....	89
3.3.2 Size-exclusion chromatography and cross-linking.....	92
3.3.3 Circular dichroism .....	94
3.3.4 Vesicle fusion.....	100
3.4 Discussion .....	103
3.4.1 Models of protein structure and stability .....	103
3.4.2 Monomers and trimers .....	107
3.4.3 Correlations between WT-HA2 fusion and electrostatic energies.....	108
3.4.4 Reduced fusion with mutants.....	109
3.4.5 Common enveloped virus fusion model with significant functions for hairpin trimers and monomers.....	111
3.5 Summary .....	114
APPENDIX.....	115
REFERENCES .....	120
Chapter 4 Structure-Function Relationship of HA2 by Hydrogen-Deuterium Exchange Mass Spectrometry .....	128
4.1 Introduction.....	129
4.2 Materials and Methods.....	137
4.2.1 Materials .....	137
4.2.2 Protein constructs.....	138
4.2.3 Hydrogen-Deuterium exchange, online pepsin digestion and LC-MS analysis .....	138
4.3 Results.....	141
4.3.1 WT- HA2, FHA2, and SHA2_TM .....	142
4.3.2 WT-, G1E-, and I173E- HA2.....	146
4.4 Discussion .....	150
4.4.1 Comparison with previous HDX data.....	150
4.4.2 Structural Model of WT-HA2.....	152
4.4.3 G1E and I173E mutations destabilize the HA2 SE .....	156
4.4.4 Fusion implications of the HDX results.....	158
4.5 Summary .....	160
APPENDIX.....	162
REFERENCES .....	184
Chapter 5 HA2 structure by protein crystallography .....	191
5.1 Introduction.....	192
5.1.1 Membrane protein crystallography .....	193
5.1.2 Protein crystallization methods.....	194

5.1.3 The proposed mechanism of crystallization by LCP method .....	196
5.2 HA2 crystallization - part I .....	198
5.2.1 Materials .....	198
5.2.2 Mixing the protein with lipid .....	198
5.2.3 Mixing the mesophase with the precipitant solution .....	201
5.2.4 Incubation and tracking crystallogenesis .....	202
5.2.5 Results and Discussion .....	202
5.2.5.1 Initial screening.....	202
5.2.5.2 Condition optimization .....	204
5.2.5.3 Other modifications .....	207
5.3 HA2 crystallization - part II .....	209
5.3.1 Crystallization screening with LCP method .....	209
5.3.1.1 Experimental methods .....	209
5.3.1.2 Results.....	211
5.3.1.3 Conclusion .....	213
5.3.2 Crystallization screening with sitting-drop method .....	213
5.3.2.1 Experimental methods .....	213
5.3.2.2 Results.....	213
5.3.2.3 Conclusion .....	214
5.3.3 Future directions .....	214
REFERENCES .....	215
Chapter 6 Cell-cell fusion studies with WT, G1E and I173E HA2 .....	218
6.1 Introduction.....	219
6.2 Methods, Results and Discussion .....	220
6.2.1 Cell fusion assay .....	220
6.2.2 Comparative CD and SEC for FHA2 and HA2 in different detergents.....	221
REFERENCES .....	226
Chapter 7 Studies on HIV gp41 membrane fusion protein .....	228
7.1 Introduction .....	229
7.2 Structure-Function relationship of HIV gp41 by site-specific mutagenesis .....	233
7.2.1 Introduction.....	233
7.2.2 Site-specific mutation .....	234
7.2.3 Protein expression and purification .....	235
7.2.4 CD and SEC .....	238
7.2.4.1 WT- and V2E- FPHM.....	238
7.2.4.2 WT- and V2E- FPHMTM.....	242
7.2.5 Future directions .....	245
7.3 Native Chemical Ligation of FPHP .....	246
7.3.1 Introduction.....	246
7.3.2 Methods, Preliminary Results and Discussion.....	247
REFERENCES .....	251
Chapter 8 Summary and future directions .....	256
8.1 Summary .....	256

8.2 Future Directions .....	262
REFERENCES .....	263



## LIST OF TABLES

<b>Table 3.1</b> Analysis of CD spectra.....	97
<b>Table 3.2</b> Thermodynamic parameters from Van't Hoff analysis .....	99
<b>Table 4.1.</b> $\langle D_{\%} \rangle$ for peptides from FHA2, SHA2_TM, and HA2.....	145
<b>Table 4.2</b> $\langle D_{\%} \rangle$ for peptides from WT-, G1E-, and I173E- HA2 .....	148
<b>Table 4A.</b> Percent deuterium incorporation vs. HDX time for peptides from FHA2, SHA2_TM, and HA2. Each value is the average of triplicate measurements, and the standard deviation is given in parentheses.....	170
<b>Table 4B.</b> Percent deuterium incorporation vs. HDX time for peptides from digested WT-, G1E-, and I173E- HA2. Each value is the average of triplicate measurements, and the standard deviation is given in parentheses .....	177
<b>Table 4C.</b> Percent deuterium incorporation vs. HDX time for the 39-52 peptide of WT-, I173E-, and G1E- HA2. WT and I173E exhibit approximately unimodal distributions in the $m/z = 496-501$ range, and G1E exhibits a bimodal distribution for $t_{HDX} = 120$ min, with the modes approximately separable into $m/z = 496-498$ and $498-501$ ranges. Each value is the average of triplicate measurements, and the RMSD is given in parentheses. $\langle D_{\%} \rangle$ is the average of the five $D_{\%}$ , and the associated uncertainty in parentheses is calculated using the $D_{\%}$ RMSD's.....	183
<b>Table 5.1</b> The precipitant solution conditions that gave “crystal-like” particles in trial 1 .....	202
<b>Table 5.2</b> Condition that gave reproducible crystal-like particles .....	206
<b>Table 5.3</b> The screen kits (precipitant solution kits) .....	210
<b>Table 5.4</b> Results from the LCP method.....	211
<b>Table 5.5</b> The precipitant solution conditions that gave crystals .....	211

## LIST OF FIGURES

<b>Figure 1.1</b> Basic structure of the influenza virion.....	5
<b>Figure 1.2</b> Influenza viral replication cycle: (a) the virus enters the host cell in an endosome; (b) fusion of viral and endosomal membranes and subsequent release of the viral genetic material in to the cytoplasm; (c) replication of viral RNA in the host cell nucleus; (d) newly produced viral RNAs associate with nucleoproteins and exported to the cytoplasm; and (e) viral budding .....	7
<b>Figure 1.3</b> Schematic representation of HA (top) and HA2 subunit (bottom), with domain colored: fusion peptide (FP), pink; soluble ectodomain (SE), blue; transmembrane domain (TM), green; and endodomain (Endo), orange .....	11
<b>Figure 1.4</b> Crystal structures of the: (A) initial complex of HA1/HA2 at pH 7.5 (PDB 1RD8); and (B) HA2 soluble ectodomain at pH 5 (PDB 1QU1). HA1 and HA2 are colored in cyan and blue respectively .....	12
<b>Figure 1.5</b> The structure of a pre-fusion state (initial state) monomer at neutral pH (left) is compared to the corresponding region of a final state monomer at low pH (right). Labels A-H represents the consecutive regions of HA2 (residues 38-175).....	15
<b>Figure 1.6</b> (A) Proposed steps in the protein-free lipid bilayer fusion: (a) lipid membranes at pre-fusion state; (b) a point like protrusion, which minimizes the repulsion associated with hydration force; (c) hemifusion stalk, in which the outer leaflets of the two membranes mix while inner leaflets remain separated. The optimized packing of the hydrocarbon chains (lipid tails) in the hemifusion stalk minimizes the energy associated with it; (d) hemifusion diaphragm formed as a result of the radial expansion of the stalk; and (e) fusion pore formation. It is hypothesized that the fusion pore can be formed either from the hemifusion stalk or from the hemifusion diaphragm. (B) The free energy reaction coordinate diagram. The blue transitions represent the energy barriers for the single step transition from pre-fusion to hemifusion state and from there to pore formation. The free-energy barriers between each intermediate steps (shown in colored arrows) in panel A are indicated in respectively colored curves. Comparison of the barrier heights are facilitated by aligning the absolute free energy values of all intermediate states at 0 kT .....	17
<b>Figure 1.7</b> Pictorial representations of how the lipid shape determines the membrane curvature: conical shaped lipids protrude towards the polar head and induce positive membrane curvature, e.g. lysophosphatidylcholine (LPC); inverted conical shaped lipids protrude towards lipid tails and induce negative curvature, e.g. phosphatidylethanolamine (PE); and cylindrical lipids induce an almost flat membrane, e.g. phosphatidylcholine (PC). The blue and brown triangles and cylinders represent the respective lipid shape stated in each panel.....	19
<b>Figure 1.8</b> A proposed mechanism of the HA-mediated membrane fusion [51]; (A) HA2 in the prefusion state (initial state). The protein traverses the viral membrane through the TM, and the FP (green) is sequestered near the viral membrane. HA1 is not shown for clarity. (B) Upon acidification, (HA1 moves away from HA2 - not shown), liberating FP from its sequestered position. The HA2 SE extends, and the FP insert into the target membrane forming an extended	

intermediate that bridges the two membranes (prehairpin intermediate). (C) Subsequently, the extended intermediate collapses moving the C-terminal region of the ectodomain (blue strands) next to the N-terminal helix (red), thus pulling the two membranes towards each other. (D) The outer leaflets of the two membranes merge leading to a hemifusion stalk. (E) The inner leaflets are merged forming a fusion pore as a result of the HA2 hairpin closure .....22

**Figure 1.9** An alternate fusion model; (A) Initial state HA2 SE (HA1 is not shown and only a HA2 monomer is shown for clarity). (B) FP binds to the target membrane during HA2 structural transformations, forming a prehairpin intermediate. (C) The HA2 hairpin formation brings the two interacting membranes in to a close proximity prior to fusion. (D) Subsequent fusion (hemifusion and pore formation) occurs when HA2 is in the final hairpin conformation.....28

**Figure 1.10** NMR structures of the FP in the absence of rest of the HA2 .....30

**Figure 1.11** The negative curvature in a bicelle induced by inverted conical shaped FP .....32

**Figure 1.12** A model representing the inhibition of low pH-induced conformational changes by cyclic peptides. The FP, HA2 and HA1 are respectively represented by yellow, grey and green colors.....36

**Figure 2.1** The basic steps of PCR .....49

**Figure 2.2** Typical bacterial growth curve .....53

**Figure 2.3** Initial steps of HA2 purification by inclusion bodies and membrane fraction.....55

**Figure 2.4** Schematic representation of the affinity purification (picture credit goes to Dr. Shuang Liang).....56

**Figure 2.5** (A) The structure of bis (sulfosuccinimidyl) suberate (BS<sub>3</sub>); (B) Schematic representation of the reaction of an ester group of BS<sub>3</sub> with a primary amine group of the protein. R represents one end of the cross-linker, and P represent the protein .....59

**Figure 2.6** Size-based fractionating of molecules in a gel filtration column .....60

**Figure 2.7** (A) Size exclusion chromatogram for the A<sub>280</sub> detection of MW standards loaded on Superdex-200 column and (B) the SEC calibration curve.....61

**Figure 2.8** Characteristics shapes of CD spectra for different protein secondary structures .....64

**Figure 2.9** A typical protein melting curve .....65

**Figure 2.10** The principle of fluorescence quenching assay .....67

**Figure 2.11** Labeled vesicles containing both fluorescent and quenching lipids are mixed with unlabeled vesicles which results in fusion associated dilution of the labelled lipids .....68

**Figure 2.12** Different types of hydrogen atoms in a protein.....70

<b>Figure 2.13</b> Mechanism of HDX.....	70
<b>Figure 2.14</b> The pH effect of amide HX .....	71
<b>Figure 2.15</b> Temperature effect for base catalyzed HX .....	72
<b>Figure 2.16</b> The classical HDX experiment.....	73
<b>Figure 3.1</b> Amino acid sequences of the FHA2 and HA2 constructs with domains colored: fusion peptide (FP), pink; soluble ectodomain (SE), blue; transmembrane domain (TM), green; and endodomain, orange. The non-native C-terminal regions in black include a H <sub>6</sub> tag for affinity chromatography. The underlined residues denote the sites of the G1E and I173E mutations. (B) Ribbon diagrams of the FP in the closed conformation (PDB ID 2KXA) and trimeric SE in the final hairpin conformation (PDB ID 1QU1).....	80
<b>Figure 3.2</b> (A) DNA sequences of HA2 and FHA2 constructs. Each line is 60 nucleotides. (B) primer sequences for the G1E and I173E site specific mutations .....	85
<b>Figure 3.3</b> SDS-PAGE of the purification eluents. The MW <sub>FHA2</sub> = 22.4 kDa and MW <sub>HA2</sub> = 26.7 kDa. The gel was made using 12% acrylamide .....	90
<b>Figure 3.4</b> SEC traces prior to cross-linking (left) and SDS-PAGE after cross-linking (right) and in (A) SRC and (B) DM detergents. SEC parameters include [loading protein] = 1 mg/mL, [protein running] $\approx$ 0.1 mg/mL, and A <sub>280</sub> detection. Arrows mark the elution volumes of the void (V <sub>0</sub> ) and MW standards in kDa. Cross-linking conditions include [protein] = 0.5 mg/mL, fifty-fold molar excess of bis(sulfosuccinimidyl) suberate cross-linking agent, and 1 hour time. The gels were made using 12% acrylamide.....	93
<b>Figure 3.5</b> CD spectra of the (A) FHA2 and (B) HA2 proteins at ambient temperature. Samples contained 15 $\mu$ M protein in 10 mM Tris buffer at pH 7.4 with 0.17% DM .....	94
<b>Figure 3.6</b> Comparison of ambient-temperature CD spectra of I173E-FHA2 without or with 150 mM NaCl. The buffers also contained 10 mM Tris-HCl at pH 7.4 and 0.17% DM. The larger $ \theta_{222} $ values for [I173E-FHA2] = 1 $\mu$ M vs. higher protein concentrations likely reflect error in the 1 $\mu$ M concentration .....	95
<b>Figure 3.7</b> Temperature-series CD spectra. Protein and buffer conditions are the same as Fig. 3.5.....	96
<b>Figure 3.8</b> Plots of $\theta_{222}$ values vs. temperature .....	97
<b>Figure 3.9</b> Van't Hoff plots of the unfolding $\ln K_{eq}$ vs. $1/T$ of the (A) G1E-FHA2 and (B) I173E-FHA2 proteins based on $\theta_{222}$ data for the temperature range around $T_m$ . Best-fit parameters are given with uncertainties in parentheses .....	99
<b>Figure 3.10</b> HA2-induced vesicle fusion with protein:lipid = 1:300. (A) Time-courses of POPC vesicle fusion at pH 5.0. (B) Percent vesicle fusion at 600 s after addition of protein for different	

lipid compositions and pH's. Each bar represents the average of three replicates. There is typically  $\pm 1\%$  variation in percent vesicle fusion among the replicates.....102

**Figure 3.11** Structural model for a monomer unit of the SE of mutant HA2's. There is partial dissociation of the C-terminal strand and partial unfolding of the N-terminal helix.....105

**Figure 3A** SEC traces prior to cross-linking (left) of replicate samples and SDS-PAGE after cross-linking (right) of replicate samples and in (A) SRC and (B) DM detergents (compare to Fig. 3.4) .....116

**Figure 3B** CD spectra of replicates samples (compare to Fig. 3.5) .....117

**Figure 3C** van't Hoff plots of the unfolding  $\ln K_{eq}$  vs.  $1/T$  based on  $\theta_{222}$  data for the temperature range around  $T_m$ . Best-fit parameters are given with uncertainties in parentheses.....118

**Figure 3D** Extents of HA2-induced vesicle fusion using different preparations of proteins and vesicles (compare to Fig. 3.10) .....119

**Figure 4.1** (A) Schematic diagrams of full-length HA2 and truncated constructs with domains colored: fusion peptide (FP), pink; soluble ectodomain (SE), blue; transmembrane domain (TM), green; and endodomain (Endo), orange. (B) Ribbon diagrams of the SE in the final trimer-of-hairpins structure (PDB 1QU1) and the FP with closed helical hairpin structure (PDB 2KXA). A monomer subunit is highlighted for the SE, as well as the terminal residues of regular secondary structure elements, and residue 173 .....130

**Figure 4.2** Structural models for WT and mutant HA2 constructs that are based on residue 1-22 FP closed helical hairpin (PDB 2KXA), 36-175 SE hairpin (PDB 1QU1), 186-211 TM continuous  $\alpha$  helix that traverses the membrane, and 212-221 endodomain that is disordered. Residues 23-35 and 176-185 are represented as squiggly lines, because there isn't electron density for these residues in PDB 1QU1 for all or some of the monomers of the HA2<sub>23-185</sub> construct. The WT SE is the trimer-of-hairpins like in Fig. 4.1B, but a single monomer SE is displayed for clarity. Two WT-HA2 models are shown that either have a FP with helical hairpin structure at the membrane-interface, or FP with continuous  $\alpha$  helical structure in complex with TM and with membrane traversal. I173E-HA2 is shown with dissociated C-terminal SE strands and reduced helicity in the N-terminal SE. G1E-HA2 is shown as an equilibrium between the WT structure and the I173E structure with unfolded FP's that bind the C-terminal SE strands.....136

**Figure 4.3.** Protein sequences with peptides observed by mass spectrometry are underlined with domains colored: fusion peptide (FP), pink; soluble ectodomain (SE), blue; transmembrane domain (TM), green; and endodomain, orange. The non-native C-terminal regions are in black.....140

**Figure 4.4** Plots of percent deuterium incorporation ( $D\%$ ) vs.  $\log_{10}[t_{HDX}/\text{min}]$  for all analyzed peptides of FHA2, SHA2\_TM, and HA2. Each point is the average of triplicate measurements, with typical RMSD of 1%. Tables B1 list all  $D\%$  data.....144

**Figure 4.5** Plots of percent deuterium incorporation ( $D_{\%}$ ) vs.  $\log_{10}[\text{HDX time/min}]$  for all analyzed peptides of WT-, G1E-, and I173E- HA2. Each point is the average of triplicate measurements, with typical RMSD of 1%. Tables B2 list all  $D_{\%}$  data.....147

**Figure 4.6** Presentation of peptide  $\langle D_{\%} \rangle$  using color-coding and the structural models of Fig. 4.2. The colors for WT-HA2 are based on the  $\langle D_{\%} \rangle$  of peptides 1-9, 10-21, 22-38, 39-52, 53-69, 70-87, 86-91, 92-98, 99-110, 110-115, 120-138, 139-141, 142-150, 151-167, 168-171, 172-178, 179-187, 188-191, 192-197, 198-202, 200-205, and 206-217. The colors for other constructs are based on the  $\langle D_{\%} \rangle$  of subsets of these peptides with a few substitutions that include 3-9 for G1E-HA2 and 1-8 and 12-22 for I173E-HA2.....149

**Figure 4.7** Structural models for WT-, I173E-, and G1E- HA2 bound to a single membrane. WT exhibits trimer-of-hairpins structure for the SE as in Fig. 4.1B, but only one monomer is shown for clarity. The approximate positions of specific residues are indicated by the numbers. The WT model combines residue 1-22 closed FP helical hairpin (PDB 2KXA), 23-35 antiparallel  $\beta$  sheet of the HA2<sub>1-175</sub>/HA1 complex (PDB 2HMG), 36-178 hairpin SE structure (PDB 1QU1), 186-211 continuous  $\alpha$  helix for TM, and 212-221 unstructured endodomain. The 179-185 extended structure is found in one monomer in PDB 1QU1, but there isn't electron density for these residues in the other two monomers. The small positive membrane curvature accommodates the respective membrane interfacial and traversal locations of the FP and TM. The I173E model shows dissociation of the C-terminal strands from the trimer-of-hairpins and reduced helicity for the N-helices. The G1E model shows binding of the FP to the C-terminal strands, and the displayed structure is in equilibrium with the WT structure, like in Fig. 4.2 .....153

**Figure 4.8** Pictorial representations of membrane apposition prior to fusion for WT-, I173E-, and G1E-HA2. For WT, the TM traverses the virus membrane, the FP is interfacially-bound to the target membrane, and the membranes are held in close apposition by the semi-rigid SE trimer-of-hairpins and the 23-35 antiparallel  $\beta$  sheet. For I173E, dissociation of the C-terminal strands from the trimer-of-hairpins results in larger average distance between the membranes. G1E also exhibits more-distant membrane apposition because of partial FP/SE binding and concurrent weaker FP binding to the target membrane .....160

**Figure 4A** HDX-MS data for peptides of FHA2, SHA2\_TM, and HA2. Plots of mass spectral signal intensity vs.  $m/z$  are displayed for all  $t_{\text{HDX}}$ . The identity and charge of each peptide is given in the left column .....163

**Figure 4B** HDX-MS data for peptides of WT-, G1E-, and I173E- HA2. Plots of mass spectral signal intensity vs.  $m/z$  are displayed for all  $t_{\text{HDX}}$ . The identity and charge of each peptide is given in the left column .....166

**Figure 4C** Comparison between  $D_{\%}$  of the present study and an earlier study for  $t_{\text{HDX}} = 1$  min and  $t_{\text{HDX}} = 30$  min (*Structure* (2015) **23**, 665-676). Data are displayed for peptides respectively from FHA2  $\equiv$  HA2<sub>1-185</sub> and from the HA2<sub>1-175</sub>/HA1 complex after incubation at pH 4.9 for three hours.....169

**Figure 5.1** Schematic representation of the Bragg's law .....193

<b>Figure 5.2</b> Temperature-composition phase diagram of monoolein .....	196
<b>Figure 5.3</b> Cartoon representation of the proposed mechanism of crystallization of a membrane protein from lipidic cubic mesophase .....	197
<b>Figure 5.4</b> (A) The chemical structure of monoolein; and (B) A schematic representation of bicontinuous cubic phase of monoolein.....	199
<b>Figure 5.5</b> The lipid mixing device: (a) disassembled coupled syringe mixture; and (b,c) assembled coupled syringe mixture.....	200
<b>Figure 5.6</b> Adding the protein-lipid mesophase to the crystallization plates using the robot tip.....	201
<b>Figure 5.7</b> The crystallization screening image (left) and the diffraction pattern for the growth condition LCP-7-G4 (right) .....	212
<b>Figure 5.8</b> The diffraction patterns for the growth conditions: Memstart-E2 (left); and MemGoldII-E10 (right) .....	214
<b>Figure 6.1</b> Fusogenic activity of FHA2 $\equiv$ HA2* is greatly reduced by G1E and I173E mutations that are known to reduce the HA-mediated fusion. Fusion between HA0 cells and bound RBCs is induced by the medium acidification (pH $\sim$ 5) in the presence of FHA2, G1E-FHA2, or I173E-FHA2. Fusion is assayed as redistribution of lipid probe PKH26 (gray bars) or the small aqueous probe CF (black bars).....	219
<b>Figure 6.2</b> CD spectra of FHA2 and HA2 proteins at ambient temperature. Samples contained 16 $\mu$ M protein in (1) 0.17% DM, 10 mM Tris buffer at pH 7.4; (2) 0.17% DM, 150 mM NaCl, 10 mM Tris buffer at pH 7.4; and (3) 0.1 % (v/v) Triton X-100, 150 mM NaF, 10 mM NaH <sub>2</sub> PO <sub>4</sub> buffer at pH 7.0.....	222
<b>Figure 6.3</b> SEC of FHA2 and HA2 proteins: (A) in (1) 0.1% SRC, 150 mM NaCl, 10 mM Tris buffer at pH 7.4; (2) 0.17% DM, 150 mM NaCl, 10 mM Tris buffer at pH 7.4; and (3) 0.1 % (v/v) Triton X-100, 150 mM NaF, 10 mM NaH <sub>2</sub> PO <sub>4</sub> buffer at pH 7.0; and (B) in buffer (3) for replicate samples. SEC parameters include [loading protein] = 1 mg/mL, [protein running] $\approx$ 0.1 mg/mL, and A <sub>280</sub> detection. Arrows mark the elution volumes of the void (V <sub>0</sub> ) and MW standards in kDa.....	223
<b>Figure 6.4</b> Size exclusion chromatogram for the A <sub>280</sub> detection of MW standards (top), and calibration curves (bottom) using either 150 mM NaCl (pink) or NaF (blue) in the buffer solution.....	224
<b>Figure 7.1</b> Schematic diagram of an HIV virion. There are $\sim$ 15 gp41/gp120 trimers per virion. This is much lower compared to the $\sim$ 400 HA2/HA1 trimers per influenza virion .....	229
<b>Figure 7.2</b> The schematic diagram of full-length gp41 in the final state with domain colored: fusion peptide (FP), red; N-terminal helix, (N-helix), blue; loop, grey; C-terminal helix (C-helix), green; membrane-proximal external region (MPER), pink; transmembrane domain (TM), yellow; and endodomain (Endo), white. The residue numbering is according to gp160 sequence numbering for	

the HXB2 laboratory strain of HIV .....	230
<b>Figure 7.3</b> The ribbon diagrams of the: (A) initial complex of gp41 ectodomain trimer with gp120 (PDB 4NCO). N-linked glycans are shown as gray spheres; (B) close view of the hetero-dimer of gp120 and gp41 in the trimer in (A); and (C) gp41 ectodomain without gp120 in the final state (PDB 2X7R). The FP is absent in this structure .....	231
<b>Figure 7.4</b> Structural model of a gp41 construct containing the FP and TM (FPHMTM). A monomer is shown for the clarity .....	232
<b>Figure 7.5</b> The schematic diagrams and amino acid sequences of gp41 constructs studied. The residues 582-627 is replaced with a non-native SGGRGG loop. FPHM $\equiv$ gp41 <sub>512-581/SGGRGG/628-683</sub> and FPHMTM $\equiv$ gp41 <sub>512-581/SGGRGG/628-705</sub> . The non-native C-terminal regions are in black .....	233
<b>Figure 7.6</b> (A) The DNA sequence of WT-FPHMTM; (B) Primer sequences for the V2E mutation; and (C) Amino acid sequence after the mutation based on DNA sequence data .....	234
<b>Figure 7.7</b> SDS gels of (A) V2E-FPHM (MW = 16.5 kDa); and (B) V2E-FPHMTM (MW = 18.9 kDa) after the affinity purification with initial PBS washing; and (C) V2E-FPHMTM without PBS washing, which also gave an impurity band at ~25 kDa .....	237
<b>Figure 7.8</b> CD spectra of the FPHM proteins at ambient temperature for 7-day dialyzed sample. Sample contained 10 $\mu$ M protein in either 0.2% SDS/ 150 mM NaCl/ 10 mM Tris at pH 7.4 or 0.2% SDS/ 150 mM NaCl/ 20 mM CH <sub>3</sub> COONa at pH 4.0 .....	239
<b>Figure 7.9</b> SEC traces of V2E-FPHM at ambient temperature for 7-day dialyzed sample in 0.2% SDS/ 150 mM NaCl/ 20 mM CH <sub>3</sub> COONa at pH 4.0 (top), and 0.2% SDS/ 150 mM NaCl/ 10 mM Tris at pH 7.4 (bottom) .....	240
<b>Figure 7.10</b> CD spectra of the FPHM proteins at ambient temperature for 7-day dialyzed sample. Sample contained 10 $\mu$ M protein in either 0.2% SDS/ 10 mM Tris at pH 7.4 with and without 150 mM NaCl; or 0.2% SDS/ 20 mM CH <sub>3</sub> COONa at pH 4.0 with and without 150 mM NaCl.....	241
<b>Figure 7.11</b> CD spectra of the FPHM proteins at ambient temperature for 7-day dialyzed sample. Sample contained 10 $\mu$ M protein in 1% SDS/ 10 mM Tris at pH 7.4/ without (top) or with (bottom) 150 mM NaCl .....	242
<b>Figure 7.12</b> CD spectra of the FPHMTM proteins at ambient temperature for (A) 1-week and (B) 3-weeks dialyzed samples. Sample contained 10 $\mu$ M protein in 0.2% SDS/ 150 mM NaCl/ 10 mM Tris at pH 7.4 .....	243
<b>Figure 7.13</b> SEC traces of V2E-FPHMTM at ambient temperature for 7-day dialyzed sample in 0.2% SDS/ 150 mM NaCl/ 10 mM Tris at pH 7.4 (purple), or 0.2% SDS/ 150 mM NaCl/ 20 mM CH <sub>3</sub> COONa at pH 4.0 (pink) .....	244
<b>Figure 7.14</b> SEC traces of FPHMTM proteins at ambient temperature for 4-week dialyzed sample in 0.2% SDS/ 150 mM NaCl/ 10 mM Tris at pH 7.4 or 0.2% SDS/ 150 mM NaCl/ 20 mM CH <sub>3</sub> COONa at pH 4.0.....	245



<b>Figure 7.15</b> The mechanism of native chemical ligation .....	246
<b>Figure 7.16</b> (A) Schematic diagram, (B) amino acid sequence, and (C) DNA sequence of gp41 hairpin (HP) .....	247
<b>Figure 7.17</b> The SDS gel of HP (MW = 12.8 kDa) after the affinity purification.....	248
<b>Figure 7.18</b> (A) The SDS-PAGE; and (B) the MALDI-MS spectrum of the NCL product.....	250
<b>Figure 8.1</b> The structural model for full length HA2 in the fused membrane. HA2 exhibits trimer-of-hairpins structure for the SE, but only a monomer is shown for clarity. The model combines residue 1-22 closed FP helical hairpin (PDB 2KXA), 23-35 antiparallel $\beta$ sheet of the initial HA2/HA1 complex (PDB 2HMG), 36-178 hairpin SE structure (PDB 1QU1), 186-211 continuous $\alpha$ helix for TM, and 212-221 unstructured endodomain. The 179-185 extended structure is found in one monomer in PDB 1QU1. The small positive membrane curvature accommodates the FP and TM.....	260
<b>Figure 8.2</b> A mechanistic model for the HA2-mediated membrane fusion. The FP binds to the target membrane during the HA2 structural transformation from the initial state to the final hairpin state. The final trimer-of-hairpins SE structure holds the two membranes in close apposition prior to fusion. The clustering of hairpin trimers and therefore FP domains in the target membrane perturbs this membrane and lowers the activation barrier to achieve hemifusion and pore formation. For this model, the SE hairpin is a fusogenic state rather than a post-fusion state ....	261

## KEY TO ABBREVIATIONS

ACN	Acetonitrile
AIDS	Acquired immune deficiency syndrome
CD	Circular dichroism
CDC	Center for disease control and prevention
DM	Decylmaltoside
DOTAP	1,2-dioleoyl-3-trimethylammonium-propane (chloride salt)
FA	Formic acid
FP	Fusion peptide
HA	Hemagglutinin
HDX	Hydrogen deuterium exchange
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HIV	Human immunodeficiency virus
IPTG	Isopropyl $\beta$ -D-1-thiogalactopyranoside
LB	Lauria-Bertani
LC	Liquid chromatography
LCP	Lipidic cubic phase
LSP	Lipidic sponge phase
MES	2-(N-morpholino) ethanesulfonic acid
MS	Mass spectrometry
NA	Neuraminidase
N-NBD-DPPE	N-(7-nitro-2, 1, 3-benzoxadiazol-4-yl) (ammonium salt) dipalmitoyl-phosphatidylethanolamine
N-Rh-DPPE	N-(lissamine rhodamine B sulfonyl) (ammonium salt) dipalmitoyl-phosphatidylethanolamine

PAGE	Poly acrylamide gel electrophoresis
PBS	Phosphate-buffered saline
PCR	Polymerase chain reaction
PEG	polyethylene glycol
PHI	Pre hairpin intermediate
POPC	1-palmitoyl-2-oleoyl- <i>sn</i> -glycero-3-phosphocholine
POPG	1-palmitoyl-2-oleoyl- <i>sn</i> -glycero-3-[phosphor-rac-(1-glycerol)]
RBC	Red blood cell
SDS	Sodium dodecyl sulfate
SHB	Six helical bundle
SE	Soluble ectodomain
SEC	Size exclusion chromatography
SRC	<i>N</i> -lauroylsarcosine
TM	Transmembrane domain
WHO	World health organization
WT	Wild-type

## **Chapter 1**

# **Introduction to influenza virus fusion protein hemagglutinin II and viral-membrane fusion**

## 1.1 Introduction

Influenza, which is popular by the name flu, is an acute respiratory infection caused by influenza viruses. Scientific research related to influenza was started right after the 1918 pandemic that caused ~50 million deaths globally.<sup>1</sup> This viral infection can result in mild to severe illnesses including some cases of hospitalization or death. According to the WHO (World Health Organization) the epidemics are estimated to cause 3 - 5 million cases of severe illness and 250,000 - 500,000 deaths annually.<sup>2</sup> For the 2017–18 influenza season, the CDC (Center for disease control and prevention - USA) has estimated 49 million cases of illnesses, 960,000 hospitalizations and 79,000 deaths related to influenza.<sup>3</sup> According to the CDC estimates, the current influenza vaccination has prevented approximately 84,700 influenza-associated hospitalizations in 2016-17 influenza season. Given the estimated 960,000 hospitalizations, the 84,700 prevention implies that the vaccine lacks efficacy, and improvements to vaccine effectiveness is highly demanded.

Influenza virus is an enveloped virus with a lipid membrane surrounding the viral capsid. Influenza viruses are sorted into 4 types: A, B, C and D. Influenza type A viruses are circulating among avian species and infect several mammalian species including humans, whereas influenza type B infects only humans. The influenza A and B cause seasonal epidemics in almost every winter in USA. Influenza A viruses are divided into subtypes based on the antigenic properties of the two viral membrane proteins, Hemagglutinin (H) and Neuraminidase (N). There are 18 hemagglutinin subtypes (H1-H18) and 11 neuraminidase subtypes (N1-N11). The influenza subtypes A(H1N1) and A(H3N2) are currently circulating among humans. The emergence of new influenza A viruses due to rapid mutations can cause a substantial health risk despite having vaccines for influenza. Type C influenza viruses are known to cause mild respiratory infections to

humans and pigs and cause less health risks, whereas influenza type D are not known to infect human, and primarily affect cattle.<sup>2, 3</sup>

Fusion (joining) of viral and cellular membranes is a critical step in viral entry into the host cell. Understanding the structural details of the influenza virus fusion protein (possible drug target) and the mechanism by which the influenza virus membrane fuse with the host cell membrane is important to develop vaccines with increased effectiveness.

## **1.2 Influenza (flu)**

There were four severe influenza pandemics in the past century; Spanish flu in 1918, Asian flu in 1957, Hong Kong flu in 1968 and swine flu in 2009. Influenza is an infectious disease that spreads easily from person to person and it can affect people in any age group. However, the children under 5 years, the elderly above 65 years, pregnant women, individuals with specific chronic medical conditions such as HIV/AIDS, chronic heart or lung diseases and health care workers are more vulnerable. The seasonal flu is characterized by fever, headache, cough, sore throat, runny nose, muscle and joint pain and fatigue. Most people recover within a week without medical attention but sometimes it can cause severe illness or death. When an infected person coughs or sneezes, the droplets containing viruses can disperse into the air which can spread onto a person with closer proximity. The virus can also be spread through contaminated hands. By covering mouth and nose when coughing and sneezing and by washing hands regularly can control the transmission.<sup>2, 3</sup>

Vaccination is the most effective way to prevent the disease and the vaccine is more effective when the circulating viruses are matched well with the viruses in the vaccine. As the influenza virus is constantly evolving, the WHO Global Influenza Surveillance and Response

System (GISRS) monitors the circulating viruses among humans and update the composition of influenza vaccine twice a year. Anti-influenza drugs are available and there are two classes of such medicines; the influenza envelope protein neuraminidase inhibitors and M2 ion channel inhibitors (discussed in sections 1.3.1 and 1.7). However, the inhibitor-resistant viruses are emerging constantly and new therapeutics are required (some specific examples of resistant-viruses are discussed in section 1.7). Numerous small molecules have also been investigated that have the ability to inhibit influenza envelope protein hemagglutinin (discussed in section 1.7).

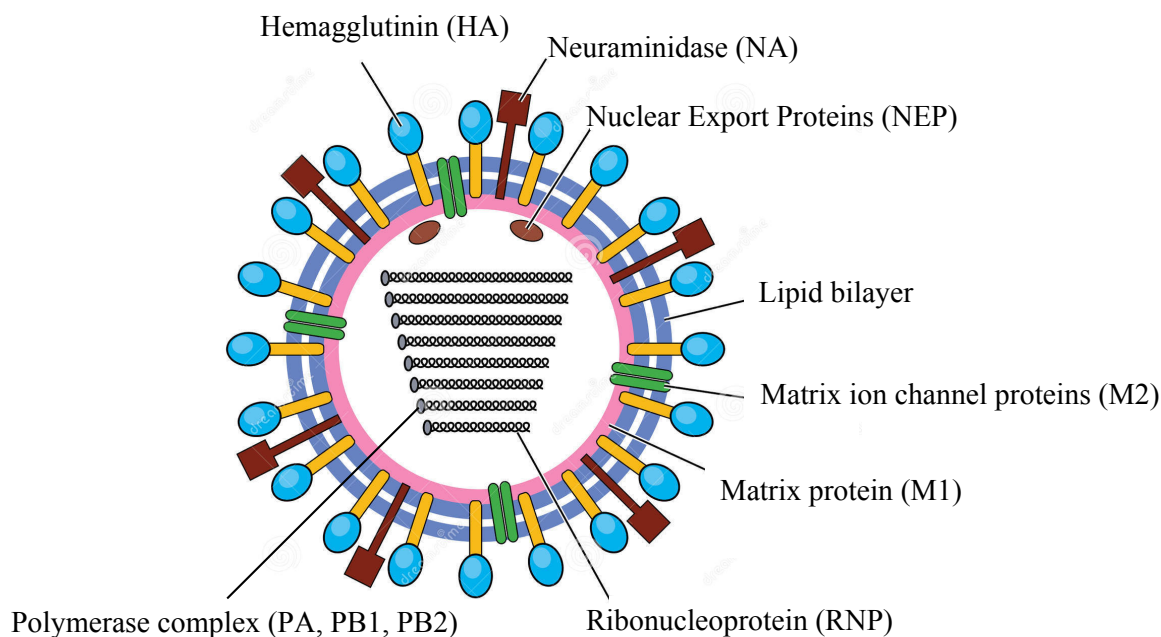
### **1.3 Influenza virus**

#### **1.3.1 Influenza type A virus structure**

Influenza virus is a RNA virus in the family *Orthomyxoviridae*. They are either spherical in shape with the diameter ranging from 80 to 120 nm or filamentous with the length  $\sim 300$  nm.<sup>4, 5</sup> Influenza is an enveloped virus with a lipid membrane surrounding the viral capsid. It gets its envelop through a process called budding, which is a method for releasing new viruses from the infected host cells. Enveloped viruses use this method to acquire host-derived membrane. The lipid composition of the viral membrane has been studied by the application of mass spectrometry on influenza virus infected MDCK cells (widely used as host cells for influenza virus).<sup>6</sup> According to this study, the influenza viral membrane consists of 29 mol% phospholipids, 19 mol% sphingolipids and 52 mol% cholesterol.

There are  $\sim 500$  spike-like projections on the viral external surface (Fig. 1.1). These spikes represent the envelope glycoproteins, hemagglutinin (HA) and neuraminidase (NA) which exist in a ratio of  $\sim 4:1$ . A smaller number of matrix ion channel proteins (M2) also traverse the membrane,

and there are ~20-60 molecules per virion.<sup>7</sup> These three proteins, HA, NA, and M2 ion channels are integral membrane proteins in the influenza virus.



**Figure 1.1** Basic structure of the influenza virion.

The most abundant envelope glycoprotein HA is synthesized in the infected host cell as a single polypeptide chain (HA0) with ~550 amino acid residues. Afterward, HA0 is proteolytically cleaved into two subunits HA1 and HA2 which remain covalently bound by a disulfide linkage.<sup>8</sup> Both these subunits are assembled as trimers, and there are ~400 trimers per virion.<sup>9</sup> HA1 is responsible for the binding of the virus to the host cell via sialic acid receptors present in the host cell glycoproteins or glycolipids. HA2 is responsible for the viral membrane fusion, thus referred to as the fusion protein. Depending on the antigenic properties of HA, there are two major



phylogenetic groups of influenza A viruses: group 1 (includes subtypes H1, H2, H5, H6, H8, H9, H11, H12, H13, H16, H17, H18); and group 2 (includes subtypes H3, H4, H7, H10, H14, H15).<sup>10</sup>

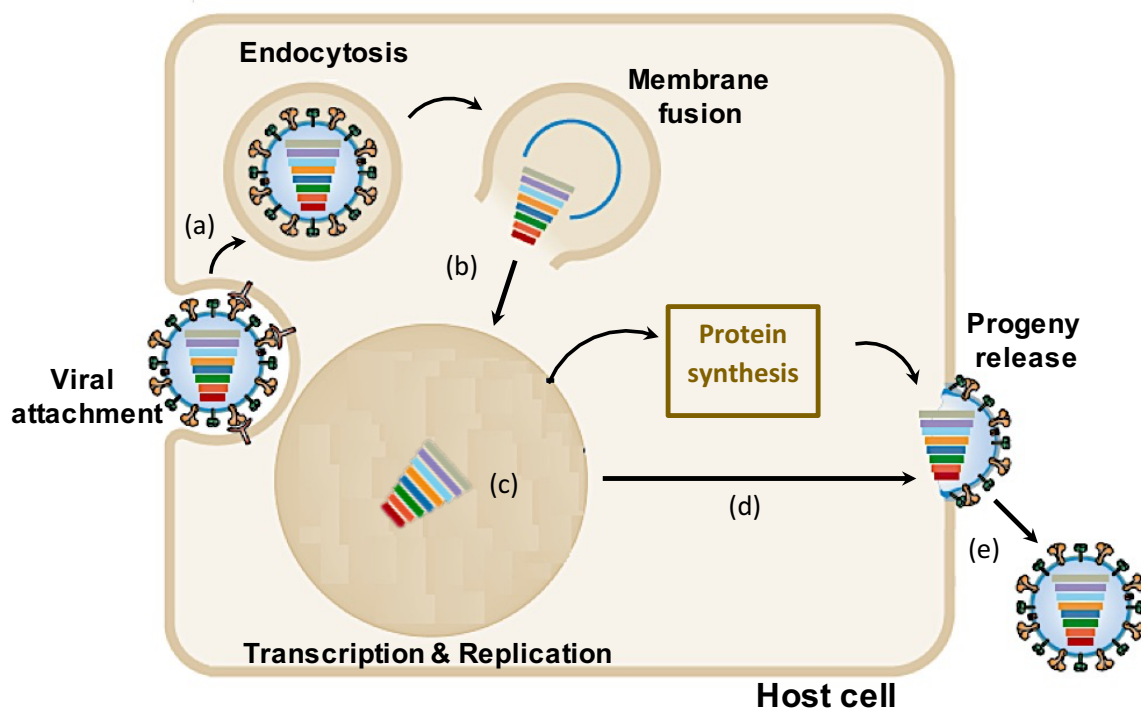
The second envelope glycoprotein NA is a tetramer.<sup>11</sup> NA act as an enzyme to cleave the terminal sialic acid receptors on the host cell membrane, thus releasing the newly formed viruses from the infected cell.<sup>12</sup> The antiviral drugs, oseltamivir (Tamiflu®) and zanamivir (Relenza®), are sialic acid analogues, and inhibit the enzymatic activity of NA. There are 11 NA subtypes based on antigenic properties (N1-N11) [CDC 2017]. The third integral membrane protein M2 is also a tetramer with the ion channel activity.<sup>13</sup> M2 involves in modulating the pH within the virus by proton channeling, thus weakening the interaction between ribonucleoproteins (RNP) and matrix protein (M1) (discussed in section 1.3.2). The antiviral drugs amantadine and rimantadine inhibit M2 proteins in type A influenza, but not effective against type B (discussed in section 1.7).<sup>14</sup>

The lipid envelope and its three integral membrane proteins overlay a matrix protein (M1) which encloses the viral core in some influenza virions. Based on Cryo-EM imaging data it has been found that ~90% of the influenza A virions have the M1 matrix underlying the viral envelope.<sup>5</sup> The viral genetic material is in the viral core, which consists of eight segments of negative-sense single-stranded RNA. The viral RNAs are associated with nucleoproteins (NP) and RNA polymerase complex (PA, PB1, PB2), forming the ribonucleoprotein (RNP) complex.<sup>15</sup> Each of these RNA is responsible for coding one or two proteins. In the viral core, there is another protein called nuclear export protein (NEP), which connects the RNP complex to the nuclear export machinery of the host cell. This aids the transfer of the viral RNAs from the infected host cell to progeny viruses.<sup>16</sup>

Influenza type B has a similar organization with four membrane proteins, whereas type C influenza has only one glycoprotein in the membrane: the hemagglutinin-esterase fusion (HEF), and it has functions of both HA and NA.

### 1.3.2 Influenza type A viral life cycle

There are five distinctive stages in the influenza viral replication cycle: (a) the viral entry into the host cell; (b) release of viral RNPs to the host cell nucleus; (c) replication of the viral genome; (d) export of viral RNPs from the nucleus; and (e) viral budding (Fig. 1.2).



**Figure 1.2** Influenza viral replication cycle<sup>17</sup>: (a) the virus enters the host cell in an endosome; (b) fusion of viral and endosomal membranes and subsequent release of the viral genetic material in

to the cytoplasm; (c) replication of viral RNA in the host cell nucleus; (d) newly produced viral RNAs associate with nucleoproteins and exported to the cytoplasm; and (e) viral budding.

The virus enters the host cell through the endocytic pathway. First, the HA1 subunit of the viral envelope protein HA binds to the terminal sialic acid receptors present on the surface glycoproteins or glycolipids on the host cell membrane. A receptor-mediated endocytosis takes place and the virus enters the host cell in an endosome (Fig. 1.2a). After the endocytosis, the host cell attempts to digest the virus particle by acidifying the endosomal interior. For this purpose, acidic lysosomes fuse with the endosome and as a result, the pH of the endosome drops to pH ~6. This is known as the early endosome and it migrates towards the host cell nucleus from the plasma membrane. When the endosome migrates towards the nucleus, the endosomal pH is further reduced to pH ~5.5 and these are known as late endosomes. The late endosome is formed after ~8 min of the binding of the virus.<sup>18</sup>

Other than the acidic pH, the late endosome is also characterized by: increased size (due to the fusion with lysosomes and with early endosomes themselves); and formation of luminal vesicles inside the endosome (due to the sorting of molecules that enter the endosome). Fusion of lysosomes with the endosome may also result in some changes in the lipid composition of the endosomal membrane.<sup>19, 20</sup> These lysosomes typically have a higher density compared to the endosomes and therefore, upon lysosome binding the endosomal density increases. The different densities of the early and late endosomes can be used to separate them for lipid composition analysis. Studies show widely varying lipid compositions of the chromatographic fractions of endosomal membranes,<sup>6, 19-21</sup> and the chromatographic fractions of the late endosomal subfraction

also show different lipid compositions. For example, some fractions are rich in phosphatidylcholine (PC), and other fractions are rich in lysobisphosphatidic acid (LBPA).<sup>19</sup>

Fusion of donor and acceptor vesicles containing LBPA shows pH dependency with higher fusion at pH 5.6, lower fusion at pH 6.0 and abolished fusion at pH 6.5 and 7.3.<sup>19</sup> The pH dependent fusogenic properties of LBPA may help in the fusion of endosomal membrane with the viral membrane (discussed below).

The late endosomes fuse with the lysosomes containing hydrolase enzymes, and the pH of these lysosomes is <5. The low pH of the late endosome leads to conformational changes in trimeric HA2 subunit of HA (discussed in section 1.4.2), followed by the joining (fusion) of the viral and the endosomal membranes. However, there are no experimental data about whether the virus first fuses with the late endosomal membrane or with a vesicle membrane inside the endosome.

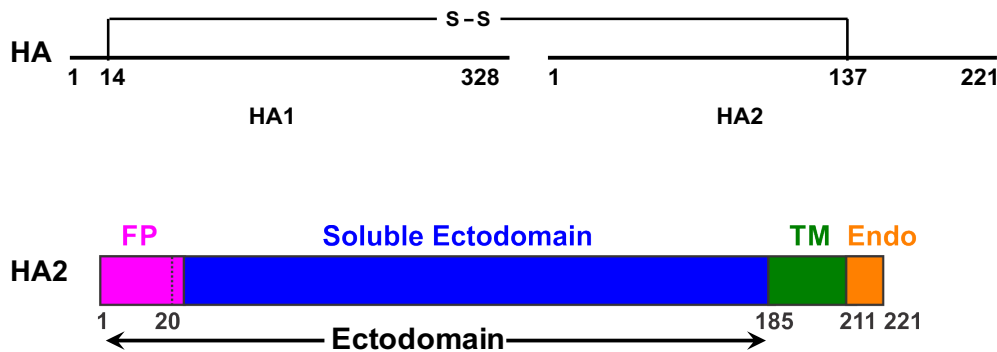
The acidification of the endosome is not only important in fusion but also it opens the M2 proton channel, thus acidifying the viral core. The viral core acidification is important in “uncoating” of the viral genome. Initially the viral RNPs are associated with M1 protein via electrostatic interactions and therefore the RNPs cannot enter to the host cell cytoplasm although the two membranes (viral and endosomal) are fused together. When the viral core pH is reduced, these electrostatic interactions are weakened, and the RNPs dissociate from M1 protein (uncoating), releasing the viral RNPs into the host cell cytoplasm.<sup>22</sup> The released RNPs must enter to the host cell nucleus as the viral transcription and replication occurs in the nucleus. All the viral proteins associated with RNP complex (NP, PA, PB1 and PB2) have nuclear localization signals that helps in binding the host cell nuclear import machinery,<sup>23</sup> thus enter to the nucleus (Fig. 1.2b).

The influenza viral genome consists of negative sense RNA that cannot be translated into proteins directly. Conversion of these RNA into positive sense RNA, which can act as mRNA, is governed by viral RNA polymerase enzyme. These mRNAs leave the host cell nucleus and the translation of the mRNA into viral proteins occurs in the endoplasmic reticulum (ER) in the host cell cytosol. The viral RNAs required for the progeny viruses are replicated in the host cell nucleus (Fig. 1.2c). The viral RNAs produced in the nucleus associate with nucleoprotein (NP) to form viral RNPs, which are then exported to the cytoplasm (Fig. 1.2d). The RNPs along with other viral proteins produced in the ER migrate towards the host cell membrane and new virus particles are formed. The host cell plasma membrane becomes the viral envelope, and this process is known as the viral budding (Fig. 1.2e). Finally, the release of progeny viruses from the host cell is mediated by NA, and the infection spreads to the other host cells.

#### **1.4 The influenza virus fusion protein**

The fusion of viral and cellular membranes with subsequent transfer of viral genome into the host cell is a characteristic process in the infection of enveloped viruses. The influenza virus HA is responsible for the viral binding to the host cell and the viral membrane fusion. HA is biosynthesized as an uncleaved precursor HA0, which is subsequently cleaved into two disulfide linked subunits HA1 with ~330 residues, and HA2 with ~220 residues. The cleavage of HA0 generates a hydrophobic N-terminus of HA2, termed the fusion peptide, which plays an important role in the fusion activity (discussed in section 1.6.1). The structure of cleaved HA (HA1+HA2) at neutral pH differs from HA0 only near the cleavage site. The disulfide linkage is in between the residue 14 of HA1 and the residue 137 of HA2.<sup>24</sup>

The influenza virus fusion is catalyzed by HA2, which contains the fusion peptide (FP), soluble ectodomain (SE), transmembrane domain (TM) and Endodomain (Endo) (Fig. 1.3). The FP and SE are outside the virus and commonly known as the ectodomain. The HA2 protein traverse the membrane through the single pass TM, and there is a ~10-residue endodomain inside the virus.

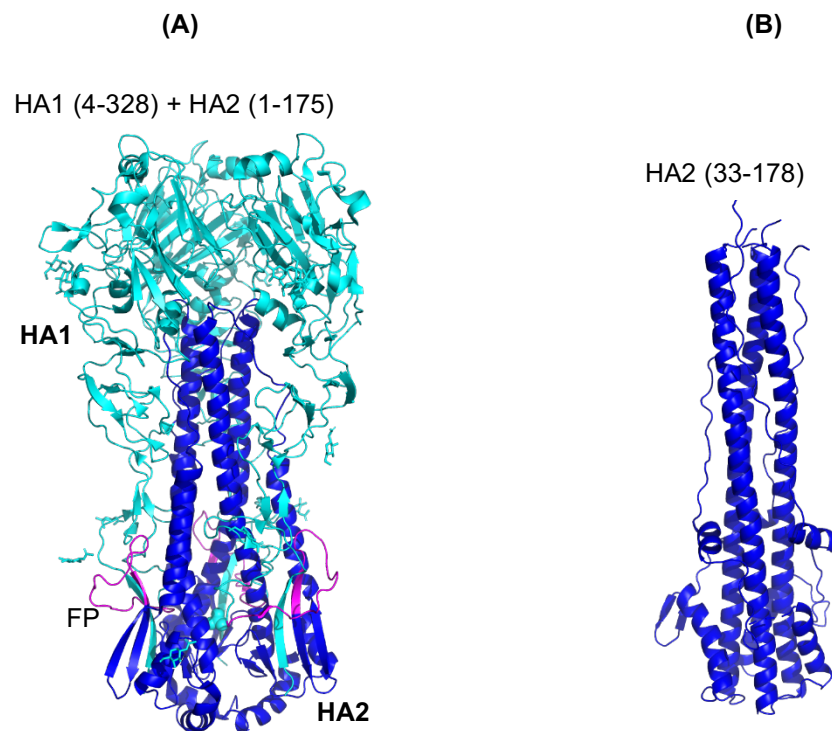


**Figure 1.3** Schematic representation of HA (top) and HA2 subunit (bottom), with domain colored: fusion peptide (FP), pink; soluble ectodomain (SE), blue; transmembrane domain (TM), green; and endodomain (Endo), orange.

#### 1.4.1 Structural studies of hemagglutinin

Viral fusion proteins can be categorized into three classes based on the structural arrangement of the fusion subunit: class I fusion proteins have a triple-stranded  $\alpha$ -helical coiled-coil structure; class II proteins typically contain  $\beta$ -sheets; and class III proteins contain both helices and sheets.<sup>9</sup> Influenza HA is a prototype of class I fusion proteins.

There are high-resolution structures for the initial complex of the proteolytically-cleaved HA2 ectodomain (residues 1-175) with HA1, and these structures typically contain trimers of HA2 with three bound HA1's (Fig. 1.4A).<sup>8, 25</sup> However, there are some recent reports of the folded monomer HA1/HA2 complex, e.g. A/Korea/01/2009 strain.<sup>25, 26</sup> The latter structure is nearly identical with the HA1/HA2 subunit structure of the trimers, which implies that the HA1/HA2 heterodimer is a fundamental folding unit.



**Figure 1.4** Crystal structures of the: (A) initial complex of HA1/HA2 at pH 7.5 (PDB 1RD8); and (B) HA2 soluble ectodomain at pH 5 (PDB 1QU1). HA1 and HA2 are colored in cyan and blue respectively.<sup>8, 27</sup>

The typical initial trimeric HA1/HA2 complex includes a stem region containing residues primarily from HA2 (blue in Fig.1.4A), and a globular region containing residues primarily from HA1 (cyan in Fig. 1.4A).<sup>8</sup> Each HA2 monomer contains two antiparallel helices: a short  $\alpha$ -helix (residues 38-56) connecting to a long  $\alpha$ -helix (residues 76-125) via an extended loop. In the trimer, the long helices form the internal coiled-coil, and the short helices pack antiparallel with the lower half of the coiled-coil. The irregular fusion peptide (residues 1-22), which is connected to the outer short helix by a  $\beta$ -sheet hairpin (residues 23-35), is protected within the trimer interface. There is a small  $\beta$ -sheet hairpin (residues 130-141) C-terminal to the internal long helix, followed by a compact association of two short helices (residues 146-153 and 159-170).<sup>8</sup>

Understanding membrane fusion with this structure has several difficulties.<sup>28</sup> First, the 135 Å length of the molecule prevent any closer approach of the membranes. Next, the fusion peptide (N-terminus of HA2), which is believed to interact with the target membrane (discussed in section 1.6.1), is located ~100 Å from the distal tip and ~35 Å from the viral membrane.

The acidic environment of the late endosome triggers major conformational changes of HA2 that facilitates membrane fusion. A three-dimensional structure of most of the HA2 soluble ectodomain (residues 33-178) at acidic pH (~5) is also available with 1.9 Å resolution.<sup>27</sup> This structure is also a trimer, and each monomer is 110 Å long hairpin, where both C-terminus and N-terminus are at one end (Fig. 1.4B). The N-terminal helices (residues 38-105) form a triple stranded coiled-coil, followed by a 180° turn. Then there are short antiparallel  $\alpha$ -helices (residues 110-128) outside the helical bundle, which pack in the grooves between the long helices. In each monomer, the short helix is followed by a short hairpin and a two turn  $\alpha$ -helix (residues 146-153), which crosses the long helix almost perpendicular to its axis. Then the rest of the C-terminal strands (residues 154-178) in the trimer, extend antiparallely in the grooves between the long helices.

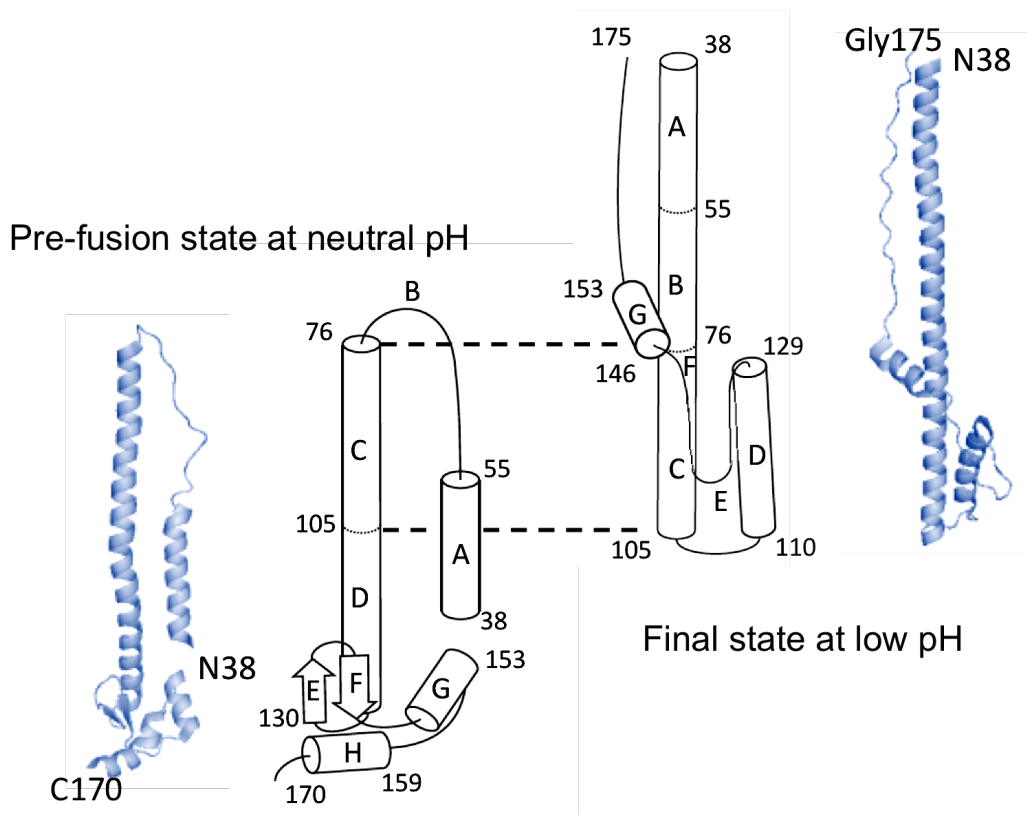


This structure is likely the final HA2 state in fusion based on its hyperthermostability with the melting temperature  $>90\text{ }^{\circ}\text{C}$ .<sup>29</sup>

#### **1.4.2 Low pH induced conformational changes of HA2**

The low pH triggers an enhanced protonation of HA leading to electrostatic repulsion of trimeric globular HA1 subunits.<sup>30</sup> When the endosomal pH drops into fusogenic pH (pH 5-6), some amino acid residues in HA1 and HA2 protonate initiating fusion-related conformational changes. Residues such as histidine ( $\text{pK}_a \sim 6$ ), glutamates ( $\text{pK}_a \sim 4$ ) and aspartates ( $\text{pK}_a \sim 4$ ) may be protonated as their  $\text{pK}_a$  values are in the acidic range. Although the  $\text{pK}_a$ s of free Glu and Asp side chains are lower compared to the fusogenic pH, the carboxylate group can still protonate depending on their chemical environment in the protein. Most of the residues that are known to be protonated at acidic pH are located surrounding the fusion peptide and HA1-HA2 interface.<sup>31</sup> The ionizable residues His17 in HA1 and Asp109 and Asp112 in HA2 are in a solvent accessible environment in the uncleaved HA structure (H3 subtype) and become completely buried in the trimer interior in the cleaved structure, which imply a role of these residues in the initiating of low pH-induced HA structural changes. When one residue is protonated, it can influence the protonation equilibrium of neighboring residues, and therefore it is difficult to identify the specific residues causing the conformational changes.<sup>24, 32, 33</sup>

The acidic environment of the late endosome results in conformational rearrangements throughout the HA2 protein. The overall effect of the rearrangement is to release the fusion peptide (FP) from its initially buried position, and to bend the molecule in half so that the FP and the transmembrane (TM) domain are at the same end.



**Figure 1.5** The structure of a pre-fusion state (initial state) monomer at neutral pH (left) is compared to the corresponding region of a final state monomer at low pH (right).<sup>28</sup> Labels A-H represents the consecutive regions of HA2 (residues 38-175).

Three major rearrangements can be observed (Fig. 1.5). First, the extended loop and short  $\alpha$ -helix in the pre fusion state (A and B) become part of the internal long helix, which form the coiled-coil in the final HA2 state. This releases the FP from the previous buried position. Second, few residues in the middle of the long  $\alpha$ -helix (in between C and D) unfold to form a 180° turn, and rest of the helix D becomes antiparallel to the long  $\alpha$ -helix. Third, the residues C-terminal to the  $\beta$ -sheet hairpin in the initial state (G and H) become mostly extended, and runs antiparallely

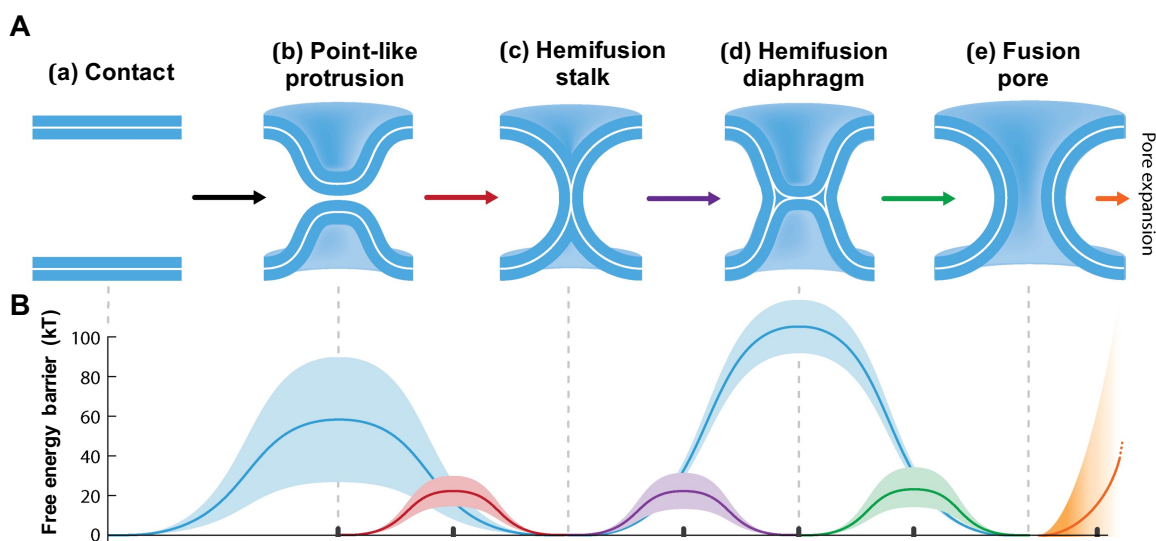
within the groove between the adjacent  $\alpha$ -helices in the central coiled coil. The short  $\alpha$ -helix H is completely extended in the final state.

## **1.5 Membrane fusion**

Membrane fusion involves the joining of the viral and cellular membranes to form pores through which their aqueous contents can interchange. The membrane fusion is characterized by multistep membrane rearrangements that include: (1) hemifusion, in which the outer leaflets of the fusing compartments are merged while inner leaflets remain separate; (2) small pore formation, which allows the passage of <1 kDa species; and (3) pore expansion, which allows the complete mixing of the viral and cellular contents. Membrane fusion does not occur spontaneously because the fusion of two lipid bilayers has to overcome several kinetic barriers, such as hydration force (a repulsive force between polar surfaces separated by a thin layer of water).<sup>34-36</sup>

### **1.5.1 Fusion of protein-free lipid bilayers**

The biological membranes are made of lipids which contain a non-polar tail and a polar head group. The lipid tails make the hydrophobic interior, whereas the head groups make the hydrophilic membrane exterior. Major forces involved in membrane rearrangements include: the hydrophobic effect, Van der Waals forces and hydration force.<sup>37</sup> These forces together with membrane geometry/curvature play a major role in membrane fusion (discussed below).

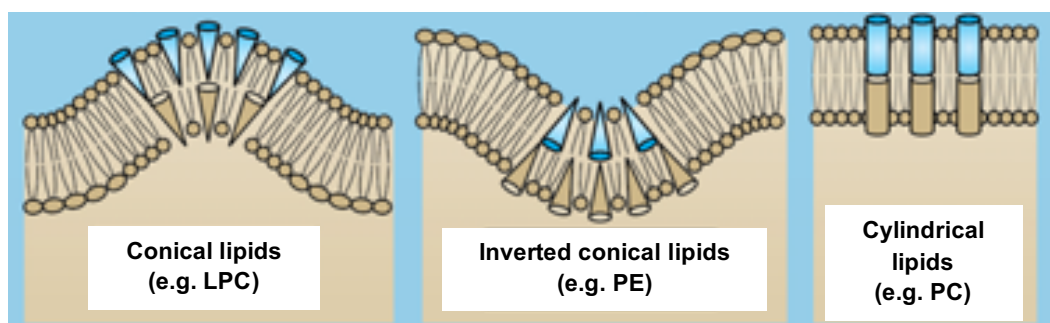


**Figure 1.6** (A) Proposed steps in the protein-free lipid bilayer fusion<sup>35</sup>: (a) lipid membranes at pre-fusion state; (b) a point like protrusion, which minimizes the repulsion associated with hydration force; (c) hemifusion stalk, in which the outer leaflets of the two membranes mix while inner leaflets remain separated. The optimized packing of the hydrocarbon chains (lipid tails) in the hemifusion stalk minimizes the energy associated with it; (d) hemifusion diaphragm formed as a result of the radial expansion of the stalk; and (e) fusion pore formation. It is hypothesized that the fusion pore can be formed either from the hemifusion stalk or from the hemifusion diaphragm. (B) The free energy reaction coordinate diagram. The blue transitions represent the energy barriers for the single step transition from pre-fusion to hemifusion state and from there to pore formation. The free-energy barriers between each intermediate steps (shown in colored arrows) in panel A are indicated in respectively colored curves. Comparison of the barrier heights are facilitated by aligning the absolute free energy values of all intermediate states at 0 kT.

Prior to fusion, it is important to establish a contact between the two membranes. The bilayers made of neutral lipids like phosphatidylcholine (PC) can come into closer contact (2-3 nm). However, the intermembrane distance is much larger (10-20 nm) for biologically relevant membranes because of the electrostatic repulsion between lipid membranes and inherent steric interactions of the membrane proteins and carbohydrates. In general, the protein-free lipid bilayers do not fuse even when kept in close proximity for days. This has been observed with bilayers made of PC, a major lipid found in mammalian cell membranes. The membranes can fuse when dehydrate the contact zones with concentrated polyethylene glycol.<sup>38</sup> In biological membranes, the action of fusion proteins can bring the two membranes into a closer proximity (discussed in section 1.5.2).

The polar lipid head groups in the membrane are tightly interacting with surrounding water molecules. Both viral and host cell membranes are stripped from this water layer (hydration layer), and removing these water molecules required energy, which is defined as the hydration repulsion.<sup>39</sup> When the two membranes come to a closer proximity there will be a strong hydration repulsion. A point like protrusion of the membrane will minimize this hydration barrier because of the reduced area of close contact (Fig. 1.6b). Next, the two outer leaflets merge giving rise to a hemifusion stalk (Fig. 1.6c). Hemifusion has been observed experimentally by lipid mixing without content mixing or lipid mixing of outer leaflets but not the inner leaflets of the interacting membranes.<sup>38</sup> Transition from pre-fusion state to the hemifusion state (Fig. 1.6a→c) is a high energy demanding process (Fig. 1.6B), because of the dehydration of polar lipid head groups and subsequent exposure of non-polar lipid tails to the aqueous environment during the transition. The point-like protrusion reduces the area of dehydration as well as the hydrophobic energy associated with outer leaflet rupture.

The hemifusion stalk can either expand to a hemifusion diaphragm or to a fusion pore. The fusion pore connects the two aqueous environments, which was initially separated by the apposed membranes. This suggests the formation of a hydrophilic pore, covered with lipid head groups.<sup>36</sup> After the pore is formed, it needs to expand to allow the complete transfer of virus genome into the host cell. Some studies report that the pore expansion is the highest energy demanding step (discussed in section 1.5.2).<sup>38, 40</sup>



**Figure 1.7** Pictorial representations of how the lipid shape determines the membrane curvature: conical shaped lipids protrude towards the polar head of the outer monolayer and induce positive membrane curvature, e.g. lysophosphatidylcholine (LPC); inverted conical shaped lipids protrude towards lipid tails of the outer monolayer and induce negative curvature, e.g. phosphatidylethanolamine (PE); and cylindrical lipids induce an almost flat membrane, e.g. phosphatidylcholine (PC). The blue and brown triangles and cylinders represent the respective lipid shape stated in each panel.

The direction of the bending of bilayer plane (the plane between the lipid monolayers in a bilayer) determines its membrane curvature (Fig. 1.7). A positive curvature of the bilayer of a vesicle or a cell refers to the bending of the bilayer plane towards the outer environment, and a

negative curvature refers to the bending of the bilayer plane towards inside the vesicle or cell.<sup>38</sup> Biologically relevant membranes typically have a net positive curvature prior to fusion. In the hemifusion and final fusion states, the lipid bilayers form negative curvatures.<sup>41</sup> Therefore, the lipid composition and the initial curvature of the interacting membranes play a major role in the energetics of membrane fusion. Vesicles made of conical shaped lipids in the outer monolayer that stabilize the positive curvatures, such as lysophosphatidylcholine, facilitate pore formation. On the other hand, lipids like phosphatidylethanolamine that stabilize negative curvature promote hemifusion, and impair pore formation.<sup>42</sup>

### **1.5.2 Protein-mediated fusion and proposed mechanisms of HA-mediated fusion**

Fusion proteins are considered as catalysts, which can lower the kinetic barriers associated with the membrane fusion. However, the mechanism of protein-mediated fusion is not well-understood. It has been proposed that the membrane proteins may contribute to the local bending of lipid bilayers that point towards the adjacent membrane, which is a critical step in membrane rearrangements.<sup>36, 43</sup>

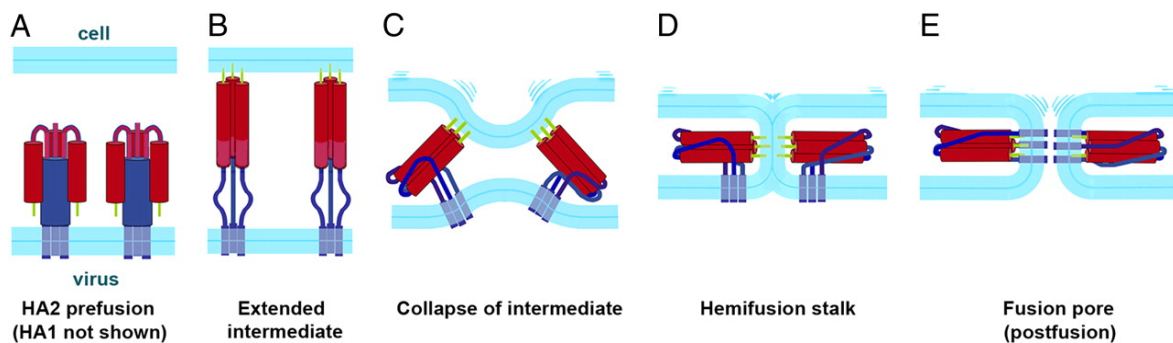
The influenza virus/endosome fusion is catalyzed by the HA2 protein. Numerous fusion models have been proposed to describe the HA2-mediated membrane fusion. These models are primarily relying on the structure of the initial complex of HA1/HA2 ectodomain and the final hairpin structure of the HA2 soluble ectodomain (Fig. 1.4). There are limited data about the transformation between the initial and final state HA2 structures and the time-correlations of changes in HA structure relative to changes in membrane structure. Much of our understanding on virus/endosome fusion are based on vesicle fusion and cell/cell fusion studies, and will be discussed in details in chapter 3.

Several models have been proposed to explain the enveloped virus fusion. In the “helix bundle formation mechanism”, helical protein residues attached to the two fusing membranes unite, thereby pulling the two membranes towards each other. The N-terminal three-stranded coiled coil in the final state is covered with the C-terminal helices forming a six helix bundle (SHB).<sup>9</sup> The HA2 SHB is at its membrane-distal end, and it is relatively shorter compared to other class I fusion proteins such as HIV gp41, which forms a SHB that extends to membrane-proximal end of the central coiled coil. The short SHB of HA2 is followed by a C-terminal “leash” (an extended chain) that packs into the “grooves” of the central coiled-coil. Several studies using mutants in the leash region (discussed below) suggest that the packing of the leash into the grooves is essential in fusion, and this may have an analogous role to SHB of other fusion proteins.<sup>44</sup> The mechanistic hypotheses of this model include: (1) the membranes protrude towards one another by early HA2 conformational changes (collapsing of the prehairpin intermediate in Fig. 1.8B - discussed in details under *Model 1* below); (2) the lipid mixing associated with hemifusion is induced by the packing of the C-terminal leash into the grooves; and (3) the fusion pore formation is mediated by the actions of FP and TM. The data that support some of these hypotheses are as follows. HA2 C-terminal region mutations, either deletion or substitution, reduce the fusion between HA-expressing cells (HA-cells) and red blood cells (RBCs). The deletion of segments C-terminal to the helical bundle (residues 176-180 and 181-184) does not affect lipid or content mixing, however mutations in the residues 154-175 region (e.g. deletion of residues 154-175, 171-175, or I173E mutation) drastically reduce the fusion activity (both hemifusion and pore formation).<sup>44, 45</sup> The data support that the packing of the C-terminal 154-175 strand into the grooves of the central coiled-coil play a key role in fusion.



### 1.5.2.1 Model 1

Calorimetric data support that the initial trimeric HA2/HA1 complex is in the lowest free-energy state at neutral pH.<sup>46, 47</sup> The lower pH (<6) leads to a metastable HA2/HA1 complex, and this dissociates to HA1 with similar structure as in the initial complex and metastable HA2. It is often proposed that some of the free energy released during the HA2 restructuring to the final hairpin structure is converted to the activation energy of fusion,<sup>48, 49</sup> and perhaps also thermodynamic energy needed for membrane intermediate states. The pictures usually show the membrane intermediate formation as time-correlated with the evolution of hairpin closure (Fig. 1.8).<sup>9, 50</sup> The final step of fusion pore expansion is sometimes pictured as requiring a transmembrane bundle of three TM segments and sometimes together with three FP segments. The SE hairpin is commonly termed the “postfusion state”, which implies that the final-state trimer-of-hairpins structure is formed after the fusion is completed. Since the pictures show that the hairpin closure and the fusion pore formation occur at the same time, calling the hairpin a postfusion state is a misnomer.



**Figure 1.8** A proposed mechanism of the HA-mediated membrane fusion<sup>50</sup>; (A) HA2 in the prefusion state (initial state). The protein traverses the viral membrane through the TM, and the

FP (green) is sequestered near the viral membrane. HA1 is not shown for clarity. (B) Upon acidification, (HA1 moves away from HA2 - not shown), liberating FP from its sequestered position. The HA2 SE extends, and the FP insert into the target membrane forming an extended intermediate that bridges the two membranes (prehairpin intermediate). (C) Subsequently, the extended intermediate collapses moving the C-terminal region of the ectodomain (blue strands) next to the N-terminal helix (red), thus pulling the two membranes towards each other. (D) The outer leaflets of the two membranes merge leading to a hemifusion stalk. (E) The inner leaflets are merged forming a fusion pore as a result of the HA2 hairpin closure.

*(A) Pre-fusion state*

HA1 binds to the sialic acid receptors on the target cell membrane and the virus is endocytosed. The low pH of the mature endosome triggers the dissociation of HA1 from HA2. In the initial neutral pH structure of HA2/HA1 complex,<sup>8</sup> the hydrophobic N-terminal region of HA2 (FP) is inserted between the HA2 stem and HA1 globular head, near the viral envelope. The low pH induced conformational changes unlock HA2 by relocating HA1, allowing HA2 to refold to its final low energy conformation.

Although the initial and the final hairpin structures are known, little is known about the conformational transition between these two structures. There are some evidences for the presence of intermediate states by studies involving mutagenesis, antibody binding, proteolysis and structural and biophysical methods (discussed below).<sup>31, 44, 51-54</sup>

*(B) The prehairpin intermediate (PHI)*

The antiparallel helices in each HA2 monomer in the initial trimer complex are connected by an unstructured loop (B-loop in Fig. 1.5) that has the tendency to form a trimeric helical coiled coil.<sup>48, 55</sup> The conversion of this loop into a helix forms a long  $\alpha$ -helical rod in the N-terminus of the molecule. This loop-to-helix transition relocates the FP towards the target membrane forming a hypothetical extended intermediate commonly known as the prehairpin intermediate (PHI), that bridges the viral and target cell membranes (Fig. 1.8B). It should be noted that the PHI relies on the loop to helix transition occurring faster than the helix to loop transition associated with the folding of the molecule in half (discussed under *Collapse of PHI*).

Structural characterization of the PHI is difficult given the transient nature, however, recent cryo-ET data have been interpreted to support its existence during influenza membrane fusion.<sup>56</sup> There are also some studies showing the existence of the PHI for other class 1 fusion proteins such as paramyxovirus PIV5 F<sup>57</sup> and HIV gp41.<sup>58</sup> The time-of-addition experiments with peptides that can inhibit HIV fusion show that the peptides are effective for a few minutes after initiation of fusion and this time could be interpreted to be the lifetime of the PHI.<sup>59</sup>

The fusion of paramyxovirus PIV5 F has been studied by measuring the distance between viral membrane and nano-bead supported bilayers using electron microscopy.<sup>57</sup> The lipid bilayer distance of the two membranes was calculated using computer modeling. In the PHI state, the calculated distance was 21 nm whereas in the pre-fusion state the value was 12 nm. To get the experimental distances, an inhibitory peptide C-1 was used that specifically bind to the PHI of the F protein, preventing it from proceeding to the final fusion state. The distance between viral and nano-bead supported bilayers was measured in the presence and absence of the C-1 peptide. In the absence of the peptide, the fusion occurs, hence a shorter distance ( $\sim 1$  nm) was observed, whereas

in the presence of the peptide the distance was 20 nm. As the calculated distance (21 nm) is close to the experimentally determined distance, it is assumed that this study was able to capture the PHI of paramyxovirus PIV5 F protein.

### *(C) Collapse of PHI*

Few residues in the middle of the internal long helix in the initial HA2 structure converts to a loop (in between helix C and D in Fig. 1.5), forming a turn, allowing the C-terminal leash to zipper into a groove in the N-terminal coiled-coil.<sup>44</sup> This draws the target-membrane-embedded FP and the viral-membrane-embedded TM into closer proximity to each other (Fig. 1.8C).<sup>41</sup>

### *(D) Hemifusion*

If only the outer leaflets of the membrane bilayers fuse while the inner leaflets remain separated, the resulting membrane state is known as the hemifusion state (Fig. 1.8D). Fusion between cells expressing glycosylphosphatidylinositol-anchored HA (without TM or endodomain) and red blood cells with fluorescent-labelled lipids and fluorescent soluble content probes shows only the lipid mixing, and not the content mixing, suggesting the hemifusion state as a membrane intermediate in fusion.<sup>60, 61</sup> However, there are no experimental data about the HA2 structure in the hemifusion state.

### *(E) Fusion pore formation*

At the end of membrane fusion, a fusion pore is formed through which the viral genome can enter to the host cell cytoplasm. The *model 1* suggests that the formation of this fusion pore is caused by the HA2 hairpin closure, merging the inner leaflets.

In the final state, HA2 has a hairpin conformation in which the FP and TM domains are positioned at the same end of a rod-like molecule. The FP and TM play a major role in the fusion process (discussed in section 1.6). Some point mutants in the FP domain impair HA-mediated fusion (discussed in section 1.6.1).<sup>62</sup> The observation of exchange of only the smaller fluorescent probes (without exchanging the larger probes) between the two fusion bodies (e.g from RBCs to HA-cells) provides evidence for the presence of a small fusion pore before the final pore expansion.<sup>63, 64</sup>

The HA2 I173E mutant shows greatly-reduced lipid mixing and pore formation for the HA-cell/RBC fusion.<sup>45, 63</sup> The nonpolar I173 side chain is inserted into a hydrophobic pocket in the final-state trimer-of-hairpins structure. The charged E173 side chain may prefer the aqueous environment because of more favorable Born energy of the charged Glu side chain in the high-dielectric aqueous environment vs. the low-dielectric protein interior. This may be made possible by dissociating the C-terminal strand of I173E from the internal helical bundle. Therefore, the I173E mutant may not have a tight hairpin structure as the wildtype. The observation of greatly-reduced lipid mixing and pore formation of I173E mutant implies that the hairpin is required for both lipid mixing and pore formation stages of fusion. This contradicts the fusion *model 1*, which suggests that the hairpin is required only for the pore formation and not for the lipid mixing associated with hemifusion.

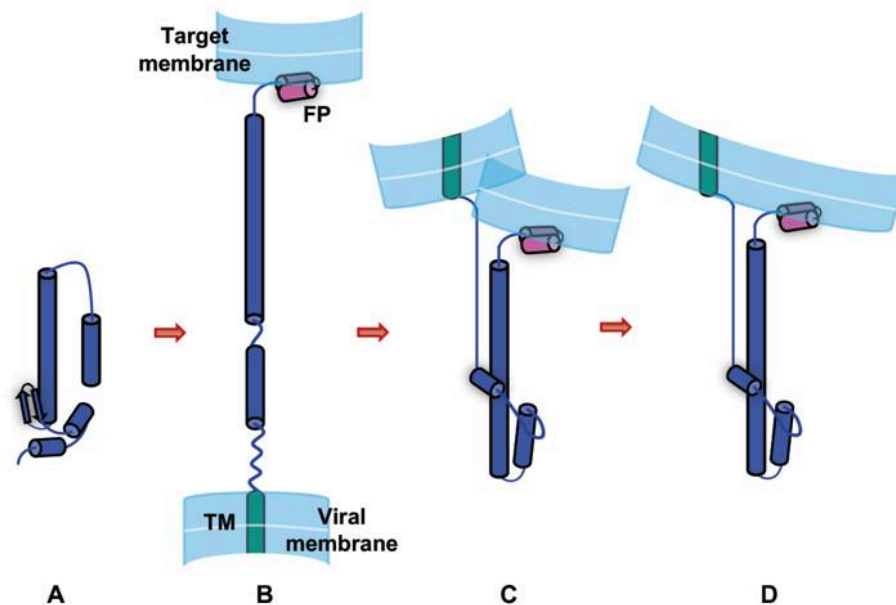
It has been found that the TM domain (with or without the endodomain) is necessary for the fusion pore expansion but not for the hemifusion and small pore formation<sup>60, 61, 63, 65</sup> (discussed in section 1.6.2). An increased perturbation of the membrane bilayers through an association between the FP and TM during fusion is also suggested.<sup>66, 67</sup>

Fusion involves the bending of lipid membranes (changing the curvature), which requires energy. Theoretical calculations suggest that the energy required for the fusion pore expansion is larger ( $>50$  kT) than hemifusion and pore formation.<sup>40</sup> Experimental evidence that support this theoretical consideration include: decreasing the HA surface density shifts the fusion phenotype of HA-mediated cell/cell fusion from complete fusion to lipid mixing in the absence of a fusion pore, and then to neither pore nor lipid mixing<sup>68</sup>; and requirement of the membrane spanning TM for pore expansion but not for the pore formation.<sup>63</sup> To our knowledge, the mechanism of pore expansion is not well understood.

#### **1.5.2.2 Model 2**

An alternate fusion model is the FP binding to the target membrane during the HA2 structural transformation to the final hairpin state, and subsequent membrane fusion with HA2 in the final state (Fig. 1.9). This model suggests that the HA2 final hairpin structure is fusogenic and not a post fusion state.<sup>63</sup>

Experimental support for this alternate model includes vesicle fusion induced by FP peptides, and much greater fusion induced by larger constructs for which the SE has the final trimer-of-hairpins structure, such as HA2(1-127), HA2(1-185) and full-length HA2 (discussed in Chapter 3).<sup>29, 63, 69-71</sup>



**Figure 1.9** An alternate fusion model; (A) Initial state HA2 SE (HA1 is not shown and only a HA2 monomer is shown for clarity). (B) FP binds to the target membrane during HA2 structural transformations, forming a prehairpin intermediate. (C) The HA2 hairpin formation brings the two interacting membranes in to a close proximity prior to fusion. (D) Subsequent fusion (hemifusion and pore formation) occurs when HA2 is in the final hairpin conformation.

To our knowledge, the mechanism of viral membrane fusion by influenza HA is not well understood. Understanding the mechanism of membrane fusion will create new ways of controlling the ubiquitous phenomenon of viral infection.

## 1.6 The membrane interacting domains of HA2

The influenza FP and TM are the only HA2 domains that are deeply inserted in the fused membrane.<sup>60, 65, 72-74</sup> Accumulating evidence from the mutagenesis and the domain swapping experiments suggests that they play an active role in the fusion process rather than just connecting the two interacting membranes.<sup>60, 61, 65, 72, 75</sup>

When determining the HA2 final hairpin structure, a HA2 construct containing residues 23-185 was used, which does not contain the FP and TM domains. In this structure the N- and C-termini are at the same end of the molecule opposite to the hairpin turn, which implies that the FP and TM domains are positioned at the same end of the molecule.

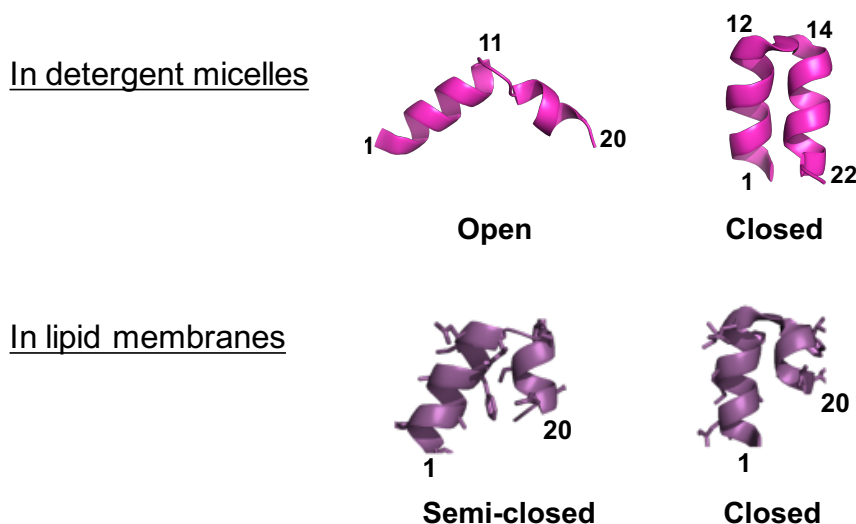
### 1.6.1 Influenza fusion peptide (FP)

The N-terminal, reasonably hydrophobic, ~23 residue FP, is the most highly conserved region of HA.<sup>9</sup> In the initial complex of HA2/HA1, the FP is inserted between the HA2 stem and HA1 globular head near the viral membrane.<sup>8</sup> In this structure, a considerable amount of FP residues form hydrogen bonds to highly conserved ionizable residues in HA2. For example, FP residues 2-6 form hydrogen bonds to aspartic acid 109 and 112. This suggests that most of the N-terminal residues are conserved not only for the functional association with target membrane but also to maintain the neutral pH, cleaved HA structure.<sup>62</sup>

Several NMR structures of the monomeric FP without the rest of the HA2 are available in detergent-rich media. In particular, a 20-residue FP from H3 viral subtype shows an open (V-shaped) N-helix/turn/C-helix structure at pH 5.0, and an open N-helix/turn/C-coil structure at pH 7.0.<sup>76</sup> In some contrast, a 23-residue FP from H1 viral subtype shows a closed (tightly packed) N-helix/turn/C-helix hairpin structure at both low and neutral pH (Fig. 1.10).<sup>77</sup> In membranes, both



peptides adopt a population mixture of closed and semi-closed (less tightly packed) N-helix/turn/C-helix structures at both pH 5.0 and 7.0.<sup>70</sup> The open structure is favored in detergent micelles may be due to a locally curved detergent micelle surface vs. locally flat membrane surface in lipid vesicles.<sup>70</sup>



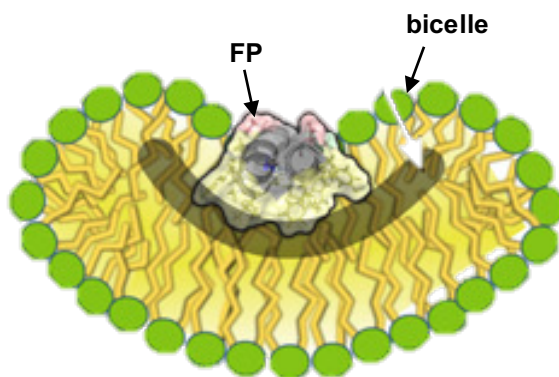
**Figure 1.10** NMR structures of the FP in the absence of rest of the HA2.

Several research groups have studied the membrane interaction of the FP by FTIR studies on short synthetic peptides ( $\leq 25$  residues). These studies support that the FP binds to vesicles in tilted orientation, however the calculated tilt angles between the horizontal plane of the membrane and the helical axis widely vary between  $20^\circ$ - $60^\circ$ .<sup>78-80</sup> The membrane topology of the FP has also been studied with Spin-labelling EPR on a truncated HA2 construct (residues 1-127).<sup>81</sup> Although the construct was trimeric, the FP region was locally monomeric in the membrane with an  $\alpha$ -helical structure, which was tilted  $\sim 25^\circ$  with respect to the membrane phosphorous layer at both

pH 5 and 7. The maximum depth of insertion is  $\sim 10$  Å (for residue 6) from the membrane phosphorous layer.

Functional studies with mutated FP shows that the influenza FP plays an important role in membrane fusion, even a single residue mutation in the FP can impair its fusion activity. For example, the HA G1E (HA2 numbering) mutant in the FP greatly reduce the fusion of HA-expressing cells (HA-cells) and red blood cells (RBCs).<sup>32</sup> These studies suggest that the FP is not just a membrane anchor. The FP and TM domains may facilitate fusion by perturbing the bilayer structure, rather than just anchoring the rest of the protein to the membrane.<sup>36</sup>

The FP has amphiphilic properties hence it cannot penetrate the membrane deeply, instead the insertion will be shallow with its hydrophobic side chains facing towards the lipid hydrocarbon chains, and hydrophilic residues remain close to the lipid head group-water interface.<sup>77</sup> This will generate a stress in the monolayer, because the shallow insertion will expand the upper region of the monolayer (close to the membrane phosphorous layer) while keeping the lower region as it is. Bending will partially release this stress,<sup>82</sup> and the resulting curvature of the membrane will depend on the location and the shape of the FP. Experimental and MD simulation studies with FP peptides without rest of the HA2 report different membrane locations for the FP. For example, MD simulation studies support: insertion of the FP in a single membrane leaflet<sup>83</sup>; membrane traversal<sup>84</sup>; and FP at the membrane surface.<sup>85</sup> Some studies report that the inverted conical shaped helical hairpin structure of the FP induces a negative curvature in the membrane (Fig. 1.11), and thereby stabilizing the early fusion states such as hemifusion stalk, which has a strong negative curvature.<sup>86</sup>



**Figure 1.11** The negative curvature in a bicelle induced by inverted conical shaped FP.<sup>86</sup>

### 1.6.2 Influenza hemagglutinin transmembrane domain (TM)

Influenza HA spans the viral membrane through its single span TM that consists of ~25 amino acid residues at the C-terminus of the HA2 subunit. There is no high resolution structure for the TM. CD studies with synthetic peptides show an  $\alpha$ -helical conformation in detergent micelles as well as in phospholipid bilayers.<sup>27, 87</sup> There are little detailed studies on sequence requirement of the TM on fusion. Replacing the HA TM with foreign TMs from both viral and non-viral proteins leads to complete fusion.<sup>88</sup> When both the TM and endodomain are derived from the same protein they are more effective in promoting fusion than when they are derived from different proteins. The oligomeric state of the TM is not known. Monomer TM is consistent with low sequence conservation in this domain among HA subtypes, and with robust HA-cell/RBC fusion after substitution of the native TM and endodomain sequences with those of an unrelated protein.<sup>88</sup> The fluorescence and SDS-PAGE data for HA2 TM peptides support a fraction of oligomeric peptides,<sup>67</sup> and recent Cryo-EM data of the HA2/HA1 initial trimer complex in detergent are interpreted to support a trimeric TM bundle.<sup>89</sup>

Fusion experiments with recombinant HA constructs suggest that the TM is required for later steps of membrane fusion, as the GPI-anchored HA is arrested at hemifusion with small, non-expanding pores.<sup>61, 75, 90</sup> This is also supported by the cell fusion studies with exogenously added HA2 ectodomain (FHA2), in which small fusion pores are formed indicating that the TM and cytosolic domains are required for fusion pore expansion but not for the pore formation.<sup>63</sup> Studies have also demonstrated a minimal length requirement of the TM.<sup>65</sup> It is found that there should be at least 17 residues in the TM to promote complete fusion, and a 15-residue TM combined with a 10-residue cytoplasmic tail or with a single arginine also efficiently promotes the complete fusion. The cytoplasmic tail or arginine is most likely help the shortened TM to span the bilayer, and therefore this suggests that the TM should span the membrane bilayer for fusion pore expansion.<sup>65</sup>

Accumulating data support the importance of a membrane-spanning TM in the final fusion pore expansion step. As there is efficient fusion when the HA2 TM is replaced with a TM sequence of a non-homologous protein, a specific TM sequence may not be important in fusion, which explains the low sequence conservation of this domain.<sup>91</sup>

## **1.7 Towards a universal flu vaccine**

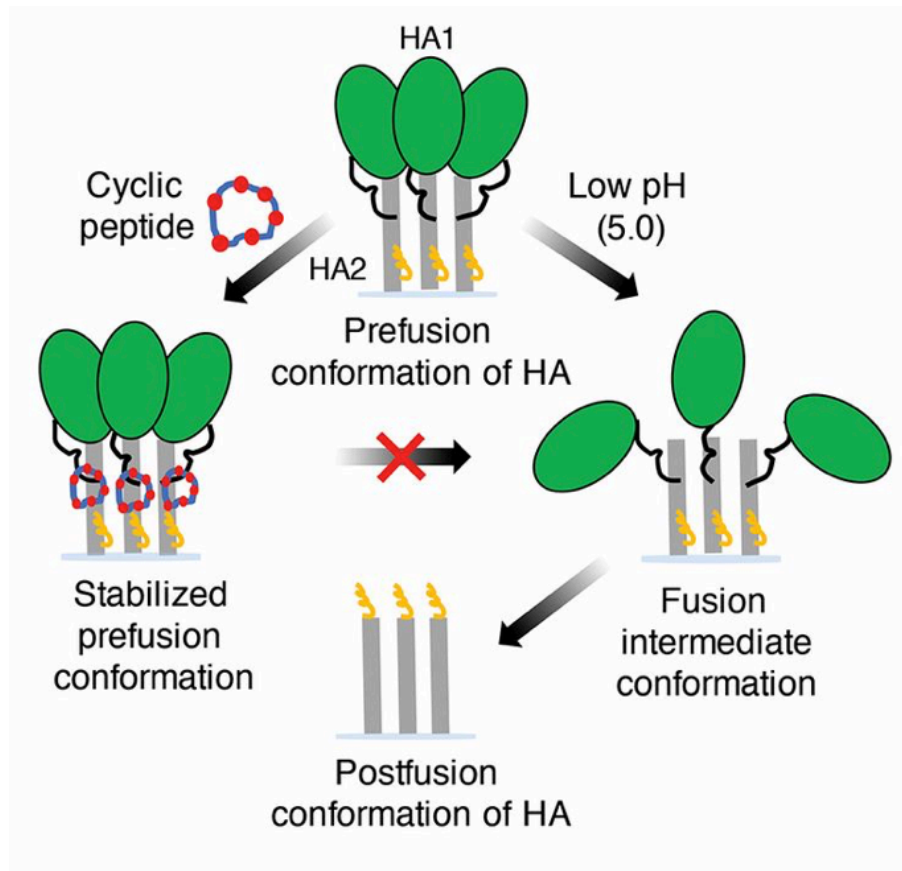
The idea of a universal flu vaccine, a vaccine that provide protection against all influenza virus strains or at least multiple strains, is challenging given the high diversity among these viruses. The current seasonal flu vaccines cannot provide a complete protection against the circulating viral strains, and the estimated protection remains within 50-60% among worldwide human population.<sup>10</sup> These vaccines are annually reformulated to match the circulating strains.

Current anti-influenza treatments target only two viral envelope proteins, the NA and M2 ion channels.<sup>2, 92, 93</sup> The drugs adamantanes, namely amantadine and rimantadine, are M2 inhibitors

and they block the virus uncoating process, which is necessary to release the viral RNPs into the host cell cytoplasm. As M2 ion channels exist only in influenza A virus, the drugs are not effective for influenza B and C. On the other hand, the NA inhibitors, drugs zanamivir and oseltamivir, inhibit efficient release of the progeny viruses from the host cell (discussed in section 1.3.1). These drugs are effective for influenza A and B viruses but resistant strains are constantly emerging.<sup>94-96</sup> For example: influenza A (H3N2) viruses with S31N mutation in M2 protein and A(H1N1)pdm09 (the last flu pandemic started in North America in 2009) are resistant to adamantanes; A(H1N1) viruses with H275Y mutation in NA protein are resistant to oseltamivir; and avian influenza viruses like A(H5N1) and A(H7N9) (potentially pandemic) are developing resistance to oseltamivir.<sup>94</sup>

New antiviral drugs that act on additional targets is urgently needed to fight against emerging viruses and constantly mutating circulating strains. The fusion protein HA is a promising target because of its critical roles during early stages of infection; the receptor binding (HA1) and viral endosome fusion (HA2). The HA2 subunit of HA is highly conserved compared to HA1, and therapeutic strategies targeting HA2 will be highly effective to reduce the possibility of generating escape mutants.<sup>92, 93</sup> Numerous small molecules have been identified that can inhibit the low pH-induced conformational changes associated with HA-mediated membrane fusion.<sup>92, 97-100</sup> For example, from crystal studies it has been found that tert-butyl hydroquinone (TBHQ) inhibits the HA conformational changes by binding to a hydrophobic pocket formed between the two monomers of the HA pre-fusion state trimer (Fig. 1.4).<sup>98</sup> However, sensitivity of TBHQ to only group 2 HAs (subtypes H3, H4, H7, H10, H14, H15) and not group 1 HAs (subtypes H1, H2, H5, H6, H8, H9, H11, H12, H13, H16, H17, H18) is a limitation (group 1 and 2 are the two major phylogenetic groups of influenza A viruses - briefly discussed in section 1.3.1).

Recent characterization of human broadly neutralizing antibodies (bnAbs), acting against HA, has raised the possibility for designing a more universal vaccine.<sup>101</sup> These antibodies bind to the HA stem (primarily the HA2 subunit) in the pre-fusion structure and block the pH induced rearrangements, which prevents HA-mediated membrane fusion. Structural characterization of bnAbs has identified a highly conserved region at the HA1/HA2 interface in the HA stem, present in both group 1 and 2 of influenza A (H1-H18) and both heredities of influenza B viruses. Several cyclic peptides have been prepared recently, that can bind to this conserved region, inhibiting conformational changes effectively.<sup>92</sup> Some optimized peptides have been recognized that can neutralize group 1 influenza A viruses H1N1 (2009 pandemic) and avian H5N1 strains with a nanomolar affinity.



**Figure 1.12** A model representing the inhibition of low pH-induced conformational changes by cyclic peptides.<sup>92</sup> The FP, HA2 and HA1 are respectively represented by yellow, grey and green colors.

## REFERENCES



## REFERENCES

1. MacNeal, W. J. (1919) The influenza epidemic of 1918 in the American Expeditionary Forces in France and England, *Archives of Internal Medicine* 23, 657-688.
2. WHO: Influenza. <http://www.who.int/immunization/diseases/influenza/en/>. (accessed April 12, 2018)
3. Centers for disease control and prevention (CDC). <https://www.cdc.gov/flu/about/burden/index.html> (accessed Nov 21, 2018).
4. Lamb, R. A. (2001) Orthomyxoviridae: the viruses and their replication, In *Fields virology*.
5. Fontana, J., and Steven, A. C. (2015) Influenza virus-mediated membrane fusion: structural insights from electron microscopy, *Archives of Biochemistry and Biophysics* 581, 86-97.
6. Gerl, M. J., Sampaio, J. L., Urban, S., Kalvodova, L., Verbavatz, J.-M., Binnington, B., Lindemann, D., Lingwood, C. A., Shevchenko, A., and Schroeder, C. (2012) Quantitative analysis of the lipidomes of the influenza virus envelope and MDCK cell apical membrane, *J Cell Biol* 196, 213-221.
7. Zebedee, S. L., and Lamb, R. A. (1988) Influenza A virus M2 protein: monoclonal antibody restriction of virus growth and detection of M2 in virions, *Journal of virology* 62, 2762-2772.
8. Wilson, I., Skehel, J., and Wiley, D. (1981) Structure of the haemagglutinin membrane glycoprotein of influenza virus at 3 Å resolution, *Nature* 289, 366.
9. White, J. M., Delos, S. E., Brecher, M., and Schornberg, K. (2008) Structures and mechanisms of viral membrane fusion proteins: multiple variations on a common theme, *Critical reviews in biochemistry and molecular biology* 43, 189-219.
10. Sautto, G. A., Kirchenbaum, G. A., and Ross, T. M. (2018) Towards a universal influenza vaccine: different approaches for one goal, *Virology journal* 15, 17.
11. Varghese, J., Laver, W., and Colman, P. M. (1983) Structure of the influenza virus glycoprotein antigen neuraminidase at 2.9 Å resolution, *Nature* 303, 35.
12. Gubareva, L. V., Kaiser, L., and Hayden, F. G. (2000) Influenza virus neuraminidase inhibitors, *The Lancet* 355, 827-835.
13. Hay, A. J. (1998) Functional properties of the virus ion channels, In *Textbook of Influenza* (Nicholson, K. G., Webster, R. G., and A.J., H., Eds.), pp 74-81, Blackwell Science.

14. Hay, A. (1992) The action of adamantanamines against influenza A viruses: Inhibition of the M<sub>2</sub> ion channel protein, *Sem. Virol* 3, 21-30.
15. Noda, T., Sagara, H., Yen, A., Takada, A., Kida, H., Cheng, R. H., and Kawaoka, Y. (2006) Architecture of ribonucleoprotein complexes in influenza A virus particles, *Nature* 439, 490.
16. O'Neill, R. E., Talon, J., and Palese, P. (1998) The influenza virus NEP (NS2 protein) mediates the nuclear export of viral ribonucleoproteins, *The EMBO journal* 17, 288-296.
17. Shi, Y., Wu, Y., Zhang, W., Qi, J., and Gao, G. F. (2014) Enabling the 'host jump': structural determinants of receptor-binding specificity in influenza A viruses, *Nature Reviews Microbiology* 12, 822.
18. Lakadamyali, M., Rust, M. J., Babcock, H. P., and Zhuang, X. (2003) Visualizing infection of individual influenza viruses, *Proceedings of the National Academy of Sciences* 100, 9280-9285.
19. Kobayashi, T., Beuchat, M.-H., Chevallier, J., Makino, A., Mayran, N., Lebrand, C., Cosson, P., Kobayashi, T., and Gruenberg, J. (2002) Separation and characterization of late endosomal membrane domains, *Journal of Biological Chemistry* 277, 32157-32164.
20. Ivanova, P. T., Myers, D. S., Milne, S. B., McClaren, J. L., Thomas, P. G., and Brown, H. A. (2015) Lipid Composition of the Viral Envelope of Three Strains of Influenza Virus Not All Viruses Are Created Equal, *ACS infectious diseases* 1, 435-442.
21. Möbius, W., Van Donselaar, E., Ohno-Iwashita, Y., Shimada, Y., Heijnen, H., Slot, J., and Geuze, H. (2003) Recycling compartments and the internal vesicles of multivesicular bodies harbor most of the cholesterol found in the endocytic pathway, *Traffic* 4, 222-231.
22. Bui, M., Whittaker, G., and Helenius, A. (1996) Effect of M1 protein and low pH on nuclear transport of influenza virus ribonucleoproteins, *Journal of virology* 70, 8391-8401.
23. Boulo, S., Akarsu, H., Ruigrok, R. W., and Baudin, F. (2007) Nuclear traffic of influenza virus proteins and ribonucleoprotein complexes, *Virus research* 124, 12-21.
24. Wiley, D. C., and Skehel, J. J. (1987) The structure and function of the hemagglutinin membrane glycoprotein of influenza virus, *Annual review of biochemistry* 56, 365-394.
25. Ekiert, D. C., Kashyap, A. K., Steel, J., Rubrum, A., Bhabha, G., Khayat, R., Lee, J. H., Dillon, M. A., O'Neil, R. E., and Faynboym, A. M. (2012) Cross-neutralization of influenza A viruses mediated by a single antibody loop, *Nature* 489, 526.
26. Cho, K. J., Lee, J.-H., Hong, K. W., Kim, S.-H., Park, Y., Lee, J. Y., Kang, S., Kim, S., Yang, J. H., and Kim, E.-K. (2013) Insight into structural diversity of influenza virus haemagglutinin, *Journal of General Virology* 94, 1712-1722.

27. Chen, J., Skehel, J. J., and Wiley, D. C. (1999) N-and C-terminal residues combine in the fusion-pH influenza hemagglutinin HA2 subunit to form an N cap that terminates the triple-stranded coiled coil, *Proceedings of the National Academy of Sciences* 96, 8967-8972.
28. Bullough, P. A., Hughson, F. M., Skehel, J. J., and Wiley, D. C. (1994) Structure of influenza haemagglutinin at the pH of membrane fusion, *Nature* 371, 37.
29. Ratnayake, P. U., Ekanayaka, E. P., Komanduru, S. S., and Weliky, D. P. (2016) Full-length trimeric influenza virus hemagglutinin II membrane fusion protein and shorter constructs lacking the fusion peptide or transmembrane domain: Hyperthermostability of the full-length protein and the soluble ectodomain and fusion peptide make significant contributions to fusion of membrane vesicles, *Protein expression and purification* 117, 6-16.
30. Huang, Q., Opitz, R., Knapp, E.-W., and Herrmann, A. (2002) Protonation and stability of the globular domain of influenza virus hemagglutinin, *Biophysical journal* 82, 1050-1058.
31. Xu, R., and Wilson, I. A. (2011) Structural characterization of an early fusion intermediate of influenza virus hemagglutinin, *Journal of virology*.
32. Gething, M.-J., Doms, R. W., York, D., and White, J. (1986) Studies on the mechanism of membrane fusion: site-specific mutagenesis of the hemagglutinin of influenza virus, *The Journal of cell biology* 102, 11-23.
33. Zhou, Y., Wu, C., Zhao, L., and Huang, N. (2014) Exploring the early stages of the pH-induced conformational change of influenza hemagglutinin, *Proteins: Structure, Function, and Bioinformatics* 82, 2412-2428.
34. Harrison, S. C. (2015) Viral membrane fusion, *Virology* 479, 498-507.
35. Boonstra, S., Blijleven, J. S., Roos, W. H., Onck, P. R., van der Giessen, E., and van Oijen, A. M. (2018) Hemagglutinin-mediated membrane fusion: A biophysical perspective, *Annual review of biophysics* 47, 153-173.
36. Chernomordik, L. V., and M.M., K. (2008) Mechanics of membrane fusion *Nat. Struct. Mol. Biol.* 15, 675-683.
37. Blijleven, J. S., Boonstra, S., Onck, P. R., van der Giessen, E., and van Oijen, A. M. (2016) Mechanisms of influenza viral membrane fusion, In *Seminars in cell & developmental biology*, pp 78-88, Elsevier.
38. Chernomordik, L. V., and Kozlov, M. M. (2003) Protein-lipid interplay in fusion and fission of biological membrane, *Annu. Rev. Biochem* 72, 175-207.
39. Rand, R. P. (1981) Intercating phospholipid bilayers: Measured forces and induced structural change, *Annual Review of Biophysics and Bioengineering* 10, 277.

40. Cohen FS, and GB, Melikyan. (2004) The energetics of membrane fusion from binding, through hemifusion, pore formation, and pore enlargement, *J. Membr. Biol* 199, 1-14.
41. Skehel, J. J., and Wiley, D. C. (2000) Receptor binding and membrane fusion in virus entry: the influenza hemagglutinin, *Annual review of biochemistry* 69, 531-569.
42. Chernomordik, L. V., Frolov, V. A., Leikina, E., Bronk, P., and Zimmerberg, J. (1998) The pathway of membrane fusion catalyzed by influenza hemagglutinin: restriction of lipids, hemifusion, and lipidic fusion pore formation, *The Journal of cell biology* 140, 1369-1382.
43. Kozlov, M. M., and Chernomordik, L. V. (1998) A mechanism of protein-mediated fusion: coupling between refolding of the influenza hemagglutinin and lipid rearrangements, *Biophysical Journal* 75, 1384-1396.
44. Park, H. E., Gruenke, J. A., and White, J. M. (2003) Leash in the groove mechanism of membrane fusion, *Nature Structural and Molecular Biology* 10, 1048.
45. Borrego-Diaz, E., Peeples, M., Markosyan, R., Melikyan, G., and Cohen, F. (2003) Completion of trimeric hairpin formation of influenza virus hemagglutinin promotes fusion pore opening and enlargement, *Virology* 316, 234-244.
46. Epand, R. M., and Epand, R. F. (2002) Thermal denaturation of influenza virus and its relationship to membrane fusion, *Biochemical Journal* 365, 841.
47. Epand, R. F., and Epand, R. M. (2003) Irreversible unfolding of the neutral pH form of Influenza hemagglutinin demonstrates that it is not in a metastable state *Biochemistry* 42, 5052-5057.
48. Carr, C. M., and Kim, P. S. (1993) A spring-loaded mechanism for the conformational change of influenza hemagglutinin, *Cell* 73, 823-832.
49. Carr, C. M., Chaudhry, C., and Kim, P. S. (1997) Influenza hemagglutinin is spring-loaded by a metastable native conformation, *Proceedings of the National Academy of Sciences* 94, 14306-14313.
50. Floyd, D. L., Ragains, J. R., Skehel, J. J., Harrison, S. C., and van Oijen, A. M. (2008) Single-particle kinetics of influenza virus membrane fusion, *Proceedings of the National Academy of Sciences* 105, 15382-15387.
51. Leikina, E., Ramos, C., Markovic, I., Zimmerberg, J., and Chernomordik, L. V. (2002) Reversible stages of the low-pH-triggered conformational change in influenza virus hemagglutinin, *The EMBO journal* 21, 5701-5710.
52. Böttcher, C., Ludwig, K., Herrmann, A., van Heel, M., and Stark, H. (1999) Structure of influenza haemagglutinin at neutral and at fusogenic pH by electron cryo-microscopy, *FEBS letters* 463, 255-259.

53. Das, D. K., Govindan, R., Nikić-Spiegel, I., Krammer, F., Lemke, E. A., and Munro, J. B. (2018) Direct visualization of the conformational dynamics of single influenza hemagglutinin trimers, *Cell* 174, 926-937. e912.
54. Garcia, N. K., Guttman, M., Ebner, J. L., and Lee, K. K. (2015) Dynamic changes during acid-induced activation of influenza hemagglutinin, *Structure (London, England : 1993)* 23, 665-676.
55. Ward, C. W., and Dopheide, T. A. (1980) Influenza virus haemagglutinin. Structural predictions suggest that the fibrillar appearance is due to the presence of a coiled-coil, *Australian journal of biological sciences* 33, 441-448.
56. Calder, L. J., and Rosenthal, P. B. (2016) Cryomicroscopy provides structural snapshots of influenza virus membrane fusion, *Nature Structural and Molecular Biology* 23, 853.
57. Kim, Y. H., Donald, J. E., Grigoryan, G., Leser, G. P., Fadeev, A. Y., Lamb, R. A., and DeGrado, W. F. (2011) Capture and imaging of a prehairpin fusion intermediate of the paramyxovirus PIV5, *Proceedings of the National Academy of Sciences* 108, 20992-20997.
58. Furuta, R. A., Wild, C. T., Weng, Y., and Weiss, C. D. (1998) Capture of an early fusion-active conformation of HIV-1 gp41, *Nature Structural and Molecular Biology* 5, 276.
59. Muñoz-Barroso, I., Durell, S., Sakaguchi, K., Appella, E., and Blumenthal, R. (1998) Dilation of the human immunodeficiency virus-1 envelope glycoprotein fusion pore revealed by the inhibitory action of a synthetic peptide from gp41, *The Journal of cell biology* 140, 315-323.
60. Durell, S. R., Martin, I., Ruyschaert, J.-M., Shai, Y., and Blumenthal, R. (1997) What studies of fusion peptides tell us about viral envelope glycoprotein-mediated membrane fusion, *Molecular membrane biology* 14, 97-112.
61. Melikyan, G. B., White, J. M., and Cohen, F. S. (1995) GPI-anchored influenza hemagglutinin induces hemifusion to both red blood cell and planar bilayer membranes, *The Journal of Cell Biology* 131, 679-691.
62. Cross, K. J., Langley, W. A., Russell, R. J., Skehel, J. J., and Steinhauer, D. A. (2009) Composition and functions of the influenza fusion peptide, *Protein and peptide letters* 16, 766-778.
63. Kim, C. S., Epand, R. F., Leikina, E., Epand, R. M., and Chernomordik, L. V. (2011) The final conformation of the complete ectodomain of the HA2 subunit of influenza hemagglutinin can by itself drive low pH-dependent fusion, *Journal of Biological Chemistry* 286, 13226-13234.
64. Blumenthal, R., Durell, S., and Viard, M. (2012) HIV entry and envelope glycoprotein-mediated fusion *Biol. Chem.* 287, 40841-40849.

65. Armstrong, R. T., Kushnir, A. S., and White, J. M. (2000) The transmembrane domain of influenza hemagglutinin exhibits a stringent length requirement to support the hemifusion to fusion transition, *The Journal of cell biology* 151, 425-438.
66. Lai, A. L., and Freed, J. H. (2015) The interaction between influenza HA fusion peptide and transmembrane domain affects membrane structure, *Biophysical journal* 109, 2523-2536.
67. Chang, D.-K., Cheng, S.-F., Kantchev, E. A. B., Lin, C.-H., and Liu, Y.-T. (2008) Membrane interaction and structure of the transmembrane domain of influenza hemagglutinin and its fusion peptide complex, *BMC biology* 6, 2.
68. Leikina, E., and Chernomordik, L. V. (2000) Reversible merger of membranes at the early stage of influenza hemagglutinin-mediated fusion, *Molecular biology of the cell* 11, 2359-2371.
69. Leikina, E., LeDuc, D. L., Macosko, J. C., Epand, R., Epand, R., Shin, Y.-K., and Chernomordik, L. V. (2001) The 1– 127 HA2 construct of influenza virus hemagglutinin induces cell– cell hemifusion, *Biochemistry* 40, 8378-8386.
70. Ghosh, U., Xie, L., Jia, L., Liang, S., and Weliky, D. P. (2015) Closed and semiclosed interhelical structures in membrane vs closed and open structures in detergent for the influenza virus hemagglutinin fusion peptide and correlation of hydrophobic surface area with fusion catalysis, *Journal of the American Chemical Society* 137, 7548-7551.
71. Ranaweera, A., Ratnayake, P. U., and Weliky, D. P. (2018) The Stabilities of the Soluble Ectodomain and Fusion Peptide Hairpins of the Influenza Virus Hemagglutinin Subunit II Protein Are Positively Correlated with Membrane Fusion, *Biochemistry* 57, 5480-5493.
72. Qiao, H., Armstrong, R. T., Melikyan, G. B., Cohen, F. S., and White, J. M. (1999) A specific point mutant at position 1 of the influenza hemagglutinin fusion peptide displays a hemifusion phenotype, *Molecular biology of the cell* 10, 2759-2769.
73. Durrer, P., Galli, C., Hoenke, S., Corti, C., Glück, R., Vorherr, T., and Brunner, J. (1996) H<sup>+</sup>-induced membrane insertion of influenza virus hemagglutinin involves the HA2 amino-terminal fusion peptide but not the coiled coil region, *Journal of Biological Chemistry* 271, 13417-13421.
74. Apellaniz, B., Huarte, N., Largo, E., and Nieva, J. L. (2014) The three lives of viral fusion peptides., *Chem. Phys. Lipids*. 181, 40-55.
75. Kemble, G. W., Danieli, T., and White, J. M. (1994) Lipid-anchored influenza hemagglutinin promotes hemifusion, not complete fusion, *Cell* 76, 383-391.
76. Han, X., Bushweller, J. H., Cafiso, D. S., and Tamm, L. K. (2001) Membrane structure and fusion-triggering conformational change of the fusion domain from influenza hemagglutinin, *Nature Structural and Molecular Biology* 8, 715.

77. Lorieau, J. L., Louis, J. M., and Bax, A. (2010) The complete influenza hemagglutinin fusion domain adopts a tight helical hairpin arrangement at the lipid: water interface, *Proceedings of the National Academy of Sciences* 107, 11341-11346.
78. Gray, C., Tatulian, S. A., Wharton, S. A., and Tamm, L. K. (1996) Effect of the N-terminal glycine on the secondary structure, orientation, and interaction of the influenza hemagglutinin fusion peptide with lipid bilayers, *Biophysical journal* 70, 2275-2286.
79. Ishiguro, R., Kimura, N., and Takahashi, S. (1993) Orientation of fusion-active synthetic peptides in phospholipid bilayers: determination by Fourier transform infrared spectroscopy, *Biochemistry* 32, 9792-9797.
80. Lüneberg, J., Martin, I., Nüßler, F., Ruyschaert, J.-M., and Herrmann, A. (1995) Structure and topology of the influenza virus fusion peptide in lipid bilayers, *Journal of Biological Chemistry* 270, 27606-27614.
81. Macosko, J. C., Kim, C.-H., and Shin, Y.-K. (1997) The membrane topology of the fusion peptide region of influenza hemagglutinin determined by spin-labeling EPR1, *Journal of molecular biology* 267, 1139-1148.
82. Campelo, F., McMahon, H. T., and Kozlov, M. M. (2008) The hydrophobic insertion mechanism of membrane curvature generation by proteins, *Biophys. J.* 95, 2325-2339.
83. Brice, A. R., and Lazaridis, T. (2014) Structure and dynamics of a fusion peptide helical hairpin on the membrane surface: comparison of molecular simulations and nmr, *The Journal of Physical Chemistry B* 118, 4461-4470.
84. Victor, B. L., Lousa, D., Antunes, J. M., and Soares, C. u. M. (2015) Self-assembly molecular dynamics simulations shed light into the interaction of the influenza fusion peptide with a membrane bilayer, *Journal of chemical information and modeling* 55, 795-805.
85. Baylon, J. L., and Tajkhorshid, E. (2015) Capturing spontaneous membrane insertion of the influenza virus hemagglutinin fusion peptide, *The Journal of Physical Chemistry B* 119, 7882-7893.
86. Smrt, S. T., Draney, A. W., and Lorieau, J. L. (2015) The influenza hemagglutinin fusion domain is an amphipathic helical hairpin that functions by inducing membrane curvature *J. Biol. Chem.* 290, 228-238.
87. Tatulian, S. A., and Tamm, L. K. (2000) Secondary structure, orientation, oligomerization, and lipid interactions of the transmembrane domain of influenza hemagglutinin, *Biochemistry* 39, 496-507.
88. Melikyan, G. B., Lin, S., Roth, M. G., and Cohen, F. S. (1999) Amino acid sequence requirements of the transmembrane and cytoplasmic domains of influenza virus hemagglutinin for viable membrane fusion, *Molecular biology of the cell* 10, 1821-1836.

89. Benton, D. J., Nans, A., Calder, L. J., Turner, J., Neu, U., Lin, Y. P., Ketelaars, E., Kallewaard, N. L., Corti, D., and Lanzavecchia, A. (2018) Influenza hemagglutinin membrane anchor, *Proceedings of the National Academy of Sciences* 115, 10112-10117.
90. Markosyan, R. M., Cohen, F. S., and Melikyan, G. B. (2000) The lipid-anchored ectodomain of influenza virus hemagglutinin (GPI-HA) is capable of inducing nonenlarging fusion pores, *Molecular Biology of the Cell* 11, 1143-1152.
91. Nobusawa, E., Aoyama, T., Kato, H., Suzuki, Y., Tateno, Y., and Nakajima, K. (1991) Comparison of complete amino acid sequences and receptor-binding properties among 13 serotypes of hemagglutinins of influenza A viruses, *Virology* 182, 475-485.
92. Rameshwar U. Kadam, Jarek Juraszek, Boerries Brandenburg, Christophe Buyck, Wim B.G. Schepens, Bart Kesteleyn, Bart Stoops, Rob Vreeken, Jan Vermond, Wouter Goutier, Chan Tang, Ronald Vogels, Robert H. E. Friesen, Jaap Goudsmit, Dongen, M. J. P. v., and Wilson, I. A. (2017) Potent peptidic fusion inhibitors of influenza virus, *Science* 10, 1126.
93. X., S., Zhang, X., and Liu, S. (2013) Novel hemagglutinin-based influenza virus inhibitors, *Journal of Thoracic Disease* 5 (Suppl 2), S149-S159.
94. Hurt, A. C. (2014) The epidemiology and spread of drug resistant human influenza viruses, *Current opinion in virology* 8, 22-29.
95. Ward, P., Small, I., Smith, J., Suter, P., and Dutkowski, R. (2005) Oseltamivir (Tamiflu) and its potential for use in the event of an influenza pandemic, *J Antimicrob Chemother* 55(Suppl 1), i5-i21.
96. Hayden, F. G., and Hay, A. J. (1992) Emergence and transmission of influenza A viruses resistant to amantadine and rimantadine *Curr Top Microbiol Immunol* 176, 119-130.
97. Combrink KD , Gulgeze HB , Yu KL , Pearce BC , Trehan AK , Wei J , Deshpande M , Krystal M , Torri A , Luo G , Cianci C , Danetz S , Tiley L , and NA, M. (2000) Salicylamide inhibitors of influenza virus fusion *Bioorganic & Medicinal Chemistry Letters* 10, 1649-1652.
98. Russell, R. J., Kerry, P. S., Stevens, D. J., Steinhauer, D. A., Martin, S. R., Gamblin, S. J., and Skehel, J. J. (2008) 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* 105, 17736-17741.
99. Hoffman, L. R., Kuntz, I. D., and White, J. M. (1997) Structure-based identification of an inducer of the low-pH conformational change in the influenza virus hemagglutinin: Irreversible inhibition of infectivity, *Virology* 71, 8808-8820.
100. L., D., Bodian, R., Bryan Yamasaki, L., R., Buswell, Jay F. Stearns, Judith M. White, and Kuntz, I. D. (1993) Inhibition of the fusion-inducing conformational change of influenza hemagglutinin by benzoquinones and hydroquinones *Biochemistry* 32, 2967-2978.



101. Lee, P. S., and Wilson, I. A. (2015) Structural Characterization of Viral Epitopes Recognized by Broadly Cross-Reactive Antibodies *Current topics in microbiology and immunology* 386, 323-341.

## **Chapter 2**

### **Methodology**

## **2.1 Introduction**

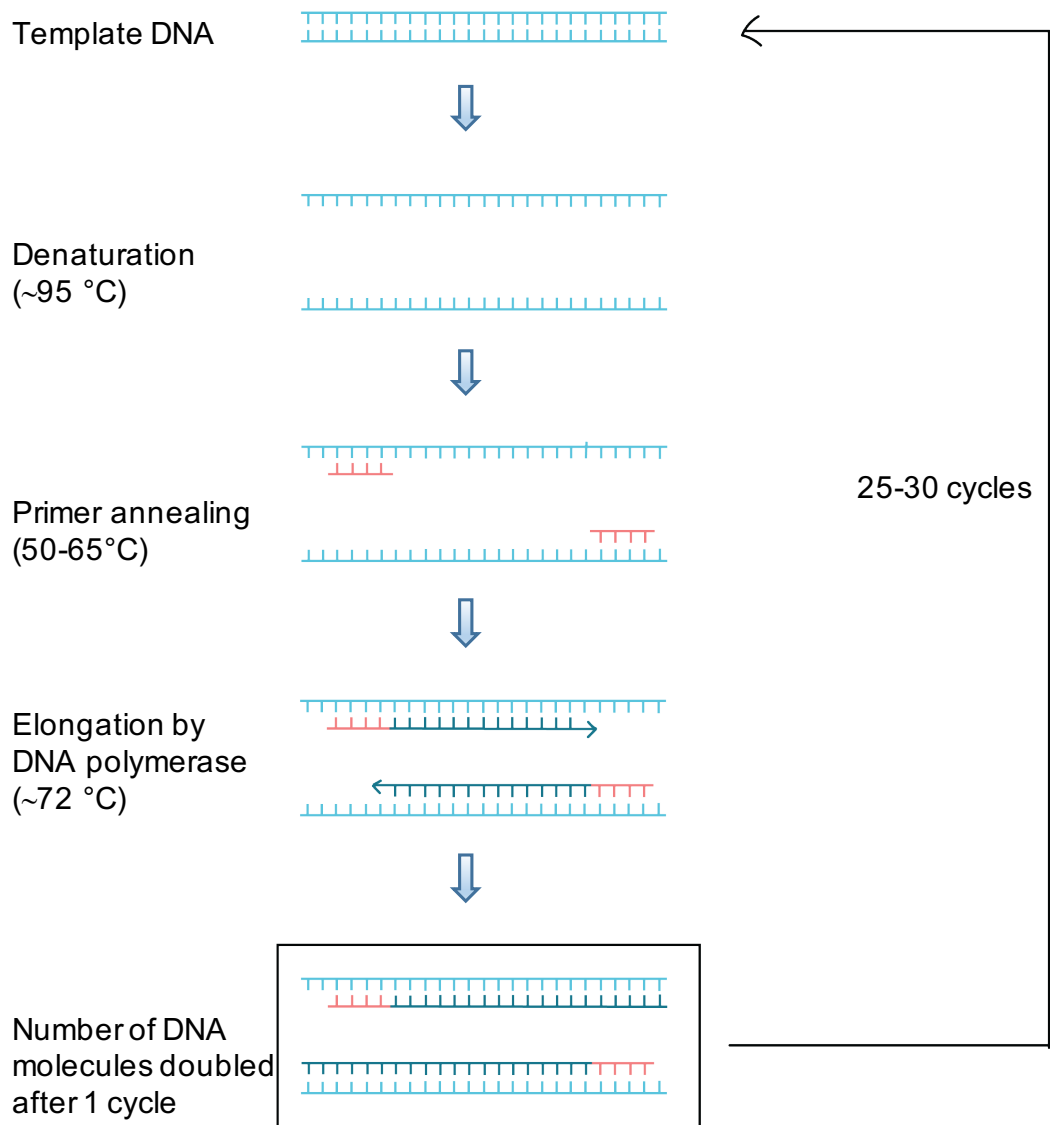
This chapter introduces the basic methods that were followed in the key research projects discussed in this dissertation (Chapter 3 and 4). This includes: the polymerase chain reaction that was used to introduce site specific mutations to proteins; and protein expression, solubilization, purification, identification, and structural and functional characterization. The materials and specific experimental procedures will be discussed in each chapter.

## **2.2 Polymerase chain reaction (PCR)**

Polymerase chain reaction (PCR) is a widely used technique to amplify a small amount of DNA, and also for the addition, deletion and mutation of a nucleotide sequence. There are several components of a PCR experiment including: the DNA template; DNA primers; a DNA amplifying enzyme (e.g. Taq polymerase); and free nucleotides. The template is the DNA molecule that is going to be amplified or modified. Primers are short nucleotide sequences that are complementary to the template, and they provide a starting point for the DNA synthesis. Two types of primers are used, the forward primer and the reverse primer, which are complementary to the 3' ends of the two strands in the DNA template.

PCR experiment is performed by employing a temperature program in an instrument known as the thermal cycler. First, the template DNA molecule is denatured by using a high temperature (~95 °C), and in this process the double stranded DNA molecule is converted to two single strands. The melting temperature of the DNA, which is the temperature at which the double stranded DNA is converted to two single strands, should be below the maximum temperature operated in PCR. The primers then bind to the complementary bases on single DNA strands, and this process is known as annealing. Then the DNA polymerase enzyme adds the complementary

nucleotides to the newly formed DNA strands via a process called elongation<sup>l</sup> (Fig. 2.1). The temperature of the elongation step is based on the optimum temperature of the DNA polymerase used. A temperature of 72 °C is typically used with Taq polymerase.



**Figure 2.1** The basic steps of PCR.

***Primer designing*** - The primers used in PCR should be carefully designed, because the final product of the PCR is mainly based on the primers. Several factors should be considered in primer designing, which include: the length of the primer; melting temperature; specificity, and GC (Guanine/Cytosine) content. Longer primers cause inefficient annealing and shorter primers result in nonspecific binding. The primer length is typically in between 20 - 30 bases in lengths. The annealing temperature of the primer is important because lower annealing temperature causes nonspecific binding, whereas higher annealing temperature causes inefficient primer-template binding. The annealing temperature used is usually 3-5 °C lower than the melting temperature of the primer, which is typically in between 55-65 °C. Generally, the GC content should be 50-60%, which contributes significantly to the primer melting temperature, as G and C form three hydrogen bonding compared to Adenine (A) and Thymine (T), which form only two hydrogen bonding in between when annealing.

***DPN1 digestion (removal of unreacted template DNA)*** - After the PCR experiment there can be a small portion of non-reacted template DNA molecules together with the newly formed DNA strands in the reaction mixture. Although this concentration may be low, it can affect in later steps, therefore getting rid of the template DNA is important. This is done by an enzymatic digestion involving DPN1 restriction endonuclease enzyme, which specifically digest the nucleotide sequence at GATC where adenine is methylated at N6 position. Usually the methylation is done by an enzyme (DAM enzyme) in bacteria during replication, however this is not present when PCR is performed. During DPN1 digestion the template is digested, leaving only the required PCR product.

***DNA gel electrophoresis*** - The DNA gel electrophoresis is used to verify whether the desired PCR product has been obtained. In the gel electrophoresis, DNA molecules are separated based on their size by running the PCR product through a DNA gel (Agarose gel) in the presence of an applied electric current. The DNA backbone phosphates are negatively charged and therefore the DNA fragments will move towards the positively charged electrode. Usually, a DNA standard (DNA ladder) is run together with the DNA sample to identify the size of the PCR product. Both absence of bands and unexpected bands suggest errors in the PCR experiment that should be corrected before moving into next steps.

***Transformation*** - The introduction of foreign DNA into living cells is known as the transformation. Once the desired DNA is obtained from PCR, they should be transformed to living cells to get the sequence expressed in large scale. *E. Coli* cells are commonly used for this purpose and their cell wall is modified so that foreign DNA can penetrate easily. These modified cells are known as competent cells, and for the present experiments *E. Coli* BL21(DE3) strain is used. Transformation is a multistep procedure that include: thawing the competent cells that are stored in the -80 °C freezer (typically 50 µL in an Eppendorf tube) on ice; gently mixing the PCR product (~5 µL) with the competent cells; incubation of the mixture on ice for 20-30 min; applying a “heat shock” by placing the mixture on a 42 °C water bath for ~45 s (it is believed that the heat shock aids the cell to take up a DNA plasmid); incubation on ice for another 2 min; combining the mixture with a growth medium such as Luria-Bertani (LB) (450 µL); and incubation the resulting solution at 37 °C and 225 rpm for 1h. Then the transformed cells are streaked on an agar plate containing appropriate antibiotics, and they are incubated overnight. The isolated colonies can be further grown in liquid media (e.g. LB) for larger scale studies.

**DNA sequencing** - Before the protein expression (discussed in section 2.3), it is important to confirm whether the expected DNA sequence is present in the living cells. For this purpose, a DNA sequencing experiment can be performed. This can be done using a Promega DNA extraction kit and the Sanger sequencing technique.<sup>2</sup>

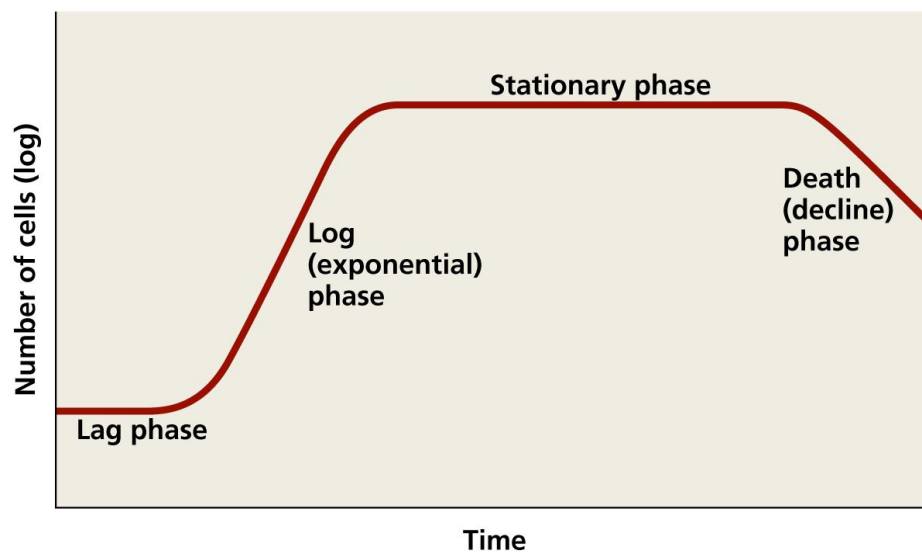
## 2.3 Protein Expression

The process of protein synthesizing and regulating in living organisms is known as protein expression. Once the protein is biosynthesized, it is modified in several ways to complete its final structure, to relocate the protein and to regulate its function. Some of these modifications are: formation of intramolecular interactions like disulfide bonds between cysteine residues; and modification of functional groups present in the protein or introducing new functional groups, e.g. phosphorylation, glycosylation.

In protein research, the techniques used for the manufacturing of the protein of interest is known as the protein expression or recombinant protein expression. Bacterial protein expression systems are widely used for recombinant protein expression because bacteria are easy to grow and result in high yields. After the PCR experiment, the plasmid containing the DNA sequence that encodes the desired protein is transformed into *E. coli* cells for protein synthesis. The cells are grown in a culture medium (typically LB), and the cell growth is estimated by measuring the optical density at 600 nm (OD<sub>600</sub>).

Typical bacterial growth curve, including *E. coli*, has four phases (Fig. 2.2): (1) the lag phase with limited number of cells; (2) the log phase in which the cells have a rapid growth rate due to the nutrients present in the culture medium; (3) the stationary phase in which the nutrients become limited and metabolic byproducts accumulated; and (4) the death phase where the cell

death occurs due to the accumulated byproducts such as ethanol and lactate. The protein expression is typically induced at the log phase ( $OD_{600} = 0.5-0.8$ ) by adding an inducer, e.g. isopropyl  $\beta$ -D-thiogalactopyranoside (IPTG). However, the optimal  $OD_{600}$  highly depends on the protein being expressed, expression system and media used, and therefore needs to be optimized.<sup>3, 4</sup>



**Figure 2.2** Typical bacterial growth curve

When recombinant proteins are expressed in high level in *E. coli*, highly aggregated proteins are formed. These are referred to as inclusion bodies (discussed in section 2.4.1).



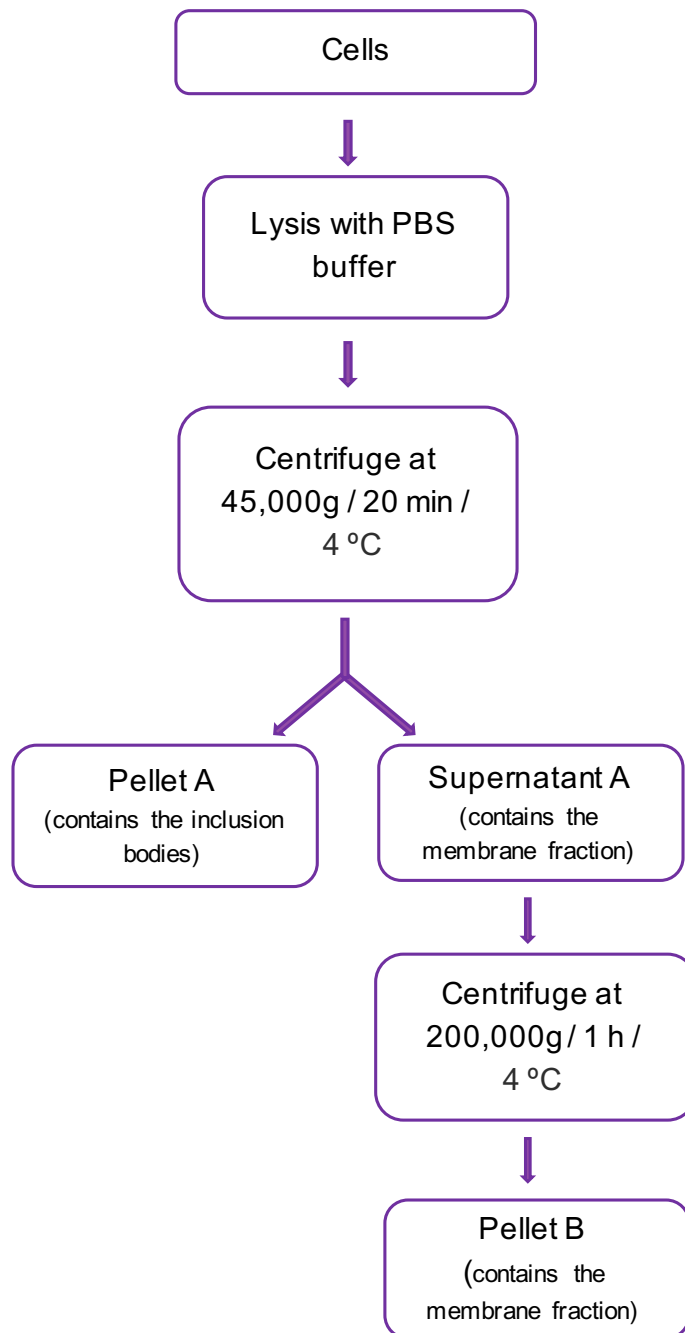
## **2.4 Solubilization and purification of the expressed protein**

### **2.4.1 Protein solubilization**

When solubilizing a protein, the location of the desired protein in the *E. Coli* cell should be taken into account. The membrane proteins can be in the cell membrane and/or in the inclusion bodies. In general, when the protein is in the membrane, it can be purified in the folded form of the protein. On the other hand, in the inclusion bodies the protein is usually unfolded, and an additional refolding step is required. However, there are some advantages of the inclusion body purification which include: increased protein yield as ~15-25% of the total cell mass are inclusion bodies; and easy isolation of the protein as the density of the inclusion bodies are higher than most of the other cellular components.<sup>5</sup>

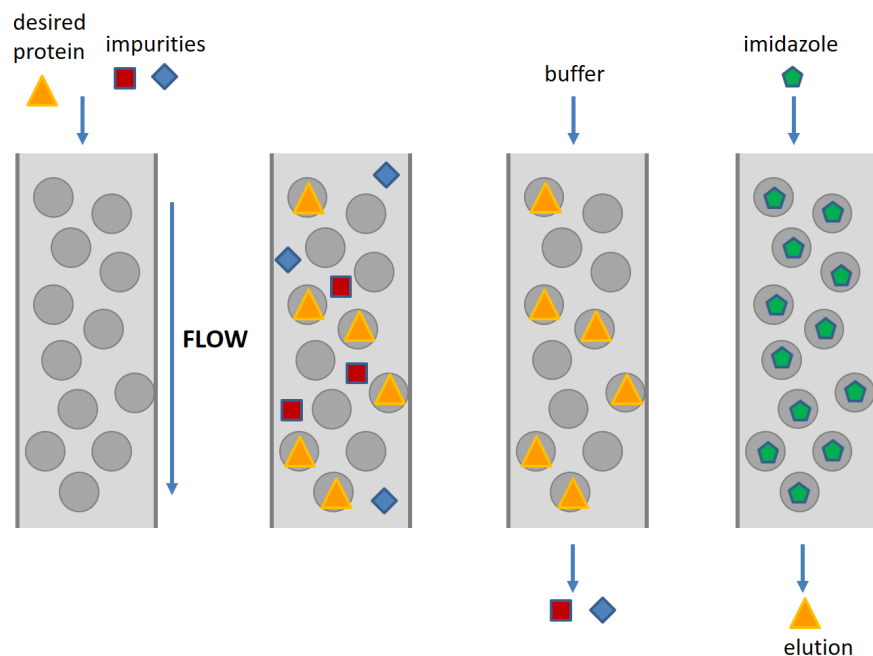
When the cells containing the desired protein are lysed with PBS buffer (Phosphate-Buffered Saline), both the membrane fraction and the inclusion bodies make a suspension. Modest centrifugation (45,000g) of this suspension gives a pellet (pellet A) which contains mostly the inclusion bodies and other cell debris, and a supernatant (supernatant A) which contains soluble proteins, other soluble molecules and the suspended membrane fraction. The pellet A is used for the inclusion body purification. In membrane fraction purification, the supernatant A is recentrifuged at >200,000g, and the resultant transparent pellet (pellet B) is used (Fig. 2.3).

The desired protein can be in the inclusion bodies (pellet A) and/or in the membrane fraction (pellet B). Detergents (e.g. Sarkosyl, Decyl maltoside) and/or denaturing agents (e.g. Urea, Guanidium hydrochloride) are often used to solubilize the membrane protein present in this pellet.<sup>3, 6, 7</sup> Once the protein is solubilized, the affinity chromatography can be used to purify the protein.



**Figure 2.3** Initial steps of HA2 purification by inclusion bodies and membrane fraction.

### 2.4.2 Affinity chromatography



**Figure 2.4** Schematic representation of the affinity purification (picture credit goes to Dr. Shuang Liang).

Affinity chromatography is a widely used method to purify proteins from its crude cell lysate. The cell lysate is a complex mixture of molecules that includes the desired protein and other components of the protein expression cell. For example, if the protein is expressed in the bacterium *E. coli*, the cellular components of the *E. coli*, including its proteins, can be in the cell lysate. The affinity chromatographic method purifies the desired protein based on its specific binding properties to a ligand (e.g.  $\text{Co}^{2+}$ ,  $\text{Ni}^{2+}$ ), typically immobilized on a resin. The protein of interest is genetically modified (typically with a multi-histidine tag) prior to the expression, to increase its binding affinity to the resin. The resin is placed in a column, and when the cell lysate solution is passed through the column, the modified protein binds to the resin with greater affinity than other molecules. A wash buffer is then poured through the column which removes the weakly bound

molecules from the resin. Then an elution buffer, which can disrupt the interaction between the protein and the resin, is poured through the column, and the recombinant protein is eluted (Fig 2.4). The elution buffer typically contains a high concentration of a molecule that has a strong binding affinity with the resin (e.g. imidazole).

## **2.5 Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE)**

Polyacrylamide gel electrophoresis (PAGE) is an analytical method used to separate macromolecules such as proteins and nucleic acids based on their size. The technique uses the principle of migration of a charged molecule towards an electrode with an opposite sign in an electric field. The mobility of a substance in the gel depends on both charge and size of the macromolecule. Therefore, the sample needs to be treated so that it acquires uniform charge, then the separation depends primarily on size. As different shapes of the molecule can also affect the mobility, a denaturant is added to remove the protein secondary, tertiary and quaternary structure. The most common denaturant used for protein separation is sodium dodecyl sulfate (SDS), then the electrophoresis is known as SDS-PAGE. In addition to SDS, the protein is sometimes heated to near boiling with a reducing agent such as  $\beta$ -mercaptoethanol to further denature by reducing disulfide linkages, and this is known as reducing SDS-PAGE.

SDS is an anionic detergent, which can bind to the amino acids of proteins giving the protein a uniform negative charge independent of the original charge of the protein. For soluble proteins it binds in a fairly fixed ratio, ~one molecule of SDS per two amino acids,<sup>8</sup> and for membrane proteins it has been found that this number has a high variability depending on the hydrophobicity of the protein.<sup>9, 10</sup> In an electric field, the negatively charged polypeptide chains migrate towards the positively charged electrode (anode) with different mobility. The migration

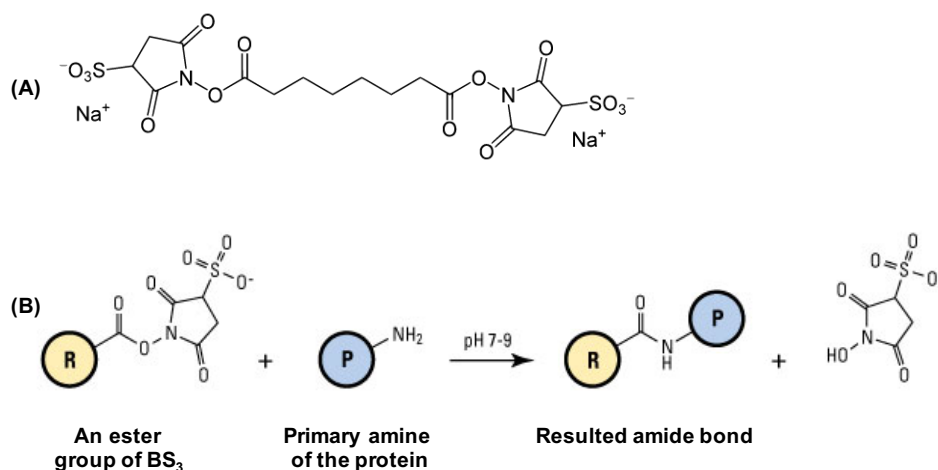
rate of a protein in the SDS gel decreases with increasing the mass of SDS-bound protein (protein mass + mass of the SDS molecules bound to the protein). Therefore, for two proteins with similar molecular weight, the migration rate in the gel can vary depending on the hydrophobicity of the protein<sup>9</sup>. When the number of SDS molecules bound to a protein increases, it also increases the overall charge of the SDS-bound protein, which may increase the migration rate. However, the observation of decreasing the migration rate with bound SDS implies that the effect of increase in mass is larger than the increase in charge.

The gels are typically made of acrylamide and bisacrylamide. Acrylamide forms linear polymers and bisacrylamide cross-links these linear polymers. The pore size of the gel is determined by the ratio of acrylamide to bisacrylamide molecules. The acrylamide concentration of the gel can be varied, typically in the range of 5-25%, depending on the protein size analyzed. The lower percentage gels form large pores and are used to resolve molecules with higher molecular weights, while molecules with lower molecular weights are better resolved with higher percentage gels that form small pores.

## **2.6 Cross-linking experiment**

When running a native SDS-PAGE, the protein is denatured due to the SDS and heat. Upon denaturation, the non-covalent interactions among the protein molecules in an oligomer are disrupted, and the protein loses its oligomerization. Therefore, with native SDS-PAGE we determine the monomer mass of the protein. To determine the mass of the oligomeric state of the protein, first the protein molecules within the oligomer should be connected to each other via covalent bonds, which cannot be disrupted during SDS-PAGE. Cross-linking experiment is used for this purpose. Here, a cross-linking agent, a molecule that has the ability to form two covalent

bonds per molecule, is used to attach the protein molecules in the oligomer to each other. Commonly used cross-linking agent bis (sulfosuccinimidyl) suberate (BS<sub>3</sub>) has two ester groups at both ends (Fig 2.5A). These ester groups can form stable amide bonds with primary amines present in the protein, e.g. the side chain of lysine and the N-terminus of the protein. The SDS-PAGE after the cross-linking can be used to determine the oligomeric state of the protein.

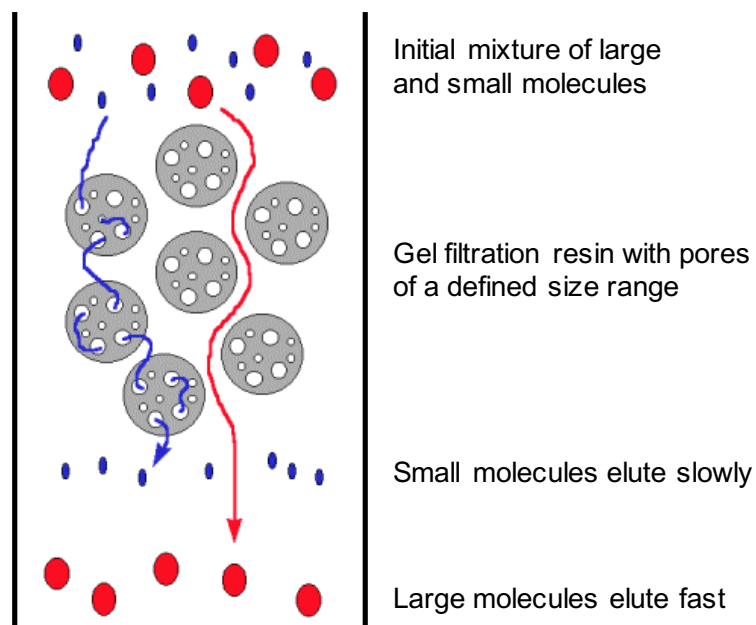


**Figure 2.5** (A) The structure of bis (sulfosuccinimidyl) suberate (BS<sub>3</sub>); (B) Schematic representation of the reaction of an ester group of BS<sub>3</sub> with a primary amine group of the protein. R represents one end of the cross-linker, and P represent the protein.

## 2.7 Size Exclusion Chromatography (SEC)

Size exclusion chromatography (SEC) is a chromatographic technique used to separate molecules such as proteins, peptides and oligonucleotides based on their size. In SEC, if the mobile phase is an organic solvent, the technique is known as the gel permeation chromatography, and if the mobile phase is an aqueous solvent, it is called the gel filtration. Gel filtration is mainly applied in fractionating proteins. The separation takes place in a column, which is a hollow tube, tightly

packed with porous polymer beads with pores of different sizes. The polymer material can be vary depending on the pore size required. The matrix of Superdex-200 column used in present studies is made of a mixture of dextran and agarose, and can be used to separate molecules in the range of 6.5 kDa - 1.3 MDa. When the proteins of different size move through the porous matrix, they diffuse into the pores to greater or lesser extents. Smaller molecules diffuse further into the pores, and therefore move slowly, while larger molecules enter into the pores to a lesser extent or not at all, thus move through the column quickly. Therefore, the largest proteins elute first and the smallest elute last (Fig 2.6).



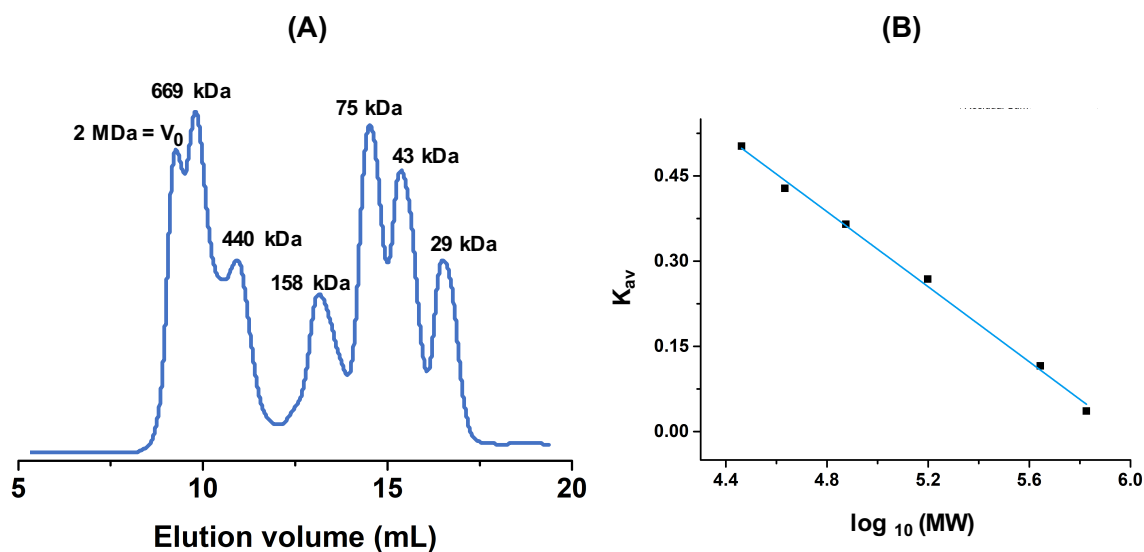
**Figure 2.6** Size-based fractionating of molecules in a gel filtration column.

The instrument is calibrated with well-defined protein standards with low and high molecular weights, typically ranging from 6.5-75 kDa and 43-669 kDa. The partition coefficient

( $K_{av}$ ) can be calculated using the following equation and is inversely proportional to the logarithm of the molecular weight of the protein.

$$K_{av} = \frac{(V_e - V_0)}{(V_c - V_0)}$$

The void volume ( $V_0$ ) is the elution volume of a molecule that does not enter into the pores of the matrix, and can be determined from a protein with a very high molecular weight (e.g. blue dextran - 2 MDa). The column volume ( $V_c$ ) is the volume of the packed gel filtration matrix and is specific to the column used.  $V_0 = 9.45$  mL and  $V_c = 25$  mL in the Superdex-200 column used in our lab. The elution volume ( $V_e$ ) depends on the molecular weight/size of the protein. More detailed SEC calibration with buffers containing either NaCl or NaF are discussed in chapter 6.



**Figure 2.7** (A) Size exclusion chromatogram for the  $A_{280}$  detection of MW standards loaded on Superdex-200 column and (B) the SEC calibration curve.



## 2.8 Circular Dichroism (CD) spectroscopy

Circular dichroism (CD) is a form of light absorption spectroscopy, and is commonly used in the analysis of secondary structure of proteins. Commonly used spectroscopic techniques measure the absorbance of isotropic light, but CD spectroscopy measures the difference in absorbance of the left and right circularly polarized light by a chiral molecule. As proteins are chiral molecules, CD can be used to find the secondary structure of protein molecules. Different secondary structures of proteins, i.e.  $\alpha$ -helix,  $\beta$ -sheet or random coil, give characteristics shapes in their CD spectra, in the region of 190 nm- 260 nm.

The differential absorption of left and right circularly polarized light absorbed by a chiral molecule can be expressed as,

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

$A_{LCP}$  - absorption of left circularly polarized light

$A_{RCP}$  - absorption of right circularly polarized light

$\Delta A$  - differential absorption

According to Beer-Lambert's law,

$$A = \epsilon Cl$$

$$\Delta A = \Delta \epsilon Cl$$

$$\Delta \epsilon = \epsilon_{LCP} - \epsilon_{RCP} = \frac{\Delta A}{C \times l}$$

$C$  - molar concentration

$L$  - path length in centimeters

$\epsilon$  - molar extinction coefficient

$\epsilon_{LCP}$  - molar extinction coefficients of left circularly polarized light

$\epsilon_{RCP}$  - molar extinction coefficients of right circularly polarized light

Then mean residue concentration ( $C_{MR}$ ) can be expressed as,

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

$$\Delta\epsilon_{MR} = \frac{\Delta A}{C_{MR} \times l}$$

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

N - number of amino acids in the protein

Differential absorption of the left and right circularly polarized light is detected by the instrument using photomultiplier and converted into ellipticity  $\theta$ , which has units of millidegrees. Ellipticity can be related to differential absorption by;

$$\Delta A = \frac{\theta}{32982}$$

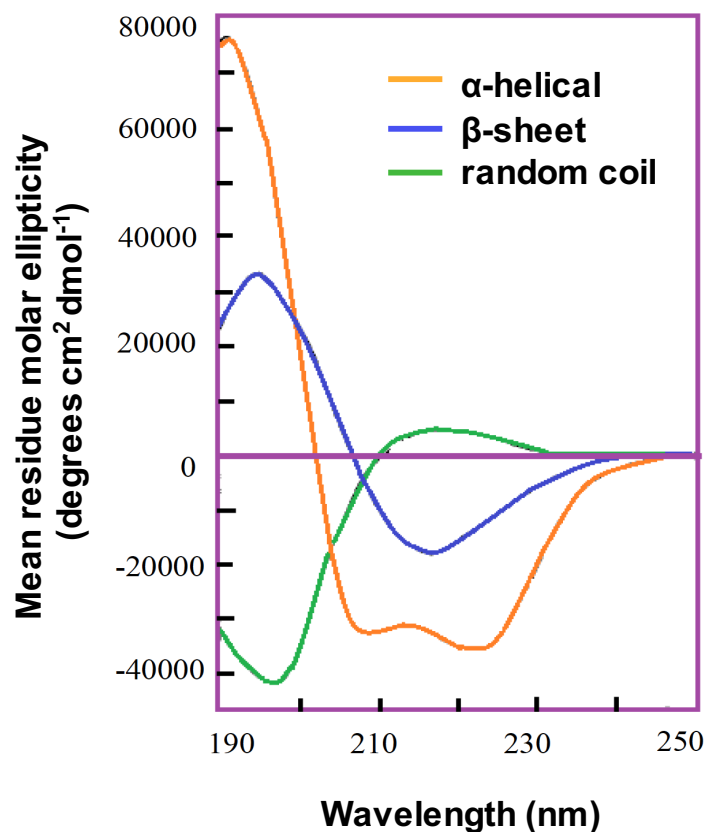
$$\theta_{MR} = \Delta A \times 32982$$

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

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

### 2.8.1 Protein secondary structure analysis

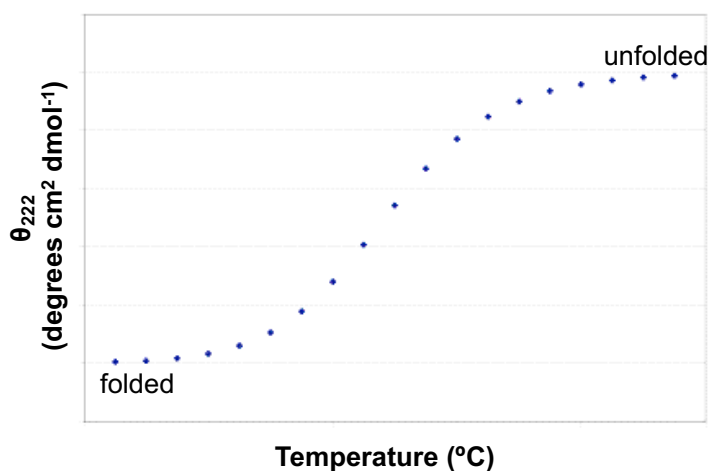
Mean residue molar ellipticity ( $\theta_{MR}$ ) of -33,000 degrees  $\text{cm}^2 \text{dmol}^{-1}$  at 222 nm wavelength is considered as the 100%  $\alpha$ -helicity. This has been determined by scanning different reference structures consisting of poly-L-lysine with completely  $\alpha$ -helical, random coil and  $\beta$ -sheet secondary structures over the wavelength range 190 - 260 nm. At 222 nm wavelength,  $\alpha$ -helix has the maximum value for  $\theta_{MR}$  (Fig. 2.8).<sup>11</sup>



**Figure 2.8** Characteristics shapes of CD spectra for different protein secondary structures.<sup>11</sup>

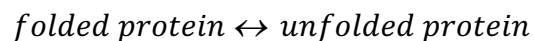
### 2.8.2 Protein thermostability analysis

Protein melting curves can be obtained by plotting the mean residue molar ellipticity value at 222 nm ( $\theta_{222}$ ) against temperature. If the protein is completely unfolded, the melting curve looks like in Fig 2.9.



**Figure 2.9** A typical protein melting curve.

Van't Hoff analysis of the  $\theta_{222}$  vs. temperature data for unfolded proteins resulted in linear plots of  $\ln K_{eq}(T)$  vs.  $1/T$  for the folded  $\leftrightarrow$  unfolded protein equilibrium. The enthalpy of melting ( $\Delta H_m$ ) and entropy of melting ( $\Delta S_m$ ) can be calculated from the slope and intercept respectively.



$$\Delta G = -RT \ln K_{eq}$$

$$\Delta G = \Delta H - T\Delta S$$

$K_{eq}$  can be obtained experimentally by the fraction of the protein molecules that are folded and unfolded.

$$K_{eq} = \frac{f_{unfold}}{f_{fold}}$$

$$K_{eq} = \frac{f_{unfold}}{(1 - f_{unfold})}$$

The observed  $\theta_{222}$  at each temperature is a linear combination of observed  $\theta_{222}$  for this sample in the fully folded state and  $\theta_{222}$  in the fully unfolded state.

$$\theta_{222_{observed}} = f_{unfold} \cdot \theta_{222_{unfold}} + f_{fold} \cdot \theta_{222_{fold}}$$

$$f_{unfold} = \frac{\{\theta_{222_{observed}} - \theta_{222_{fold}}\}}{\{\theta_{222_{unfold}} - \theta_{222_{fold}}\}}$$

Van't Hoff equation is an equation of a straight line;

$$\ln K_{eq} = - \left( \frac{\Delta H}{RT} \right) + \left( \frac{\Delta S}{R} \right)$$

At  $T_m$ ,  $\Delta G = 0$ , and therefore,

$$\Delta H = T_m \Delta S$$

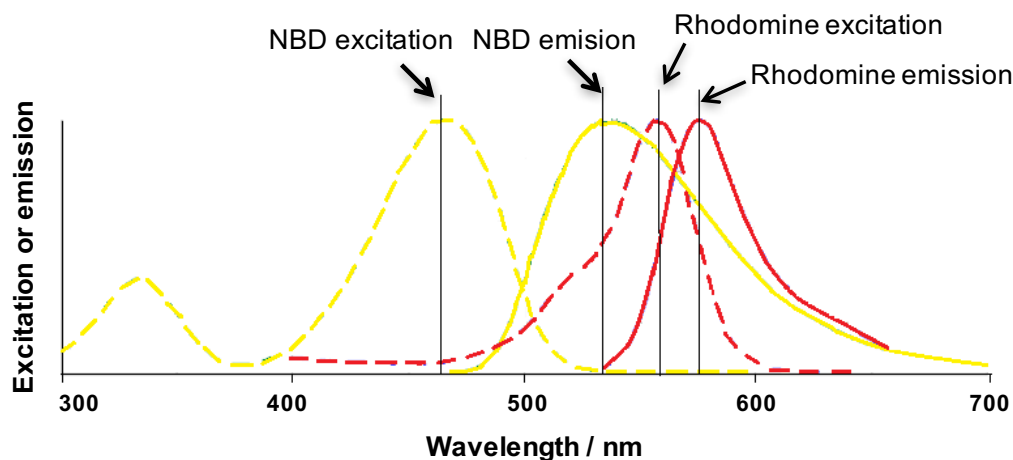
$$T_m = \frac{\Delta H}{\Delta S}$$

$\Delta H$  and  $\Delta S$  can be respectively calculated from the slope and intercept of the Van't Hoff plot.

## 2.9 Lipid mixing assay

Lipid mixing and exchange occur between cellular membranes during fusion. Fluorescence quenching is the most common technique used to assess *in vitro* lipid mixing between membranes such as vesicles. This assay can be used to compare the fusogenic activity of proteins as discussed below.

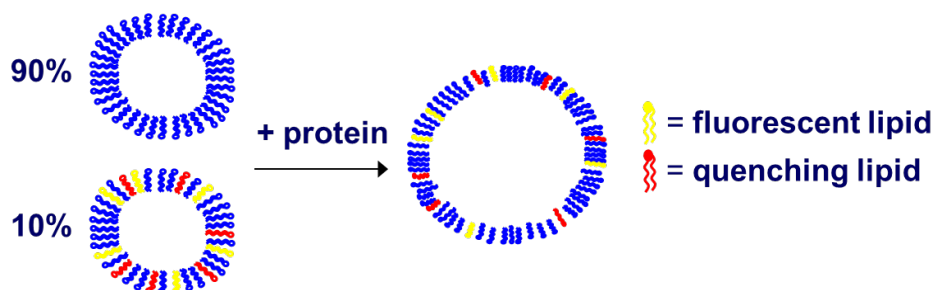
This experiment determines the fusion between a labelled and an unlabeled vesicle in the presence of an exogenously added protein. The labelled vesicle contains two types of fluorescent dyes: the donor (fluorophore) and the acceptor (quencher). The donor-acceptor pair is selected such that the emission wavelength of the donor is comparable with the excitation wavelength of the acceptor. NBD-PE {*N*-(7-nitro-2,1,3-benzoxadiazol-4-yl) (ammonium salt) dipalmitoylphosphatidylethanolamine} and Rhodamine {*N*-(lissamine rhodamine B sulfonyl) (ammonium salt) dipalmitoylphosphatidyl- ethanolamine} is a commonly used donor-acceptor pair (Fig. 2.10).<sup>12</sup> Head group of the phosphatidylethanolamine lipid can be modified to add NBD fluorophore and Rhodamine quencher in lipid mixing assays.



**Figure 2.10** The principle of fluorescence quenching assay.

Before the vesicle fusion, the donor and the acceptor molecules are in a close proximity, and the emission signal of the donor can be effectively absorbed by the acceptor. When the vesicle fusion happens in the presence of a protein, it will increase the donor-acceptor distance (Fig. 2.11)

and the acceptor will not effectively absorb the donor signal. This will lead to an increased fluorescent signal which is measured in this assay. The efficiency of the energy transfer by this process is inversely proportional to the sixth power of the distance between the donor and the acceptor.



**Figure 2.11** Labeled vesicles containing both fluorescent and quenching lipids are mixed with unlabeled vesicles which results in fusion associated dilution of the labelled lipids.

## 2.10 Liquid Chromatography-Mass Spectrometry (LC-MS)

Liquid chromatography (LC) is an analytical technique used to separate the components of a liquid sample. This involves a stationary phase (solid or liquid) and a liquid mobile phase, which carry the sample components. Reverse phase partition chromatography is a widely used LC technique, which employs a hydrophobic stationary phase, typically a matrix of long-chain alkyl groups (e.g. C<sub>18</sub>) attached to silica particles, and a polar mobile phase, which is a mixture of water and polar solvents (e.g. acetonitrile, methanol). When the mobile phase containing the sample flows through the stationary phase, the sample components are partitioned between the two phases depending on their chemical affinity. Components having stronger interactions with the stationary phase migrate slower than the components having weaker interactions. The separation is based on the differences in the migration rates.

Mass spectrometry (MS) is an analytical technique used to determine the mass of analytes (peptides or other chemical components), their isotopic distribution, and in some cases to determine their chemical structure.<sup>13</sup> MS involves the ionization of sample components into gaseous ions, with or without fragmentation. Then the resulting ions are categorized based on their mass-to-charge ( $m/z$ ) ratio and relative abundance. The instrument contains three major components: (1) the ion source produces the gaseous ions from the analyte; (2) the analyzer resolves the ions according to their  $m/z$  ratio; and (3) the detector detects and records the resolved-ions.

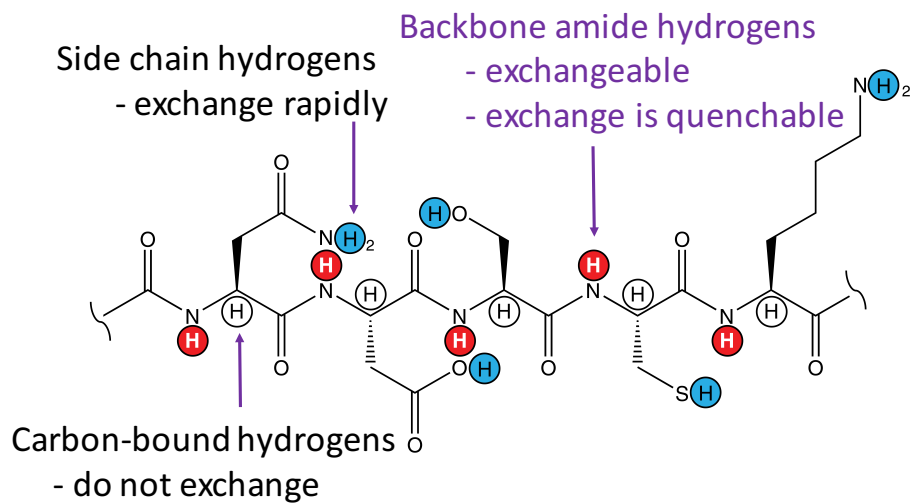
When the LC is coupled with the MS (LC-MS), it becomes a powerful analytical technique which use the resolving power of the LC and the detection specificity of the MS.

## **2.11 Hydrogen-Deuterium Exchange Mass Spectrometry**

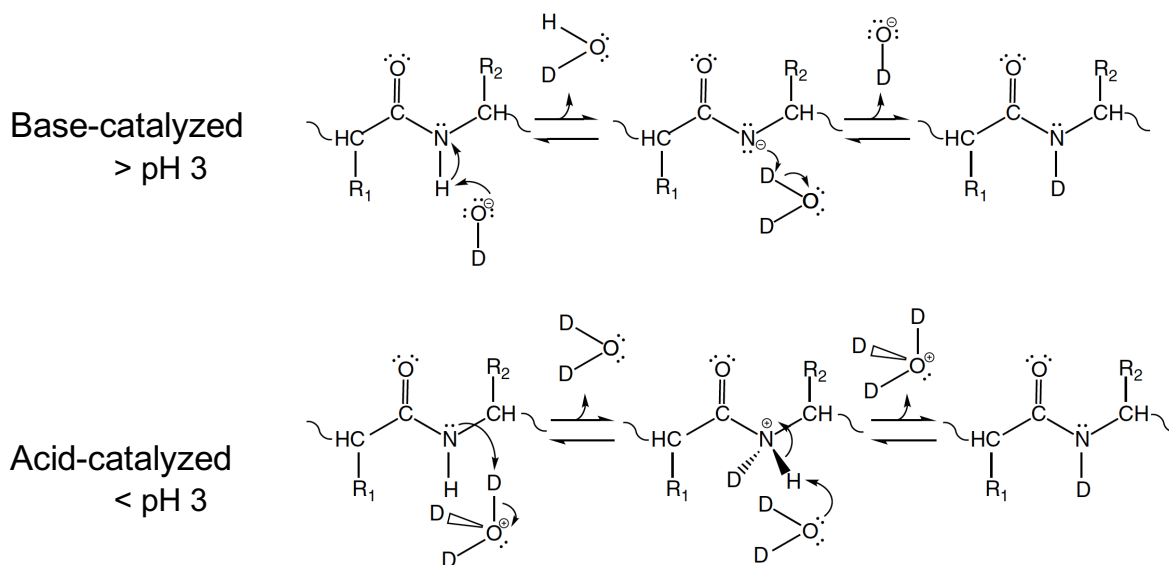
Hydrogen Deuterium Exchange - Mass Spectrometry (HDX-MS) provides details on protein conformation and dynamics, primarily the H-bonding and solvent accessibility, by monitoring the hydrogen exchange at protein backbone amides (Fig 2.12). Amide HDX can be acid-catalyzed or base-catalyzed depending on the pH of the medium (Fig. 2.13).

The proton abstraction by solvent  $\text{OH}^-/\text{OD}^-$  ions is the rate limiting step of amide HDX (base-catalyzed).<sup>14</sup> The rate of HDX is therefore depends on whether the amide hydrogen participates in an intramolecular H-bonding and on its solvent accessibility.<sup>15, 16</sup> Thus, the protein structure is the key governing factor of HDX of a native protein. HDX also depends on several physical parameters that needs to be controlled during the HDX experiment (discussed below).



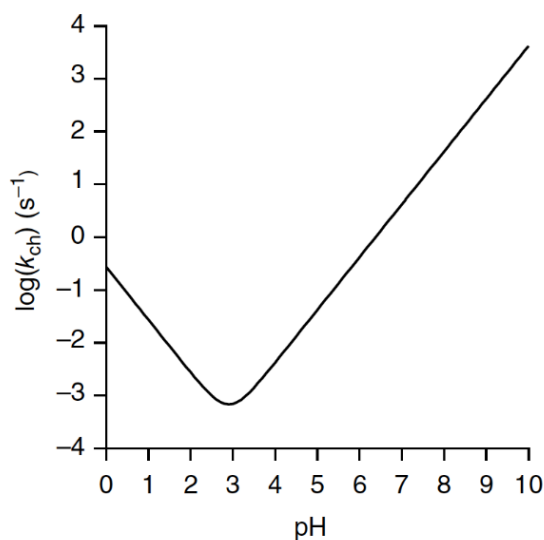


**Figure 2.12** Different types of hydrogen atoms in a protein.<sup>14</sup>



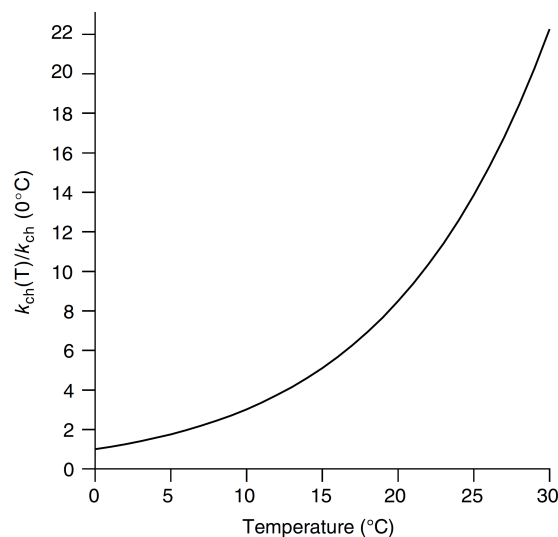
**Figure 2.13** Mechanism of HDX.<sup>14</sup>

**pH** - pH has a massive effect on the chemical HDX rate constant ( $k_{\text{ch}}$ ), and pH controlling is essential to effectively quench the HDX reaction. NMR studies using test peptides support that the HDX of amide hydrogens is minimum at pH 2.0-3.0 (Fig. 2.14). The acid-catalyzed HDX predominates at pH < 2.0, whereas base catalyzed reaction predominates pH > 3.0. For proteins, typically pH > 4 and the mechanism is base-catalyzed.



**Figure 2.14** The pH effect of amide HX.<sup>17</sup>

**Temperature** - The  $k_{\text{ch}}$  follows Arrhenius equation, and the HDX of proteins has an exponential relationship with temperature (Fig. 2.15). This is primarily because of the dependence of water ionization constant on temperature, thus affecting the  $[\text{OH}^-]$ , and also to some extent due to the increase of collision rate constant with temperature.<sup>14</sup>

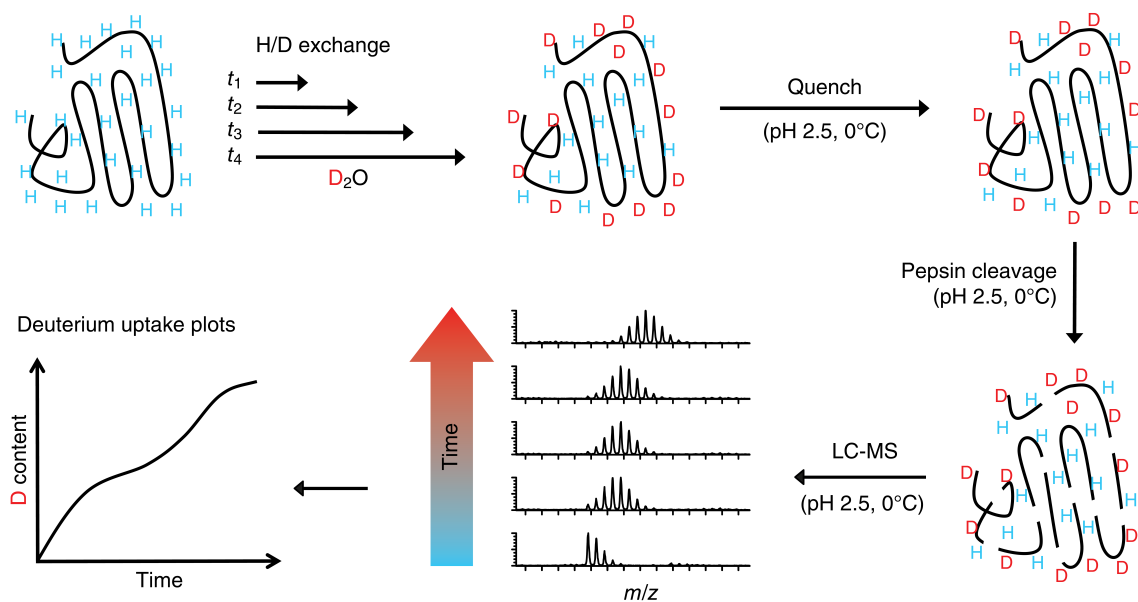


**Figure 2.15** Temperature effect for base catalyzed HX.<sup>14</sup>

**Time** - During HDX, the protein sample is diluted to a large extent (>15-fold) with the D<sub>2</sub>O containing buffer, and therefore under this condition the HDX reaction is unidirectional (minimal back exchange). Therefore, when increasing the reaction time, the number of amide hydrogens replaced by deuterium will increase until all the exchangeable amide hydrogens in the polypeptide are exchanged depending on the deuterium percentage in the reaction medium. For example, if the deuterium percentage in the reaction medium is 95% and the peptide has 20 exchangeable amide hydrogens, the long-time deuteration will be 19 amide hydrogens. The reaction time series is selected such that it covers both the fast ( $\leq 1$  min) and slow (10-100 min) exchange reactions.

A classical HDX-MS experiment is shown in Fig. 2.16, and will be discussed in details in chapter 4. Initially a concentrated protein solution in an aqueous buffer is diluted into a D<sub>2</sub>O containing buffer, which leads to exchange of backbone N-H for N-D. Then the HDX reaction is quenched by reducing the pH and the temperature of the reaction mixture (pH  $\approx$  2.5, 0 °C). The

deuterated protein is then subjected to pepsin digestion and the resulting peptides are introduced to the LC-MS analysis. The deuterium uptake of a peptide is positively-correlated with the solvent exposure, structural disorder, and motion of the corresponding segment of the intact protein.



**Figure 2.16** The classical HDX experiment.<sup>14</sup>

## REFERENCES

## REFERENCES

1. Arnheim, N., and Erlich, H. (1992) Polymerase chain reaction strategy, *Annual review of biochemistry* 61, 131-156.
2. Sanger, F., and Coulson, A. R. (1975) A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase, *Journal of molecular biology* 94, 441-448.
3. Curtis-Fisk, J., Spencer, R. M., and Weliky, D. P. (2008) Isotopically labeled expression in E. coli, purification, and refolding of the full ectodomain of the influenza virus membrane fusion protein, *Protein expression and purification* 61, 212-219.
4. Collins, T., Azevedo-Silva, J., da Costa, A., Branca, F., Machado, R., and Casal, M. (2013) Batch production of a silk-elastin-like protein in E. coli BL21 (DE3): key parameters for optimisation, *Microbial cell factories* 12, 21.
5. Singh, S. M., and Panda, A. K. (2005) Solubilization and refolding of bacterial inclusion body proteins, *Journal of bioscience and bioengineering* 99, 303-310.
6. Vogel, E. P., Curtis-Fisk, J., Young, K. M., and Weliky, D. P. (2011) Solid-state nuclear magnetic resonance (NMR) spectroscopy of human immunodeficiency 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* 50, 10013-10026.
7. Banerjee, K., and Weliky, D. P. (2014) Folded monomers and hexamers of the ectodomain of the HIV gp41 membrane fusion protein: potential roles in fusion and synergy between the fusion peptide, hairpin, and membrane-proximal external region, *Biochemistry* 53, 7184-7198.
8. Armstrong, R. T., Kushnir, A. S., and White, J. M. (2000) The transmembrane domain of influenza hemagglutinin exhibits a stringent length requirement to support the hemifusion to fusion transition, *The Journal of cell biology* 151, 425-438.
9. Liang, S., Ratnayake, P., Keinath, C., Jia, L., Wolfe, R., Ranaweera, A., and Weliky, D. P. (2018) Efficient Fusion at Neutral pH by Human Immunodeficiency Virus gp41 Trimers containing the Fusion Peptide and Transmembrane Domain, *Biochemistry*.
10. Rath, A., Glibowicka, M., Nadeau, V. G., Chen, G., and Deber, C. M. (2009) Detergent binding explains anomalous SDS-PAGE migration of membrane proteins, *Proceedings of the National Academy of Sciences* 106, 1760-1765.
11. Greenfield, N. J., and Fasman, G. D. (1969) Computed circular dichroism spectra for the evaluation of protein conformation, *Biochemistry* 8, 4108-4116.

12. Struck, D. K., Hoekstra, D., and Pagano, R. E. (1981) Use of resonance energy transfer to monitor membrane fusion, *Biochemistry* 20, 4093-4099.
13. Dass, C. (2007) *Fundamentals of contemporary mass spectrometry*, Vol. 16, John Wiley & Sons.
14. Jensen, P. F., and Rand, K. D. (2016) Hydrogen exchange: A sensitive analytical window into protein conformation and dynamics, *Hydrogen exchange mass spectrometry of proteins: Fundamentals, methods, and applications 1*.
15. Wales, T. E., and Engen, J. R. (2006) Hydrogen exchange mass spectrometry for the analysis of protein dynamics, *Mass spectrometry reviews* 25, 158-170.
16. Englander, S. W., and Kallenbach, N. R. (1983) Hydrogen exchange and structural dynamics of proteins and nucleic acids, *Quarterly reviews of biophysics* 16, 521-655.
17. Smith, D. L., Deng, Y., and Zhang, Z. (1997) Probing the non-covalent structure of proteins by amide hydrogen exchange and mass spectrometry, *Journal of mass spectrometry* 32, 135-146.

**Chapter 3**

**Structure-Function Relationship of HA2 by Site-Specific  
Mutagenesis**



### 3.1 Introduction

Influenza is an enveloped RNA virus in the family *Orthomyxoviridae*, and is either spherical with ~100 nm diameter, or filamentous with ~300 nm length.<sup>1</sup> An initial infection step is fusion (joining) of virus and cell membranes with consequent deposition of the nucleocapsid in the host cytoplasm.<sup>2-4</sup> The influenza virus hemagglutinin (HA) protein mediates virus binding to the host cell and subsequent membrane fusion. HA is translated as a single protein, and then proteolytically cleaved into the disulfide-linked subunits HA1 (328 residues) and HA2 (221 residues). Fusion is catalyzed by HA2 which contains the fusion peptide (FP), soluble ectodomain (SE), transmembrane domain (TM), and endodomain with respective lengths of ~25, 160, 25, and 10 residues (Fig. 3.1).

The initial HA complex outside the membrane includes three HA2 ectodomains and three HA1's. The complex has a mushroom-shape with stem (HA2 trimer) and head (HA1's).<sup>5, 6</sup> The stem includes an interior core with a parallel coiled coil of residue 76-125 helices. The HA2 N-terminal regions include 56-75 extended loops, 38-55 helices that pack antiparallel in exterior grooves of the coiled-coil, and 1-20 irregular FP and 21-35 antiparallel  $\beta$  sheet structures located between the stem and the head. The C-terminal HA2 regions include 130-141  $\beta$  sheet hairpins and compactly-associated 146-153 and 159-170 helices.

Host cell infection begins with HA1 binding to cell-surface sialic acids, which initiates endocytosis, and then endosome maturation with pH reduction to <6. HA1 separates from HA2 at this pH, and the HA2 SE transforms to a final-state trimer-of-hairpins structure (Fig. 3.1B).<sup>7, 8</sup> This structure includes an interior core of parallel coiled-coil 38-105 helices, 106-109 180° turns, and 110-128 helices and 154-176 strands that both pack antiparallel in the exterior grooves of the coiled-coil. Structural differences between the initial and final SE states include: (1) the 38-55

exterior helices and 56-75 loops become part of the interior helices; (2) the 106-128 interior helices become the 106-109 turns and 110-128 exterior helices; and (3) the 159-170 helices become part of the 154-176 strand regions of the hairpins.

Besides causing significant HA structural rearrangement, reduced pH is also the trigger for virus/endosome membrane fusion. Most mechanistic insight for fusion is based on the surrogate system of fusion between HA-expressing cells (HA-cells) and red blood cells (RBC's). The process includes a hemifusion step with outer leaflet membrane mixing, followed by breakage of the hemifusion diaphragm by pore formation, and then pore expansion.<sup>9-11</sup> There isn't yet high resolution imaging that temporally correlates the HA-structural changes and membrane changes during fusion, so mechanisms are primarily based on effects of mutations on fusion.<sup>10, 12, 13</sup>

The HA2 FP is sequestered in the interior of the initial HA complex, and is released during the process of HA1/HA2 separation and HA2 structural transformation into the final hairpin state. The FP is the most-conserved region of the HA sequence and is important in all steps of membrane fusion, as evidenced by both the G1E mutant (HA2 numbering), which results in no HA-cell/RBC fusion, and the G1S mutant which results in hemifusion without pore expansion.<sup>10, 14, 15</sup> The NMR structure of monomeric FP in detergent-rich media (without the rest of the HA2) shows predominant N-helix/turn/C-helix structure with tight antiparallel packing of the two helices.<sup>16, 17</sup> In membrane, the FP shows a mixture of this "closed" structure and a "semi-closed" structure in which the F9 ring is inserted between the two helices.<sup>18, 19</sup> These monomer FP structures may be relevant for the membrane-bound FP in the final hairpin state of HA2, as EPR spectral linewidths of a HA2<sub>1-127</sub> construct with trimeric SE correlate with monomeric FP domains.<sup>20</sup>

## (A) Amino acid sequence

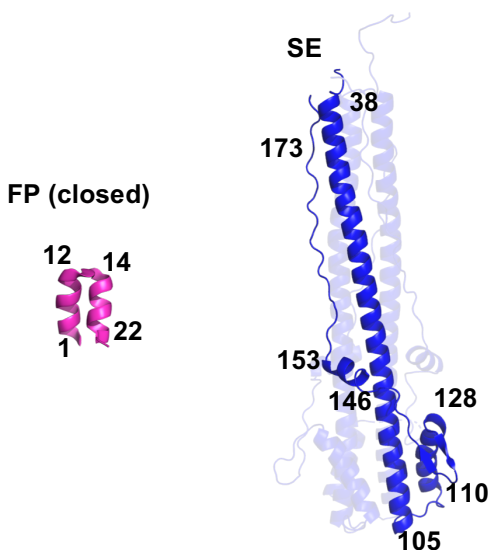
### HA2

1 GLFGAIAGFI ENGWEGMIDG WYGFRHQNSE GTGQAADLKS TQAAIDQING KLN RVIEKTN  
61 EK FHQIEKEF SEVEGRIQDL EKYVEDTKID LWSYNAELLV AL ENQHTIDL TDSEMNKLFE  
121 KTRRQLRENA EEMGNGSFKI YHKADNAAIE SIRNGTYDHD VYRDEALNNR FQIKGVELKS  
181 GYKDWILWIS FAISAFLLAV VLLGFIMWAA QRGNIRANIA IGGGGGGLEH HHHHH

### FHA2

1 GLFGAIAGFI ENGWEGMIDG WYGFRHQNSE GTGQAADLKS TQAAIDQING KLN RVIEKTN  
61 EK FHQIEKEF SEVEGRIQDL EKYVEDTKID LWSYNAELLV AL ENQHTIDL TDSEMNKLFE  
121 KTRRQLRENA EEMGNGSFKI YHKADNAAIE SIRNGTYDHD VYRDEALNNR FQIKGVELKS  
181 GYKDWLEHHH HHH

## (B) Fusion peptide and soluble ectodomain structures



**Figure 3.1** Amino acid sequences of the FHA2 and HA2 constructs with domains colored: fusion peptide (FP), pink; soluble ectodomain (SE), blue; transmembrane domain (TM), green; and endodomain, orange. The non-native C-terminal regions in black include a H<sub>6</sub> tag for affinity chromatography. The underlined residues denote the sites of the G1E and I173E mutations. (B) Ribbon diagrams of the FP in the closed conformation (PDB ID 2KXA) and trimeric SE in the final hairpin conformation (PDB ID 1QU1).

HA constructs that are truncated in the HA2 TM region result in arrested HA-cell/RBC hemifusion with non-expanding pores, which evidences the importance of a membrane-spanning TM in the final fusion pore expansion step.<sup>21, 22</sup> Fusion is also observed for constructs containing the HA2 ectodomain fused with the TM and endodomain from a different protein.<sup>23</sup> CD spectra of the HA2 TM peptide in detergent and membrane correlate with the expected  $\alpha$  helical structure.<sup>24</sup>

Earlier studies have shown that large HA2 constructs are hyperthermostable with  $T_m > 80$  °C. These constructs include HA2<sub>20-185</sub>  $\equiv$  SE  $\equiv$  SHA2, HA2<sub>1-185</sub>  $\equiv$  FP+SE  $\equiv$  FHA2, HA2<sub>20-211</sub>  $\equiv$  SE+TM  $\equiv$  SHA2-TM, and HA2<sub>1-221</sub>  $\equiv$  full-length HA2.<sup>25, 26</sup> Size-exclusion chromatography (SEC) and cross-linking data often correlate to a large trimer fraction in N-lauroylsarcosine (SRC) or *n*-decyl- $\beta$ -D-maltopyranoside (DM) detergent at pH 7.4, with visible aggregation at pH 5.0. The circular dichroism (CD)  $\theta_{222}$  values in DM at pH 7.4 correlate with ~60% helicity, which is generally consistent with a trimer-of-hairpins structure for the SE. All four constructs induce fusion of anionic vesicles at pH 5.0, with significant fusion contributions from the SE and FP. Exogenous addition of constructs like FHA2 or the shorter HA2<sub>1-127</sub> induce hemifusion and pore formation between “HA0”-cells bound to RBC’s, where HA0-cells express uncleavable HA’s that don’t induce fusion.<sup>27, 28</sup>

This chapter describes structural, biophysical, and functional studies for the G1E and I173E mutants of FHA2 and HA2, with comparison to their WT counterparts. Both mutants result in reduced HA-cell/RBC fusion and also complete loss of HA0-cell/RBC fusion mediated by exogenously-added FHA2.<sup>10, 13, 27, 28</sup> The N-terminal G1E mutant is in the FP, whereas the I173E mutant is in the SE. For the hairpin structure, I173 is in the exterior strands that bind to the grooves of the interior helical bundle. Some data are consistent with shallower membrane insertion of G1E vs. WT FP.<sup>29</sup> To our knowledge, there aren’t yet biophysical data for the I173E mutant.

## 3.2 Materials and Methods

### 3.2.1 Materials

Descriptions and commercial sources of materials include: DNA primers – Integrated DNA Technologies, Coralville, IA; *Escherichia coli* BL21(DE3) strain – Novagen, Gibbstown, NJ; Luria-Bertani (LB) medium – Dot Scientific, Burton, MI; isopropyl  $\beta$ -D-thiogalactopyranoside (IPTG) – Goldbio, St. Louis, MO; Cobalt affinity resin – Thermo Scientific, Waltham, MA; n-decyl- $\beta$ -D-maltopyranoside (DM) – Anatrace, Maumee, OH; 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine (POPC), 1-palmitoyl-2-oleoyl-sn-glycero-3-[phospho-*rac*-(1-glycerol)] (sodium salt) (POPG), 1,2-dioleoyl-3-trimethylammonium-propane (DOTAP), *N*-(7-nitro-2,1,3-benzoxadiazol-4-yl) (ammonium salt) dipalmitoylphosphatidylethanolamine (*N*-NBD-DPPE), *N*-(lissaminerhodamine B sulfonyl) (ammonium salt) dipalmitoylphosphatidylethanolamine (*N*-Rh-DPPE) – Avanti Polar Lipids, Alabaster, AL. Most other materials were purchased from Sigma-Aldrich, St. Louis, MO.

### 3.2.2 Protein constructs, expression, and purification

The plasmids containing the DNA sequences of full-length HA2 and FHA2 were extracted from previously sub-cloned *E. coli* cell cultures.<sup>26</sup> The G1E and I173E mutants of FHA2 and HA2 were generated by PCR using DNA primers and confirmed by DNA sequencing (Fig. 3.2). Each pET24a(+) plasmid was transformed into BL21(DE3) competent cells that were then streaked on an agar plate. After overnight growth in agar, a single colony was transferred to 50 mL Luria-Bertani (LB) medium and grown overnight. All media contained 50 mg/L kanamycin antibiotic, and growth cultures were shaken at 180 rpm and 37 °C. Stock aliquots for future growths were made by mixing 1 mL culture and 0.5 mL 50% glycerol, and were stored at -80 °C.

(A)

## FHA2

**GGT**CTGTTTCGGTGCTATCGCTGGCTTTATTGAAAATGGCTGGGAAGGCATGATTGATGGC  
TGGTACGGCTTTTCGTCACCAAACTCGGAAGGCACCGGTCAGGCGGCGGATCTGAAAAGC  
ACGCAGGCAGCTATTGACCAAATCAACGGCAAACCTGAATCGTGTCATCGAAAAAACCAAC  
GAAAAATTCATCAGATCGAAAAAGAATTTAGCGAAGTGGAAGGTCGCATTCAAGATCTG  
GAAAAATACGTCTGAAGATACGAAAATCGACCTGTGGAGTTACAACGCAGAACTGCTGGTT  
GCGCTGGAAAATCAGCACACCATTTGATCTGACGGACTCCGAAATGAACAACTGTTTCGAA  
AAAACCCGTCGCCAACTGCGTGAAAACGCAGAAGAAATGGGCAATGGTAGCTTCAAAATC  
TATCATAAAGCCGATAACGCGGCCATTCGCTCTATCCGCAATGGCACGTATGATCACGAC  
GTTTACCGCGACGAAGCCCTGAACAATCGCTTCCAA**ATC**AAAGGTGTCTGAACTGAAAAGT  
GGCTACAAAGACTGGCTCGAGCACCACCACCACCACCAC

## G1E – FHA2

**GAG**CTGTTTCGGTGCTATCGCTGGCTTTATTGAAAATGGCTGGGAAGGCATGATTGATGGC  
TGGTACGGCTTTTCGTCACCAAACTCGGAAGGCACCGGTCAGGCGGCGGATCTGAAAAGC  
ACGCAGGCAGCTATTGACCAAATCAACGGCAAACCTGAATCGTGTCATCGAAAAAACCAAC  
GAAAAATTCATCAGATCGAAAAAGAATTTAGCGAAGTGGAAGGTCGCATTCAAGATCTG  
GAAAAATACGTCTGAAGATACGAAAATCGACCTGTGGAGTTACAACGCAGAACTGCTGGTT  
GCGCTGGAAAATCAGCACACCATTTGATCTGACGGACTCCGAAATGAACAACTGTTTCGAA  
AAAACCCGTCGCCAACTGCGTGAAAACGCAGAAGAAATGGGCAATGGTAGCTTCAAAATC  
TATCATAAAGCCGATAACGCGGCCATTCGCTCTATCCGCAATGGCACGTATGATCACGAC  
GTTTACCGCGACGAAGCCCTGAACAATCGCTTCCAA**ATC**AAAGGTGTCTGAACTGAAAAGT  
GGCTACAAAGACTGGCTCGAGCACCACCACCACCACCAC

## I173E-FHA2

**GGT**CTGTTTCGGTGCTATCGCTGGCTTTATTGAAAATGGCTGGGAAGGCATGATTGATGGC  
TGGTACGGCTTTTCGTCACCAAACTCGGAAGGCACCGGTCAGGCGGCGGATCTGAAAAGC  
ACGCAGGCAGCTATTGACCAAATCAACGGCAAACCTGAATCGTGTCATCGAAAAAACCAAC  
GAAAAATTCATCAGATCGAAAAAGAATTTAGCGAAGTGGAAGGTCGCATTCAAGATCTG  
GAAAAATACGTCTGAAGATACGAAAATCGACCTGTGGAGTTACAACGCAGAACTGCTGGTT  
GCGCTGGAAAATCAGCACACCATTTGATCTGACGGACTCCGAAATGAACAACTGTTTCGAA  
AAAACCCGTCGCCAACTGCGTGAAAACGCAGAAGAAATGGGCAATGGTAGCTTCAAAATC  
TATCATAAAGCCGATAACGCGGCCATTCGCTCTATCCGCAATGGCACGTATGATCACGAC  
GTTTACCGCGACGAAGCCCTGAACAATCGCTTCCAA**GAG**AAAGGTGTCTGAACTGAAAAGT  
GGCTACAAAGACTGGCTCGAGCACCACCACCACCACCAC

## HA2

**GGT**CTGTTTCGGTGCTATCGCTGGCTTTATTGAAAACGGTTGGGAAGGCATGATCGACGGC  
TGGTACGGCTTTTCGCCATCAAACTCAGAAGGCACCGGTCAGGCGGCGGATCTGAAAAGC  
ACGCAGGCAGCTATTGACCAAATCAACGGCAAACCTGAATCGTGTGATCGAAAAAACCAAC  
GAAAAATTCCATCAGATCGAAAAAGAATTTTCTGAAGTCGAAGGTCGCATTCAAGATCTG  
GAAAAATATGTGGAAGATACGAAAATCGACCTGTGGTCATACAACGCGGAACTGCTGGTT  
GCCCTGGAAAATCAGCACACCATTTGATCTGACGGACTCGGAAATGAACAACTGTTTCGAA  
AAAACCCGTCGCCAACTGCGTGAAAACGCAGAAGAAATGGGCAACGGTAGTTTCAAAATC  
TACCATAAAGCTGATAACGCGGCCATTGAATCCATCCGCAATGGCACGTATGATCACGAC  
GTTTACCGTGACGAAGCGCTGAACAATCGCTTTTCAG**ATT**AAAGGCGTCGAACTGAAATCC  
GGTTACAAAGATTGGATTCTGTGGATCAGCTTTGCAATTTCTGCTTTCCTGCTGGCCGTG  
GTTCTGCTGGGTTTCATCATGTGGGCGGCGCAGCGTGGCAACATTTCGTGCAAACATCGCA  
ATCGGTGGCGGTGGCGGCGGCCCTCGAGCACCACCACCACCACCAC

## G1E-HA2

**GAG**CTGTTTCGGTGCTATCGCTGGCTTTATTGAAAACGGTTGGGAAGGCATGATCGACGGC  
TGGTACGGCTTTTCGCCATCAAACTCAGAAGGCACCGGTCAGGCGGCGGATCTGAAAAGC  
ACGCAGGCAGCTATTGACCAAATCAACGGCAAACCTGAATCGTGTGATCGAAAAAACCAAC  
GAAAAATTCCATCAGATCGAAAAAGAATTTTCTGAAGTCGAAGGTCGCATTCAAGATCTG  
GAAAAATATGTGGAAGATACGAAAATCGACCTGTGGTCATACAACGCGGAACTGCTGGTT  
GCCCTGGAAAATCAGCACACCATTTGATCTGACGGACTCGGAAATGAACAACTGTTTCGAA  
AAAACCCGTCGCCAACTGCGTGAAAACGCAGAAGAAATGGGCAACGGTAGTTTCAAAATC  
TACCATAAAGCTGATAACGCGGCCATTGAATCCATCCGCAATGGCACGTATGATCACGAC  
GTTTACCGTGACGAAGCGCTGAACAATCGCTTTTCAG**ATT**AAAGGCGTCGAACTGAAATCC  
GGTTACAAAGATTGGATTCTGTGGATCAGCTTTGCAATTTCTGCTTTCCTGCTGGCCGTG  
GTTCTGCTGGGTTTCATCATGTGGGCGGCGCAGCGTGGCAACATTTCGTGCAAACATCGCA  
ATCGGTGGCGGTGGCGGCGGCCCTCGAGCACCACCACCACCACCAC

## I173E-HA2

**GGT**CTGTTTCGGTGCTATCGCTGGCTTTATTGAAAACGGTTGGGAAGGCATGATCGACGGC  
TGGTACGGCTTTTCGCCATCAAACTCAGAAGGCACCGGTCAGGCGGCGGATCTGAAAAGC  
ACGCAGGCAGCTATTGACCAAATCAACGGCAAACCTGAATCGTGTGATCGAAAAAACCAAC  
GAAAAATTCCATCAGATCGAAAAAGAATTTTCTGAAGTCGAAGGTCGCATTCAAGATCTG  
GAAAAATATGTGGAAGATACGAAAATCGACCTGTGGTCATACAACGCGGAACTGCTGGTT  
GCCCTGGAAAATCAGCACACCATTTGATCTGACGGACTCGGAAATGAACAACTGTTTCGAA  
AAAACCCGTCGCCAACTGCGTGAAAACGCAGAAGAAATGGGCAACGGTAGTTTCAAAATC  
TACCATAAAGCTGATAACGCGGCCATTGAATCCATCCGCAATGGCACGTATGATCACGAC  
GTTTACCGTGACGAAGCGCTGAACAATCGCTTTTCAG**GAG**AAAGGCGTCGAACTGAAATCC  
GGTTACAAAGATTGGATTCTGTGGATCAGCTTTGCAATTTCTGCTTTCCTGCTGGCCGTG  
GTTCTGCTGGGTTTCATCATGTGGGCGGCGCAGCGTGGCAACATTTCGTGCAAACATCGCA  
ATCGGTGGCGGTGGCGGCGGCCCTCGAGCACCACCACCACCACCAC

**(B)**

Sequence – FHA2/HA2 G1E forward

**5' – CAT ATG GAG CTG TTC GGT GC – 3'**

Sequence – FHA2/HA2 G1E reverse

**5' – GCA CCG AAC AGC TCC ATA TG – 3'**

Sequence – FHA2 I173E forward

**5' – CGC TTC CAA GAG AAA GGT GTC – 3'**

Sequence – FHA2 I173E reverse

**5' – GAC ACC TTT CTC TTG GAA GCG – 3'**

Sequence – HA2 I173E forward

**5' – CGC TTT CAG GAG AAA GGC GTC – 3'**

Sequence – HA2 I173E reverse

**5' – GAC GCC TTT CTC CTG AAA GCG – 3'**

**Figure 3.2** (A) DNA sequences of HA2 and FHA2 constructs. Each line is 60 nucleotides. (B) primer sequences for the G1E and I173E site specific mutations.

An expression culture was prepared by adding 50  $\mu$ L stock to 50 ml LB, followed by overnight growth, and then addition of 1 L fresh LB and growth until  $OD_{600} \approx 0.5$ . Protein expression was induced by addition of 1 mM IPTG and then continued for 5 hours at 37 °C. The culture was centrifuged at 9000g and 4 °C for 10 min, followed by harvesting the cell pellet, and storage at -20 °C.

Purifications from cells expressing WT-, G1E-, and I173E-FHA2 were based on earlier work, and is denoted **protocol A**.<sup>26</sup> Buffers were prepared at 4 °C. Wet cells (5 g) were subjected



to four rounds of sonication in 40 mL of buffer A (50 mM sodium phosphate at pH 8.0, 0.5% *N*-lauroylsarcosine (SRC), 300 mM NaCl, and 10 mM imidazole) in an ice bath. Each round lasted for 1 min, with 80% amplitude intervals for 0.8 s followed by 0.2 s off. The lysate was centrifuged at 48000g and 4 °C for 20 min, and subsequent purification was done at ambient temperature using cold buffers. The lysate supernatant and 1 mL of Co<sup>2+</sup> affinity resin were stirred for 1 h, and the protein-coated resin was then collected by pouring the suspension through a fritted column. Weakly-bound proteins were removed from the resin by washes (4×) with buffer A, with wash volumes of 1 mL (WT-FHA2) or 2 mL (G1E- and I173E-FHA2). Subsequent elutions were done with buffer B (50 mM sodium phosphate at pH 8.0, 0.5% SRC, 300 mM NaCl, and 250 mM imidazole). The elutions (4 × 0.5 mL) were individually characterized by SDS-PAGE, and those with highest purities and quantities of FHA2 were pooled.

Low yield- and low-purity protein was obtained when protocol A was applied to cells expressing WT-HA2, so a new **protocol B** was developed for HA2 that began with sonication of 5 g wet cells in 40 mL PBS (10 mM sodium phosphate and 2 mM potassium phosphate at pH 7.4, with 137 mM NaCl and 3 mM KCl) using the same sonication parameters as above, followed by removal of the supernatant. The remaining pellet was subjected to sonication in fresh PBS and removal of the supernatant, and the new pellet was also subjected to this procedure. The final pellet was then subjected to a purification protocol similar to that used for FHA2-expressing cells, including sonication in buffer A, centrifugation of the suspension, and affinity purification of the supernatant. The eluent was mixed with an equal volume of ice-cold buffer that contained 10 mM Tris at pH 8.0, 0.17% DM, 2 mM EDTA, and 1 M *L*-arginine, followed by mixing overnight at 4 °C.<sup>26, 30, 31</sup>

### 3.2.3 CD, SEC, Cross-linking, and vesicle fusion

Protein samples for biophysical experiments were prepared by dialysis of FHA2 purification eluent or HA2 in buffer with 1 M *L*-arginine. For CD, SEC, and vesicle fusion, the dialysis buffer contained 10 mM Tris at pH 7.4, and for cross-linking, 20 mM HEPES at pH 7.4. The dialysis buffer also contained either 0.10% SRC or 0.17% DM, and for SEC and vesicle fusion, 150 mM NaCl. Dialysis buffer was replaced each day for the 3-day total dialysis time (5-day for CD samples). The protein concentration was then adjusted by dilution or by a concentrator.

CD spectra of protein (15  $\mu$ M) in buffer with DM were acquired with a CD instrument (Chirascan, Applied Photophysics), temperature-controlled cuvette with 1 mm pathlength, 260-190 nm wavelength range, 0.5 nm wavelength increments, and 1.5 s averaging-time per wavelength point. Each displayed spectrum is the difference between the protein + buffer and buffer-only spectra.

SEC was done with a FPLC (DuoFlow Pathfinder 20, Bio-Rad), semi-preparative column (Tricorn Superdex 200 Increase 10/300 GL, GE Technologies), 0.3 mL/min flow rate, and  $A_{280}$  detection. The column was equilibrated with an initial run with dialysis buffer without protein. Subsequent SEC was done with [loading protein]  $\approx$  1 mg/mL which corresponds to [running protein]  $\approx$  0.1 mg/mL.

Cross-linking between lysine  $\epsilon$ -NH<sub>2</sub> groups was achieved by stirring protein (20  $\mu$ M) with bis(sulfosuccinimidyl) suberate (1 mM) for 1 hour. The reaction was quenched by addition of Tris-HCl buffer at pH 6.8, with final [Tris] = 50 mM, and products were assayed by SDS-PAGE.

Vesicles for fusion studies contained POPC, POPC:POPG (4:1), or POPC:DOTAP (4:1) lipids. PC is a common lipid headgroup of the membranes of host cells infected by influenza virus.<sup>32, 33</sup> The PG lipids are included to reflect the negatively-charged lipids in the cell membrane.

The DOTAP headgroup has a charge of +1 and provides an additional probe of electrostatic effects on fusion. The role of protein charge is probed by comparison of fusion at physiological pH 7.4 vs. pH 5.0, where the latter is in the 5-6 range of influenza fusion with the endosome. The protein charge is  $\sim -10$  at pH 7.4 and  $+2$  at pH 5.0, whereas the lipid charges don't change with pH. Lipids were dissolved in chloroform:methanol (9:1 v/v), followed by solvent evaporation with nitrogen gas and vacuum pumping. The resulting lipid films were suspended in 5 mM HEPES and 10 mM MES at pH 7.4 or 5.0, with 0.01% NaN<sub>3</sub>. Fluorescently labeled vesicles were similarly prepared, with additional 2 mole% fluorescent lipid (*N*-NBD-DPPE) and 2 mole% quenching lipid (*N*-Rh-DPPE). The lipid suspensions were subjected to 10 freeze/thaw cycles followed by multiple extrusions through a filter with 100 nm diameter pores which produced unilamellar vesicles with  $\sim 200$  nm diameter. Labeled and unlabeled vesicles were mixed in 1:9 ratio with [total lipid] = 150  $\mu$ M. HA2-induced fusion between labeled and unlabeled vesicles was detected as an increase in fluorescence due to larger average distance between fluorescent and quenching lipids.

Assay conditions included vesicle stirring at 37 °C, continuous excitation with 467 nm radiation, and gated detection of fluorescence at 530 nm with 1 s time increments. The initial fluorescence of the vesicle solution is denoted  $F_0$ , and addition of HA2 stock marks  $t = 0$ , with  $F(t)$  then measured for  $\sim 600$  s. The stock contained 40  $\mu$ M HA2 and 0.17% DM in pH 7.4 buffer with 150 mM NaCl, and the mixed solution had HA2:lipid mole ratio = 1:300 and 0.002% DM. Subsequent addition of 12  $\mu$ l of 20 % Triton X-100 detergent solubilized the vesicles and provides  $F_{max}$ . Percent vesicle fusion is calculated as  $[(F(t) - F_0)/(F_{max} - F_0)] \times 100$ . There is typical  $\pm 1\%$  variation in long-time fusion extent between assay replicates, and there is negligible fusion after addition of an aliquot of stock buffer without protein. Although our conversion of fluorescence increases to percent vesicle fusion is standard practice, the calculated fusion is not a quantitative

measure of percent fused vesicles. The fluorescence varies as  $[1 + (R_0/R_{FQ})^6]^{-1}$ , where  $R_0$  and  $R_{FQ}$  are the Forster and average fluorophore-quencher distances, so fluorescence changes and percent lipid mixing are sensitive to the initial fluorophore and quencher concentrations. The most meaningful fusion comparisons are between different proteins with the same vesicle stock, and between different vesicle compositions using the same protein stocks.

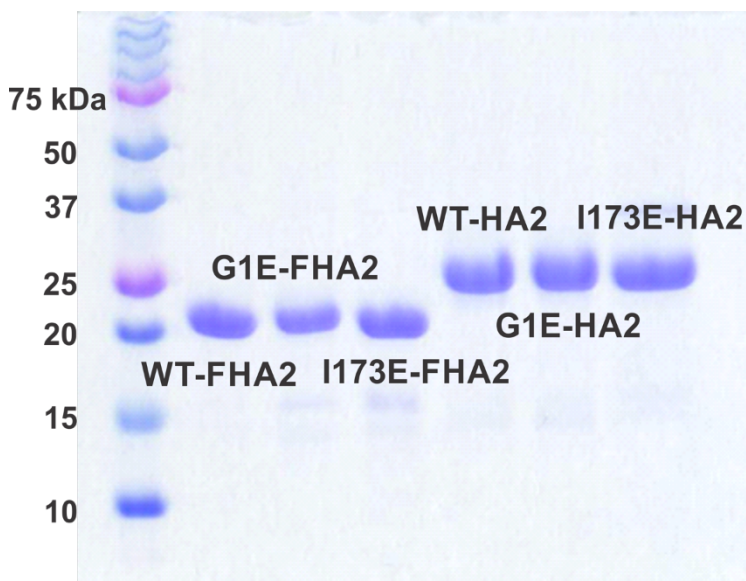
### 3.3 Results

#### 3.3.1 Protein preparation and folding

Fig. 3.3 displays SDS-PAGE of the six proteins. The purified yields of WT-, G1E-, and I173E-FHA2 are ~10, 5, and 3 mg/L culture, whereas the yields of the corresponding HA2 variants are ~2, 1, and 0.7 mg/L culture. High-purity FHA2 proteins were obtained by protocol A with affinity chromatography of the supernatant of the SRC-solubilized cell lysate. The proteins were well-folded after dialysis into buffer with DM, as evidenced by prominent minima in the CD spectra at 208 and 222 nm, which are characteristic of the expected predominant  $\alpha$  helical structure. Quantitative analysis of the  $|\theta_{222}|$  values is presented in the CD results.

Application of protocol A to cells that expressed HA2 resulted in low-yield and low-purity HA2. Reasonable yields of high-purity HA2 proteins were obtained by protocol B for which there was an initial step of removal of PBS -soluble and -suspendable material, followed by solubilization of the remaining cell pellet with sonication in buffer with SRC, and then affinity purification. The HA2 product was pure but not well-folded, based on a small  $|\theta_{222}|$  value of the CD spectrum obtained after direct dialysis into buffer with DM. Well-folded HA2 was obtained if there was intermediate dialysis into buffer with DM and 1 M *L*-arginine, followed by dialysis into buffer with DM and no *L*-arginine. These data support folded HA2 as the lowest free-energy state

in buffer with DM. The *L*-arginine may break up partially-misfolded HA2 aggregates.



**Figure 3.3** SDS-PAGE of the purification eluents. The  $MW_{FHA2} = 22.4$  kDa and  $MW_{HA2} = 26.7$  kDa. The gel was made using 12% acrylamide.

We describe a model that explains the underlying bases of the different protocols to obtain pure and well-folded FHA2 and HA2, as well as the different yields. We first consider the protocol B initial sonications in PBS, with discarding of the PBS supernatants obtained after centrifugation with 48000g. The HA2-content of the initial PBS supernatant was investigated by centrifugation at 220000g, and then subjecting the pellet to protocol A. The resulting eluent contained very pure HA2 that was well-folded, as judged by the CD spectrum after direct dialysis into buffer with DM. However, the HA2 yield from this “220000g pellet” was much lower than the yield of partially-folded HA2 from the “48000g pellet” of protocol B. We interpret our results as supporting FHA2/HA2 locations both in the bacterial membrane and in inclusion bodies, with respective well-

folded and partially-folded molecular structures. The protocol A supernatant from SRC sonication of cells expressing FHA2 contains mostly membrane-located FHA2, whereas the protocol B supernatant from SRC sonication of cells expressing HA2 contains mostly inclusion body HA2. The discarded supernatants from the protocol B initial sonications in PBS contain a small quantity of membrane-suspended and well-folded HA2. Overall, these data evidence that SRC is an effective detergent for both membrane- and inclusion body- FHA2/HA2, and solubilizes more rapidly the folded membrane fractions of these proteins. The latter idea is consistent with the results of protocol A for FHA2, including folded FHA2 and a substantial pellet which is not purified and contains predominantly partially-folded inclusion body FHA2. The latter point is supported by previous purification of FHA2 from this pellet, with resulting ~10 mg FHA2/L culture yield that is comparable to the protocol A yield.<sup>31</sup> This inclusion body FHA2 was only well-folded when there was intermediate dialysis into buffer with 1 M *L*-arginine and DM.

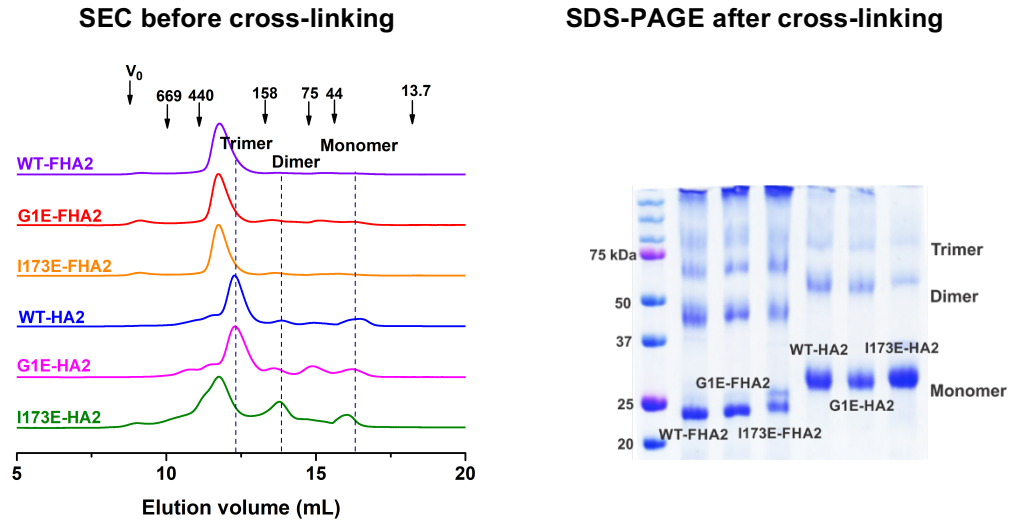
We estimate >10× higher expression of FHA2 vs. HA2 based on ~5× greater purified yields of WT-, G1E-, and I173E- FHA2 via protocol A vs. their HA2 counterparts via protocol B, ~2× greater purified FHA2 yield from complete solubilization of the cell pellet vs. the protocol A partial solubilization, and detection of FHA2 but not HA2 in the washes of the Co<sup>2+</sup>-resin with bound protein. The lower purified yields of G1E- and I173E- vs. WT proteins probably has some contribution from different expression levels, as the ratio of purified yields for FHA2 vs. HA2 is ~5 for the WT-, G1E-, and I173E- variants. All of the reported yields are from sonications with 40 mL buffer. For protocol B, there was complete solubilization in 40 mL buffer with SRC for the pellet from cells expressing WT-HA2, whereas 50 mL was required for the pellets from cells expressing G1E- or I173E- HA2. Use of 50 vs. 40 mL resulted in a ~0.3 mg/L greater purified yields of G1E- and I173E- HA2.

### 3.3.2 Size-exclusion chromatography and cross-linking

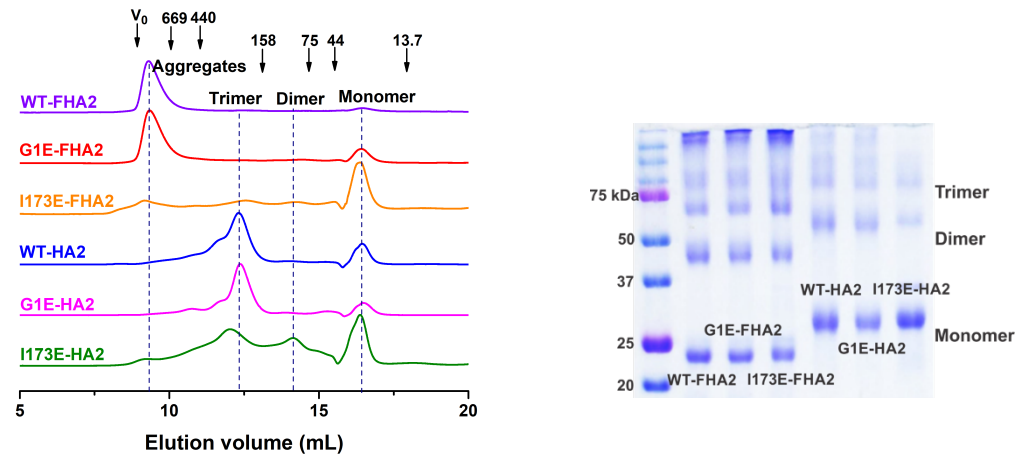
Fig. 3.4 displays (left) SEC before cross-linking and (right) SDS-PAGE after cross-linking. Data were obtained in (A) SRC and (B) DM detergents. Data for replicate samples are presented in Fig. 3A. Individual SEC elution peaks and cross-linking bands are assigned to monomer, dimer, trimer, and aggregate. Data for WT proteins are similar to previous results.<sup>26</sup> For all proteins in SRC and HA2 proteins in DM, there is a large SEC peak at ~250 kDa which is assigned to trimer with ~75 kDa protein and ~175 kDa detergent contributions. This SEC assignment is consistent with earlier observations, including a ~1:120 protein:detergent mole ratio for a HA2 fusion protein trimer, and dominant trimer for the HA2 soluble ectodomain in aqueous solution.<sup>25, 30</sup> Several of the SEC traces exhibit a peak at ~40 kDa which is assigned to a monomer with ~25 kDa protein and ~15 kDa detergent contributions. This is a major peak for the I173E mutants in DM and a minor peak in other traces. Some traces also exhibit a minor peak at ~150 kDa which is assigned to dimer with ~50 kDa protein and ~100 kDa detergent contributions. Both WT- and G1E- FHA2 proteins in DM have a dominant peak in the void volume corresponding to oligomers/aggregates with MW's > 1 MDa.

After cross-linking, the monomer is the most intense band for all constructs in both detergents, followed by the dimer and trimer (Fig. 3.4). Higher MW bands are weaker, except for FHA2 constructs which exhibit a band for large aggregates that doesn't migrate in the gel. The cross-linking data are generally consistent with the monomer, dimer, and trimer peaks observed in SEC, with the caveat that the cross-linking reaction is not complete so that the band intensities do not reflect populations.

(A) 0.10% SRC detergent



(B) 0.17% DM detergent

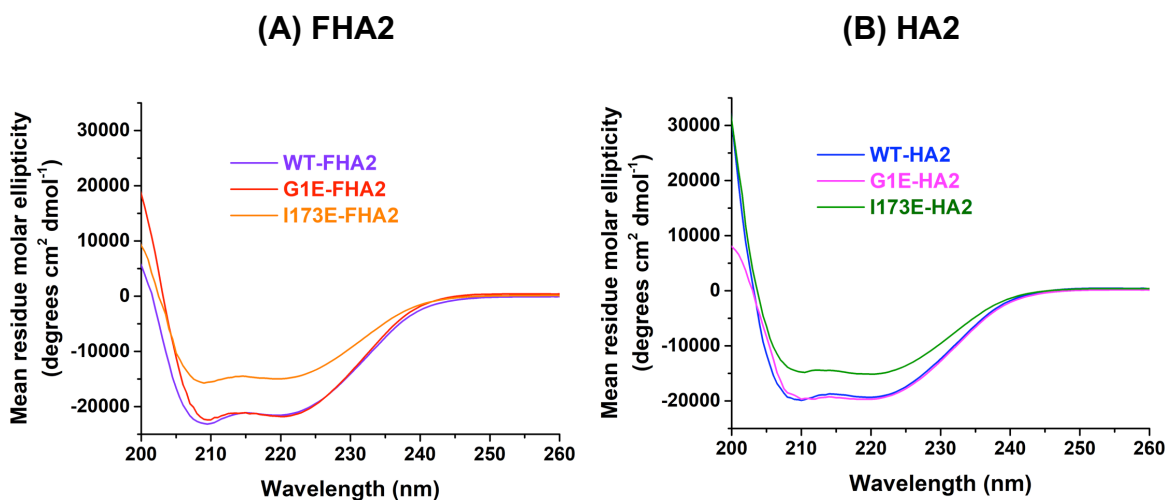


**Figure 3.4** SEC traces prior to cross-linking (left) and SDS-PAGE after cross-linking (right) and in (A) SRC and (B) DM detergents. SEC parameters include [loading protein] = 1 mg/mL, [protein running]  $\approx$  0.1 mg/mL, and  $A_{280}$  detection. Arrows mark the elution volumes of the void ( $V_0$ ) and MW standards in kDa. Cross-linking conditions include [protein] = 0.5 mg/mL, fifty-fold molar excess of bis(sulfosuccinimidyl) suberate cross-linking agent, and 1 hour time. The gels were made using 12% acrylamide.

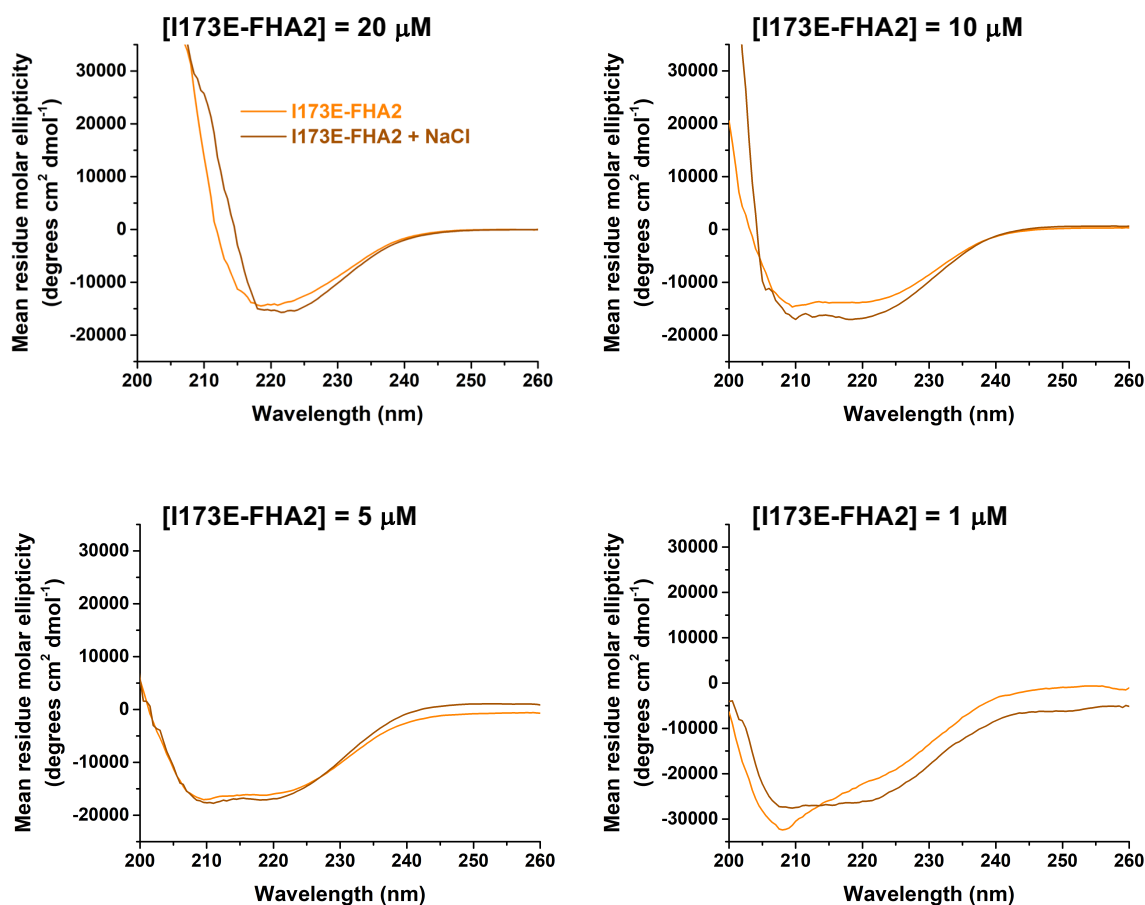


### 3.3.3 Circular dichroism

Fig. 3.5 presents CD spectra of the six constructs in DM at pH 7.4 and ambient temperature. These data were taken on the same day using the same CD spectrometer. Data for replicate samples are presented in Fig. 3B and typically exhibit  $<\pm 5\%$  variation in  $\theta_{222}$  values among replicates. All constructs exhibit spectra characteristic of  $\alpha$  helical structure, with minima at 208 and 222 nm. Table 1 reports average helicities calculated from  $\theta_{222}$  values, with higher helicity for FHA2 vs. HA2, and for WT and G1E vs. I173E. The similar helicities of G1E and WT proteins are consistent with helical FP structure for G1E, whereas reduced helicity of I173E proteins supports some disruption of the SE structure in this mutant. There isn't NaCl in the sample buffers for the Fig. 3.5 CD spectra, but Fig. 3.6 shows that for I173E-FHA2, similar spectra are obtained in the absence vs. presence of 150 mM NaCl. Fig. 3.6 also presents spectra with [I173E-FHA2] = 1, 5, 10, and 20  $\mu\text{M}$ , and all the  $\theta_{222}$  values correlate to a significant fraction of  $\alpha$  helical structure.

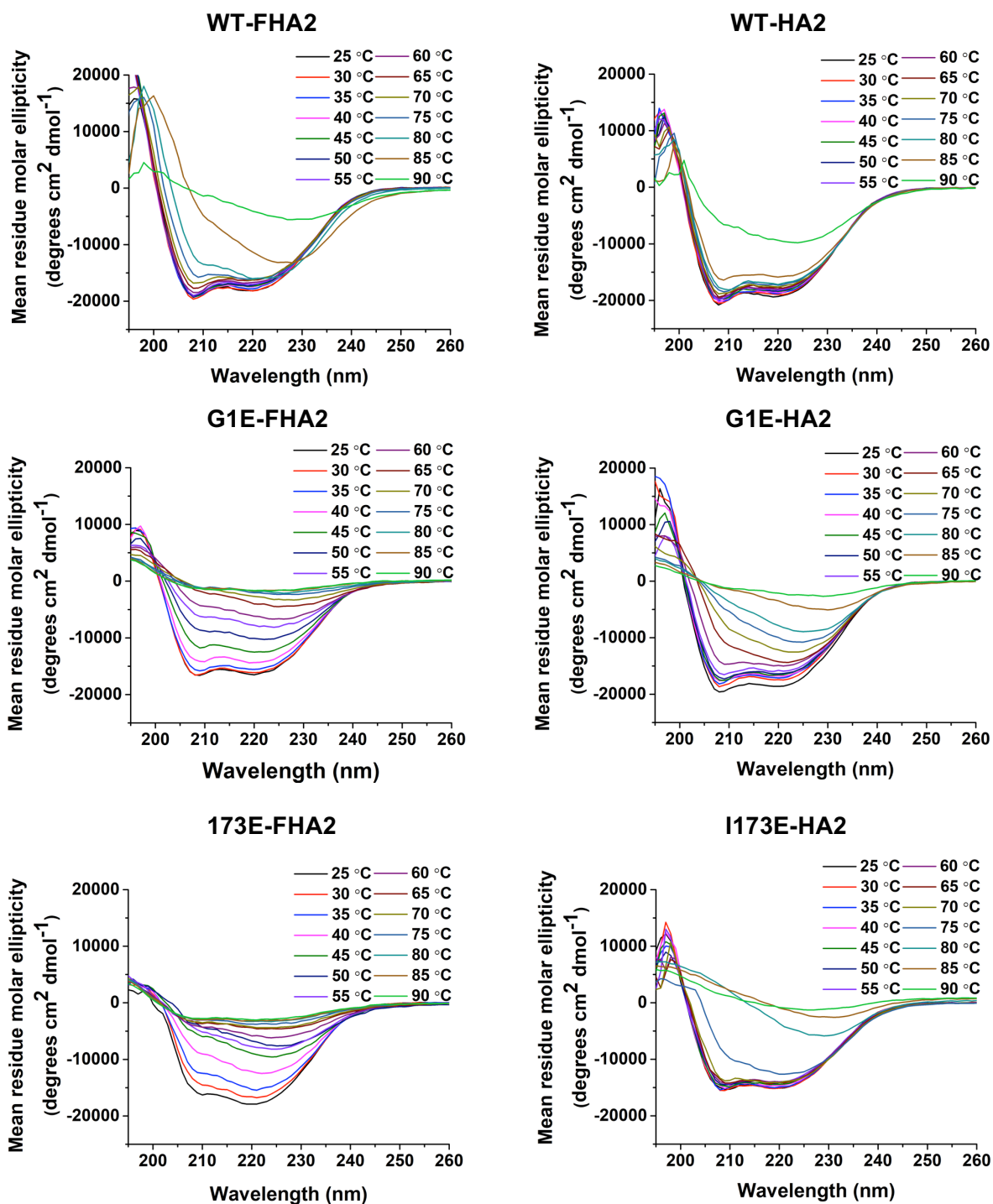


**Figure 3.5** CD spectra of the (A) FHA2 and (B) HA2 proteins at ambient temperature. Samples contained 15  $\mu\text{M}$  protein in 10 mM Tris buffer at pH 7.4 with 0.17% DM.

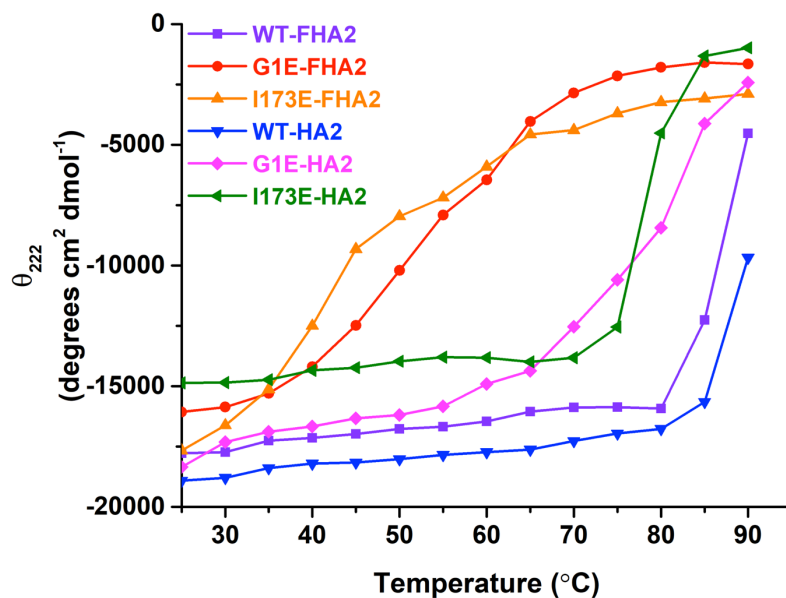


**Figure 3.6** Comparison of ambient-temperature CD spectra of I173E-FHA2 without or with 150 mM NaCl. The buffers also contained 10 mM Tris-HCl at pH 7.4 and 0.17% DM. The larger  $|\theta_{222}|$  values for [I173E-FHA2] = 1  $\mu$ M vs. higher protein concentrations likely reflect error in the 1  $\mu$ M concentration.

Fig. 3.7 displays CD spectra obtained between temperatures of 25 and 90  $^{\circ}$ C in 5  $^{\circ}$ C increments and Fig. 3.8 displays plots of  $\theta_{222}$  vs. temperature. All spectra for a particular construct were acquired on the same day. Thermostability of constructs are qualitatively ordered as follows: WT-HA2 > WT-FHA2 > G1E-HA2, I173E-HA2 > G1E-FHA2 > I173E-FHA2 (Table 3.1).



**Figure 3.7** Temperature-series CD spectra. Protein and buffer conditions are the same as Fig. 3.5.



**Figure 3.8** Plots of  $\theta_{222}$  values vs. temperature.

**Table 3.1** Analysis of CD spectra.

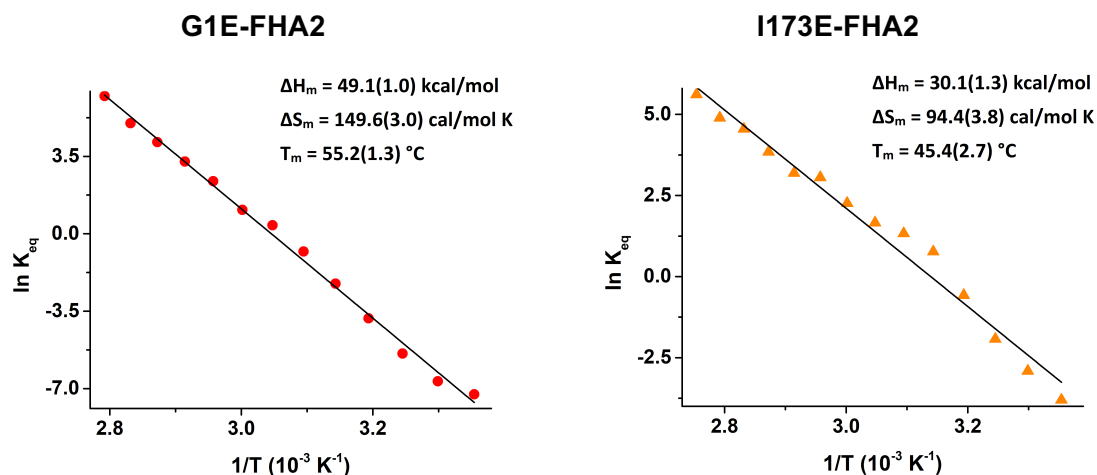
Construct	Average percent helicity <sup>a</sup>	$T_m$ <sup>b</sup>
WT-FHA2	64	88
G1E-FHA2	65	58
I173E-FHA2	44	51
WT-HA2	58	>90
G1E-HA2	58	80
I173E-HA2	45	79

<sup>a</sup> Calculated using the  $\theta_{222}$  of ambient temperature CD spectra (Fig. 3.5), with 100% helicity taken as  $\theta_{222} = -33000$  degrees  $\text{cm}^2 \text{dmol}^{-1}$ .

<sup>b</sup> Calculated using  $|\theta_{222}(T_m)|/|\theta_{222}(25^\circ\text{C}) - 2000| = 0.5$

The  $\theta_{222}$  value at each temperature data is used to calculate the fraction folded protein via the equation  $f_{folded} = |\theta_{222} - \theta_{unfolded}| / |\theta_{folded} - \theta_{unfolded}|$ , and then  $f_{unfolded} = 1 - f_{folded}$ . The model equilibrium is folded trimer  $\leftrightarrow$  3 unfolded monomers, so that  $K_{eq} = (f_{unfolded})^3 / (1 - f_{folded})/3$ . The unfolded state is likely a monomer rather than trimer because: (1) the favorable hydrophobic effect of the trimer is achieved with folded monomer units but not unfolded units; and (2) the translational entropy is greater for three independent monomers than for a trimer. The van't Hoff analysis of  $\ln K_{eq}$  vs.  $1/T$  is linear-fitted, and the best-fit slope and  $\ln K_{eq}$ -intercept are used to calculate  $\Delta H_m = -R \times \text{slope}$ ,  $\Delta S_m = R \times \ln K_{eq}\text{-intercept}$ , and  $T_m = \Delta H_m / \Delta S_m$  that corresponds to  $\ln K_{eq} = 0$  (Table 3.2).

The  $T_m$  values for G1E-FHA2 and I173E-FHA2 are near the middle of the 25-90 °C experimental temperature range, and the  $\ln K_{eq}$  vs.  $1/T$  plots are reasonably linear over the whole range (Fig. 3.9). The  $T_m$  values are higher for the other proteins, and the  $\ln K_{eq}$  vs.  $1/T$  plots are only linear for the temperature range around  $T_m$  over which  $\ln K_{eq}$  changes sign and exhibits significant changes in magnitude (Fig. 3C). The numbers in parentheses in Table 3.2 are the fitting-associated uncertainties of the parameter values. There are also uncertainties in the  $\theta_{folded}$  and  $\theta_{unfolded}$  values which are chosen based on extrapolations from the respective  $\theta_{25C}$  and  $\theta_{90C}$ . These  $\theta$  uncertainties result in uncertainties in the calculated  $f_{unfolded}$ ,  $\ln K_{eq}$ , and fitted parameter values. Changes in  $\theta_{folded}$  and  $\theta_{unfolded}$  typically lead to <2 °C changes in  $T_m$ .



**Figure 3.9** Van't Hoff plots of the unfolding  $\ln K_{eq}$  vs.  $1/T$  of the (A) G1E-FHA2 and (B) I173E-FHA2 proteins based on  $\theta_{222}$  data for the temperature range around  $T_m$ . Best-fit parameters are given with uncertainties in parentheses.

**Table 3.2** Thermodynamic parameters from Van't Hoff analysis.

Construct	$T_m^{c,d}$ (°C)	$\Delta H_m$ (kcal/mole)	$\Delta S_m$ (cal/mole-K)
WT-FHA2	86.8(0.2)	186.2(10.5)	517 (29)
G1E-FHA2	55.2(1.3)	49.1(1.0)	149.6(3.0)
I173E-FHA2	45.4(2.7)	30.1(1.3)	94.4(3.8)
WT-HA2	90.7(1.2)	126 (36)	345 (99)
G1E-HA2	75.0(2.4)	43.6(3.6)	125 (11)
I173E-HA2	79.2(1.2)	165 (17)	469 (48)

<sup>c</sup> The  $T_m$ ,  $\Delta H_m$ , and  $\Delta S_m$  parameters are calculated using van't Hoff analysis of  $\ln K_{eq}$  vs.  $1/T$ , where  $K_{eq} = (f_{unfolded})^3 / (1 - f_{unfolded})/3$ ,  $f_{folded} = |\theta_{222} - \theta_{unfolded}| / |\theta_{folded} - \theta_{unfolded}|$ ,  $f_{unfolded} = 1 - f_{folded}$ , and  $\theta_{folded}$  and  $\theta_{unfolded}$  values are based on extrapolations from the respective  $\theta_{25C}$  and  $\theta_{90C}$ . Fitting is done over the temperature range for which  $\ln K_{eq}$  changes sign and exhibits significant changes in magnitude, and over which  $\ln K_{eq}$  vs.  $1/T$  is reasonably linear. The  $\Delta H_m = -R \times \text{slope}$ ,  $\Delta S_m = R \times \ln K_{eq}\text{-intercept}$ , and  $T_m = \Delta H_m / \Delta S_m$ . Specific  $\theta_{folded}$  and  $\theta_{unfolded}$  in degrees  $\text{cm}^2 \text{ dmole}^{-1}$ , and fitting temperature ranges in  $^{\circ}\text{C}$  are as follows: WT-FHA2, -18000, -2000, 80-90; G1E-FHA2, -17000, -1500, 25-85; I173E-FHA2, -21000, -2700, 25-90; WT-HA2, -19000, -2000, 80-90; G1E-HA2, -19500, -500, 55-90; I173E-HA2, -15000, -900, 70-90.

<sup>d</sup> The numbers in parentheses are fitting-associated uncertainties. The uncertainties in the  $\Delta H_m$  and  $\Delta S_m$  parameters are based on statistically-reasonable variations in the sum of residuals  $\sum_j (\ln K_j^{exp} - \ln K_j^{calc})^2$  where  $exp \equiv$  experimental,  $calc \equiv$  calculated, and  $j$  is the data index. Variations in  $\Delta H_m$  and  $\Delta S_m$  are highly correlated with one another. The uncertainty in  $T_m \equiv \delta_{Tm}$  was calculated using the variation in  $1/T \equiv \delta_{1/T}$  associated with the average residual  $\equiv \delta_{lnK}$ . The  $\delta_{1/T} = \delta_{lnK} / (\Delta H_m / R)$  and the  $\delta_{1/T} / (1/T_m) = \delta_{Tm} / T_m$  were the basis for the calculated  $\delta_{Tm} = \delta_{1/T} \times T_m^2$ .

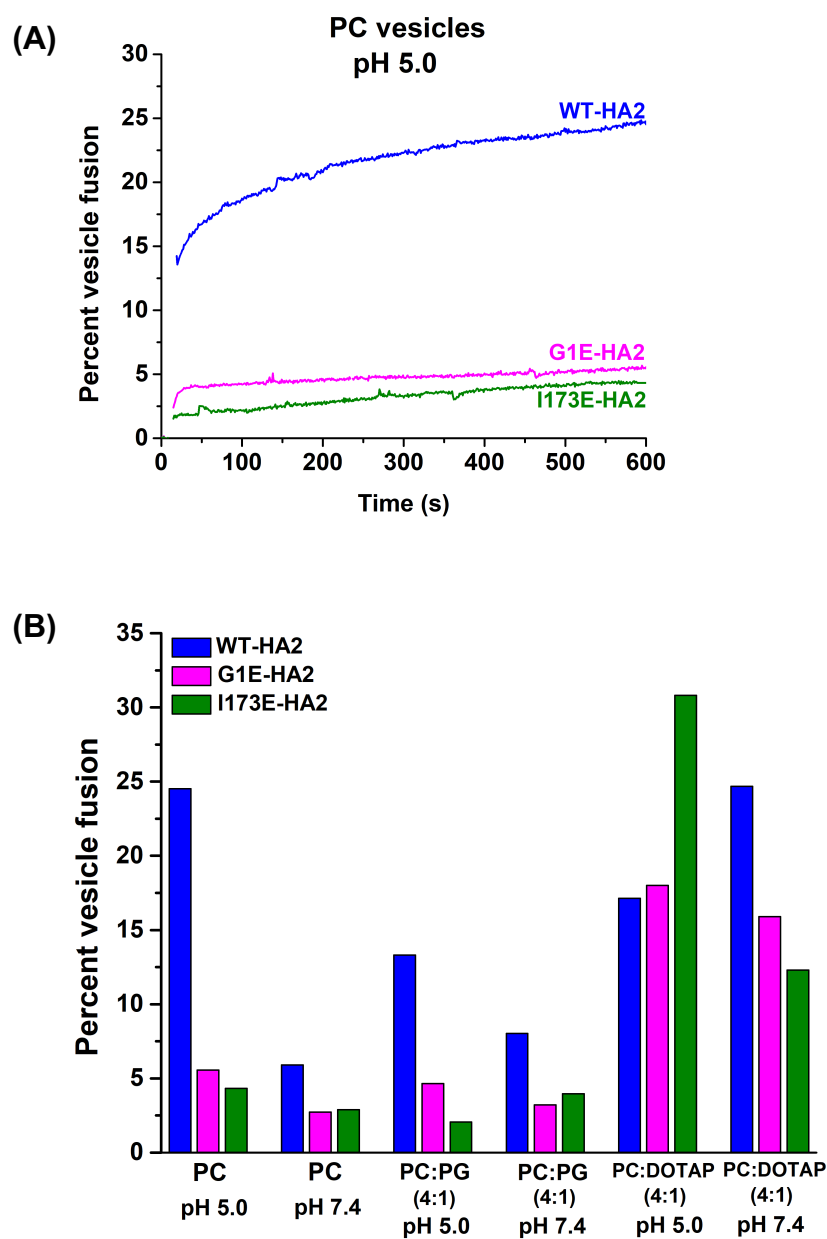
### 3.3.4 Vesicle fusion

Fig. 3.10A displays traces of HA2-induced PC vesicle fusion vs. time at pH 5.0, and Fig. 3.10B displays long-time fusion extents for PC, PC:PG (4:1), and PC:DOTAP (4:1) vesicles at pH's 5.0 and 7.4. Fig. 3D displays fusion extents from replicate assays with different protein and vesicle preparations. The trends for each protein are typically similar between the Figures for pH 5.0 vs. 7.4 and for PC:PG vs. PC:DOTAP vesicles. Influenza virus fuses with the late endosome at  $\text{pH} < 6$ . At pH's 5.0 and 7.4, PC has a neutral zwitterionic headgroup, and PG and DOTAP

respectively carry charges of -1 and +1. The membranes of host cells of influenza contain ~25 mole% PC and ~15 mole% anionic lipid including PG, and there are widely-varying lipid compositions of the chromatographic fractions of endosomal membranes.<sup>26, 32-35</sup> The influenza virus also fuses with vesicles with a variety of lipid compositions.<sup>36</sup> Although there aren't physiological positively-charged lipids, PC:DOTAP vesicles are studied to probe the effect on fusion of the sign of lipid charge. The HA2:lipid mole ratio = 1:300 of the assay is smaller than the ~1:50 ratio of the virus.<sup>37, 38</sup> HA-mediated fusion includes lipid mixing as well as pore formation, leakage, and contents mixing.<sup>12, 13, 36</sup> Protein-mediated vesicle fusion exhibits faster leakage than contents mixing, and in the present study, vesicle fusion is assayed by lipid mixing, which is a common feature with HA-fusion.<sup>39, 40</sup>

Fig. 3.10A shows much more extensive PC vesicle fusion at pH 5.0 for WT- vs. mutant-HA2's, with similar reduced fusion for both G1E and I173E. Fig. 3.10B shows that PC fusion is reduced at pH 7.4, but still higher for WT vs. mutant. At pH 5.0, there is lower fusion of PC:PG vs. PC vesicles, while at pH 7.4, there is similar reduced fusion for both vesicle-types. At pH 5.0, the calculated WT-HA2 charge is +3, so the higher PC vs. PC:PG fusion evidences that attractive HA2/vesicle electrostatic energy is not required for fusion catalysis. This conclusion is similarly supported by significant fusion of PC:DOTAP vesicles at both pH's, even though there are respective repulsive vs. attractive HA2/vesicle electrostatic energies at pH 5.0 vs. 7.4. Detailed interpretations of the fusion extents are presented in the Discussion (Section 3.4).





**Figure 3.10** HA2-induced vesicle fusion with protein:lipid = 1:300. (A) Time-courses of POPC vesicle fusion at pH 5.0. (B) Percent vesicle fusion at 600 s after addition of protein for different lipid compositions and pH's. Each bar represents the average of three replicates. There is typically  $\pm 1\%$  variation in percent vesicle fusion among the replicates.

### 3.4 Discussion

The present study describes a structural and functional comparison between WT- full-length HA2 and the truncated construct lacking the TM, FHA2, and the G1E and I173E point mutants that are known to inhibit HA-mediated fusion. Significant findings of the present study include: (1) predominant trimer fraction in SRC detergent at pH 7.4 for all protein constructs *vs.* mixtures of trimer, monomer, and oligomers/aggregates in DM detergent; (2) similar helicities of WT- and G1E- proteins *vs.* reduced helicity of I173E- proteins; (3) hyper-thermostable WT- FHA2 and HA2 and less-stable FHA2 and HA2 mutants with respective reductions of ~40 and ~15 °C in  $T_m$ ; and (4) efficient HA2-induced vesicle fusion of neutral and anionic vesicles at pH 5.0 for WT- HA2 *vs.* reduced fusion with mutants.

#### 3.4.1 Models of protein structure and stability

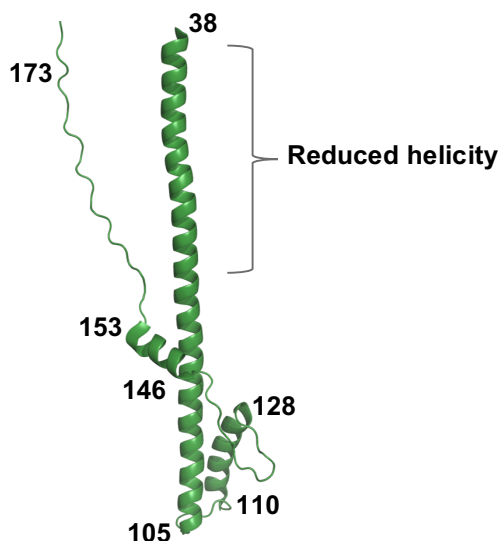
The CD spectra and analyses of Fig. 3.5 and Table 3.1 support a 65% average helicity for WT- and G1E- FHA2 and a 58% average helicity for WT- and G1E- HA2. The proteins are well-folded, based on reasonable agreements between these average helicities and the fractions  $\alpha$  helical residues calculated for a model in which the only  $\alpha$  helical residues are those in high-resolution structures of the FP in detergent and the SE in aqueous solution, and in the TM (residues 2-12, 14-22, 38-105, 110-128, 146-153, 186-210, Fig. 3.1B).<sup>8, 16</sup> The calculated helical fractions from this model are 115/193 residues for FHA2 and 140/235 residues for HA2. The helicities of WT- FHA2 and HA2 determined in the present study are similar to the helicities reported in some earlier studies.<sup>26, 31, 41</sup> However, the 65% helicity for WT- and G1E- FHA2 in the present study is higher than the respective ~25% and ~35% helicities of a 2011 study.<sup>28</sup> The origin of this discrepancy isn't known, but we note that our high *vs.* their low helicities correlates with the presence *vs.*

absence of detergent in the samples. The present study also shows that the I173E mutants exhibit only 45% helicities and a model explaining this reduced helicity is presented below.

The hyperthermostabilities of WT- FHA2 and HA2 are evidenced by the CD spectra and  $\theta_{222}$  vs. T plots (Figs. 3.7 and 3.8), and the accompanying van't Hoff analyses to determine  $T_m$  values (Table 3.2). The  $T_m \approx 90$  °C for FHA2 and HA2, and the respective  $T_m \approx 80$  °C for HA2<sub>20-185</sub> (SE) and  $>85$  °C for HA2<sub>20-211</sub> (SE+TM) support a major contribution to stability by the SE, with smaller contributions from the FP and TM, where the latter two HA2 domains are the only ones that are deeply-inserted in the fused membrane.<sup>25, 26, 42</sup> The hairpin structure of the SE in the final HA2 state allows for proximity between residues 1-37 and 177-210 that could be the basis of the higher  $T_m$  of full-length HA2 vs. truncated constructs. In previous papers, this possibility has been imagined as a bundle complex between the three FP's (e.g. HA2<sub>1-23</sub>) and three TM's (e.g. HA2<sub>187-209</sub>).<sup>2, 11</sup>

The I173E mutation exerts significant destabilization of the SE of FHA2 as evidenced by loss of 30% of the overall helicity and reduction of  $T_m$  by 40 °C (Table 3.1). There is smaller but still significant SE destabilization for I173E- vs. WT- HA2. The WT SE structure in Fig. 3.1B has C-terminal strands in the external grooves of the N-helix bundle, with the I173 sidechain inserted in a hydrophobic pocket of the bundle. Fig. 3.11 displays a structural model for the I173E SE in which there is dissociation of the strands from the bundle, with accompanying destabilization and loss of helicity near the bundle terminus. Strand dissociation is due in part to more favorable Born energy of the charged Glu sidechain in the high-dielectric aqueous environment vs. the low-dielectric protein interior. This I173E model is supported by data for shorter HA2 constructs with C-terminal truncation. The HA2<sub>38-89</sub> peptide forms a trimer of parallel helices at lower temperatures and unfolded monomers at higher temperatures, and its  $T_m = 51$  °C is comparable to

the  $T_m$  of I173E-FHA2.<sup>43</sup> Ambient-temperature CD spectra of HA2<sub>1-127</sub> correlate with 50% helicity, which is much lower than the 85% value (106/127 residues) calculated using helical regions of the FP and SE structures.<sup>44</sup>



**Figure 3.11** Structural model for a monomer unit of the SE of mutant HA2's. There is partial dissociation of the C-terminal strand and partial unfolding of the N-terminal helix.

G1E- and WT- FHA2 both exhibit 65% ambient-temperature helicity but the  $T_m$  for G1E-FHA2 is reduced by 32 °C vs. WT-FHA2, which indicates significant destabilization of the SE by the G1E mutation. There is smaller but still significant destabilization of G1E- vs. WT- HA2. Destabilization of the SE is surprising because the G1E mutation is more than 35-residues from the SE structure. We propose that the SE destabilization for G1E is similar to that for I173E, and is a result of loss of binding of C-terminal strands with the N-terminal helical bundle (Fig. 3.11). This loss is due to competition with the G1E FP's for binding to the strands. Binding is not

expected for the WT-FP which adopts tight helical hairpin structure that is stabilized by the favorable G1-NH<sub>3</sub><sup>+</sup>/(14-22) helix-dipole interaction (Fig. 3.1B).<sup>45</sup> This interaction may be disrupted for zwitterionic E1, with consequent separation of the two FP helices that can then bind with the SE strands.

G1E-FHA2, G1E-HA2, and I173E-FHA2 have respective  $\Delta H_m$  values of 49, 44, and 30 kcal/mole that are determined from van't Hoff analyses over broad temperature ranges for which there are clear changes from mostly-folded to mostly-unfolded protein (Figs. 3.9 and A3 and Table 3.2). There are likely similar unfolded monomer states for all of the proteins, so the  $\Delta H_m$  values likely reflect more favorable enthalpies for folded G1E- FHA2 and HA2 vs. I173E-FHA2. This interpretation is consistent with the previously-described direct I173E destabilization of binding between the C-terminal strands and the N-terminal helical bundle of the SE vs. less-direct G1E destabilization of the SE by weakening FP interhelical interactions, with consequent competition by the FP for binding to the strands.

There are less accurate  $\Delta H_m$  values for WT- FHA2 and HA2, and I173E-HA2, because these values are determined from van't Hoff analyses from limited temperature ranges, and unfolding is sometimes incomplete at the highest temperature. The  $\Delta H_m$  values are larger for WT- vs. G1E- and I173E- FHA2 and WT- vs. G1E- HA2, which evidences more favorable enthalpies of the WT proteins. The WT  $\Delta H_m$  are similar to those of papain and trypsin, which have masses similar to the monomer HA2 constructs.<sup>46</sup> The  $\Delta S_m$  and  $\Delta H_m$  values are correlated in Table 3.2, as expected from  $\Delta S_m = \Delta H_m/T_m$ .

### 3.4.2 Monomers and trimers

The SEC data for all constructs in SRC show predominant trimers, with minor populations of monomers and dimers (Figs. 3.4A and 3A). The relative populations of different species are more varied between the constructs in DM, and the monomer fractions are generally higher, particularly for I173E- FHA2 and HA2 (Figs. 3.4B and 3A). SEC was done with [protein]  $\approx 4 \mu\text{M}$  and with [NaCl] = 150 mM, and CD spectra of I173E-FHA2 evidence well-folded protein for these conditions and also for different protein concentrations and no NaCl (Figs. 3.5 and 3.6). The greater monomer fraction of I173E- vs. WT- proteins is correlated with the lower stability of the I173E trimer.

To our knowledge, folded monomer HA2 has not been previously reported, but there are reports of folded monomer HA1+HA2 ectodomain, including complexes with antibodies, and folded monomer gp41.<sup>47-57</sup> Interestingly, both HA2 and HIV gp41 SE constructs can form folded hyperthermostable monomers and trimers, with similar structures for the monomers and monomer subunits of the trimers. Monomer fraction is correlated with inter-protein electrostatic repulsion, as evidenced by higher monomer fractions for HA2 constructs at neutral pH (charge  $\approx -10$ ) and for gp41 constructs at low pH (charge  $\approx +10$ ). There are higher populations of trimer and larger oligomers for HA2 at low pH and gp41 at neutral pH, with respective charges of  $\sim +2$  and  $-2$ . Another similarity between HA2 and gp41 is predominant trimers in anionic detergents, SRC and SDS, respectively, and higher monomer fractions in neutral detergents, DM and DPC, respectively. We hypothesize that the protein trimer may carry less than three times the monomer charge because of protonation of the Asp and Glu side chains in the trimer interior. The trimer may therefore be favored in anionic detergent because the electrostatic repulsion between the trimer and micelle is smaller than between three monomers and micelles.

Although the SEC data for I173E-FHA2 in DM show predominant monomers, the cross-linking data are consistent with a significant fraction of trimers. This difference may be due to the presence (SEC) *vs.* absence (cross-linking) of NaCl, as other proteins also exhibit smaller oligomer and greater monomer fractions with increased [NaCl].<sup>58, 59</sup> This behavior is sometimes attributed to disruption of salt bridges between monomer units of the oligomer. The HA1+HA2 ectodomain of influenza strains A/Korea/01/2009 (KR01), A/California/04/2009 (CA04) and A/Darwin/2001/2009 (DA01) are monomers in a pH 8.0 solution with 200 mM NaCl, but only the latter two are trimers in their crystal structures, whereas KR01 shows a V-shaped arrangement of two monomers connecting through the HA1 globular head domains.<sup>49, 54</sup> The monomer structure is similar in all three strains. The sequence of KR01 differs at only three residues from those of CA04 and DA01, and the crystallization conditions only differ in 100 mM NaCl for KR01 *vs.* no NaCl for the strains that adopt trimer structures. The presence *vs.* absence of NaCl may therefore favor folded monomer *vs.* folded trimer HA2.

### 3.4.3 Correlations between WT-HA2 fusion and electrostatic energies

Fig. 3.10B displays comparative vesicle fusion extents for WT- *vs.* mutant- HA2's, pH 5.0 *vs.* 7.4, and vesicle compositions of zwitterionic PC lipids *vs.* mixtures of PC and anionic PG lipids, or PC and cationic DOTAP lipids. We first focus on WT-HA2 fusion under different conditions, and discuss how differences in fusion extents correlate with electrostatic energies. The charge of WT-HA2, excluding the tag, is  $\sim +3$  at pH 5.0 and  $\sim -9$  at pH 7.4, whereas the lipid charges are pH-invariant, and are 0 for PC, -1 for PG, and +1 for DOTAP.

Higher PC fusion at pH 5.0 *vs.* 7.4 correlates with smaller inter-trimer electrostatic repulsion at pH 5.0 *vs.* 7.4. This also correlates with visible aggregation of the HA2 constructs at

pH 5.0 that may contribute to fusion via increased local membrane concentrations and perturbations of the FP and perhaps TM.<sup>26</sup> These effects are also evidenced by higher PC:PG fusion at pH 5.0 vs. 7.4, with an additional contribution from attractive vs. repulsive HA2/vesicle electrostatic energies. This latter effect is also evident from higher PC:DOTAP fusion at pH 7.4 vs. 5.0, and corresponding attractive vs. repulsive HA2/vesicle electrostatic energies. This energy contribution is also manifested by higher fusion of PC:DOTAP at pH 7.4 vs. PC:PG at pH 5.0, and the corresponding greater attraction between the -9 charge HA2 and DOTAP vs. the +3 charge HA2 and PG. Similarly, higher PC:DOTAP fusion at pH 5.0 vs. PC:PG fusion at pH 7.4 correlates with smaller HA2/DOTAP vs. larger HA2/PG electrostatic repulsion. Vesicle/vesicle electrostatics are manifested by higher fusion for PC vs. PC:PG at pH 5.0 and the corresponding neutral vs. repulsive inter-vesicle electrostatic energies. In addition, higher PC:DOTAP fusion at pH 5.0 vs. PC fusion at pH 7.4 correlates with the previously-described fusion contribution of HA2 aggregation at pH 5.0.

#### **3.4.4 Reduced fusion with mutants**

G1E or I173E- HA-cells do not fuse with RBC's (i.e. negligible hemifusion and pore formation), with very similar losses-of-function of cell/RBC fusion induced by exogenous FHA2 with either of these mutations.<sup>10, 13, 28</sup> There are qualitatively-similar but less dramatic effects of these mutants on HA2-induced PC and PC:PG vesicle fusion (Fig. 3.10). The ~4-fold reduction in fusion extent for G1E- vs. WT- HA2 with PC vesicles matches well with the reduction previously observed for G1E- vs. WT- FHA2 with PC:cholesterol vesicles.<sup>28</sup> Some mutant HA2-induced vesicle fusion vs. no mutant HA-cell/RBC fusion is ascribed to the ~10× higher density of HA2-trimers/vesicle vs. HA-trimers/cell.<sup>37, 38</sup> The importance of HA density for fusion is supported by



the correlation between comparable fusion of G1S- and WT- virions with vesicles *vs.* reduced G1S- relative to WT- HA-cell/RBC fusion, and the  $\sim 50\times$  higher density of HA-trimers/virion *vs.* HA-trimers/cell.<sup>60</sup>

For PC:DOTAP vesicles at pH 7.4, there is reduced fusion of mutant- *vs.* WT- HA2, which is qualitatively-similar to the PC and PC:PG results. The opposite trend is observed for PC:DOTAP at pH 5.0, with respective comparable and higher fusion extents of G1E- and I173E- *vs.* WT- HA2. This trend correlates with mutant- *vs.* WT- HA2 total charges of +2 *vs.* +3, and corresponding weaker mutant- *vs.* WT- HA2/DOTAP electrostatic repulsion.

We focus on PC and PC:PG vesicle fusion at pH 5.0 because these lipid compositions and this pH reflect some of the conditions of HA-mediated virus/endosome fusion. Relative to WT- HA2, G1E- and I173E- HA2 exhibit similar reductions in fusion that correlate with their lower SE hairpin stabilities. A more-open hairpin could reduce fusion via the larger separation of vesicles bound to the terminal FP and TM segments. In addition, the mutant HA2's may have less FP/membrane interaction due to higher FP interaction with the dissociated C-terminal SE region (Fig. 3.11). Reduced I173E-HA2 fusion may also be associated with its larger monomer:trimer ratio (Figs. 3.4 and 3A). Relative to trimer, monomer protein has lower local concentrations of the FP and TM in the membrane and consequently less membrane perturbation. This hypothesis is supported by earlier observations of reduced vesicle fusion for monomer *vs.* trimer hairpin constructs of the HIV gp41 fusion protein.<sup>47, 52, 61</sup> Gp41 is non-homologous with HA2 but the two proteins share some structural similarities, including a final hairpin SE state. Reduced fusion of the G1E FP of HA2 may also be correlated to shallower membrane insertion of the N-terminal FP.<sup>29</sup>

### **3.4.5 Common enveloped virus fusion model with significant functions for hairpin trimers and monomers**

We integrate the present and earlier results into a common model for HA-mediated virus/endosome fusion, as well as fusion by other enveloped viruses. The initial trimeric HA1/HA2 complex is the lowest free-energy state at pH 7.4, whereas at pH 5-6 of the mature endosome, the complex is metastable.<sup>62</sup> The HA1 subunits move away from the HA2 subunits, which then rearrange to the final trimer-of-hairpins state, which is the predominant structural state of the WT-HA2 in our experiments. There are accompanying membrane intermediates during fusion that include the hemifusion diaphragm, pore formation, and pore expansion. In some contrast, there isn't yet imaging that temporally correlates the HA-structural changes and membrane changes during fusion.

It is sometimes proposed that some of the free energy released during the HA2 transformation to the final hairpin state is converted to activation energy and perhaps also thermodynamic energy needed for membrane intermediate states.<sup>2, 4</sup> The pictures usually show a HA2 “pre-hairpin intermediate” with fully-extended SE and with FP inserted in the target membrane that then folds into the final hairpin state, with membrane intermediate formation time-correlated with the evolution of hairpin closure. The final step of fusion pore expansion is sometimes pictured as requiring a transmembrane bundle of three TM segments and sometimes an additional three FP segments. The SE hairpin is commonly referred to as the “post-fusion” state which implies that fusion is completed before the hairpin is formed. This mechanical picture of coupling of HA2 structural changes to lipid rearrangements is visually-appealing, but is disfavored by the entropic penalties for such a coordinated mechanism.

We propose an alternate model in which FP binds to the target membrane during the HA2

structural transformation to the final hairpin state, and subsequent membrane fusion occurs with HA2 in this final state. The SE hairpin maintains membrane apposition during the ~1 minute fusion time, and clustering of hairpin trimers and therefore FP domains in the target membrane perturbs this membrane and lowers the activation barrier to achieve hemifusion and pore formation.<sup>11, 26, 47,</sup>  
<sup>52</sup> For this model, the SE hairpin is a fusogenic state rather than a post-fusion state.

This model is supported by the WT-HA2-induced vesicle fusion at pH 5.0, which is observed with <100 HA2 trimers/vesicle, less than the ~400 HA trimers/influenza virion.<sup>37, 38</sup> The SE hairpins may be more stable at pH 5.0 vs. 7.4, based on the higher  $T_m$  for the HA2<sub>38-89</sub> peptide at pH 5.0 vs. 7.0, and greater HA2 clustering at low pH is evidenced by visible aggregation of the SE constructs.<sup>26, 43</sup> The fusion relevance of the final hairpin state is also supported by observation of cell/RBC fusion after addition of exogenous FHA2. Such fusion is very similar to HA-cell/RBC fusion, with both fusion types requiring low-pH and exhibiting nearly-identical losses-of-function for the G1E mutant and for the I173E mutant.<sup>10, 28, 63</sup> These mutants also show reduced vesicle fusion (Fig. 3.10). Reduced cell/cell and vesicle fusion for I173E- and G1E- vs. WT- proteins correlates with lower SE hairpin stability for the mutants and consequent impaired membrane apposition. There may also be less membrane perturbation by the mutants because of FP binding with the C-terminal strand of the SE, and for G1E, because of shallower insertion of the FP in the target membrane.

This model can also describe gp41-mediated fusion between the HIV membrane and the plasma membrane of a host-cell. Gp41 forms an initial trimeric gp160 complex with a receptor-binding protein (gp120) that is analogous with the trimeric HA1/HA2 complex.<sup>64, 65</sup> Fusion occurs via gp120-binding to host-cell receptors, separation of gp120 from gp41, and structural transformation of gp41 into a final trimer-of-hairpins state.<sup>66, 67</sup> A FP+SE+TM gp41 construct in

this final state catalyzes fusion at pH 7.4 between vesicles composed of PC:Cholesterol or PC:PG:Cholesterol.<sup>52</sup> This is the pH of HIV/host-cell fusion, and there are significant fractions of PC and anionic lipids in the plasma membranes of cells infected by HIV.<sup>68</sup> Gp160-cells also fuse with receptor-bearing cells, and the V2E mutation at the N-terminus of gp41 results in highly-impaired V2E- vs. WT- gp160-cell/receptor-cell fusion.<sup>69</sup> This observation correlates with highly-impaired G1E- vs. WT- HA-cell/RBC fusion.

A monomer fraction for HA2 in SEC correlates with monomer fractions of the initial HA1/HA2 complex under some conditions, as well as large monomer fractions for gp41 under some conditions.<sup>47, 48, 50, 52, 56, 57, 70</sup> For both HA2 and gp41, there is a significant structural rearrangement between the initial trimeric HA1/HA2 or gp160 complex, and the final trimer-of-hairpins without the receptor proteins. Transient dissociation of HA2 and gp41 into monomers may be functionally important because monomer rearrangement into a hairpin followed by association into a trimer-of-hairpins may be topologically more straightforward than a concerted rearrangement of trimeric protein.

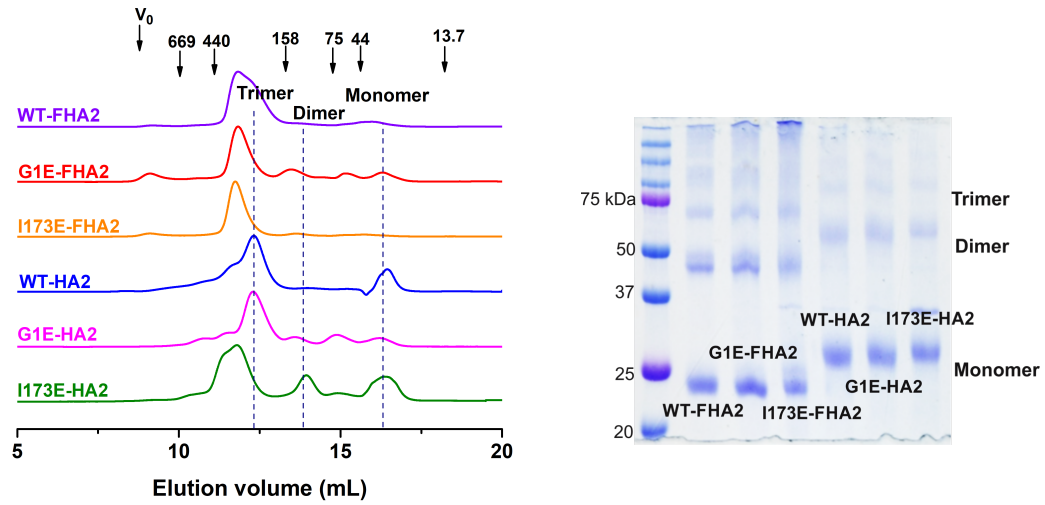
Monomer intermediates have also been postulated for the structural transformations of class II and class III enveloped virus fusion proteins.<sup>3, 70</sup> These hypotheses are evidenced by: (1) the initial class II protein complex is a single heterodimer of E1+E2 subunits, whereas the final state is a trimer-of-hairpins of E1 subunits; and (2) one class III structure contains two different monomer units, with one monomer similar to the initial state and the other similar to the final state. Monomer gp41 may also be the binding target of peptides which inhibit gp160-mediated fusion and HIV infection, and whose sequences correspond to segments of the C-terminal region of the gp41 SE.<sup>52, 63</sup>

### **3.5 Summary**

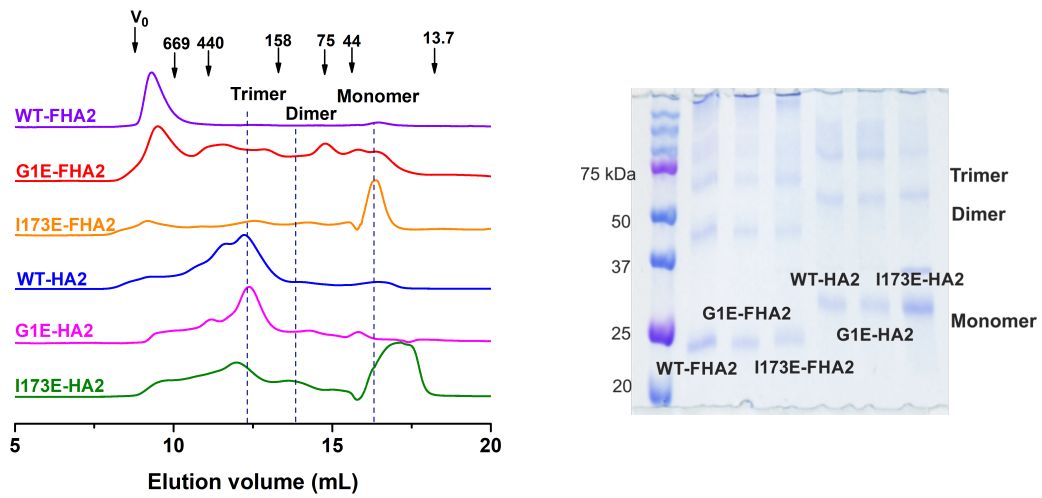
The present study reports protein-induced vesicle fusion by the full-length HA2 containing the FP and the TM, and reduced fusion by the G1E and I173E mutants. These mutants also result in reduced HA-mediated cell fusion. The present and previous studies support that the HA2 final trimer-of-hairpins is fusogenic. This implies common features of the influenza virus/endosome fusion and HA2/vesicle fusion that include: (1) apposing the two fusing compartments by folded hairpin HA2 proteins and maintaining this apposition with a stable SE; (2) binding of the FP to the target membrane; and (3) HA2 aggregation at low pH and consequent clustering of membrane-perturbing FP's.

## APPENDIX

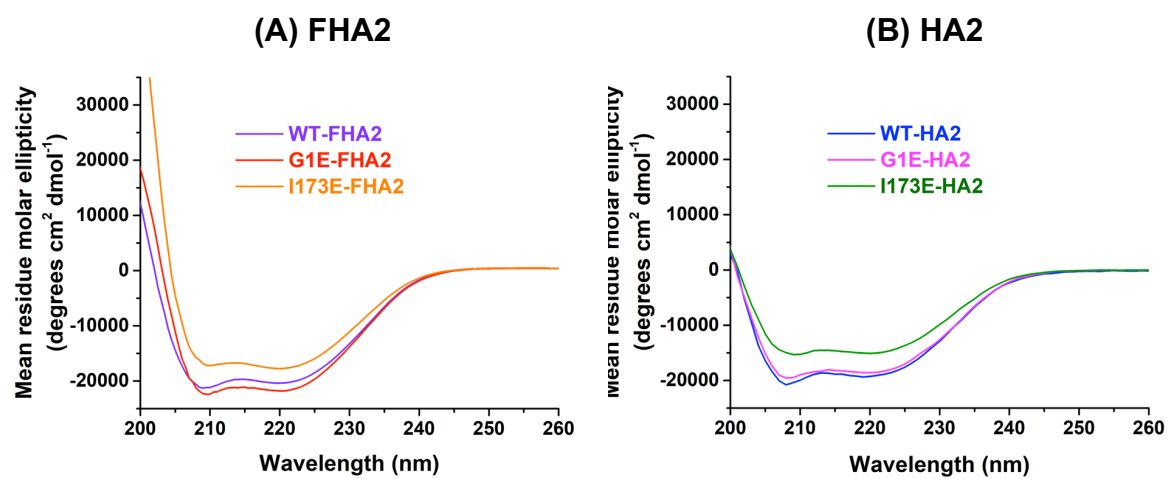
**(A) 0.10% SRC detergent**



**(B) 0.17% DM detergent**

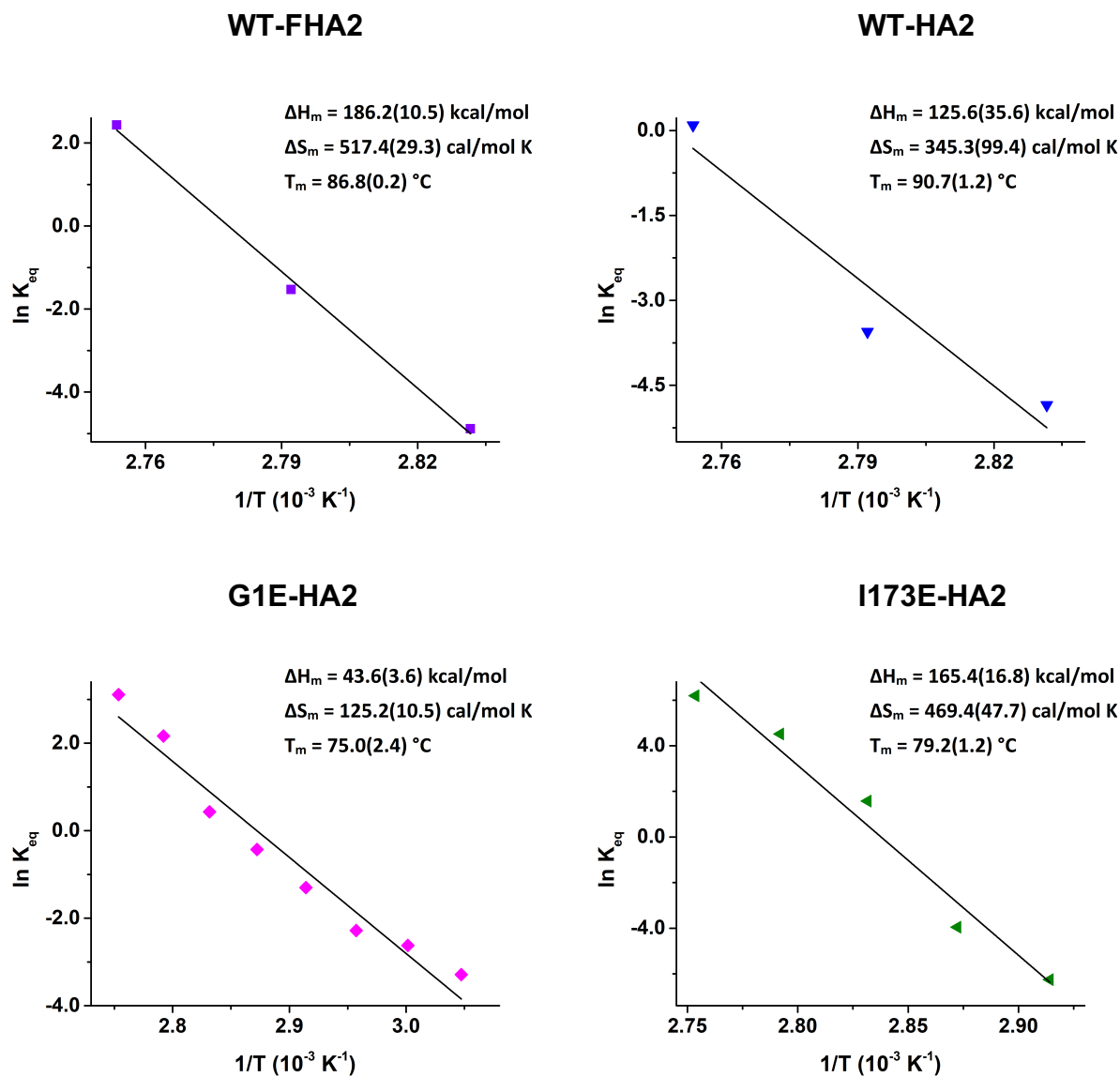


**Figure 3A** SEC traces prior to cross-linking (left) of replicate samples and SDS-PAGE after cross-linking (right) of replicate samples and in (A) SRC and (B) DM detergents (compare to Fig. 3.4).

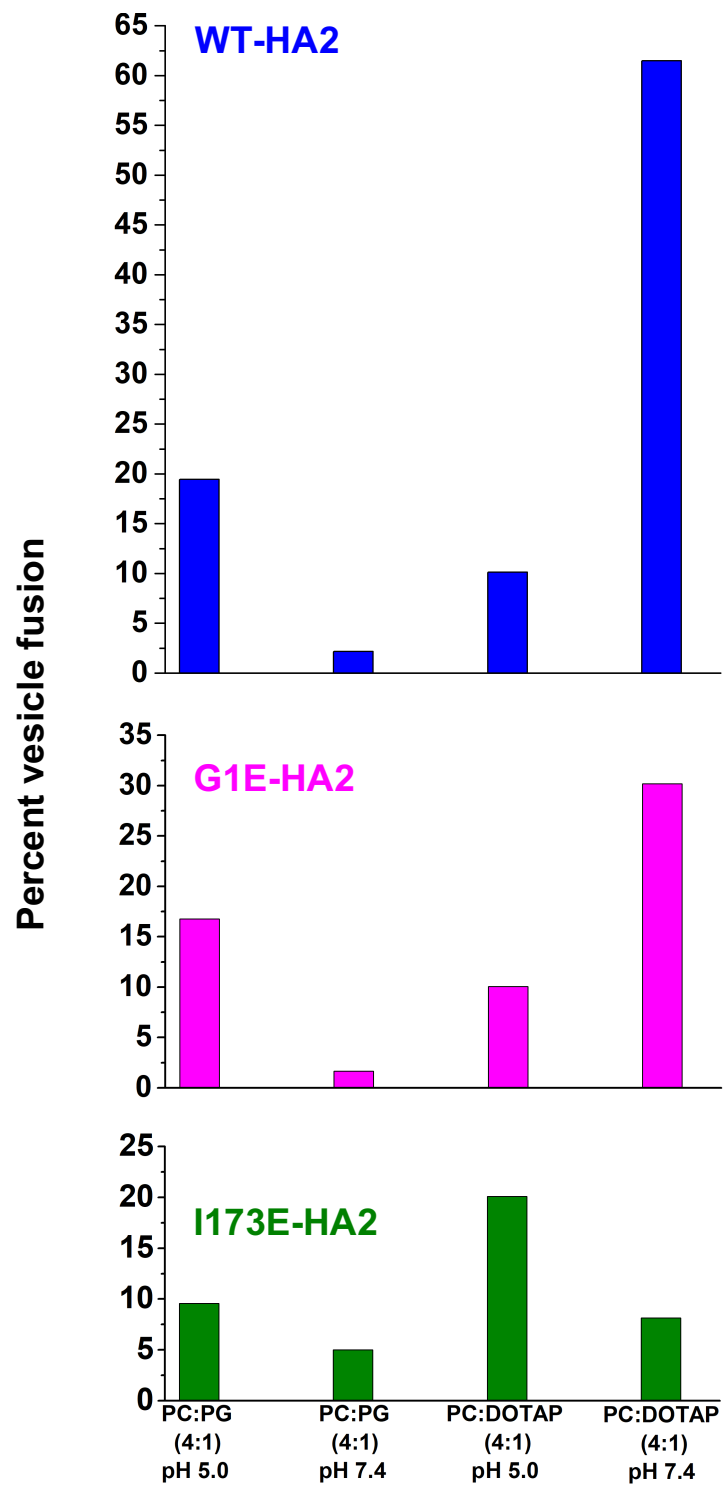


**Figure 3B** CD spectra of replicates samples (compare to Fig. 3.5).





**Figure 3C** van't Hoff plots of the unfolding  $\ln K_{eq}$  vs.  $1/T$  based on  $\theta_{222}$  data for the temperature range around  $T_m$ . Best-fit parameters are given with uncertainties in parentheses.



**Figure 3D** Extents of HA2-induced vesicle fusion using different preparations of proteins and vesicles (compare to Fig. 3.10).

## REFERENCES

## REFERENCES

1. Fontana, J., and Steven, A. C. (2015) Influenza virus-mediated membrane fusion: Structural insights from electron microscopy, *Arch. Biochem. Biophys.* 581, 86-97.
2. White, J. M., Delos, S. E., Brecher, M., and Schornberg, K. (2008) Structures and mechanisms of viral membrane fusion proteins: Multiple variations on a common theme, *Crit. Rev. Biochem. Mol. Biol.* 43, 189-219.
3. Kielian, M. (2014) Mechanisms of virus membrane fusion proteins, *Annual Rev. Virol.* 1, 171-189.
4. Harrison, S. C. (2015) Viral membrane fusion, *Virology* 479, 498-507.
5. Wilson, I. A., Skehel, J. J., and Wiley, D. C. (1981) Structure of the haemagglutinin membrane glycoprotein of influenza virus at 3 Å resolution, *Nature* 289, 366-373.
6. Ekiert, D. C., Kashyap, A. K., Steel, J., Rubrum, A., Bhabha, G., Khayat, R., Lee, J. H., Dillon, M. A., O'Neil, R. E., Faynboym, A. M., Horowitz, M., Horowitz, L., Ward, A. B., Palese, P., Webby, R., Lerner, R. A., Bhatt, R. R., and Wilson, I. A. (2012) Cross-neutralization of influenza A viruses mediated by a single antibody loop, *Nature* 489, 526-532.
7. Bullough, P. A., Hughson, F. M., Skehel, J. J., and Wiley, D. C. (1994) Structure of influenza haemagglutinin at the pH of membrane fusion, *Nature* 371, 37-43.
8. Chen, J., Skehel, J. J., and Wiley, D. C. (1999) N- and C-terminal residues combine in the fusion-pH influenza hemagglutinin HA<sub>2</sub> subunit to form an N cap that terminates the triple-stranded coiled coil, *Proc. Natl. Acad. Sci. U.S.A.* 96, 8967-8972.
9. Chernomordik, L. V., Frolov, V. A., Leikina, E., Bronk, P., and Zimmerberg, J. (1998) The pathway of membrane fusion catalyzed by influenza hemagglutinin: Restriction of lipids, hemifusion, and lipidic fusion pore formation, *J. Cell Biol.* 140, 1369-1382.
10. Qiao, H., Armstrong, R. T., Melikyan, G. B., Cohen, F. S., and White, J. M. (1999) A specific point mutant at position 1 of the influenza hemagglutinin fusion peptide displays a hemifusion phenotype, *Mol. Biol. Cell* 10, 2759-2769.
11. Blijleven, J. S., Boonstra, S., Onck, P. R., van der Giessen, E., and van Oijen, A. M. (2016) Mechanisms of influenza viral membrane fusion, *Sem. Cell Dev. Biol.* 60, 78-88.
12. Park, H. E., Gruenke, J. A., and White, J. M. (2003) Leash in the groove mechanism of membrane fusion, *Nature Struct. Biol.* 10, 1048-1053.

13. Borrego-Diaz, E., Peeples, M. E., Markosyan, R. M., Melikyan, G. B., and Cohen, F. S. (2003) Completion of trimeric hairpin formation of influenza virus hemagglutinin promotes fusion pore opening and enlargement, *Virology* 316, 234-244.
14. Nobusawa, E., Aoyama, T., Kato, H., Suzuki, Y., Tateno, Y., and Nakajima, K. (1991) Comparison of complete amino acid sequences and receptor binding properties among 13 serotypes of hemagglutinins of influenza A viruses, *Virology* 182, 475-485.
15. Cross, K. J., Langley, W. A., Russell, R. J., Skehel, J. J., and Steinhauer, D. A. (2009) Composition and functions of the Influenza fusion peptide, *Prot. Pep. Lett.* 16, 766-778.
16. Lorieau, J. L., Louis, J. M., and Bax, A. (2010) The complete influenza hemagglutinin fusion domain adopts a tight helical hairpin arrangement at the lipid:water interface, *Proc. Natl. Acad. Sci. U.S.A.* 107, 11341-11346.
17. Du, T. P., Jiang, L., and Liu, M. L. (2014) NMR structures of fusion peptide from influenza hemagglutinin H3 subtype and its mutants, *J. Peptide Sci.* 20, 292-297.
18. Ghosh, U., Xie, L., and Weliky, D. P. (2013) Detection of closed influenza virus hemagglutinin fusion peptide structures in membranes by backbone  $^{13}\text{CO}$ - $^{15}\text{N}$  rotational-echo double-resonance solid-state NMR, *J. Biomol. NMR* 55, 139-146.
19. Ghosh, U., Xie, L., Jia, L. H., Liang, S., and Weliky, D. P. (2015) Closed and semiclosed interhelical structures in membrane vs closed and open structures in detergent for the Influenza Virus hemagglutinin fusion peptide and correlation of hydrophobic surface area with fusion catalysis, *J. Am. Chem. Soc.* 137, 7548-7551.
20. Macosko, J. C., Kim, C. H., and Shin, Y. K. (1997) The membrane topology of the fusion peptide region of influenza hemagglutinin determined by spin-labeling EPR, *J. Mol. Biol.* 267, 1139-1148.
21. Armstrong, R. T., Kushnir, A. S., and White, J. M. (2000) The transmembrane domain of influenza hemagglutinin exhibits a stringent length requirement to support the hemifusion to fusion transition, *J. Cell Biol.* 151, 425-437.
22. Markosyan, R. M., Cohen, F. S., and Melikyan, G. B. (2000) The lipid-anchored ectodomain of influenza virus hemagglutinin (GPI-HA) is capable of inducing nonenlarging fusion pores, *Mol. Biol. Cell* 11, 1143-1152.
23. Melikyan, G. B., Lin, S. S., Roth, M. G., and Cohen, F. S. (1999) Amino acid sequence requirements of the transmembrane and cytoplasmic domains of influenza virus hemagglutinin for viable membrane fusion, *Mol. Biol. Cell* 10, 1821-1836.
24. Tatulian, S. A., and Tamm, L. K. (2000) Secondary structure, orientation, oligomerization, and lipid interactions of the transmembrane domain of influenza hemagglutinin, *Biochemistry* 39, 496-507.

25. Chen, J., Skehel, J. J., and Wiley, D. C. (1998) A polar octapeptide fused to the N-terminal fusion peptide solubilizes the influenza virus HA(2) subunit ectodomain, *Biochemistry* 37, 13643-13649.
26. Ratnayake, P. U., Ekanayaka, E. A. P., Komanduru, S. S., and Weliky, D. P. (2016) Full-length trimeric influenza virus hemagglutinin II membrane fusion protein and shorter constructs lacking the fusion peptide or transmembrane domain: Hyperthermostability of the full-length protein and the soluble ectodomain and fusion peptide make significant contributions to fusion of membrane vesicles, *Protein Expression Purif.* 117, 6-16.
27. Leikina, E., LeDuc, D. L., Macosko, J. C., Epand, R., Shin, Y. K., and Chernomordik, L. V. (2001) The 1-127 HA2 construct of influenza virus hemagglutinin induces cell-cell hemifusion, *Biochemistry* 40, 8378-8386.
28. Kim, C. S., Epand, R. F., Leikina, E., Epand, R. M., and Chernomordik, L. V. (2011) The final conformation of the complete ectodomain of the HA2 subunit of Influenza Hemagglutinin can by itself drive low pH-dependent fusion, *J. Biol. Chem.* 286, 13226-13234.
29. Wu, C. W., Cheng, S. F., Huang, W. N., Trivedi, V. D., Veeramuthu, B., Kantchev, A. B., Wu, W. G., and Chang, D. K. (2003) Effects of alterations of the amino-terminal glycine of influenza hemagglutinin fusion peptide on its structure, organization and membrane interactions, *Biochim. Biophys. Acta* 1612, 41-51.
30. Swalley, S. E., Baker, B. M., Calder, L. J., Harrison, S. C., Skehel, J. J., and Wiley, D. C. (2004) Full-length influenza hemagglutinin HA<sub>2</sub> refolds into the trimeric low-pH-induced conformation, *Biochemistry* 43, 5902-5911.
31. Curtis-Fisk, J., Spencer, R. M., and Weliky, D. P. (2008) Isotopically labeled expression in *E. coli*, purification, and refolding of the full ectodomain of the Influenza virus membrane fusion protein, *Protein Expression Purif.* 61, 212-219.
32. Gerl, M. J., Sampaio, J. L., Urban, S., Kalvodova, L., Verbavatz, J. M., Binnington, B., Lindemann, D., Lingwood, C. A., Shevchenko, A., Schroeder, C., and Simons, K. (2012) Quantitative analysis of the lipidomes of the influenza virus envelope and MDCK cell apical membrane, *J. Cell Biol.* 196, 213-221.
33. Ivanova, P. T., Myers, D. S., Milne, S. B., McClaren, J. L., Thomas, P. G., and Brown, H. A. (2015) Lipid composition of the viral envelope of three strains of Influenza virus-Not all viruses are created equal, *ACS Infect. Dis.* 1, 435-442.
34. Kobayashi, T., Beuchat, M. H., Chevallier, J., Makino, A., Mayran, N., Escola, J. M., Lebrand, C., Cosson, P., Kobayashi, T., and Gruenberg, J. (2002) Separation and characterization of late endosomal membrane domains, *J. Biol. Chem.* 277, 32157-32164.
35. Mobius, W., van Donselaar, E., Ohno-Iwashita, Y., Shimada, Y., Heijnen, H. F. G., Slot, J. W., and Geuze, H. J. (2003) Recycling compartments and the internal vesicles of

- multivesicular bodies harbor most of the cholesterol found in the endocytic pathway, *Traffic* 4, 222-231.
36. Shangguan, T., Alford, D., and Bentz, J. (1996) Influenza-virus-liposome lipid mixing is leaky and largely insensitive to the material properties of the target membrane, *Biochemistry* 35, 4956-4965.
  37. Harris, A., Cardone, G., Winkler, D. C., Heymann, J. B., Brecher, M., White, J. M., and Steven, A. C. (2006) Influenza virus pleiomorphy characterized by cryoelectron tomography, *Proc. Natl. Acad. Sci. U.S.A.* 103, 19123-19127.
  38. Yamaguchi, M., Danev, R., Nishlyama, K., Sugawara, K., and Nagayama, K. (2008) Zernike phase contrast electron microscopy of ice-embedded influenza A virus, *J. Struct. Biol.* 162, 271-276.
  39. Epand, R. F., Macosko, J. C., Russell, C. J., Shin, Y. K., and Epand, R. M. (1999) The ectodomain of HA2 of influenza virus promotes rapid pH dependent membrane fusion, *J. Mol. Biol.* 286, 489-503.
  40. Yang, J., Gabrys, C. M., and Weliky, D. P. (2001) Solid-state nuclear magnetic resonance evidence for an extended beta strand conformation of the membrane-bound HIV-1 fusion peptide, *Biochemistry* 40, 8126-8137.
  41. Curtis-Fisk, J., Preston, C., Zheng, Z. X., Worden, R. M., and Weliky, D. P. (2007) Solid-state NMR structural measurements on the membrane-associated influenza fusion protein ectodomain, *J. Am. Chem. Soc.* 129, 11320-11321.
  42. Durrer, P., Galli, C., Hoenke, S., Corti, C., Gluck, R., Vorherr, T., and Brunner, J. (1996) H<sup>+</sup>-induced membrane insertion of influenza virus hemagglutinin involves the HA2 amino-terminal fusion peptide but not the coiled coil region, *J. Biol. Chem.* 271, 13417-13421.
  43. Carr, C. M., and Kim, P. S. (1993) A spring-loaded mechanism for the conformational change of influenza hemagglutinin, *Cell* 73, 823-832.
  44. Kim, C. H., Macosko, J. C., Yu, Y. G., and Shin, Y. K. (1996) On the dynamics and confirmation of the HA2 domain of the influenza virus hemagglutinin, *Biochemistry* 35, 5359-5365.
  45. Lorieau, J. L., Louis, J. M., and Bax, A. (2011) Helical hairpin structure of Influenza Hemagglutinin fusion peptide stabilized by charge-dipole interactions between the N-terminal amino group and the second helix, *J. Am. Chem. Soc.* 133, 2824-2827.
  46. Murphy, K. P., and Freire, E. (1992) Thermodynamics of structural stability and cooperative folding behavior in proteins, *Adv. Prot. Chem.* 43, 313-361.
  47. Banerjee, K., and Weliky, D. P. (2014) Folded monomers and hexamers of the ectodomain of the HIV gp41 membrane fusion protein: Potential roles in fusion and synergy between

- the fusion peptide, hairpin, and membrane-proximal external region, *Biochemistry* 53, 7184-7198.
48. Roche, J., Louis, J. M., Grishaev, A., Ying, J. F., and Bax, A. (2014) Dissociation of the trimeric gp41 ectodomain at the lipid-water interface suggests an active role in HIV-1 Env-mediated membrane fusion, *Proc. Natl. Acad. Sci. U.S.A.* 111, 3425-3430.
  49. Cho, K. J., Lee, J.-H., Hong, K. W., Kim, S.-H., Park, Y., Lee, J. Y., Kang, S., Kim, S., Yang, J. H., Kim, E.-K., Seok, J. H., Unzai, S., Park, S. Y., Saelens, X., Kim, C.-J., Lee, J.-Y., Kang, C., Oh, H.-B., Chung, M. S., and Kim, K. H. (2013) Insight into structural diversity of influenza virus haemagglutinin, *J. Gen. Virol.* 94, 1712-1722.
  50. Dai, Z., Tao, Y. S., Liu, N. N., Brenowitz, M. D., Girvin, M. E., and Lai, J. R. (2015) Conditional trimerization and lytic activity of HIV-1 gp41 variants containing the membrane-associated segments, *Biochemistry* 54, 1589-1599.
  51. Joyce, M. G., Wheatley, A. K., Thomas, P. V., Chuang, G.-Y., Soto, C., Bailer, R. T., Druz, A., Georgiev, I. S., Gillespie, R. A., Kanekiyo, M., Kong, W.-P., Leung, K., Narpala, S. N., Prabhakaran, M. S., Yang, E. S., Zhang, B., Zhang, Y., Asokan, M., Boyington, J. C., Bylund, T., Darko, S., Lees, C. R., Ransier, A., Shen, C.-H., Wang, L., Whittle, J. R., Wu, X., Yassine, H. M., Santos, C., Matsuoka, Y., Tsybovsky, Y., Baxa, U., Mullikin, J. C., Subbarao, K., Douek, D. C., Graham, B. S., Koup, R. A., Ledgerwood, J. E., Roederer, M., Shapiro, L., Kwong, P. D., Mascola, J. R., McDermott, A. B., and Sequencing, N. C. (2016) Vaccine-induced antibodies that neutralize group 1 and group 2 Influenza A viruses, *Cell* 166, 609-623.
  52. Liang, S., Ratnayake, P. U., Keinath, C., Jia, L., Wolfe, R., Ranaweera, A., and Weliky, D. P. (2018) Efficient fusion at neutral pH by Human Immunodeficiency Virus gp41 trimers containing the fusion peptide and transmembrane domains, *Biochemistry* 57, 1219-1235.
  53. Xiong, X., Corti, D., Liu, J., Pinna, D., Foglierini, M., Calder, L. J., Martin, S. R., Lin, Y. P., Walker, P. A., Collins, P. J., Monne, I., Suguitan, A. L., Jr., Santos, C., Temperton, N. J., Subbarao, K., Lanzavecchia, A., Gamblin, S. J., and Skehel, J. J. (2015) Structures of complexes formed by H5 influenza hemagglutinin with a potent broadly neutralizing human monoclonal antibody, *Proc. Natl. Acad. Sci. U.S.A.* 112, 9430-9435.
  54. Seok, J. H., Kim, J., Lee, D. B., Cho, K. J., Lee, J.-H., Bae, G., Chung, M. S., and Kim, K. H. (2017) Conformational modulation of influenza virus hemagglutinin: characterization and in vivo efficacy of monomeric form, *Sci. Rep.* 7.
  55. Raymond, D. D., Stewart, S. M., Lee, J., Ferdman, J., Bajic, G., Do, K. T., Ernandes, M. J., Suphaphiphat, P., Settembre, E. C., Dormitzer, P. R., Del Giudice, G., Finco, O., Kang, T. H., Ippolito, G. C., Georgiou, G., Kepler, T. B., Haynes, B. F., Moody, M. A., Liao, H.-X., Schmidt, A. G., and Harrison, S. C. (2016) Influenza immunization elicits antibodies specific for an egg-adapted vaccine strain, *Nature Med.* 22, 1465-+.



56. Lakomek, N. A., Kaufman, J. D., Stahl, S. J., Louis, J. M., Grishaev, A., Wingfield, P. T., and Bax, A. (2013) Internal dynamics of the homotrimeric HIV-1 viral coat protein gp41 on multiple time scales, *Angew. Chem. Int. Ed.* 52, 3911-3915.
57. Chiliveri, S. C., Louis, J. M., Ghirlando, R., Baber, J. L., and Bax, A. (2018) Tilted, uninterrupted, monomeric HIV-1 gp41 transmembrane helix from residual dipolar couplings, *J. Am. Chem. Soc.* 140, 34-37.
58. Balakrishnan, K., Krishnan, N. M., and Rao, B. J. (2009) Salt modulates oligomerization properties of hRad51 and hRad52 proteins, *Open Biol. J.* 2, 1-9.
59. Politi, L., Chiancone, E., Giangiacomo, L., Cervoni, L., D'Abusco, A. S., Scorsino, S., and Scandurra, R. (2009) pH-, temperature- and ion-dependent oligomerization of *Sulfolobus solfataricus* recombinant amidase: a study with site-specific mutants, *Archaea* 2, 221-231.
60. Zawada, K. E., Okamoto, K., and Kasson, P. M. (2018) Influenza hemifusion phenotype depends on membrane context: Differences in cell-cell and virus-cell fusion, *J. Mol. Biol.* 430, 594-601.
61. Ratnayake, P. U., Sackett, K., Nethercott, M. J., and Weliky, D. P. (2015) pH-dependent vesicle fusion induced by the ectodomain of the human immunodeficiency virus membrane fusion protein gp41: Two kinetically distinct processes and fully-membrane-associated gp41 with predominant beta sheet fusion peptide conformation, *Biochim. Biophys. Acta* 1848, 289-298.
62. Epand, R. F., and Epand, R. M. (2003) Irreversible unfolding of the neutral pH form of influenza hemagglutinin demonstrates that it is not in a metastable state, *Biochemistry* 42, 5052-5057.
63. Markosyan, R. M., Cohen, F. S., and Melikyan, G. B. (2003) HIV-1 envelope proteins complete their folding into six-helix bundles immediately after fusion pore formation, *Mol. Biol. Cell* 14, 926-938.
64. Pancera, M., Zhou, T. Q., Druz, A., Georgiev, I. S., Soto, C., Gorman, J., Huang, J. H., Acharya, P., Chuang, G. Y., Ofek, G., Stewart-Jones, G. B. E., Stuckey, J., Bailer, R. T., Joyce, M. G., Louder, M. K., Tumba, N., Yang, Y. P., Zhang, B. S., Cohen, M. S., Haynes, B. F., Mascola, J. R., Morris, L., Munro, J. B., Blanchard, S. C., Mothes, W., Connors, M., and Kwong, P. D. (2014) Structure and immune recognition of trimeric pre-fusion HIV-1 Env, *Nature* 514, 455-461.
65. Garces, F., Lee, J. H., de Val, N., de la Pena, A. T., Kong, L., Puchades, C., Hua, Y., Stanfield, R. L., Burton, D. R., Moore, J. P., Sanders, R. W., Ward, A. B., and Wilson, I. A. (2015) Affinity maturation of a potent family of HIV antibodies is primarily focused on accommodating or avoiding glycans, *Immunity* 43, 1053-1063.
66. Caffrey, M., Cai, M., Kaufman, J., Stahl, S. J., Wingfield, P. T., Covell, D. G., Gronenborn, A. M., and Clore, G. M. (1998) Three-dimensional solution structure of the 44 kDa ectodomain of SIV gp41, *EMBO J.* 17, 4572-4584.

67. Yang, Z. N., Mueser, T. C., Kaufman, J., Stahl, S. J., Wingfield, P. T., and Hyde, C. C. (1999) The crystal structure of the SIV gp41 ectodomain at 1.47 Å resolution, *J. Struct. Biol.* 126, 131-144.
68. Lorizate, M., Sachsenheimer, T., Glass, B., Habermann, A., Gerl, M. J., Krausslich, H. G., and Brugger, B. (2013) Comparative lipidomics analysis of HIV-1 particles and their producer cell membrane in different cell lines, *Cellular Microbiology* 15, 292-304.
69. Freed, E. O., Delwart, E. L., Buchschacher, G. L., Jr., and Panganiban, A. T. (1992) A mutation in the human immunodeficiency virus type 1 transmembrane glycoprotein gp41 dominantly interferes with fusion and infectivity, *Proc. Natl. Acad. Sci. U.S.A.* 89, 70-74.
70. Abou-Hamdan, A., Belot, L., Albertini, A., and Gaudin, Y. (2018) Monomeric intermediates formed by Vesiculovirus glycoprotein during its low-pH-induced structural transition, *J. Mol. Biol.* 430, 1685-1695.

## **Chapter 4**

### **Structure-Function Relationship of HA2 by Hydrogen-Deuterium**

#### **Exchange Mass Spectrometry**

## 4.1 Introduction

Influenza is an enveloped virus enclosed by a membrane obtained during budding from an infected host cell.<sup>1-3</sup> The viral protein hemagglutinin (HA) is responsible for attachment of the virus to an uninfected host cell and subsequent membrane fusion within the cell. HA is synthesized as a single polypeptide, and then proteolytically cleaved into two disulfide-linked subunits, HA1 and HA2, with respective lengths of ~330 and ~220 residues.<sup>4</sup> HA2 is a single-pass integral membrane protein with an ectodomain (ED) that contains a N-terminal fusion peptide (FP) and soluble ectodomain (SE), a transmembrane domain (TM) and endodomain with respective lengths of ~25, 160, 25 and 10 residues (Fig. 4.1A). An initial complex is formed with a HA2 trimeric ED bundle and three HA1's bound to the outside of this bundle.<sup>5, 6</sup> The HA1's bind to sialic acids on the host cell surface which triggers endocytosis. The early endosome with virus cargo migrates to the host cell nucleus and matures with reduction of its pH to <6, which induces HA1/HA2 separation and HA2 SE rearrangement into a final trimer-of-hairpins structure (Fig. 4.1B).<sup>7, 8</sup>

Reduced pH also triggers HA-mediated membrane fusion with stages that include: (1) hemifusion in which there is a single bilayer diaphragm that separates the two bodies; (2) pore formation that allows passage of <1 kDa species; and (3) pore expansion that allows deposition of the virus capsid in the cytoplasm.<sup>3, 9</sup> Most data have been derived for pH-triggered fusion between HA-expressing cells (HA-cells) and red blood cells (RBC's).<sup>10, 11</sup> For example, fusion is arrested at step (2) for HA-cells expressing a C-terminal truncation mutant with only a membrane anchor, which evidences a functional role of the HA2 TM in step (3).<sup>12</sup> Interestingly, full fusion is observed when the HA2 TM is exchanged for a non-homologous TM sequence from another protein, which evidences that the specific TM sequence is not important for step (3).<sup>13</sup> Another important observation is that exogenously-added HA2 FP+SE constructs induce steps (1) and (2) of "HA0-

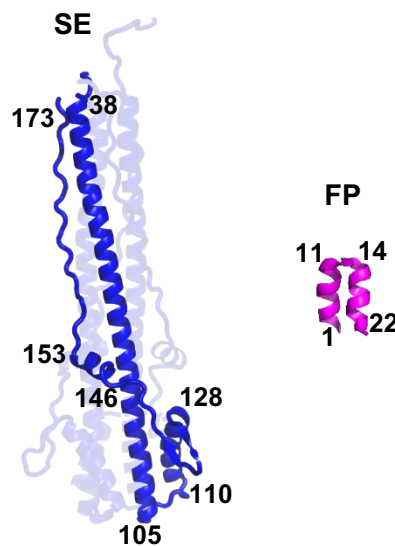
cell"/RBC fusion, where HA0 is a HA-variant that binds to RBC's but does not induce fusion.<sup>14,</sup>

15

## A Protein constructs



## B Ribbon diagrams of the soluble ectodomain and fusion peptide



**Figure 4.1** (A) Schematic diagrams of full-length HA2 and truncated constructs with domains colored: fusion peptide (FP), pink; soluble ectodomain (SE), blue; transmembrane domain (TM), green; and endodomain (Endo), orange. (B) Ribbon diagrams of the SE in the final trimer-of-hairpins structure (PDB 1QU1) and the FP with closed helical hairpin structure (PDB 2KXA). A monomer subunit is highlighted for the SE, as well as the terminal residues of regular secondary structure elements, and residue 173.

There are high-resolution structures for the initial complex of the proteolytically-cleaved HA ED which includes the HA2 ED and HA1. These structures typically show an organization of HA2 ED trimers with three bound HA1, but sometimes show a different organization of individual HA2/HA1 heterodimers.<sup>5, 6, 16, 17</sup> The geometries of the HA2 ED/HA1 units are very similar in either oligomeric organization, which supports the HA2 ED/HA1 heterodimer as a fundamental folding unit. The trimer has a mushroom shape with a stem region containing mostly HA2, and a globular head region containing mostly HA1. The stem includes three HA2<sub>76-125</sub> helices that form an internal parallel coiled-coil pointing away from the head, and three HA2<sub>38-55</sub> helices that are antiparallel to the middle of the coiled-coil. The HA2<sub>1-20</sub> FP's have irregular and extended structure buried in the stem and connected to the stem by the HA2<sub>23-35</sub>  $\beta$  sheet hairpins. The HA2<sub>130-141</sub> segments are also  $\beta$  sheet hairpins and connect the stem to the HA2<sub>146-153</sub> and HA2<sub>159-170</sub> helices.

There is a topologically-distinct structure of the HA2<sub>33-178</sub> SE without HA1 (Fig. 4.1B). This HA2 structure is hyperthermostable and likely part of the final HA2 fusion state.<sup>18, 19</sup> The structure is a trimer-of-hairpins that includes HA2<sub>38-105</sub> helices in an internal parallel coiled-coil, 180° turns, and HA2<sub>110-128</sub> helices that are antiparallel to and in the grooves of the HA2<sub>90-104</sub> region of the coiled-coil. The HA2<sub>130-141</sub>  $\beta$  sheet hairpins connect to HA2<sub>146-153</sub> helices that are perpendicular to the internal coiled-coil, and the HA2<sub>154-178</sub> segments have extended structures antiparallel with the HA2<sub>38-78</sub> segment of the coiled-coil. Differences between the HA2 SE in its initial complex with HA1 vs. final state without HA1 include: (1) the structurally distinct helices and loops of HA2<sub>38-75</sub> become part of the contiguous HA2<sub>38-105</sub> interior coiled-coil; (2) the initial HA2<sub>76-125</sub> coiled-coil breaks at residue 106 to form 180° turns and antiparallel helices; and (3) the HA2<sub>159-170</sub> helices become part of the HA2<sub>154-178</sub> extended structures. There are limited data about the transformation between the initial and final SE structures, but some data support a functional

role for the final trimer-of-hairpins state.<sup>3, 20-23</sup> In particular, exogenous addition of HA2 ED constructs in this state induce lipid mixing and small pores between the membranes of HA0-cells and bound RBC's.<sup>15</sup> There are striking similarities to HA-cell/RBC fusion, including low pH requirement and greatly reduced fusion for the G1E and I173E mutants that are respectively in the HA2 FP's and SE strands. There are similar trends with pH and mutations for vesicle fusion induced by full-length HA2 in the final hairpin structure.<sup>18, 19</sup> These integrated results support a model in which HA1 dissociation leads to FP insertion in the target membrane, SE hairpin structure and membrane apposition, and then membrane fusion. An alternate common mechanistic model is membrane fusion with HA2 in intermediate structures that are not the trimer-of-hairpins, with conversion of some of the free energy released during HA2 structural transformations into activation energy of fusion.<sup>1-3, 24</sup> For this model, the final trimer-of-hairpins state forms after fusion is completed and is therefore a “post-fusion” state.

The FP sequence is highly-conserved across viral subtypes and its importance in fusion is evidenced by it (and the TM) as the only HA segments that are deeply inserted in the membrane after virus/vesicle fusion.<sup>4, 25</sup> HA-cell/RBC fusion is also eliminated by the G1E mutation at the N-terminus of the FP.<sup>11</sup> The FP is initially in the HA2 ED/HA1 trimer in the protein interior and has extended structure, but adopts very different structures without the rest of HA2 and HA1, and in detergent-rich or lipid bilayer media.<sup>26-29</sup> These latter structures are monomers and are likely more representative of the FP's bound to membrane during fusion. For example, HA2<sub>1-23</sub> in detergent is a monomer and adopts “closed” helical hairpin structure at both low and neutral pH, with 1-12 helix/tight-turn/14-22 helix topology and antiparallel tight-packing of the two helices (Fig. 4.1B). In membrane, there is some population of the closed structure, as well as a population of a related “semi-closed” (less tightly-packed) structure, with a larger semi-closed:closed

population ratio at pH 5 vs. 7. The EPR linewidths of the spin-labeled FP segment of the trimeric HA2<sub>1-127</sub> construct bound to membrane are also consistent with non-interacting monomer FP's, and the EPR relaxation data correlate with FP helices tilted  $\sim 65^\circ$  relative to the membrane normal.<sup>30</sup> The EPR data support insertion in a single leaflet, with L2 and F3  $\sim 12$  Å from the membrane phosphorus plane.

The present study is focused on hydrogen-deuterium exchange coupled to mass spectrometry (HDX-MS), which is an approach to probe segmental structure and motion of proteins.<sup>31-34</sup> Incubation of the whole protein in D<sub>2</sub>O leads to exchange of backbone N-H for N-D, and is followed by pepsin digestion, and then MS-determination of percent HDX (D<sub>%</sub>) of the resultant peptides.<sup>35</sup> The D<sub>%</sub> of a peptide is positively-correlated with the solvent exposure, structural disorder, and motion of the corresponding segment of the intact protein. HDX-MS has previously been applied to the HA2 ED/HA1 trimer complex, with D<sub>2</sub>O incubation at pH's between 7.5 and 5.6, where the latter value corresponds to the late endosome pH prior to fusion.<sup>36</sup> The D<sub>%</sub> for incubation at pH 7.5 were generally consistent with the X-ray structures of the initial complex. For example, the FP is sequestered in this complex, which correlates with D<sub>%</sub> < 5% for HA2 peptide 4-9 with 5-minute incubation time. Dissociation of the complex at endosomal pH 5.6 correlates with much larger D<sub>%</sub>  $\approx$  50% for this peptide. In separate experiments, incubation for three hours at pH 4.9 gave D<sub>%</sub> that were consistent with the trimer-of-hairpins structure for the HA2 SE (Fig. 4.1B). For example, the expected protection of the HA2<sub>38-105</sub> internal helices in this structure correlates with D<sub>%</sub> < 2% for the HA2 53-69 and 70-87 peptides for 30-minute exchange time, and exposure of HA2<sub>106-175</sub> loops, helices, and strands correlates with larger D<sub>%</sub> > 70% of the HA2 102-109, 100-115, 139-148, 149-167, and 168-175 peptides.

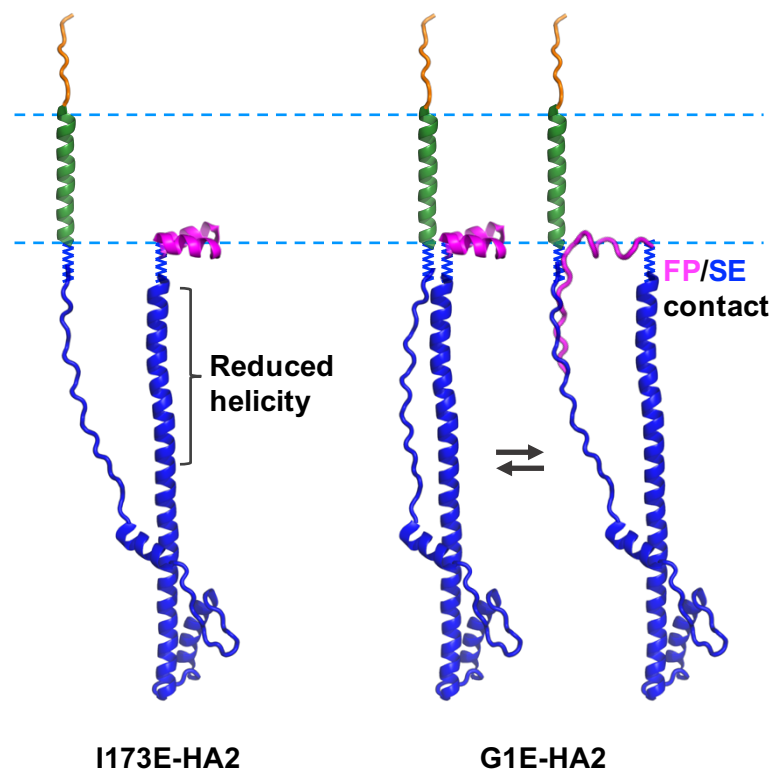
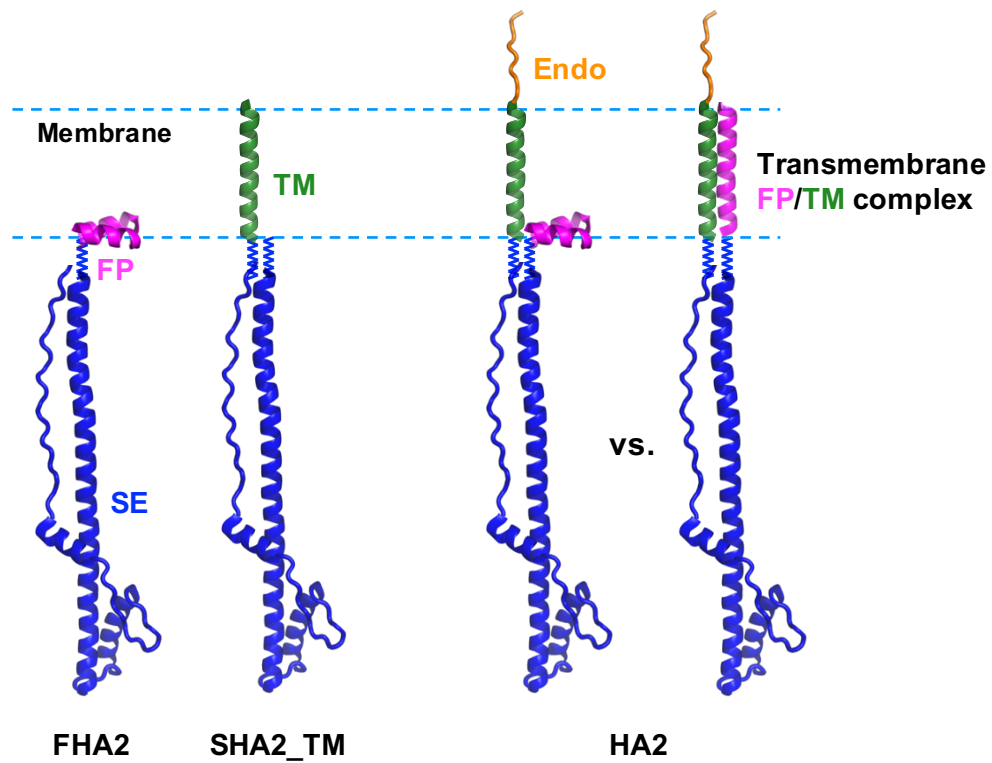


The present study describes HDX-MS results for HA2 constructs in the absence of HA1. This study builds on earlier biophysical studies, discussed in Chapter 3, of full-length HA2<sub>1-221</sub>, HA2<sub>1-185</sub>  $\equiv$  FP + SE  $\equiv$  “FHA2”, and HA2<sub>20-211</sub>  $\equiv$  SE + TM  $\equiv$  “SHA2\_TM” (Fig. 4.1A).<sup>18, 19</sup> The circular dichroism (CD)  $\theta_{222}$  values for these constructs in decylmaltoside (DM) detergent agree semi-quantitatively with the ~60% percent helicities expected with helical hairpin FP structure, hairpin SE structure, and helical TM structure. Analysis of  $\theta_{222}$  vs. temperature yielded  $T_m = 91$  °C for HA2 and 87 °C for FHA2, which supports the hyperthermostability of the SE, with additional stability provided by the FP and TM. There are major peaks in size-exclusion chromatography assigned to trimers in N-lauroylsarcosine detergent at pH 7.4, and mixtures of monomers, trimers, and larger oligomers in DM at pH 7.4. The proteins visibly aggregate at pH 5.0.

One hypothesis probed by the present study is formation for HA2 of a thermodynamically-stable bundle containing one or more TM helices and FP helix/helices, with all helices traversing the membrane (Fig. 4.2). Relative to isolated independent monomer FP's, the FP in this bundle is expected to have much smaller D%, because of membrane-traversal rather than -interfacial location of the FP, and because of TM protection of the FP. The FP/TM model would therefore be supported by smaller HDX for the FP segment of HA2 vs. FHA2. The FP/TM bundle is sometimes pictured in the later steps of models of HA2-mediated fusion, as well as fusion models for other enveloped proteins.<sup>1, 37-39</sup> The bundle, like the trimer-of-hairpins SE structure, is considered to be a final thermostable state, and is implied to be important in the final pore expansion step of membrane fusion. The bundle model is potentially supported by inhibition of this final but not earlier steps of HA-cell/RBC fusion by truncation of the TM domain that prevents its traversal of the HA-cell membrane.<sup>12</sup> Another potentially supporting observation is full HA-cell/RBC fusion

when the HA2 TM and endodomain are replaced by the TM and endodomain of a non-homologous protein.<sup>13</sup>

The present study also uses HDX-MS to probe the structural and motional bases of impaired fusion for the G1E and I173E mutants, which are respectively in the FP and SE regions. These mutants exhibit greatly-reduced lipid mixing and pore formation for HA-cell/RBC fusion, and there are similar mutant-associated reductions for HA0-cell/RBC fusion mediated by exogenous FHA2.<sup>11, 15, 40, 41</sup> There is also reduced vesicle fusion induced by G1E- and I173E- vs. WT-HA2.<sup>19</sup> Earlier measurements of CD  $\theta_{222}$  were consistent with similar fractional helicities for WT- and G1E- HA2, and for WT- and G1E- FHA2, and moderately reduced helicities for I173E- HA2 and FHA2. CD  $\theta_{222}$  vs. temperature studies showed dramatic 30-40 °C reductions in  $T_m$  for G1E- and I173E- vs. WT- FHA2, and smaller but still significant reductions in  $T_m$  for G1E- and I173E- vs. WT- HA2.<sup>19</sup> Correlation of the stability of the trimer-of-hairpins with fusion is consistent with the model of hairpin apposition of the two membranes followed by fusion. The I173 sidechain in the trimer-of-hairpins is buried in the protein near the sidechains of the T41 and I45 residues. The I173E mutation may be less stably packed against these residues, leading to the dissociation of the C-terminal strands of the SE, with consequent destabilization of the N-terminal bundle of the SE (Fig. 4.2). Similar destabilization of the SE by the G1E mutation is surprising because the mutation site is in the FP and >35 residues from the SE. We hypothesize that the FP structure is destabilized by the G1E mutation, and that the destabilized FP's compete with the N-helix bundle of the SE for binding with the C-terminal strands (Fig. 4.2). The present study probes these destabilization hypotheses using comparative HDX-MS of WT-, G1E-, and I173E- HA2.



**Figure 4.2** Structural models for WT and mutant HA2 constructs that are based on residue 1-22 FP closed helical hairpin (PDB 2KXA), 36-175 SE hairpin (PDB 1QU1), 186-211 TM continuous  $\alpha$  helix that traverses the membrane, and 212-221 endodomain that is disordered. Residues 23-35 and 176-185 are represented as squiggly lines, because there isn't electron density for these residues in PDB 1QU1 for all or some of the monomers of the HA2<sub>23-185</sub> construct. The WT SE is the trimer-of-hairpins like in Fig. 4.1B, but a single monomer SE is displayed for clarity. Two WT-HA2 models are shown that either have a FP with helical hairpin structure at the membrane-interface, or FP with continuous  $\alpha$  helical structure in complex with TM and with membrane traversal. I173E-HA2 is shown with dissociated C-terminal SE strands and reduced helicity in the N-terminal SE. G1E-HA2 is shown as an equilibrium between the WT structure and the I173E structure with unfolded FP's that bind the C-terminal SE strands.

## 4.2 Materials and Methods

### 4.2.1 Materials

Descriptions and commercial sources of reagents include: n-decyl- $\beta$ -D-maltopyranoside (DM), Anatrace, Maumee, OH; D<sub>2</sub>O (99.9 atom % D), Cambridge Isotope Laboratories – Andover, MA; Enzymate BEH pepsin column (2.1  $\times$  30 mm), XBridge BEH Vanguard pre-column/trap column (2.1  $\times$  5 mm, 2.5  $\mu$ m particle size), C18 analytical column (2.1  $\times$  50 mm), Waters, Milford, MA. Most other reagents were purchased from Sigma-Aldrich – St. Louis, MO. Solvents were HPLC grade.

#### 4.2.2 Protein constructs

Fig. 4.3 displays amino acid sequences of the constructs HA2, FHA2  $\equiv$  HA2<sub>1-185</sub>, and SHA2\_TM  $\equiv$  HA2<sub>20-211</sub>. The sequences are from the X31 strain of influenza virus with mutation of all native Cys residues, specifically C137S, C144A, C148A, C195A, C199A, C210A, and C220A. SHA2\_TM includes a N-terminal G20C mutation done for future native-chemical ligation (not part of the present study). Inter-trimer Cys cross-linking is unlikely, based on the absence of such cross-linking in a similar construct of the non-homologous HIV gp41 protein which also adopts trimer-of-hairpins structure.<sup>42</sup> There are non-native C-terminal regions LEG<sub>6</sub>H<sub>6</sub> (HA2 and SHA2\_TM) or LEH<sub>6</sub> (FHA2) for affinity chromatography. The G1E- and I173E- HA2 constructs were also studied. Preparation of the constructs, as well as expression, purification, and refolding protocols were discussed in Chapter 3.<sup>19</sup> Protein stocks contained 40  $\mu$ M protein in 10 mM Tris-HCl buffer at pH 7.4 with 0.17% DM.

#### 4.2.3 Hydrogen-Deuterium exchange, online pepsin digestion and LC-MS analysis

The HDX-MS apparatus included a temperature-controlled auto-sampler (Waters 2777), external pump (Shimadzu LC-20AD), UPLC pump system (Waters Acquity Binary Solvent Manager), and mass spectrometer (Waters Xevo G2S QTof) operating with electrospray ionization in positive ion mode, and continuum-mode  $m/z = 50 - 2000$ .<sup>35</sup> HDX was carried out in 1.6 mL capped tubes at 25 °C. Protein stock (6  $\mu$ L) was mixed with D<sub>2</sub>O buffer (114  $\mu$ L), where the latter contained 5 mM HEPES/10 mM MES at pD 7.4 with 0.17% DM. After initial manual mixing, individual tubes were kept without stirring in the auto-sampler for either 1, 5, 30, 60, or 120 min. A “0 min” sample was made by mixing with H<sub>2</sub>O buffer. After the HDX time, the tube was removed from the auto-sampler, and 120  $\mu$ L of ice-cold formic acid (FA) was added, so that pH

was reduced to <3 which inhibited additional HDX. Back “DHX” exchange was minimized by surrounding ice around the tube, and around other solutions, injection ports, valves, and columns.

The next set of liquid transfers was done with the external pump. Protein digestion was done in a pepsin column that had been equilibrated with flow of water with 0.1% FA ( $H_2O_{FA}$ ). Protein solution (100  $\mu$ L) was transferred to the column using 0.1 mL/min flow, followed by digestion in the column for 1 min. The solution was then transferred to the trap column using 0.5 mL/min flow, followed by desalting with 1 min flow of  $H_2O_{FA}$ . Flow control was transferred to the UPLC and the flow rate set to 0.3 mL/min. The flow then went from the trap column to the C18 column, and separation of the peptides was done using mixtures of  $H_2O_{FA}$  and acetonitrile with 0.1% FA ( $ACN_{FA}$ ). The separation program was 1%  $ACN_{FA}$  (1 min), a gradient to 30%  $ACN_{FA}$  (3 min), a second gradient to 99%  $ACN_{FA}$  (3 min), and 99%  $ACN_{FA}$  (1 min). The flow from the column went to the mass spectrometer for analysis.

The above protocol was first applied to the “0 min” sample prepared by mixing with  $H_2O$  buffer. Individual m/z peak distributions in the mass spectrum were assigned to specific peptide sequences by manual comparison of the average natural-abundance mass ( $M_H$ ) of the distribution with theoretical  $M_H$ 's. Assignment were confirmed by MS/MS analysis. The mass spectra of HDX samples were assigned by the same procedure. Peaks assigned to a single peptide within a user-selected m/z window were inputs for the HX-Express software, which then calculated the peak area weighted-average m/z for each  $t_{HDX}$ .<sup>43</sup> The outputs were the average  $M_H$  ( $t_{HDX} = 0$ ) and average  $M_D$  at each  $t_{HDX} > 0$ . For a particular peptide, the percent deuterium incorporation  $D\%(t) = 100 \times [M_D(t) - M_H]/N$ , where N is the number of backbone amide hydrogens in the peptide.

## FHA2

```

      10      20      30      40      50      60
GLFGAIAGFI ENGWEGMIDG WYGFRHQNSE GTGQAADLKS TQAAIDQING KLNRVIEKTN
      70      80      90     100     110     120
EKFHQIEKEF SEVEGRIQDL EKYVEDTKID LWSYNAELLV ALENQHTIDL TDSEMKNLFE
     130     140     150     160     170     180
KTRRQLRENA EEMGNFSFKI YHKADNAAIE SIRNGTYDHD VYRDEALNNR FQIKGVELKS
     190
    GYKDWLEHHHH HH
  
```

## SHA2\_TM

```

           20      30      40      50      60
           C WYGFRHQNSE GTGQAADLKS TQAAIDQING KLNRVIEKTN
      70      80      90     100     110     120
EKFHQIEKEF SEVEGRIQDL EKYVEDTKID LWSYNAELLV ALENQHTIDL TDSEMKNLFE
     130     140     150     160     170     180
KTRRQLRENA EEMGNFSFKI YHKADNAAIE SIRNGTYDHD VYRDEALNNR FQIKGVELKS
     190     200     210     220
    GYKDWILWIS FAISAFLLAV VLLGFIMWAA QGGGGGGLEH HHHHH
  
```

## HA2

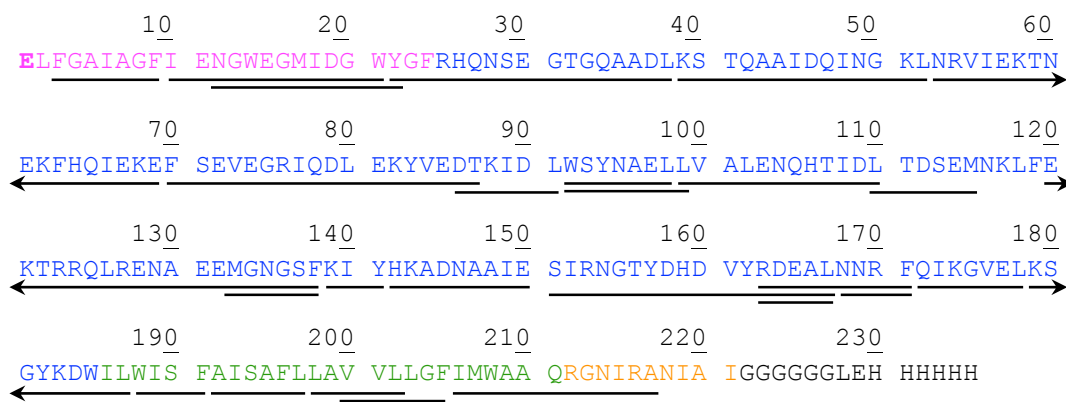
```

      10      20      30      40      50      60
GLFGAIAGFI ENGWEGMIDG WYGFRHQNSE GTGQAADLKS TQAAIDQING KLNRVIEKTN
      70      80      90     100     110     120
EKFHQIEKEF SEVEGRIQDL EKYVEDTKID LWSYNAELLV ALENQHTIDL TDSEMKNLFE
     130     140     150     160     170     180
KTRRQLRENA EEMGNFSFKI YHKADNAAIE SIRNGTYDHD VYRDEALNNR FQIKGVELKS
     190     200     210     220     230
    GYKDWILWIS FAISAFLLAV VLLGFIMWAA QRGNIRANIA IGGGGGGLEH HHHHH
  
```

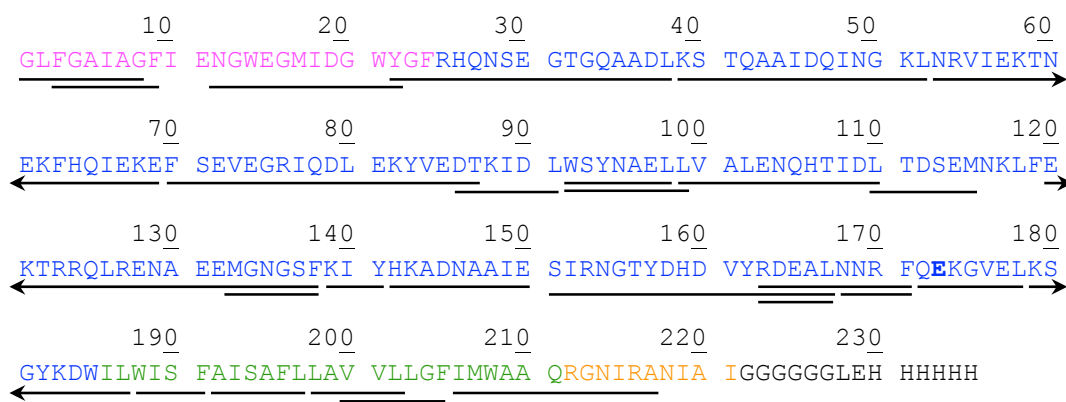
**Figure 4.3.** Protein sequences with peptides observed by mass spectrometry are underlined with domains colored: fusion peptide (FP), pink; soluble ectodomain (SE), blue; transmembrane domain (TM), green; and endodomain, orange. The non-native C-terminal regions are in black.

**Figure 4.3 (cont'd)**

### **G1E-HA2**



### **I173E-HA2**



## **4.3 Results**

Fig. 4.3 displays sequence coverages for identified peptides from the different constructs. For all constructs, there is >90% coverage of the HA2 region of the sequence. Fig. 4A displays mass spectra *vs*  $t_{HDX}$  for peptides of WT- FHA2, SHA2\_TM, and HA2, and Fig. 4.4 displays  $D_{\%}$  *vs.*  $\log_{10}(t_{HDX}/\text{min})$  for all peptides of these constructs. Figs. 4B and 4.5 display similar plots for WT-, G1E-, and I173E- HA2. The Figs. 4A, 4.4 HA2 data are the same as the Figs. 4B, 4.5 WT-



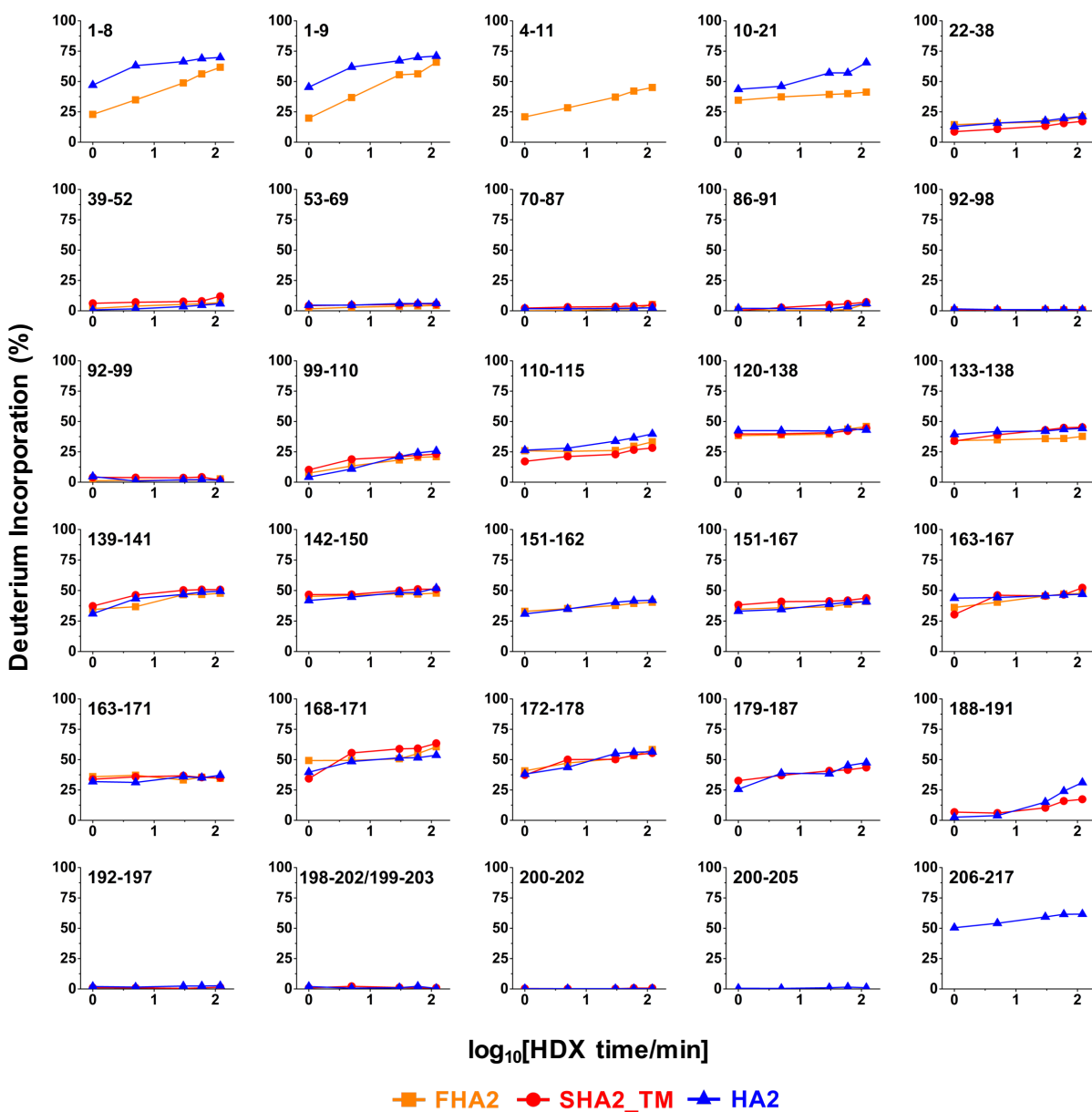
HA2 data. Use of  $\log_{10}(t_{\text{HDX}})$  for the horizontal axis is for clarity of presentation and is not based on a specific kinetic model of HDX. Tables 4A and 4B also present the  $D_{\%}$  vs.  $t_{\text{HDX}}$  data, including the RMSD's for the triplicate trials, with typical RMSD of 1%. The maximum  $D_{\%}$  values are  $\sim 70\%$  rather than 100%, which is consistent with some structure in all regions other than the HA2<sub>212-221</sub> endodomain, and may also reflect some reduction in  $D_{\%}$  due to DHX back-exchange during the  $\sim 10$ -minute digestion and chromatographic separation steps.<sup>33</sup> Informative comparison can still be made between the relative  $D_{\%}$  values of the peptides. Tables 4.1 and 4.2 present  $\langle D_{\%} \rangle$  for the peptides, where  $\langle D_{\%} \rangle$  is the average of all  $t_{\text{HDX}}$  data, i.e. 1, 5, 30, 60, and 120 min. Although  $\langle D_{\%} \rangle$  doesn't have a specific physical meaning,  $\langle D_{\%} \rangle$  is a convenient single number to describe each peptide. Fig. 4.6 displays the Fig. 4.2 models with colors based on the  $\langle D_{\%} \rangle$  from Tables 4.1 and 4.2. A  $\langle D_{\%} \rangle$  value has a typical uncertainty of 0.5% that is calculated using the RMSD's of the  $D_{\%}$  values.

#### 4.3.1 WT- HA2, FHA2, and SHA2\_TM

There are very similar  $D_{\%}$  vs.  $t_{\text{HDX}}$  and therefore similar  $\langle D_{\%} \rangle$  for the same peptide derived from the SE of HA2, FHA2, or SHA2\_TM, and for the same peptide derived from the TM of HA2 or SHA2\_TM. The data support very similar SE structures and very similar TM structures of the constructs. The SE  $D_{\%}$  are consistent with the trimer-of-hairpins structure shown in Fig. 4.1B, with typical  $\langle D_{\%} \rangle < 5\%$  for residue 39-99 peptides and typical  $\langle D_{\%} \rangle$  in the 35-50% range for residue 120-178 peptides. These regions respectively correlate with the interior and the exterior of the trimer. The  $\langle D_{\%} \rangle < 3\%$  for the 192-205 region are consistent with membrane traversal by the TM, whereas  $\langle D_{\%} \rangle = 57\%$  for the 206-217 region is consistent with a contribution from an aqueous-exposed and structurally disordered HA2<sub>212-221</sub> endodomain.

Peptides from the N-terminal FP of FHA2 and HA2 exhibit significant exchange, e.g. respective  $\langle D_{\%} \rangle_{1-9} = 47$  and 63%, and  $\langle D_{\%} \rangle_{10-21} = 38$  and 53%. There is greater exchange for the FP in the presence *vs.* absence of the TM, which contrasts with protection of TM that is FP-independent, e.g.  $\langle D_{\%} \rangle < 0.5\%$  for 200-202 of both HA2 and SHA2\_TM. The data support separated FP and TM domains for HA2, with respective membrane interfacial and traversal locations. The data do not support a transmembrane FP/TM complex, for which the  $\langle D_{\%} \rangle$  for the FP would be small and comparable with the TM (Fig. 4.2). The extensive exchange for the FP is more consistent with monomer than with oligomer FP's, whereas complete protection of the TM is consistent with either monomer or oligomer TM's that traverse the membrane.

The 22-38 region is located between the FP and the interior N-helices of the trimer-of-hairpins. The 22-38  $D_{\%}$  are comparable for FHA2 and HA2, and smaller for SHA2\_TM, with respective  $\langle D_{\%} \rangle$  of 17 and 13%. For comparison, the  $\langle D_{\%} \rangle_{39-52}$  are respectively  $\sim 4$  and 8%. Our  $\langle D_{\%} \rangle_{22-38}$  values are much lower than values expected from interpretation of previous structural data. The HA2<sub>23-185</sub> construct was used for the SE crystal structure, but electron density is not reported for  $\sim 10$  N-terminal residues, which reflects their disorder, presumed contact with the aqueous medium, and expected  $\langle D_{\%} \rangle > 50\%$ .<sup>8</sup> The observed  $\langle D_{\%} \rangle_{22-38} \approx 15\%$  supports a different and new model of defined and solvent-protected structure for the 22-38 segment. We note that our results with large HA2 constructs are also different than those from NMR of the smaller HA2<sub>1-28</sub> construct.<sup>44</sup> For the latter study, NMR signals were not observed for residues 25-28 which is consistent with disorder, aqueous exposure, and predicted large  $\langle D_{\%} \rangle$  for these residues.



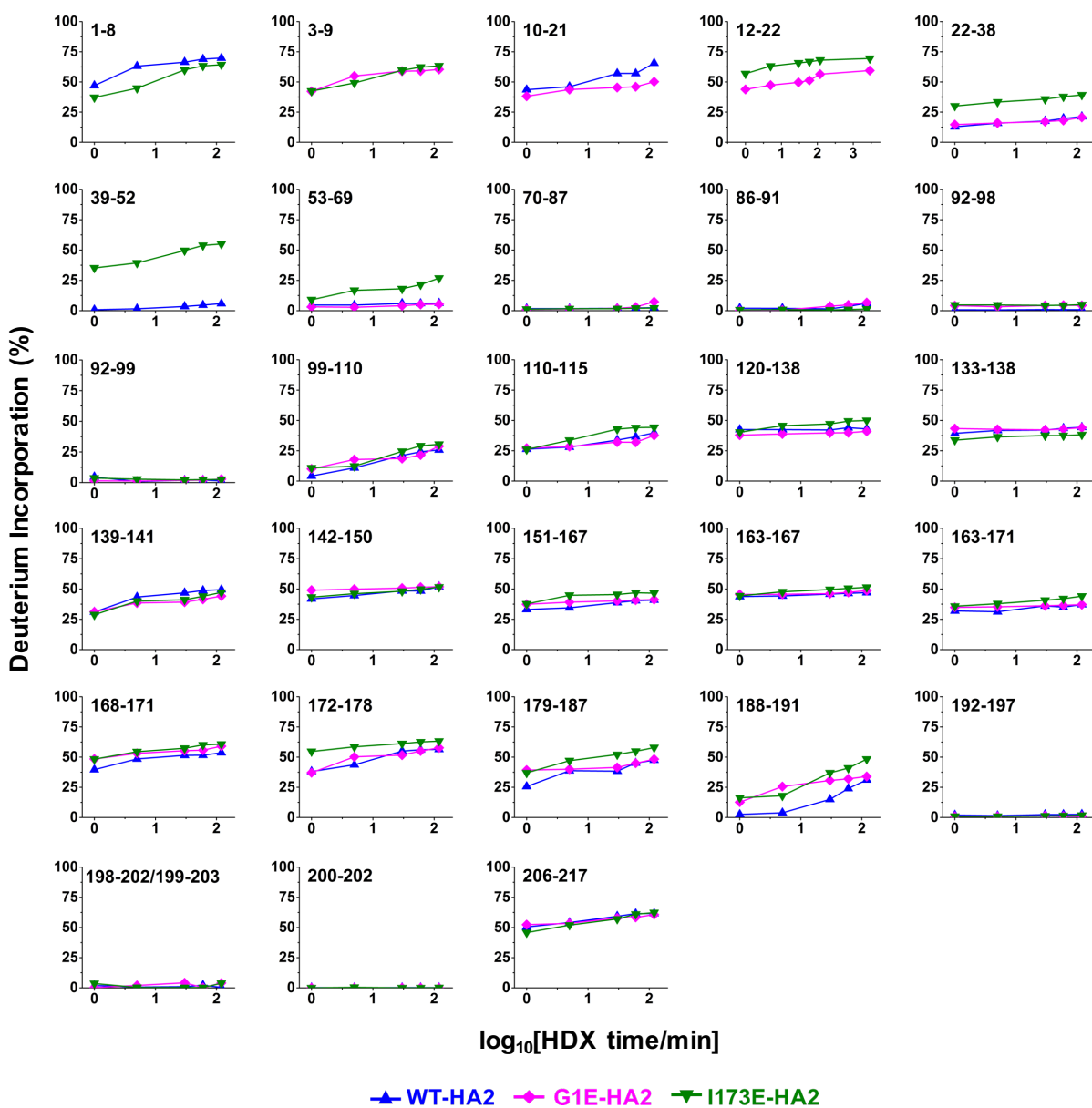
**Figure 4.4** Plots of percent deuterium incorporation (D%) vs.  $\log_{10}[\text{t}_{\text{HDX}}/\text{min}]$  for all analyzed peptides of FHA2, SHA2\_TM, and HA2. Each point is the average of triplicate measurements, with typical RMSD of 1%. Tables 4A list all D% data.

**Table 4.1.**  $\langle D_{\%} \rangle$  for peptides from FHA2, SHA2\_TM, and HA2.

Peptide	Protein		
	FHA2	SHA2_TM	HA2
1-8	44.88 (0.58)		63.02(0.52)
1-9	46.82(0.40)		63.03(0.94)
10-21	38.43(0.68)		53.88(0.31)
4-11	34.67(0.76)		
22-38	17.19(0.61)	13.13(0.55)	17.43(0.66)
39-52	4.74(0.35)	8.12(0.13)	3.26(0.23)
53-69	3.33(0.34)	5.05(0.29)	5.50(0.23)
70-87	1.66(0.78)	3.47(0.15)	2.01(0.06)
86-91	1.81(0.42)	4.16(0.55)	3.02(0.54)
92-98	0.82(0.06)	0.74(0.12)	1.09(0.39)
92-99	1.73(0.29)	3.44(0.18)	2.18(0.25)
99-110	16.10(0.35)	19.05(0.52)	17.21(0.22)
110-115	28.04(0.70)	23.19(0.65)	32.87(0.36)
120-138	41.29(0.54)	41.40(0.96)	42.81(0.87)
133-138	35.72(0.54)	41.17(0.31)	42.20(0.51)
139-141	42.43(0.49)	47.03(0.55)	43.88(0.96)
142-150	46.68(0.83)	49.13(0.54)	47.04(0.04)
151-162	37.14(0.65)		37.95(0.38)
151-167	37.41(0.76)	41.21(0.66)	37.50(0.56)
163-167	43.33(0.71)	44.20(0.68)	45.43(0.34)
163-171	35.26(1.07)	35.38(0.83)	34.27(0.83)
168-171	52.98(0.73)	54.26(0.55)	48.97(0.87)
172-178	50.03(0.45)	49.36(0.98)	49.85(0.67)
179-187		39.10(0.47)	39.07(0.54)
188-191		11.27(0.54)	15.31(0.13)
192-197		2.59(0.32)	2.19(0.17)
198-202/199-203		1.19(0.28)	1.24(0.42)
200-202		0.40(0.24)	0.13(0.11)
200-205			0.89(0.13)
206-217			57.41(1.04)

### 4.3.2 WT-, G1E-, and I173E- HA2

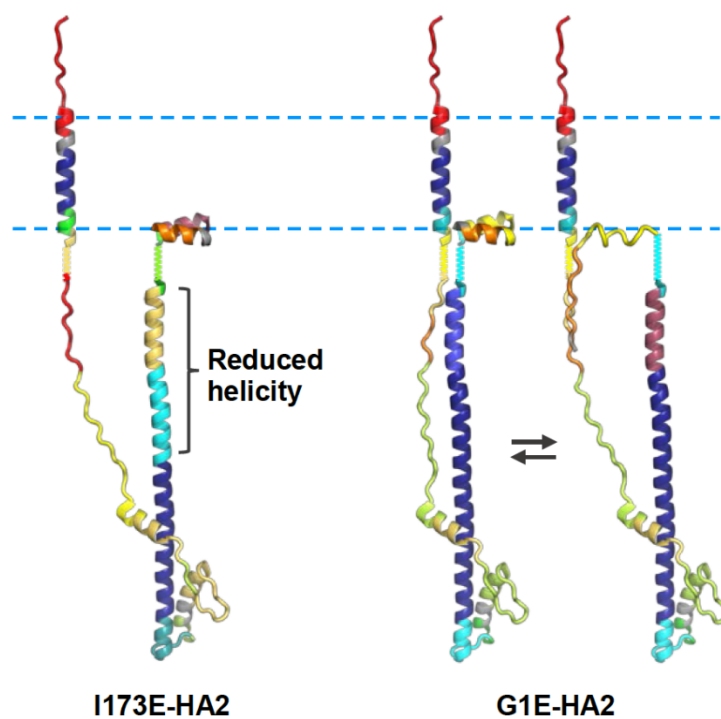
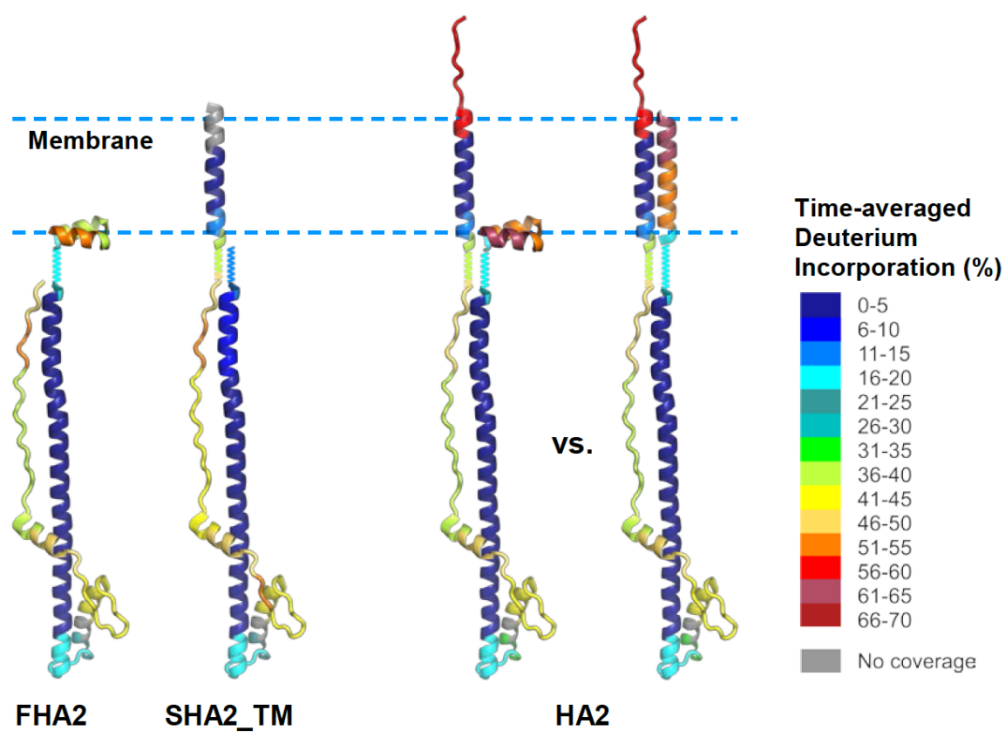
For the same peptide in the residue 70-150 and 192-217 regions, there are similar  $D_{\%}$  vs.  $t_{\text{HDX}}$  and therefore similar  $\langle D_{\%} \rangle$  for WT-, G1E-, and I173E- HA2 (Fig. 4.5, Table 4.2, and Table 4B). These data support similar folded regions of the SE for the three constructs, helical TM's which traverse the membrane, and aqueous-exposed endodomains. Peptides from the FP exhibit significant exchange for all three constructs, with smallest  $D_{\%}$  for G1E, e.g.  $\langle D_{\%} \rangle_{10-21} = 45\%$  for G1E vs. 54% for WT and  $\langle D_{\%} \rangle_{12-22} = 49\%$  for G1E vs. 63% for I173E. This result is surprising because G1E likely has a less thermostable FP and might therefore be expected to have higher rather than lower exchange. A different trend is observed for the 22-69 regions, which at least for WT- include some of the 38-105 internal N-helices in the trimer-of-hairpins SE structure (Fig. 4.1B). I173E exhibits significant aqueous exposure in this region vs. high protection for WT, e.g.  $\langle D_{\%} \rangle_{22-38} = 35$  vs. 17%,  $\langle D_{\%} \rangle_{39-52} = 47$  vs. 3%, and  $\langle D_{\%} \rangle_{53-69} = 18$  vs. 6%. WT and G1E exhibit similar  $D_{\%}$  vs.  $t_{\text{HDX}}$  profiles for 22-38 and for 53-69. G1E exhibits bimodal behavior for 39-52, with one distribution of peaks in a  $m/z$  range similar to WT and the other in a  $m/z$  range similar to I173E (Fig. 4B). For  $t_{\text{HDX}} = 120$  min, there are approximately equal sums of G1E peak intensities in the two ranges, and the calculated  $D_{\%} = 16$  and 63% are similar to the 6 and 54% values for WT and I173E, respectively. Table 4C lists the  $D_{\%}$  for 39-52 for all  $t_{\text{HDX}}$ . For all three constructs, peptides in the 150-191 region exhibit significant exchange with typical  $\langle D_{\%} \rangle > 40\%$ . For all peptides in this region, there is a consistent trend of  $\langle D_{\%} \rangle_{\text{WT}} < \langle D_{\%} \rangle_{\text{G1E}} < \langle D_{\%} \rangle_{\text{I173E}}$ . This region includes the 154-178 extended strands on the exterior of the trimer-of-hairpins of the WT SE.



**Figure 4.5** Plots of percent deuterium incorporation ( $D\%$ ) vs.  $\log_{10}[\text{HDX time/min}]$  for all analyzed peptides of WT-, G1E-, and I173E- HA2. Each point is the average of triplicate measurements, with typical RMSD of 1%. Tables 4B list all  $D\%$  data.

**Table 4.2**  $\langle D_{\%} \rangle$  for peptides from WT-, G1E-, and I173E- HA2.

Peptide	Protein		
	WT-HA2	G1E-HA2	I173E-HA2
1-8	63.02(0.52)		53.89(0.74)
3-9		55.09(1.00)	55.33(1.45)
10-21	53.88(0.31)	44.68(0.70)	
12-22		49.79(1.14)	63.18(1.19)
22-38	17.43(0.66)	17.29(0.57)	35.23(0.42)
39-52	3.26(0.23)		47.00(0.87)
53-69	5.50(0.23)	4.12(0.51)	18.44(0.41)
70-87	2.01(0.06)	2.93(0.41)	1.75(0.11)
86-91	3.02(0.54)	3.16(0.54)	0.90(0.22)
92-98	1.09(0.39)	4.12(0.21)	4.64(0.12)
92-99	2.18(0.25)	1.92(0.59)	2.74(0.58)
99-110	17.21(0.22)	19.32(0.53)	21.40(0.92)
110-115	32.87(0.36)	31.43(0.77)	38.17(0.98)
120-138	42.81(0.87)	39.47(0.44)	46.49(0.77)
133-138	42.20(0.51)	43.02(0.39)	36.69(0.40)
139-141	43.88(0.96)	38.90(0.46)	40.22(1.05)
142-150	47.04(0.04)	50.58(0.29)	47.77(0.61)
151-167	37.50(0.56)	39.77(0.16)	44.26(0.85)
163-167	45.43(0.34)	46.64(0.41)	48.70(0.71)
163-171	34.27(0.83)	35.99(0.90)	40.11(0.71)
168-171	48.97(0.87)	54.33(1.51)	56.21(0.82)
172-178	49.85(0.67)	50.34(0.53)	60.05(0.81)
179-187	39.07(0.54)	42.79(0.96)	49.78(0.94)
188-191	15.31(0.13)	27.00(1.10)	32.11(1.11)
192-197	2.19(0.17)	1.07(0.17)	1.31(0.10)
198-202/199-203	1.24(0.42)	2.13(1.12)	1.62(0.85)
200-202	0.13(0.11)	0.00(0.00)	0.13(0.06)
206-217	57.41(1.04)	56.56(0.45)	55.63(1.00)





**Figure 4.6** Presentation of peptide  $\langle D_{\%} \rangle$  using color-coding and the structural models of Fig. 4.2. The colors for WT-HA2 are based on the  $\langle D_{\%} \rangle$  of peptides 1-9, 10-21, 22-38, 39-52, 53-69, 70-87, 86-91, 92-98, 99-110, 110-115, 120-138, 139-141, 142-150, 151-167, 168-171, 172-178, 179-187, 188-191, 192-197, 198-202, 200-205, and 206-217. The colors for other constructs are based on the  $\langle D_{\%} \rangle$  of subsets of these peptides with a few substitutions that include 3-9 for G1E-HA2 and 1-8 and 12-22 for I173E-HA2.

## 4.4 Discussion

### 4.4.1 Comparison with previous HDX data

The  $D_{\%}$  data from the present study are compared to those of an earlier study in which the HA2<sub>1-175</sub>/HA1 complex was incubated in a pH 4.9 solution without detergent for three hours, neutralized, and subjected to a protocol similar to the present study.<sup>36</sup> The earlier protocol included incubation in 85% D<sub>2</sub>O for  $t_{\text{HDX}} = 0.05, 1, 30,$  or 1200 min, acidification, pepsin digestion, and mass spectrometric analysis of the peptide products. Direct comparison can be made between the  $D_{\%}$  for  $t_{\text{HDX}} = 1$  and 30 min, which are common to both studies. Comparison is made using our data for FHA2  $\equiv$  HA2<sub>1-185</sub>, as this construct is closest to the HA2<sub>1-175</sub> subunit of the complex with HA1. For the same peptide and  $t_{\text{HDX}}$ , the  $D_{\%}$  values of the two studies should agree within a factor of 1.2, based on the minor differences in how  $D_{\%}$  was calculated in the studies. For both studies, peptides in the 53-99 region exhibit  $D_{\%} < 5\%$ , which is consistent with the interior location of the 38-105 helices of the trimer-of-hairpins structure. For both studies, peptides from the 99-171 region exhibit larger  $D_{\%}$ , which correlates with the exterior location of most of this region in the structure. However, the  $D_{\%}$  for the present study are much lower than for the previous study. Fig. 4C provides some examples of this trend which holds for both the 22-52 and the 99-171 regions.

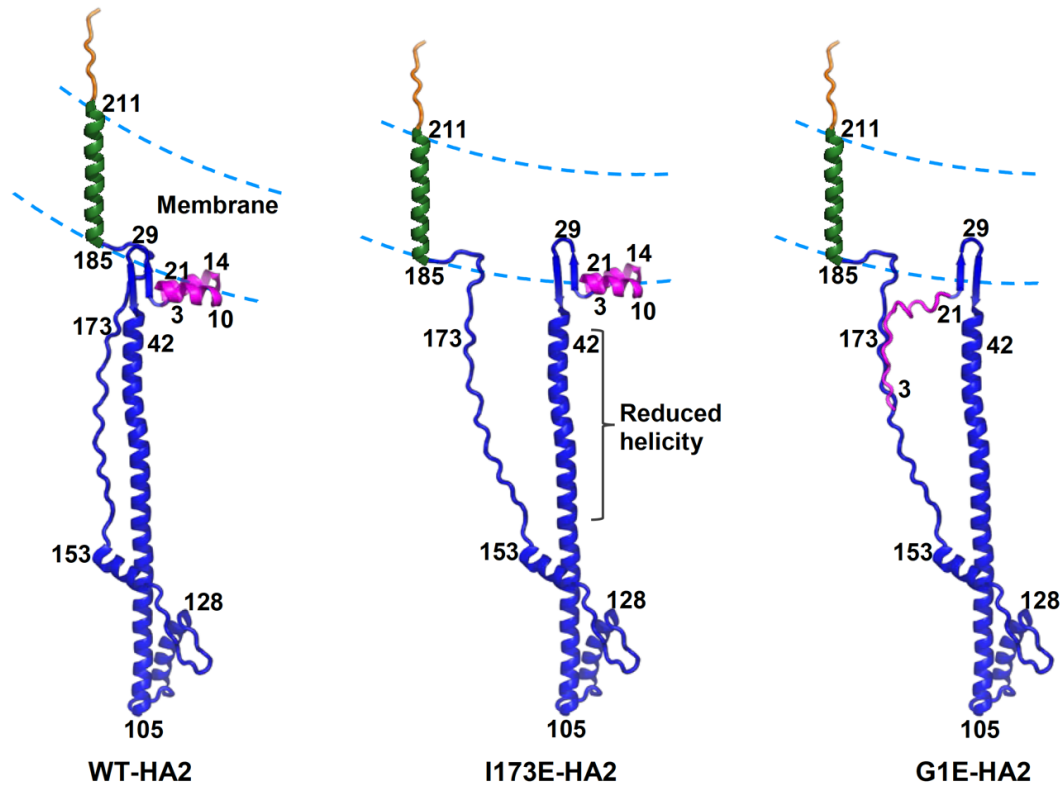
For example, comparison of 22-38 for FHA2 *vs.* HA2<sub>1-175</sub>/HA1 shows  $D_{\%} = 14$  *vs.* 75% for  $t_{\text{HDX}} = 1$  min and  $D_{\%} = 17$  *vs.* 85% for  $t_{\text{HDX}} = 30$  min. For 39-52,  $D_{\%} = 2$  *vs.* 40% for  $t_{\text{HDX}} = 1$  min and  $D_{\%} = 5$  *vs.* 55% for  $t_{\text{HDX}} = 30$  min. For the 110-115 peptide at either  $t_{\text{HDX}}$ , the comparison is  $D_{\%} = 26$  *vs.* 100%. For the 151-167 peptide, the  $D_{\%} = 35$  *vs.* 80% for  $t_{\text{HDX}} = 1$  min and  $D_{\%} = 37$  *vs.* 100% for  $t_{\text{HDX}} = 30$  min. Another difference is that peptides from the FP were straightforwardly analyzed in the present study with HA2-only constructs, but not in the earlier study with low pH incubation of the HA2<sub>1-175</sub>/HA1 complex because of significant broadening of the envelopes of peaks. Such broadening wasn't observed for the complex that hadn't undergone the incubation and consequent large-scale structural rearrangement. One experimental difference between the present *vs.* earlier study is the presence *vs.* absence of detergent. HA2 constructs form aggregates in the absence of detergent via the exposed FP's. Detergent is also important for FHA2 folding as evidenced by circular dichroism  $\theta_{222}$  values that correlate with 65 *vs.* 25% helicity in the presence *vs.* absence of detergent.<sup>15, 19</sup> The 65% helicity in detergent is consistent with folded protein as it matches the percent helicity calculated using the numbers of helical residues in the trimer-of-hairpins SE structure and in the helical hairpin FP structure in detergent. The envelope broadening of peptides from the FP and the larger  $D_{\%}$  of the earlier study may therefore reflect HA2<sub>1-175</sub> proteins that are aggregated and less-folded than the HA2 constructs in detergent of the present study.

We also compare the  $D_{\%}$  of the FP region of the present study with earlier NMR-measured HDX rates for HA2<sub>1-23</sub> in detergent.<sup>26</sup> Residues 2-6, 12-17, and 20-23 exhibit rates  $>0.1 \text{ s}^{-1}$ , and would be completely exchanged even for the shortest  $t_{\text{HDX}} = 1$  min of the present study. For  $t_{\text{HDX}} = 1$  min, the 1-9 peptide exhibits  $D_{\%} = 20\%$  for FHA2 *vs.* 45% for HA2, and 10-21 exhibits  $D_{\%} = 35\%$  for FHA2 *vs.* 44% for HA2. For a model in which FP residues 2-6, 12-17, and 20-23

completely exchange and other residues don't exchange, the calculated  $D\% = 63\%$  for 1-9 and  $D\% = 73\%$  for 10-21. The FP therefore exhibits greater protection in FHA2 and HA2 vs. HA2<sub>1-23</sub>.

#### 4.4.2 Structural Model of WT-HA2

Fig. 4.7 displays a structural model of WT-HA2 in a membrane bilayer that integrates the HDX data of the present study with previous high-resolution structures of segments of these proteins. Our data are for protein in DM detergent which is probably a reasonable membrane mimetic, as evidenced by well-folded and hyperthermostable HA2 structure ( $T_m = 91\text{ }^\circ\text{C}$ ), and protein which is fusion-active.<sup>19</sup> There are very similar HDX data for FHA2, SHA2\_TM, and HA2, which supports that FHA2 and SHA2\_TM retain the structures of segments of HA2. A monomer SE unit is shown for clarity in Fig. 4.7, but the HDX data are consistent with the trimer-of-hairpins structure for HA2<sub>36-178</sub>, e.g.  $\langle D\% \rangle < 5\%$  for the 70-87, 86-91, and 92-99 peptides of the three constructs. This is consistent with the interior protected location of the 38-105 helices in this structure. A new and unexpected result of our study is the low  $D\%$  for the 22-38 peptide, e.g.  $\langle D\% \rangle_{22-38} = 13\%$  for SHA2\_TM. The HDX data support a well-structured and protected 22-38 segment. Although much of this segment doesn't exhibit defined structure in the HA2<sub>1-28</sub> and HA2<sub>23-185</sub> constructs, the 23-35 region is a well-structured  $\beta$  sheet hairpin in the HA2<sub>1-175</sub>/HA1 complex, i.e. 23-27 strand/28-30 turn/31-35 strand.<sup>5, 22</sup> The sequence of the HA2 23-35 region is also reasonably-conserved across the subtypes of hemagglutinin.<sup>4</sup> The region is displayed as a  $\beta$  sheet hairpin in Fig. 4.7.



**Figure 4.7** Structural models for WT-, I173E-, and G1E- HA2 bound to a single membrane. WT exhibits trimer-of-hairpins structure for the SE as in Fig. 4.1B, but only one monomer is shown for clarity. The approximate positions of specific residues are indicated by the numbers. The WT model combines residue 1-22 closed FP helical hairpin (PDB 2KXA), 23-35 antiparallel  $\beta$  sheet of the HA2<sub>1-175</sub>/HA1 complex (PDB 2HMG), 36-178 hairpin SE structure (PDB 1QU1), 186-211 continuous  $\alpha$  helix for TM, and 212-221 unstructured endodomain. The 179-185 extended structure is found in one monomer in PDB 1QU1, but there isn't electron density for these residues in the other two monomers. The small positive membrane curvature accommodates the respective membrane interfacial and traversal locations of the FP and TM. The I173E model shows dissociation of the C-terminal strands from the trimer-of-hairpins and reduced helicity for the N-

helices. The G1E model shows binding of the FP to the C-terminal strands, and the displayed structure is in equilibrium with the WT structure, like in Fig. 4.2.

The FP and TM are displayed with membrane interfacial and traversal locations, respectively, that correlate with peptides with  $\langle D_{\%} \rangle > 35\%$  in the 1-21 region and  $\langle D_{\%} \rangle < 3\%$  in the 192-205 region. Either monomeric or oligomeric TM that traverses the membrane would exhibit the observed high protection from HDX. Monomer TM is consistent with low sequence conservation in this domain among HA subtypes, and with robust HA-cell/RBC fusion after substitution of the native TM and endodomain sequences with those of an unrelated protein.<sup>4, 13</sup> However, cryo-electron microscopy of the HA2/HA1 initial trimer complex in detergent were interpreted to support a trimeric TM bundle.<sup>38</sup> We note that fluorescence and SDS-PAGE data for HA2 TM peptides support a fraction of oligomeric peptides.<sup>39, 45</sup>

Peptides from the FP of HA2 exhibit the highest  $\langle D_{\%} \rangle$  of any segment of FHA2, SHA2\_TM, or HA2, e.g.  $\langle D_{\%} \rangle = 63\%$  for 1-8 and 54% for 10-21. The values are comparable to  $\langle D_{\%} \rangle = 57\%$  of 205-217 which includes part of the likely unstructured and aqueous-exposed HA2<sub>212-221</sub> endodomain. The FP data are completely inconsistent with membrane traversal by the FP and with a complex with TM. Surprisingly, the TM increases the aqueous exposure of the FP, as evidenced by much smaller  $\langle D_{\%} \rangle$  for the FP region of FHA2, e.g. 45% for 1-8 and 38% for 10-21. This may be related to the local positive membrane curvature in the Fig. 4.7 model to achieve interfacial locations for the FP and the N-terminus of the TM. Less curvature is needed without the TM and the FP could insert more deeply in the membrane, which is consistent with the FHA2 vs. HA2 data. HA2-stabilized curvature could be important for fusion catalysis, as there are likely membrane intermediates during fusion with local curvature.<sup>1, 46</sup> Earlier studies have provided

experimental data that evidence stabilization of local membrane curvature by viral fusion peptides.<sup>47-50</sup> The mechanical coupling of the TM to the FP requires at least semi-rigid connections between the TM and SE, and between the SE and FP. Rigidity in the SE/TM connection is consistent with  $\langle D_{\%} \rangle = 39$  vs. 50% for 179-187 vs. 172-178, with 172-178 as a well-structured segment on the exterior of the SE trimer-of-hairpins. Rigidity in the SE/FP connection is consistent with  $\langle D_{\%} \rangle = 17\%$  for 22-38 vs. 54% for 10-21, with 23-35 modeled as an antiparallel  $\beta$  sheet in Fig. 4.7. Protection of 22-38 may also be due to its membrane insertion and to interaction with the TM. The latter effect appears to be enhanced with truncation of the FP, as both 22-38 and 188-191 are more protected for SHA2\_TM  $\equiv$  HA2<sub>20-211</sub> than for HA2, i.e.  $\langle D_{\%} \rangle = 13$  vs. 17%, and 11 vs. 15%.

We note that the  $D_{\%}$  of the FP in HA2 could be consistent with unfolded FP in aqueous solution. However, we consider this scenario unlikely because of the stability of the FP helical hairpin in detergent and membrane, and the hydrophobic-effect free energy penalty associated with aqueous solubilization.<sup>26, 29</sup> We also note that earlier fluorescence, EPR, and calorimetric data were interpreted to support weak association in membrane between a peptide containing the TM sequence and a peptide containing the FP sequence.<sup>39, 51</sup> By contrast, there was no close contact between the FP and TM domains of a 194-residue FP+SE+TM HIV gp41 fusion protein construct.<sup>52</sup> HIV gp41 is non-homologous with HA2, but the SE's of the two proteins both adopt trimer-of-hairpins structures.<sup>42, 53-55</sup> Finally, we note that although the preponderance of simulations and experimental data support membrane interfacial location for the FP in the absence of the rest of HA2, there are some simulations and experimental data that support membrane traversal for the isolated FP.<sup>26, 56-60</sup>

#### 4.4.3 G1E and I173E mutations destabilize the HA2 SE

G1E and I173E are canonical point mutations respectively in the FP and SE of HA2 that result in greatly-reduced HA-cell/RBC fusion, FHA2-mediated cell/RBC fusion, and HA2-mediated vesicle fusion. Fig. 4.7 displays structural models for the two mutants that are similar to those in Fig. 4.2, and also include the 22-35 antiparallel  $\beta$  sheet of the WT model of Fig. 4.7. The Fig. 4.2 and 4.7 models incorporate previous observations of reduced fractional helicity for I173E, and reduced thermostability for G1E and I173E.<sup>19</sup> As detailed below, the Fig. 4.7 G1E and I173E structural models are generally supported by the HDX data.

WT, G1E and I173E all exhibit small  $D_{\%}$  for 192-202 peptides, and large  $D_{\%}$  for 206-217, with respective  $\langle D_{\%} \rangle < 3\%$  and  $\langle D_{\%} \rangle = 56\%$ . These data are respectively consistent with membrane traversal by the TM and with an unstructured and aqueous-exposed HA2<sub>212-221</sub> endodomain. WT, G1E, and I173E all exhibit  $\langle D_{\%} \rangle < 5\%$  for 70-99 peptides, and also exhibit similar  $\langle D_{\%} \rangle$  in the 40-50% range for 120-150 peptides. These data are consistent with retention for G1E and I173E of some of the trimer-of-hairpins structure that has interior N-helices and exterior C-terminal regions. Retention also correlates with  $T_m = 75\text{ }^{\circ}\text{C}$  for G1E and  $79\text{ }^{\circ}\text{C}$  for I173E, vs.  $91\text{ }^{\circ}\text{C}$  for WT.<sup>19</sup>

We now focus on I173E, which shows greater  $D_{\%}$  than WT for other parts of the SE as well as the N-terminal section of the TM. The most dramatic differences are in the 22-69 region which is well-protected in WT but not in I173E, e.g.  $\langle D_{\%} \rangle = 3$  vs. 47% for 39-52. WT exhibits significant  $D_{\%}$  for the 99-115 and 151-191 regions that are structured and on the exterior of the trimer of hairpins, and I173E exhibits even higher  $D_{\%}$  for these regions, e.g. 60 vs. 50% for the 172-178 peptide that contains the mutation site. Fig. 4.7 displays a structural model for I173E-HA2 based on these observations in which the WT trimer-of-hairpins structure is modified by dissociation of the 154-178 C-terminal exterior region from the 23-69 interior N-terminal region.

Another change is destabilization of structure in the N-terminal region which also correlates with reduced fractional helicity for I173E vs. WT, 45 vs. 58%.<sup>19</sup> The larger  $D_{\%}$  of I173E vs. WT for the 22-69 and 150-191 regions is attributed both to increased aqueous exposure of these locally dissociated segments, and to loss of regular secondary structure in these segments.

For the same peptide, G1E exhibits  $\langle D_{\%} \rangle$  similar to WT for the 22-38 and 53-150 regions, and  $\langle D_{\%} \rangle$  intermediate between WT and I173E in the 150-191 regions, e.g. 54 vs. 49 and 56% for 168-171, and 27 vs. 15 and 32% for 188-191. G1E exhibits bimodal behavior for 39-52, and at  $t_{\text{HDX}} = 120$  min, there are comparable populations of each mode with respective  $D_{\%} = 16$  or 63% that correlate with  $D_{\%} = 6\%$  for WT and 56% for I173E (Table 4C). These data are incorporated into the Fig. 4.2 and 4.7 models for G1E that include an equilibrium between the trimer-of-hairpins WT structure and the C-terminal strand-dissociated I173E structure. This model correlates with similar G1E and WT fractional helicities and with lower  $T_m$  for G1E vs. WT.

SE structural modification by FP mutation is unanticipated because the trimer-of-hairpins structure of the SE is thermostable in the absence of the FP.<sup>18</sup> The FP's of WT, I173E, and G1E all exhibit substantial  $D_{\%}$ , with smaller  $D_{\%}$  in the 1-9 regions of I173E and G1E vs. WT, and smaller  $D_{\%}$  for the 10-22 regions of G1E vs. WT and I173E, e.g.  $\langle D_{\%} \rangle_{10-21} = 45\%$  for G1E vs. 54% for WT. We synthesize these results into the Fig. 4.7 G1E model by showing binding between a non-helical hairpin FP and dissociated C-terminal SE region. This model is consistent with 150-191  $D_{\%}$  for G1E that are intermediate between WT and I173E, and with FP  $D_{\%}$  for G1E smaller than WT and I173E. The partially unfolded SE with bound FP is energetically feasible because the G1E mutation may destabilize the helical hairpin structure of the FP via reduction of the N-terminal  $\text{NH}_3^+$  charge/12-22 helix dipole interaction.<sup>61</sup>

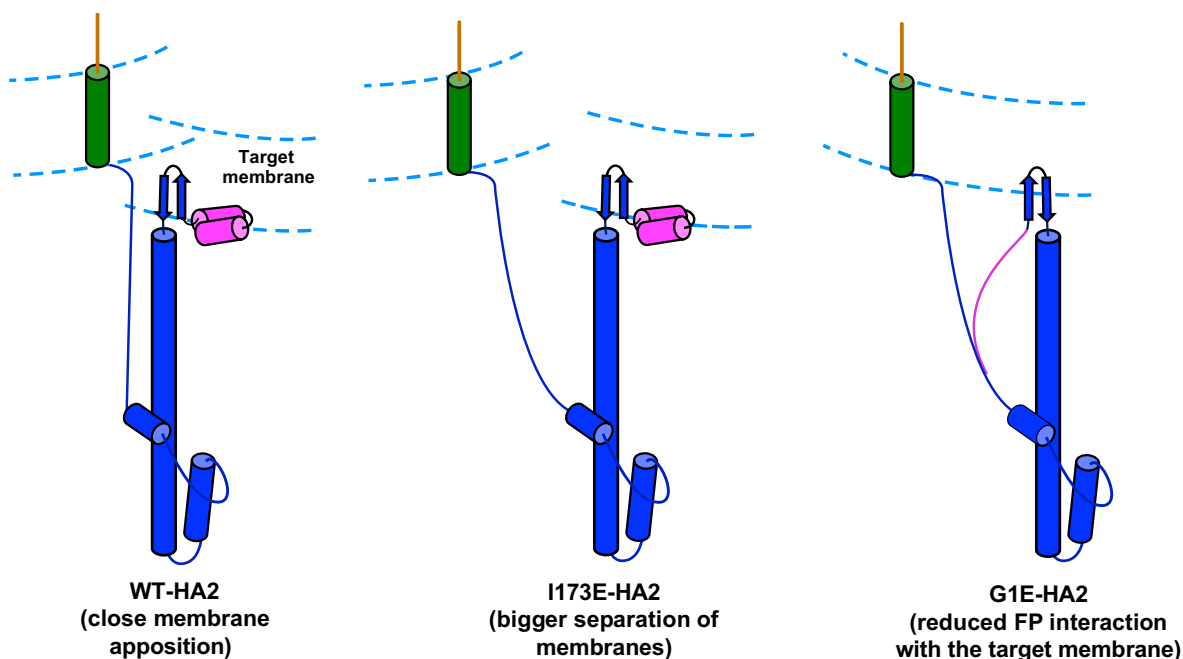


#### 4.4.4 Fusion implications of the HDX results

The HDX results for WT-HA2 support: (1) FP which doesn't form a complex with TM and which has high aqueous exposure; (2) structured and aqueous-protected 23-35 region between the FP and the trimer-of-hairpins SE; (3) membrane traversal by helical TM; and (4) aqueous-exposed and disordered endodomain. The I173E-HA2 results support dissociation of the 154-178 exterior strands from the 23-75 interior region and loss of secondary structure in these regions. The G1E-HA2 results support two distinct SE states with either the WT trimer-of-hairpins structure or the I173E partially -dissociated and -unfolded structure. The latter state also has unfolding of the FP helical hairpin and FP binding to the SE.

A FP/TM complex with membrane traversal of all segments has been commonly pictured in the final HA2 state of fusion models, and has been proposed to be important for the final membrane pore expansion step. WT finding (1) is inconsistent with this complex. It has been shown that the TM is necessary for pore expansion in HA-cell/RBC fusion, and that efficient fusion occurs with replacement of the HA2 TM with a TM sequence of a non-homologous protein.<sup>13, 62</sup> The integrated results support a model in which TM segments facilitate pore expansion by shielding acyl chains of lipids from water during their transition from the hemifusion diaphragm separating the two bodies to the final continuous bilayer enclosing the fused body. In addition, the positive membrane curvature associated with the HA2 structure in Fig. 4.7 may aid pore expansion, which has an associated decrease in negative membrane curvature. The positive curvature in Fig. 4.7 is a consequence of the constraints of the semi-rigid and hyperthermostable HA2 structure and the FP interfacial and TM traversal locations in the membrane. Curvature would probably be decreased with deletion of the TM from HA2. TM deletion would also allow deeper FP insertion which correlates with the smaller  $\langle D_{\%} \rangle$  for the FP in FHA2 vs. HA2.

The data from previous and present studies support a model in which the WT final trimer-of-hairpins SE structure holds the two membranes in close apposition prior to fusion (Fig. 4.8). The close inter-membrane distance compensates for much of the ~60 kT (computed) activation barrier between the initial two-membranes state and the first “stalk” fusion intermediate.<sup>3</sup> Subsequent hemifusion and pore formation steps have much lower computed activation barriers, and occur with the SE remaining in the hairpin state. There is also some contribution to fusion from FP-induced membrane perturbation.<sup>18, 29</sup> The close WT apposition in Fig. 4.8 is made possible by the hyperthermostability of the HA2 trimer-of-hairpins and the 23-35 antiparallel  $\beta$  sheet that provides a semi-rigid connection between the FP and the SE. Highly-impaired fusion for I173E is a consequence of a more flexible SE hairpin and more distant membranes, where the former is evidenced by lower  $T_m$  and a large increase in aqueous exposure of the SE (Fig. 4.8). G1E also exhibits increased aqueous exposure of the SE, as well as reduced aqueous exposure of the FP. The latter effect is explained by a less stable FP helical hairpin and consequent FP/SE binding. G1E fusion is therefore impaired by reduced FP interaction with the target membrane and by membrane apposition at greater distance (Fig. 4.8).



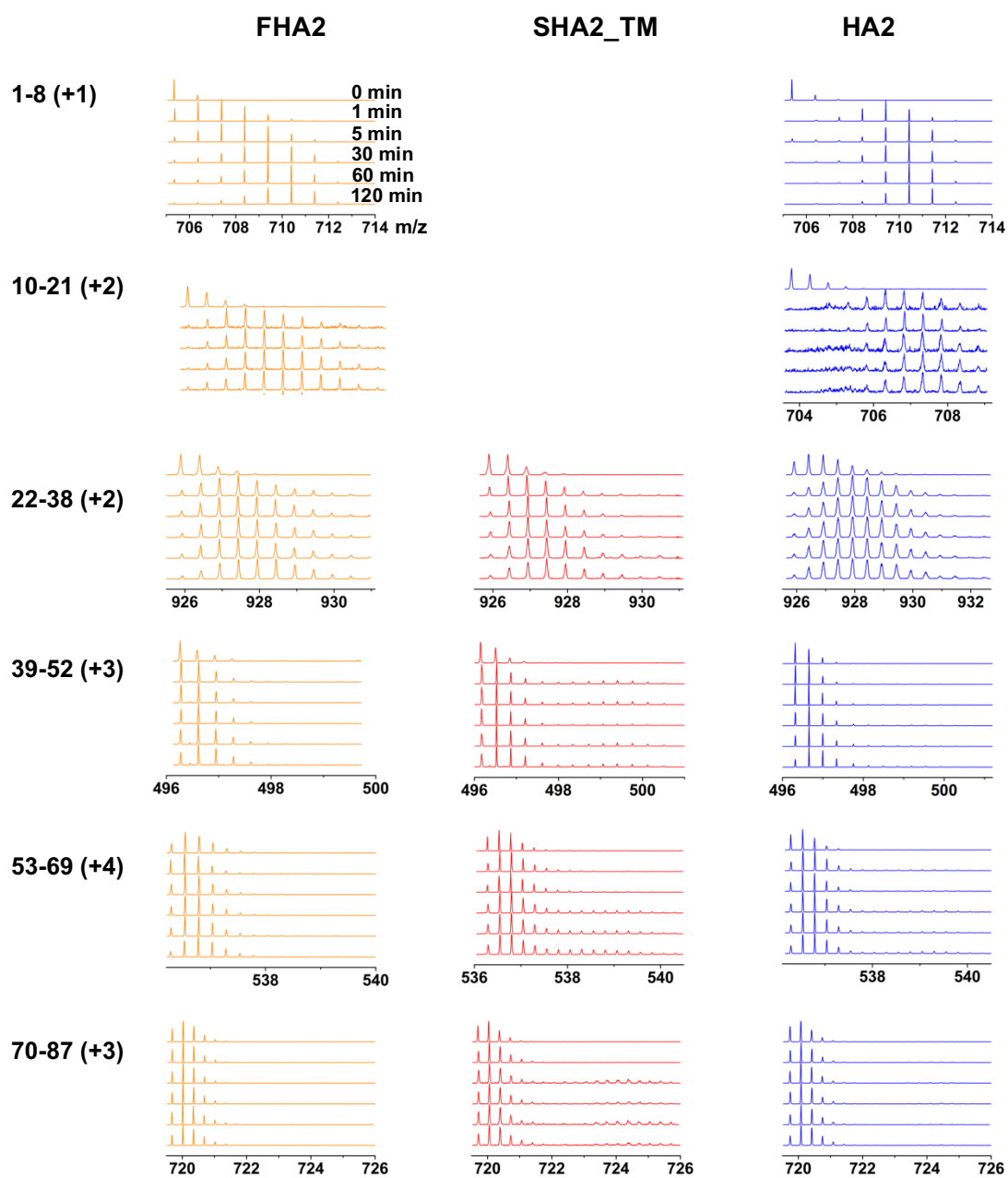
**Figure 4.8** Pictorial representations of membrane apposition prior to fusion for WT-, I173E-, and G1E-HA2. For WT, the TM traverses the virus membrane, the FP is interfacially-bound to the target membrane, and the membranes are held in close apposition by the semi-rigid SE trimer-of-hairpins and the 23-35 antiparallel  $\beta$  sheet. For I173E, dissociation of the C-terminal strands from the trimer-of-hairpins results in larger average distance between the membranes. G1E also exhibits more-distant membrane apposition because of partial FP/SE binding and concurrent weaker FP binding to the target membrane. The figure color-coded like the domains in Fig. 4.1.

#### 4.5 Summary

HDX-MS has been applied to subunit 2 of the influenza virus hemagglutinin protein that mediates fusion between the virus and endosome membranes. High aqueous exposure of the fusion peptide and high aqueous protection of the transmembrane region are consistent with separate domains that have respective membrane interfacial and traversal locations. The HDX-MS data are also consistent with a trimer-of-hairpins structure for the soluble ectodomain. High aqueous

protection of the 22-38 peptide supports a well-structured segment and are consistent with retention of the 23-35 antiparallel  $\beta$  sheet found in the initial HA2/HA1 complex. The  $\beta$  sheet provides a semi-rigid connection between the fusion peptide and soluble ectodomain, and aids close apposition of the two membranes prior to fusion. HDX-MS was also done on the G1E and I173E mutants that result in highly-impaired fusion. I173E shows much greater aqueous exposure *vs.* WT of the 22-69 and 150-191 regions, which correlates with dissociation of the C-terminal strands of the trimer-of-hairpins. G1E has aqueous exposure of these regions intermediate between WT and I173E, as well as increased protection of the fusion peptide. These data are consistent with binding of the G1E fusion peptide to the dissociated C-terminal strands. The HDX-MS data support a model in which reduced fusion by the mutants is due to larger distance between the apposed membranes, and for G1E, also loss of fusion peptide contact with the target membrane.

## APPENDIX



**Figure 4A.** HDX-MS data for peptides of FHA2, SHA2\_TM, and HA2. Plots of mass spectral signal intensity vs. m/z are displayed for all  $t_{\text{HDX}}$ . The identity and charge of each peptide is given in the left column.

Figure 4A (cont'd)

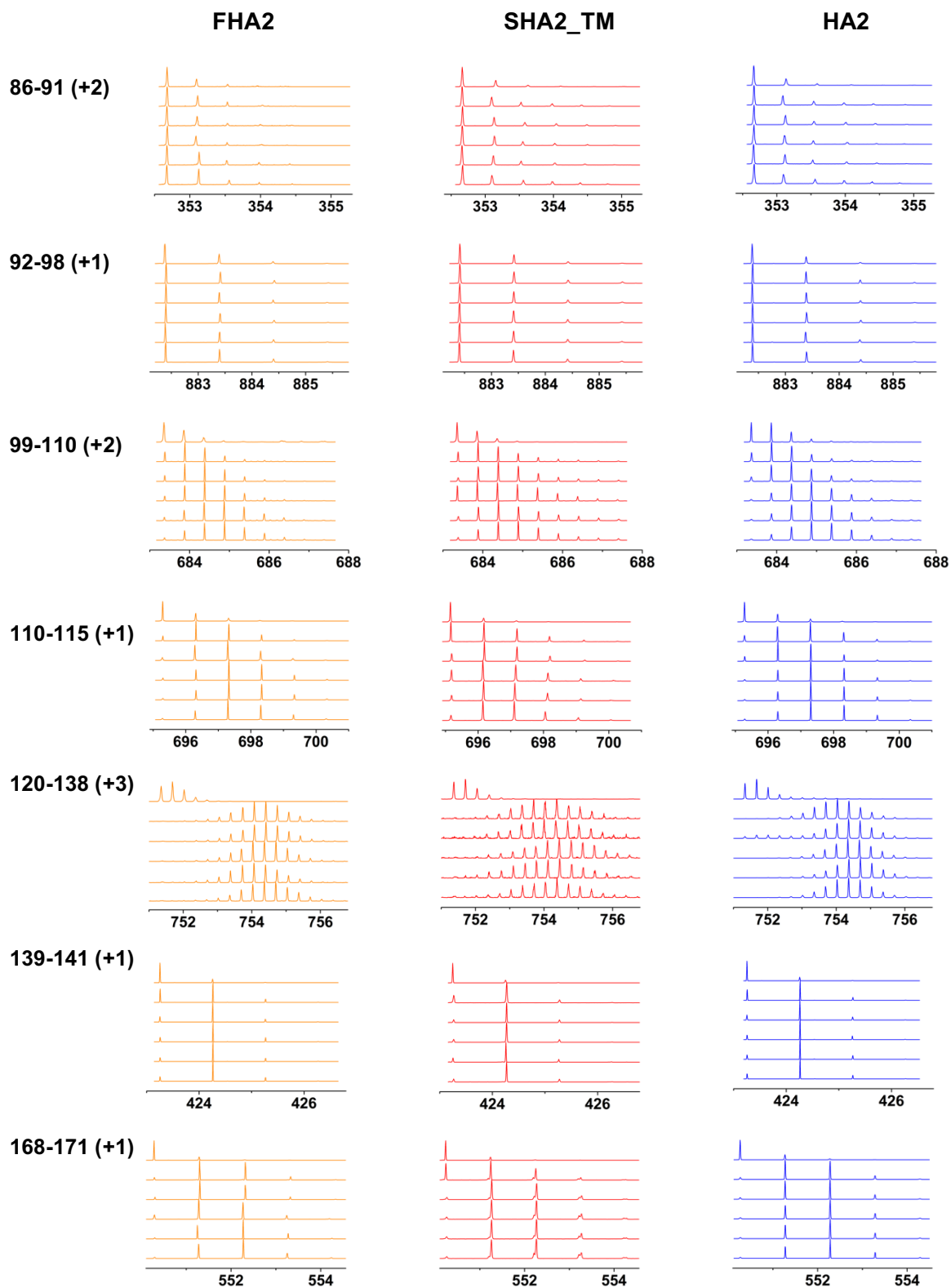
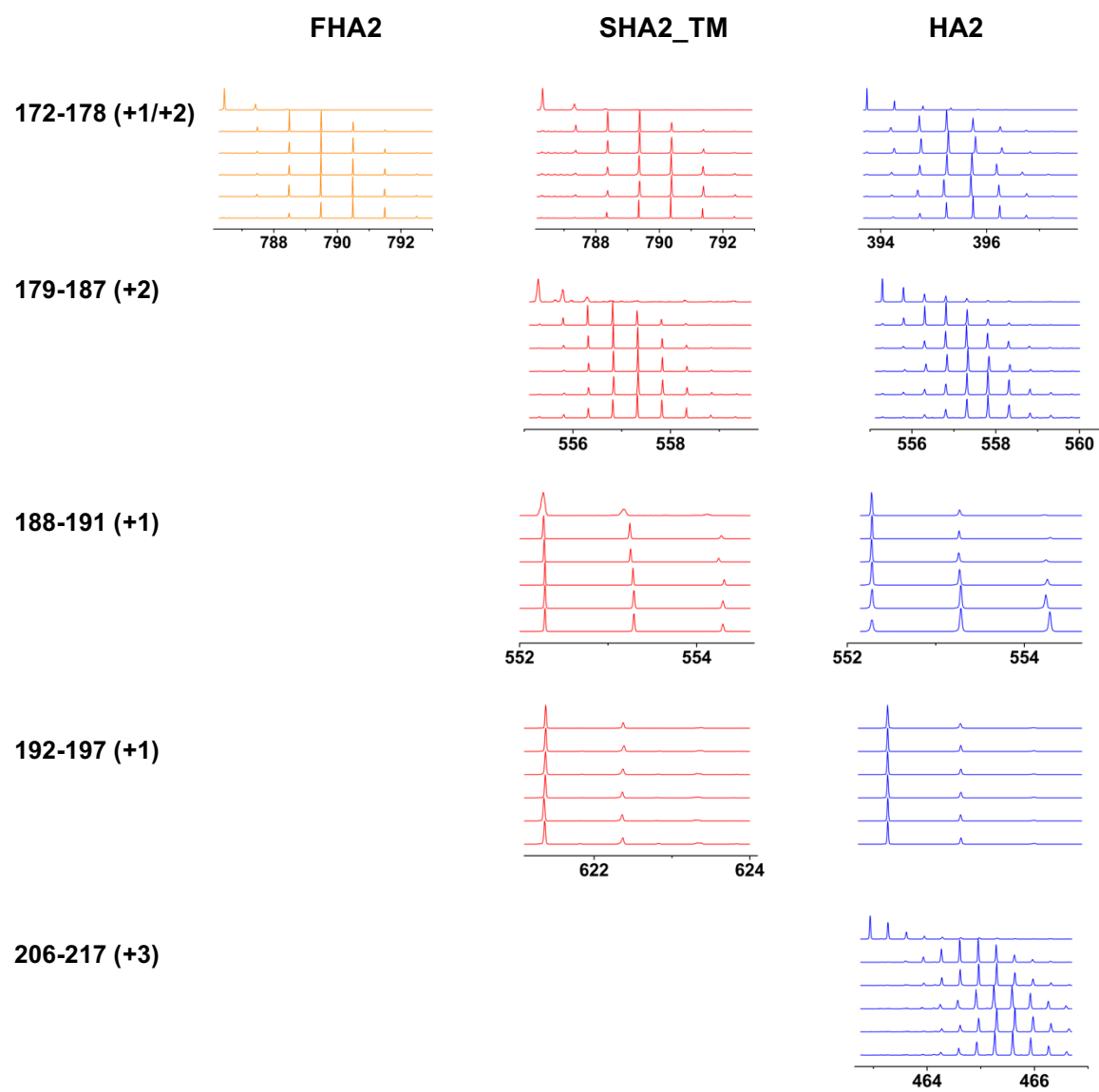
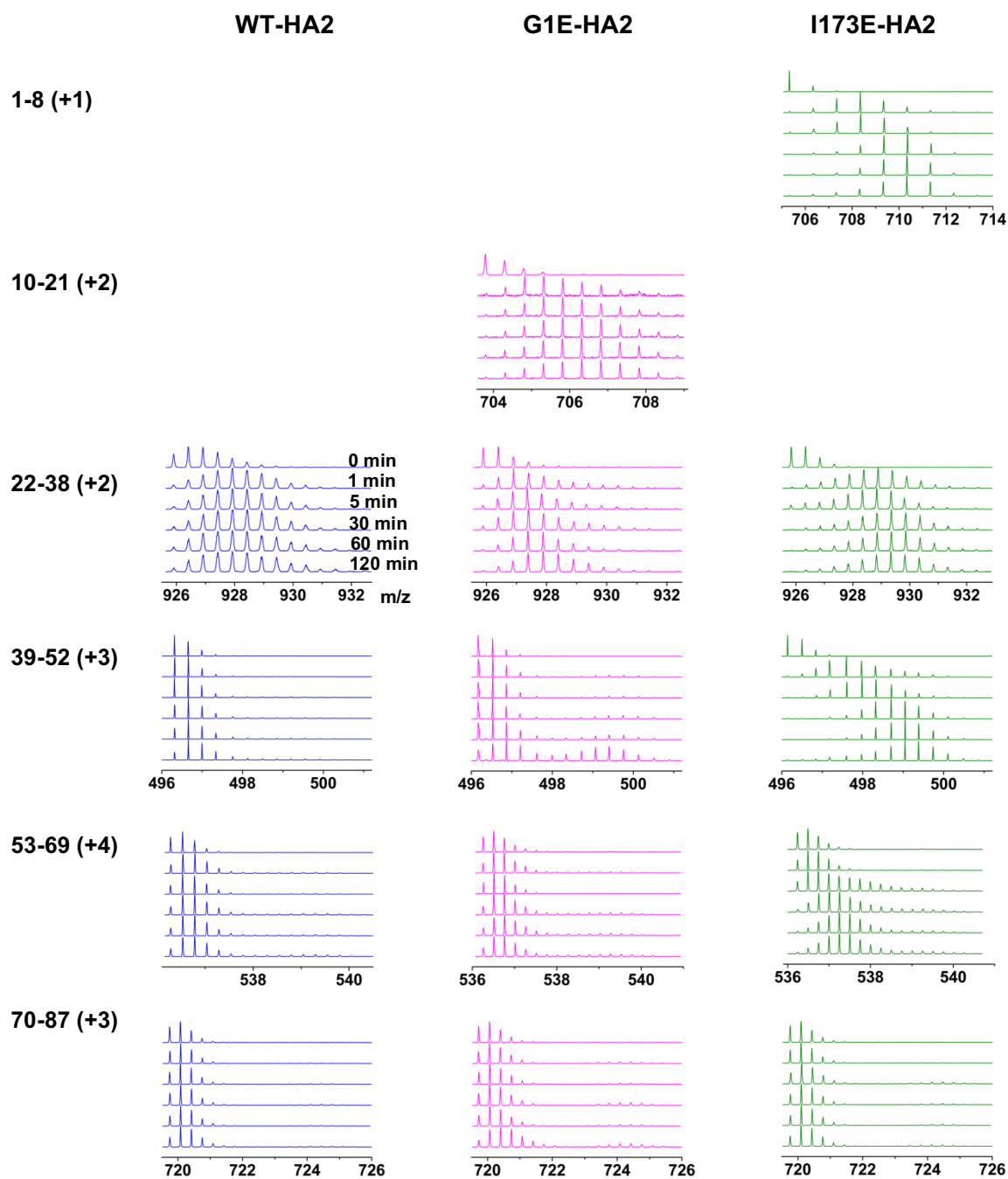


Figure 4A (cont'd)







**Figure 4B.** HDX-MS data for peptides of WT-, G1E-, and I173E- HA2. Plots of mass spectral signal intensity vs. m/z are displayed for all  $t_{\text{HDX}}$ . The identity and charge of each peptide is given in the left column.

Figure 4B (cont'd)

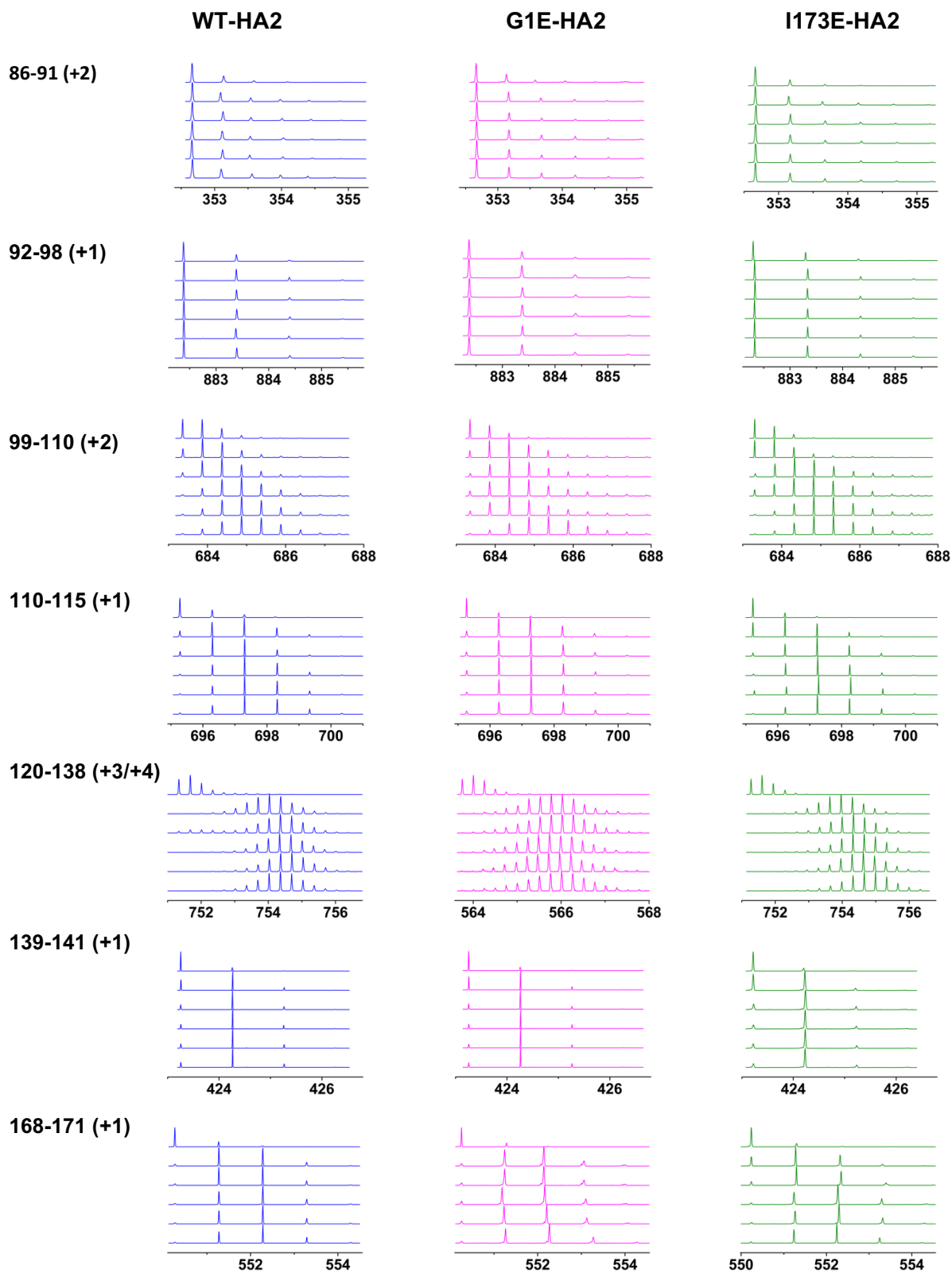
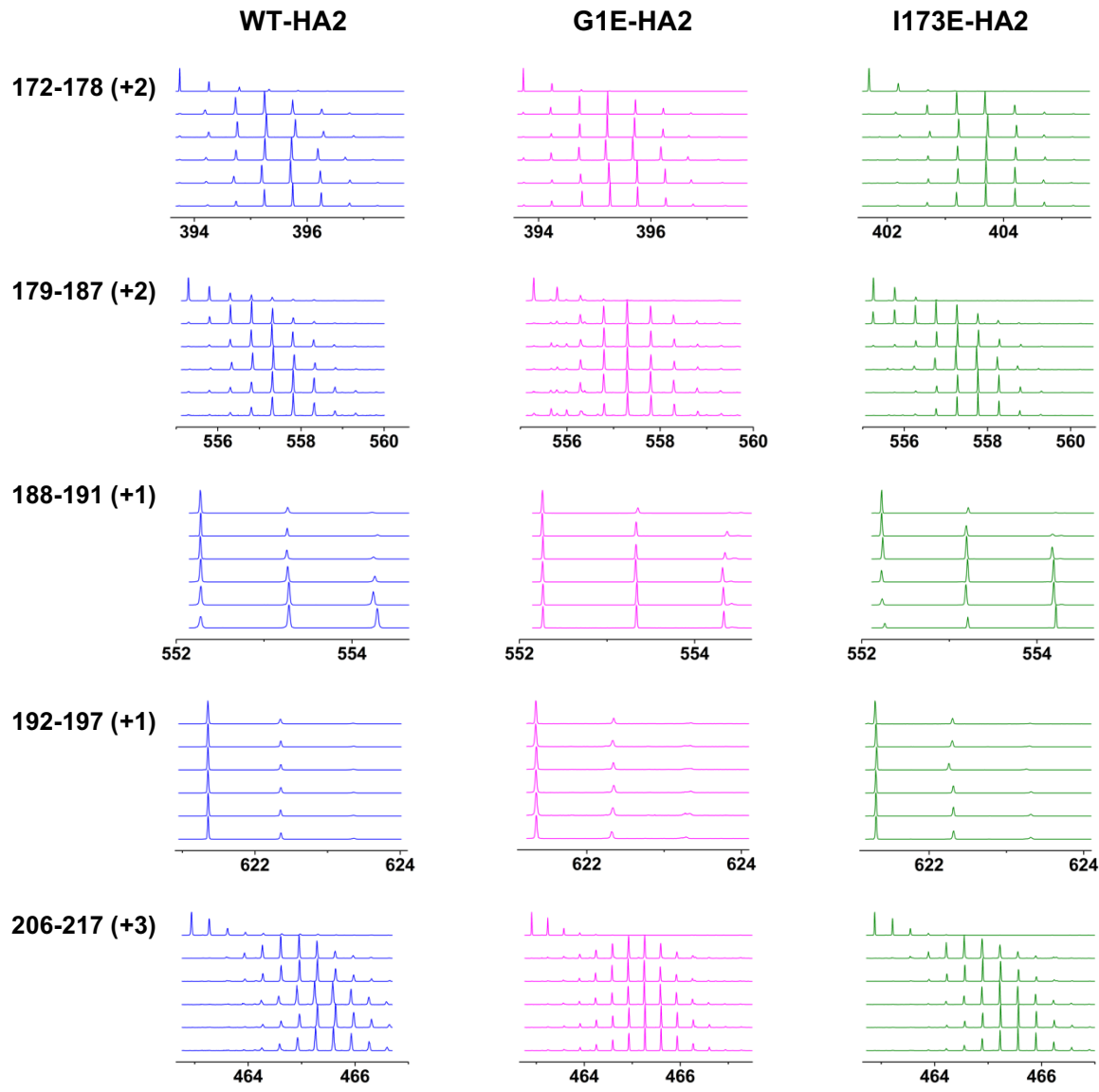
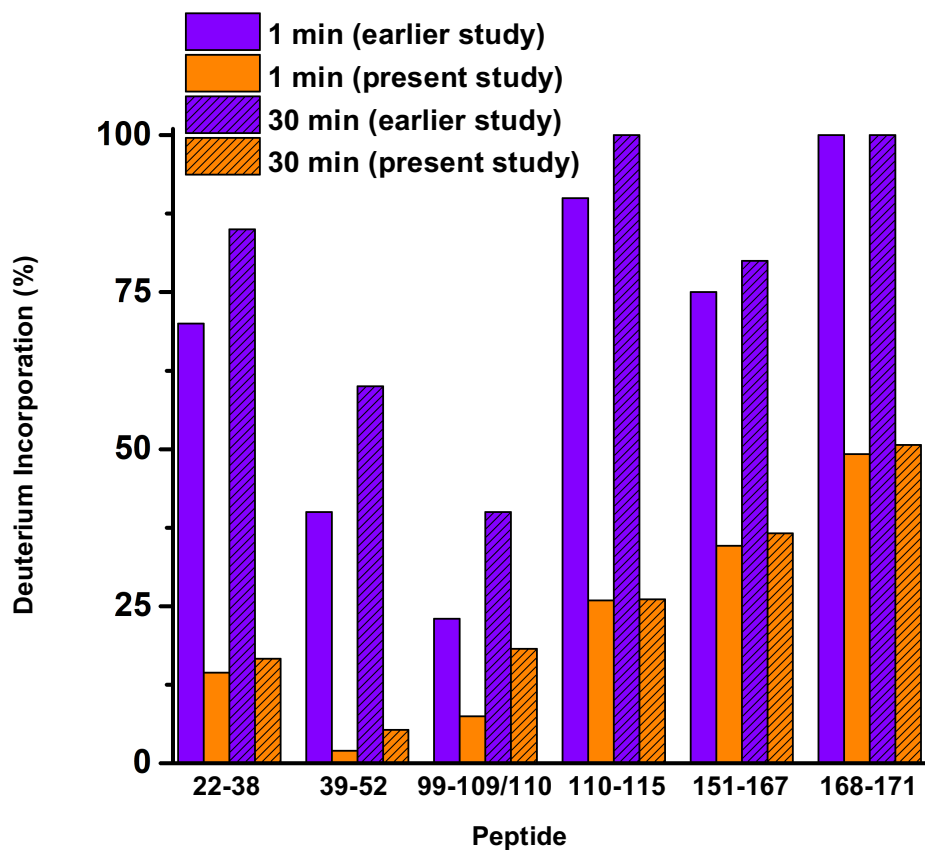


Figure 4B (cont'd)





**Figure 4C.** Comparison between  $D_{\%}$  of the present study and an earlier study for  $t_{\text{HDX}} = 1$  min and  $t_{\text{HDX}} = 30$  min (*Structure* (2015) **23**, 665-676). Data are displayed for peptides respectively from FHA2  $\equiv$  HA2<sub>1-185</sub> and from the HA2<sub>1-175</sub>/HA1 complex after incubation at pH 4.9 for three hours.

**Tables 4A.** Percent deuterium incorporation vs. HDX time for peptides from FHA2, SHA2\_TM, and HA2. Each value is the average of triplicate measurements, and the standard deviation is given in parentheses.

HDX time/min	Deuterium Incorporation (%) (1-8)	
	FHA2	HA2
1	22.86 (0.62)	47.00 (1.21)
5	34.86 (1.03)	63.05 (0.73)
30	48.77 (1.19)	66.38 (1.72)
60	56.21 (1.71)	68.90 (1.21)
120	61.71 (1.62)	69.76 (0.46)

HDX time/min	Deuterium Incorporation (%) (1-9)	
	FHA2	HA2
1	19.79 (0.85)	45.31 (2.21)
5	36.75 (0.43)	61.88 (2.12)
30	55.50 (0.82)	67.13 (2.29)
60	56.25 (0.00)	69.92 (2.74)
120	65.83 (1.57)	70.93 (0.26)

HDX time/min	Deuterium Incorporation (%) (10-21)	
	FHA2	HA2
1	34.58 (2.78)	43.58 (0.18)
5	37.27 (0.77)	46.09 (0.00)
30	39.27 (0.68)	57.05 (1.09)
60	39.91 (1.02)	57.07 (1.09)
120	41.21 (1.28)	65.59 (0.06)

HDX time/min	Deuterium Incorporation (%) (4-11) FHA2
1	20.86 (0.51)
5	28.33 (0.21)
30	37.10 (2.06)
60	42.05 (3.13)
120	45.00 (0.01)

**Tables 4A (cont'd)**

HDX time/min	Deuterium Incorporation (%) (22-38)		
	FHA2	SHA2 TM	HA2
1	14.40 (2.38)	8.69 (0.00)	12.83 (2.02)
5	15.73 (1.15)	10.79 (0.85)	15.81 (0.50)
30	16.60 (0.07)	13.33 (2.17)	17.60 (1.87)
60	18.25 (0.92)	15.63 (1.32)	19.69 (1.77)
120	20.98 (1.16)	17.19 (0.60)	21.21 (0.22)

HDX time/min	Deuterium Incorporation (%) (39-52)		
	FHA2	SHA2 TM	HA2
1	1.95 (0.22)	6.18 (0.32)	0.74 (0.24)
5	3.90 (1.35)	7.08 (0.25)	1.59 (0.51)
30	5.28 (0.58)	7.61 (0.27)	3.49 (0.85)
60	5.79 (0.54)	7.97 (0.40)	4.64 (0.45)
120	6.77 (0.75)	11.97 (0.09)	5.85 (0.23)

HDX time/min	Deuterium Incorporation (%) (53-69)		
	FHA2	SHA2 TM	HA2
1	1.63 (1.23)	4.25 (0.68)	4.68 (0.77)
5	2.91 (0.49)	4.79 (0.91)	4.63 (0.18)
30	3.72 (0.31)	5.04 (0.84)	5.99 (0.61)
60	4.00 (0.35)	5.63 (0.31)	6.00 (0.35)
120	4.41 (0.93)	5.53 (0.04)	6.21 (0.44)

HDX time/min	Deuterium Incorporation (%) (70-87)		
	FHA2	SHA2 TM	HA2
1	0.09 (0.12)	2.21 (0.04)	1.68 (0.21)
5	0.29 (0.41)	3.09 (0.29)	1.68 (0.04)
30	1.26 (1.54)	3.44 (0.04)	2.00 (0.17)
60	1.47 (0.99)	3.97 (0.37)	2.21 (0.04)
120	5.18 (3.41)	4.65 (0.58)	2.50 (0.04)

HDX time/min	Deuterium Incorporation (%) (86-91)		
	FHA2	SHA2 TM	HA2
1	1.00 (1.56)	0.20 (0.28)	2.07 (1.33)
5	0.40 (0.28)	2.70 (2.40)	1.93 (0.41)
30	0.47 (0.23)	5.00 (0.84)	1.50 (0.14)
60	1.20 (1.38)	5.70 (0.98)	3.53 (1.89)
120	6.00 (0.20)	7.20 (0.20)	6.07 (1.33)

**Tables 4A (cont'd)**

HDX time/min	Deuterium Incorporation (%) (92-98)		
	FHA2	SHA2 TM	HA2
1	0.78 (0.09)	0.61 (0.53)	1.58 (0.12)
5	0.67 (0.16)	0.89 (0.09)	0.94 (0.25)
30	0.92 (0.11)	0.83 (0.16)	0.94 (0.41)
60	0.83 (0.00)	0.78 (0.09)	1.06 (0.25)
120	0.89 (0.19)	0.61 (0.25)	0.94 (0.25)

HDX time/min	Deuterium Incorporation (%) (92-99)		
	FHA2	SHA2 TM	HA2
1	1.19 (0.41)	3.93 (0.30)	4.64 (0.10)
5	1.52 (0.43)	3.76 (0.35)	0.90 (0.08)
30	1.64 (0.70)	3.62 (0.16)	1.90 (0.86)
60	1.64 (0.90)	4.21 (0.50)	2.09 (0.82)
120	2.67 (0.70)	1.67 (0.54)	1.38 (0.36)

HDX time/min	Deuterium Incorporation (%) (99-110)		
	FHA2	SHA2 TM	HA2
1	7.48 (0.54)	10.15 (1.99)	4.15 (0.10)
5	13.33 (1.42)	18.85 (0.84)	10.91 (0.86)
30	18.23 (0.32)	21.00 (1.10)	21.15 (0.60)
60	20.50 (0.32)	22.09 (0.72)	24.12 (0.19)
120	20.97 (0.72)	23.18 (0.51)	25.73 (0.15)

HDX time/min	Deuterium Incorporation (%) (110-115)		
	FHA2	SHA2 TM	HA2
1	25.90 (1.55)	17.13 (1.22)	26.33 (0.61)
5	25.40 (1.41)	21.13 (2.51)	28.00 (1.59)
30	26.10 (0.14)	22.90 (0.99)	33.80 (0.00)
60	29.60 (2.83)	26.60 (1.25)	36.50 (0.14)
120	33.20 (0.00)	28.20 (0.60)	39.73 (0.61)

HDX time/min	Deuterium Incorporation (%) (120-138)		
	FHA2	SHA2 TM	HA2
1	38.39 (1.72)	39.72 (0.00)	42.50 (3.22)
5	38.93 (1.57)	39.83 (2.84)	42.44 (1.86)
30	39.56 (0.39)	40.72 (0.02)	42.11 (2.19)
60	43.72 (1.26)	42.17 (1.41)	43.97 (0.71)
120	45.86 (0.27)	44.56 (3.61)	43.03 (0.04)

**Tables 4A (cont'd)**

HDX time/min	Deuterium Incorporation (%) (133-138)		
	FHA2	SHA2_TM	HA2
1	34.33 (0.94)	33.87 (0.11)	39.33 (0.11)
5	34.83 (1.17)	39.00 (1.13)	41.73 (0.98)
30	35.83 (0.00)	42.93 (0.75)	42.07 (0.46)
60	35.94 (1.00)	44.80 (0.69)	43.53 (1.72)
120	37.67 (2.02)	45.27 (0.23)	44.33 (1.55)

HDX time/min	Deuterium Incorporation (%) (139-141)		
	FHA2	SHA2_TM	HA2
1	34.50 (2.17)	37.25 (1.76)	31.00 (1.80)
5	36.75 (1.06)	46.25 (1.76)	43.33 (3.78)
30	46.50 (0.00)	50.25 (1.06)	46.83 (1.60)
60	46.75 (0.35)	50.75 (0.35)	48.75 (1.06)
120	47.75 (0.35)	50.67 (0.28)	49.50 (1.32)

HDX time/min	Deuterium Incorporation (%) (142-150)		
	FHA2	SHA2_TM	HA2
1	45.13 (0.87)	46.63 (0.00)	41.83 (0.14)
5	46.04 (1.22)	46.96 (0.14)	44.63 (0.00)
30	47.29 (2.30)	49.96 (2.67)	48.46 (0.07)
60	47.08 (3.10)	51.08 (0.07)	48.38 (0.00)
120	47.88 (0.00)	51.04 (0.07)	51.92 (0.14)

HDX time/min	Deuterium Incorporation (%) (151-162)	
	FHA2	HA2
1	32.91 (0.00)	30.88 (0.57)
5	35.18 (1.41)	34.85 (0.31)
30	37.76 (0.10)	40.39 (1.63)
60	39.42 (2.93)	41.55 (0.68)
120	40.42 (0.05)	42.06 (0.37)

HDX time/min	Deuterium Incorporation (%) (151-167)		
	FHA2	SHA2_TM	HA2
1	34.62 (0.64)	38.21 (1.60)	33.09 (0.75)
5	35.87 (1.40)	40.88 (1.92)	34.50 (1.03)
30	36.60 (2.15)	41.25 (1.20)	38.75 (1.25)
60	38.89 (2.18)	41.90 (0.25)	40.33 (1.58)
120	41.06 (1.67)	43.79 (1.75)	40.83 (1.43)



**Tables 4A (cont'd)**

HDX time/min	Deuterium Incorporation (%) (163-167)		
	FHA2	SHA2_TM	HA2
1	36.12 (0.17)	30.33 (0.14)	43.75 (1.06)
5	40.37 (3.00)	46.12 (1.94)	44.25 (0.25)
30	45.50 (1.76)	45.67 (1.84)	45.75 (0.35)
60	47.08 (0.14)	46.62 (1.59)	46.42 (1.01)
120	47.58 (0.76)	52.25 (1.41)	47.00 (0.75)

HDX time/min	Deuterium Incorporation (%) (163-171)		
	FHA2	SHA2_TM	HA2
1	35.94 (1.67)	33.83 (2.52)	31.88 (1.23)
5	37.13 (1.59)	35.81 (2.20)	31.19 (0.44)
30	33.25 (3.89)	36.71 (1.94)	36.06 (0.08)
60	35.50 (2.62)	35.38 (0.79)	35.08 (0.76)
120	34.50 (1.06)	35.19 (1.32)	37.17 (0.68)

HDX time/min	Deuterium Incorporation (%) (168-171)		
	FHA2	SHA2_TM	HA2
1	49.22 (0.50)	34.33 (0.88)	39.67 (0.50)
5	49.44 (0.38)	55.50 (0.70)	48.50 (0.38)
30	50.67 (2.29)	58.83 (1.64)	51.44 (3.29)
60	55.00 (2.64)	59.22 (1.38)	51.56 (2.65)
120	60.56 (0.76)	63.44 (1.34)	53.67 (0.76)

HDX time/min	Deuterium Incorporation (%) (172-178)		
	FHA2	SHA2_TM	HA2
1	40.72 (0.85)	37.33 (2.82)	38.08 (0.82)
5	47.25 (1.29)	50.06 (1.10)	43.61 (2.89)
30	50.75 (1.06)	50.25 (1.53)	55.06 (1.10)
60	53.17 (1.17)	53.75 (3.41)	56.06 (1.00)
120	58.25 (0.35)	55.39 (0.96)	56.42 (0.35)

HDX time/min	Deuterium Incorporation (%) (179-187)	
	SHA2_TM	HA2
1	32.63 (0.35)	25.63 (0.65)
5	37.06 (1.67)	38.75 (1.76)
30	40.75 (1.11)	38.38 (1.51)
60	41.54 (1.16)	45.08 (1.01)
120	43.50 (0.33)	47.50 (0.65)

**Tables 4A (cont'd)**

HDX time/min	Deuterium Incorporation (%) (188-191)	
	SHA2 TM	HA2
1	6.78 (0.38)	2.56 (0.19)
5	6.00 (0.01)	3.89 (0.19)
30	10.33 (0.00)	15.00 (0.00)
60	15.89 (2.50)	24.00 (0.57)
120	17.33 (0.00)	31.11 (0.19)

HDX time/min	Deuterium Incorporation (%) (192-197)	
	SHA2 TM	HA2
1	1.07 (0.64)	2.00 (0.20)
5	0.93 (0.30)	1.50 (0.14)
30	0.50 (0.14)	2.40 (0.53)
60	0.87 (0.11)	2.47 (0.50)
120	9.60 (1.41)	2.60 (0.34)

HDX time/min	Deuterium Incorporation (%) (198-202/199-203)	
	SHA2 TM	HA2
1	0.88 (0.88)	2.00 (0.17)
5	2.17 (0.29)	0.63 (0.17)
30	1.17 (0.87)	1.00 (1.52)
60	0.88 (0.53)	2.17 (1.44)
120	0.83 (0.14)	0.42 (0.14)

HDX time/min	Deuterium Incorporation (%) (200-202)	
	SHA2 TM	HA2
1	0.33 (0.28)	0.00 (0.00)
5	0.17 (0.28)	0.17 (0.28)
30	0.17 (0.28)	0.17 (0.28)
60	0.67 (0.76)	0.17 (0.28)
120	0.67 (0.76)	0.17 (0.28)

HDX time/min	Deuterium Incorporation (%) HA2 (200-205)
1	0.53 (0.23)
5	0.40 (0.00)
30	1.00 (0.59)
60	1.53 (0.11)
120	1.00 (0.20)

**Tables 4A (cont'd)**

HDX time/min	Deuterium Incorporation (%) HA2 (206-217)
1	50.46 (2.76)
5	54.12 (2.09)
30	59.33 (3.22)
60	61.50 (1.80)
120	61.64 (1.10)

**Tables 4B.** Percent deuterium incorporation vs. HDX time for peptides from digested WT-, G1E-, and I173E- HA2. Each value is the average of triplicate measurements, and the standard deviation is given in parentheses.

HDX time/min	Deuterium Incorporation (%) (1-8)	
	WT-HA2	I173E-HA2
1	47.00 (1.21)	37.14 (2.45)
5	63.05 (0.73)	44.79 (0.91)
30	66.38 (1.72)	60.00 (2.62)
60	68.90 (1.21)	63.29 (0.40)
120	69.76 (0.46)	64.21 (0.10)

HDX time/min	Deuterium Incorporation (%) (3-9)	
	G1E-HA2	I173E-HA2
1	42.17 (0.94)	42.42 (2.47)
5	54.94 (1.86)	49.11 (3.26)
30	58.94 (2.77)	59.58 (3.41)
60	59.08 (0.82)	62.25 (3.89)
120	60.33 (3.49)	63.28 (3.00)

HDX time/min	Deuterium Incorporation (%) (10-21)	
	WT-HA2	G1E-HA2
1	43.58 (0.18)	38.14 (2.51)
5	46.09 (0.00)	43.67 (0.50)
30	57.05 (1.09)	45.39 (0.71)
60	57.07 (1.09)	46.03 (1.08)
120	65.59 (0.06)	50.18 (1.98)

HDX time/min	Deuterium Incorporation (%) (12-22)	
	G1E-HA2	I173E-HA2
1	43.83 (2.92)	52.25 (3.16)
5	45.10 (3.53)	63.13 (4.84)
30	49.67 (3.35)	65.73 (0.67)
60	51.30 (3.13)	66.73 (0.90)
120	59.05 (1.20)	68.07 (0.66)

**Tables 4B (cont'd)**

HDX time/min	Deuterium Incorporation (%) (22-38)		
	WT-HA2	G1E-HA2	I173E-HA2
1	12.83 (2.02)	14.67 (0.60)	30.06 (1.24)
5	15.81 (0.50)	15.96 (1.60)	33.34 (0.04)
30	17.60 (1.87)	17.15 (1.02)	35.75 (0.89)
60	19.69 (1.77)	18.04 (0.65)	37.77 (0.97)
120	21.21 (0.22)	20.65 (1.92)	39.23 (1.07)

HDX time/min	Deuterium Incorporation (%) (39-52)		
	WT-HA2	G1E-HA2	I173E-HA2
1	0.90 (0.22)		35.27 (3.86)
5	1.56 (0.54)		39.36 (1.76)
30	4.10 (0.55)	EX1 kinetics	51.58 (0.60)
60	4.85 (0.16)		53.85 (0.62)
120	5.95 (2.05)		54.95 (0.42)

HDX time/min	Deuterium Incorporation (%) (53-69)		
	WT-HA2	G1E-HA2	I173E-HA2
1	4.68 (0.77)	3.21 (0.77)	9.02 (0.12)
5	4.63 (0.18)	2.82 (1.30)	16.77 (0.03)
30	5.99 (0.61)	4.10 (0.22)	18.04 (0.82)
60	6.00 (0.35)	5.16 (1.10)	21.53 (1.76)
120	6.21 (0.44)	5.33 (1.75)	26.83 (0.64)

HDX time/min	Deuterium Incorporation (%) (70-87)		
	WT-HA2	G1E-HA2	I173E-HA2
1	1.68 (0.21)	1.03 (0.04)	1.23 (0.35)
5	1.68 (0.04)	1.32 (0.12)	1.64 (0.21)
30	2.00 (0.17)	1.91 (0.04)	1.71 (0.21)
60	2.21 (0.04)	3.03 (1.29)	2.03 (0.27)
120	2.50 (0.04)	7.35 (1.58)	2.13 (0.20)

HDX time/min	Deuterium Incorporation (%) (86-91)		
	WT-HA2	G1E-HA2	I173E-HA2
1	2.07 (1.33)	0.20 (0.34)	0.73 (0.46)
5	1.93 (0.41)	0.47 (0.42)	0.93 (0.46)
30	1.50 (0.14)	3.73 (0.11)	0.60 (0.00)
60	3.53 (1.89)	4.73 (1.28)	0.93 (0.30)
120	6.07 (1.33)	6.67 (2.33)	1.33 (0.81)

**Tables 4B (cont'd)**

HDX time/min	Deuterium Incorporation (%) (92-98)		
	WT-HA2	G1E-HA2	I173E-HA2
1	0.78 (0.09)	4.17 (0.47)	4.72 (0.19)
5	0.67 (0.16)	3.25 (0.11)	4.72 (0.09)
30	0.92 (0.11)	4.28 (0.10)	4.33 (0.44)
60	0.83 (0.00)	4.56 (0.84)	4.44 (0.10)
120	0.89 (0.19)	4.33 (0.33)	5.00 (0.29)

HDX time/min	Deuterium Incorporation (%) (92-99)		
	WT-HA2	G1E-HA2	I173E-HA2
1	4.64 (0.10)	1.48 (0.29)	3.71 (1.61)
5	0.90 (0.08)	1.29 (0.51)	2.81 (1.30)
30	1.90 (0.86)	1.86 (1.24)	2.20 (1.03)
60	2.09 (0.82)	2.09 (1.53)	2.43 (0.87)
120	1.38 (0.36)	2.86 (2.10)	2.57 (1.51)

HDX time/min	Deuterium Incorporation (%) (99-110)		
	WT-HA2	G1E-HA2	I173E-HA2
1	4.15 (0.10)	10.00 (1.41)	10.86 (0.32)
5	10.91 (0.86)	17.76 (0.58)	12.27 (1.80)
30	21.15 (0.60)	18.61 (0.32)	24.51 (2.73)
60	24.12 (0.19)	21.45 (1.96)	29.06 (2.80)
120	25.73 (0.15)	28.79 (0.93)	30.30 (1.59)

HDX time/min	Deuterium Incorporation (%) (110-115)		
	WT-HA2	G1E-HA2	I173E-HA2
1	26.33 (0.61)	27.07 (1.70)	26.13 (4.52)
5	28.00 (1.59)	28.47 (2.14)	33.60 (0.69)
30	33.80 (0.00)	32.10 (0.99)	42.80 (1.70)
60	36.50 (0.14)	31.93 (1.51)	44.07 (0.31)
120	39.73 (0.61)	37.60 (1.98)	44.27 (0.42)

HDX time/min	Deuterium Incorporation (%) (120-138)		
	WT-HA2	G1E-HA2	I173E-HA2
1	42.50 (3.22)	37.88 (0.41)	40.31 (2.87)
5	42.44 (1.86)	38.84 (0.48)	45.67 (1.81)
30	42.11 (2.19)	39.73 (1.52)	47.11 (0.63)
60	43.97 (0.71)	39.88 (1.15)	49.41 (0.12)
120	43.03 (0.04)	41.01 (0.87)	49.96 (1.67)

**Tables 4B (cont'd)**

HDX time/min	Deuterium Incorporation (%) (133-138)		
	WT-HA2	G1E-HA2	I173E-HA2
1	39.33 (0.11)	43.30 (0.14)	33.75 (0.82)
5	41.73 (0.98)	42.67 (1.20)	36.42 (1.06)
30	42.07 (0.46)	42.20 (1.00)	37.56 (1.00)
60	43.53 (1.72)	42.87 (0.83)	37.50 (1.00)
120	44.33 (1.55)	44.07 (0.81)	38.22 (0.42)

HDX time/min	Deuterium Incorporation (%) (139-141)		
	WT-HA2	G1E-HA2	I173E-HA2
1	31.00 (1.80)	31.17 (0.29)	28.75 (1.77)
5	43.33 (3.78)	38.50 (0.00)	40.00 (3.53)
30	46.83 (1.60)	39.17 (1.04)	41.17 (1.61)
60	48.75 (1.06)	41.50 (2.00)	44.00 (0.87)
120	49.50 (1.32)	44.17 (0.29)	47.17 (2.93)

HDX time/min	Deuterium Incorporation (%) (142-150)		
	WT-HA2	G1E-HA2	I173E-HA2
1	41.83 (0.14)	49.00 (0.90)	43.25 (2.65)
5	44.63 (0.00)	49.88 (1.02)	46.13 (0.35)
30	48.46 (0.07)	50.71 (0.47)	48.17 (0.85)
60	48.38 (0.00)	51.46 (0.14)	49.75 (1.19)
120	51.92 (0.14)	51.83 (0.19)	51.54 (0.29)

HDX time/min	Deuterium Incorporation (%) (152-167)		
	WT-HA2	G1E-HA2	I173E-HA2
1	33.09 (0.75)	37.41 (0.49)	37.66 (3.67)
5	34.50 (1.03)	39.10 (0.24)	44.78 (1.54)
30	38.75 (1.25)	40.25 (0.47)	45.44 (0.65)
60	40.33 (1.58)	40.81 (0.90)	46.90 (0.67)
120	40.83 (1.43)	41.29 (0.30)	46.52 (1.15)

HDX time/min	Deuterium Incorporation (%) (163-167)		
	WT-HA2	G1E-HA2	I173E-HA2
1	43.75 (1.06)	45.50 (0.50)	44.33 (2.63)
5	44.25 (0.25)	45.63 (0.53)	47.67 (2.12)
30	45.75 (0.35)	46.17 (1.25)	49.58 (0.52)
60	46.42 (1.01)	47.25 (0.71)	50.50 (0.87)
120	47.00 (0.75)	48.67 (1.28)	51.42 (0.52)

**Tables 4B (cont'd)**

HDX time/min	Deuterium Incorporation (%) (163-171)		
	WT-HA2	G1E-HA2	I173E-HA2
1	31.88 (1.23)	34.83 (2.38)	35.75 (0.53)
5	31.19 (0.44)	35.33 (2.00)	37.88 (1.56)
30	36.06 (0.08)	36.17 (1.70)	40.75 (1.94)
60	35.08 (0.76)	36.54 (2.24)	42.04 (1.25)
120	37.17 (0.68)	37.08 (1.66)	44.13 (2.14)

HDX time/min	Deuterium Incorporation (%) (168-171)		
	WT-HA2	G1E-HA2	I173E-HA2
1	39.67 (1.33)	48.44 (2.21)	48.33 (1.88)
5	48.50 (0.71)	53.11 (3.59)	54.50 (1.18)
30	51.44 (0.19)	55.33 (3.78)	57.33 (3.21)
60	51.56 (0.19)	55.78 (5.01)	60.11 (1.02)
120	53.67 (0.00)	59.00 (0.00)	60.78 (0.77)

HDX time/min	Deuterium Incorporation (%) (172-178)		
	WT-HA2	G1E-HA2	I173E-HA2
1	38.08 (0.82)	37.08 (0.35)	54.67 (0.94)
5	43.61 (2.89)	50.22 (1.51)	58.50 (3.46)
30	55.06 (1.10)	51.67 (2.02)	61.22 (1.92)
60	56.06 (1.00)	55.08 (0.59)	62.61 (0.19)
120	56.42 (0.35)	57.67 (0.29)	63.25 (0.11)

HDX time/min	Deuterium Incorporation (%) (179-187)		
	WT-HA2	G1E-HA2	I173E-HA2
1	25.63 (0.65)	39.25 (2.47)	36.96 (1.73)
5	38.75 (1.76)	40.00 (1.77)	47.04 (1.58)
30	38.38 (1.51)	41.42 (0.56)	52.13 (1.96)
60	45.08 (1.01)	44.96 (2.94)	54.88 (2.54)
120	47.50 (0.65)	48.31 (2.21)	57.88 (2.51)

HDX time/min	Deuterium Incorporation (%) (188-191)		
	WT-HA2	G1E-HA2	I173E-HA2
1	2.56 (0.19)	12.67 (3.77)	16.33 (1.41)
5	3.89 (0.19)	25.67 (1.15)	18.00 (4.58)
30	15.00 (0.00)	30.67 (1.33)	37.00 (2.02)
60	24.00 (0.58)	32.00 (1.45)	40.89 (0.84)
120	31.11 (0.19)	34.00 (3.30)	48.33 (1.76)



**Tables 4B (cont'd)**

HDX time/min	Deuterium Incorporation (%) (192-197)		
	WT-HA2	G1E-HA2	I173E-HA2
1	2.00 (0.20)	0.87 (0.30)	1.00 (0.34)
5	1.50 (0.14)	0.87 (0.64)	0.93 (0.23)
30	2.40 (0.53)	1.40 (0.20)	1.27 (0.11)
60	2.47 (0.50)	1.13 (0.30)	1.53 (0.11)
120	2.60 (0.35)	1.07 (0.23)	1.80 (0.20)

HDX time/min]	Deuterium Incorporation (%) (198-202/199-203)		
	WT-HA2	G1E-HA2	I173E-HA2
1	2.00 (0.17)	0.25 (0.00)	3.58 (2.98)
5	0.63 (0.17)	2.00 (3.03)	0.33 (0.14)
30	1.00 (1.52)	4.17 (3.39)	0.25 (0.00)
60	2.17 (1.44)	0.25 (0.00)	0.25 (0.00)
120	0.42 (0.14)	4.00 (3.30)	3.75 (3.03)

HDX time/min	Deuterium Incorporation (%) (200-202)		
	WT-HA2	G1E-HA2	I173E-HA2
1	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)
5	0.17 (0.28)	0.00 (0.00)	0.50 (0.00)
30	0.17 (0.28)	0.00 (0.00)	0.00 (0.00)
60	0.17 (0.28)	0.00 (0.00)	0.00 (0.00)
120	0.17 (0.28)	0.00 (0.00)	0.17 (0.29)

HDX time/min	Deuterium Incorporation (%) (206-217)		
	WT-HA2	G1E-HA2	I173E-HA2
1	50.46 (2.76)	52.36 (1.52)	45.80 (3.74)
5	54.12 (2.09)	53.42 (0.54)	51.99 (0.31)
30	59.33 (3.22)	57.92 (0.54)	57.04 (1.86)
60	61.50 (1.80)	58.40 (0.37)	61.03 (1.47)
120	61.64 (1.10)	60.72 (1.44)	62.31 (2.27)

**Table 4C.** Percent deuterium incorporation vs. HDX time for the 39-52 peptide of WT-, I173E-, and G1E- HA2. WT and I173E exhibit approximately unimodal distributions in the  $m/z = 496-501$  range, and G1E exhibits a bimodal distribution for  $t_{\text{HDX}} = 120$  min, with the modes approximately separable into  $m/z = 496-498$  and  $498-501$  ranges. Each value is the average of triplicate measurements, and the RMSD is given in parentheses.  $\langle D_{\%} \rangle$  is the average of the five  $D_{\%}$ , and the associated uncertainty in parentheses is calculated using the  $D_{\%}$  RMSD's.

HDX time/min	Protein ( $m/z$ range)			
	WT-HA2 (496-501)	I173E-HA2 (496-501)	G1E-HA2 (496-498)	G1E-HA2 (498-501)
1	0.90 (0.22)	35.27 (3.86)	6.89 (0.11)	66.21 (0.66)
5	1.56 (0.54)	39.36 (1.76)	7.15 (0.28)	65.18 (0.75)
30	4.10 (0.55)	51.58 (0.60)	7.23 (0.35)	65.44 (0.83)
60	4.85 (0.16)	53.85 (0.62)	15.37 (0.21)	65.52 (1.24)
120	5.95 (2.05)	54.95 (0.42)	16.23 (0.96)	62.83 (1.32)
$\langle D_{\%} \rangle$	3.26 (0.23)	47.00 (0.87)	10.58 (0.22)	65.04 (0.45)

## REFERENCES

## REFERENCES

1. White, J. M., Delos, S. E., Brecher, M., and Schornberg, K. (2008) Structures and mechanisms of viral membrane fusion proteins: Multiple variations on a common theme, *Crit. Rev. Biochem. Mol. Biol.* **43**, 189-219.
2. Harrison, S. C. (2015) Viral membrane fusion, *Virology* **479**, 498-507.
3. Boonstra, S., Blijleven, J. S., Roos, W. H., Onck, P. R., van der Giessen, E., and van Oijen, A. M. (2018) Hemagglutinin-mediated membrane fusion: A biophysical perspective, *Ann. Revs. Biophys.* **47**, 153-173.
4. Nobusawa, E., Aoyama, T., Kato, H., Suzuki, Y., Tateno, Y., and Nakajima, K. (1991) Comparison of complete amino acid sequences and receptor binding properties among 13 serotypes of hemagglutinins of influenza A viruses, *Virology* **182**, 475-485.
5. Wilson, I. A., Skehel, J. J., and Wiley, D. C. (1981) Structure of the haemagglutinin membrane glycoprotein of influenza virus at 3 Å resolution, *Nature* **289**, 366-373.
6. Ekiert, D. C., Kashyap, A. K., Steel, J., Rubrum, A., Bhabha, G., Khayat, R., Lee, J. H., Dillon, M. A., O'Neil, R. E., Faynboym, A. M., Horowitz, M., Horowitz, L., Ward, A. B., Palese, P., Webby, R., Lerner, R. A., Bhatt, R. R., and Wilson, I. A. (2012) Cross-neutralization of influenza A viruses mediated by a single antibody loop, *Nature* **489**, 526-532.
7. Lakadamyali, M., Rust, M. J., Babcock, H. P., and Zhuang, X. W. (2003) Visualizing infection of individual influenza viruses, *Proc. Natl. Acad. Sci. U. S. A.* **100**, 9280-9285.
8. Chen, J., Skehel, J. J., and Wiley, D. C. (1999) N- and C-terminal residues combine in the fusion-pH influenza hemagglutinin HA<sub>2</sub> subunit to form an N cap that terminates the triple-stranded coiled coil, *Proc. Natl. Acad. Sci. U.S.A.* **96**, 8967-8972.
9. Blijleven, J. S., Boonstra, S., Onck, P. R., van der Giessen, E., and van Oijen, A. M. (2016) Mechanisms of influenza viral membrane fusion, *Sem. Cell Dev. Biol.* **60**, 78-88.
10. Chernomordik, L. V., Frolov, V. A., Leikina, E., Bronk, P., and Zimmerberg, J. (1998) The pathway of membrane fusion catalyzed by influenza hemagglutinin: Restriction of lipids, hemifusion, and lipidic fusion pore formation, *J. Cell Biol.* **140**, 1369-1382.
11. Qiao, H., Armstrong, R. T., Melikyan, G. B., Cohen, F. S., and White, J. M. (1999) A specific point mutant at position 1 of the influenza hemagglutinin fusion peptide displays a hemifusion phenotype, *Mol. Biol. Cell* **10**, 2759-2769.

12. Armstrong, R. T., Kushnir, A. S., and White, J. M. (2000) The transmembrane domain of influenza hemagglutinin exhibits a stringent length requirement to support the hemifusion to fusion transition, *J. Cell Biol.* 151, 425-437.
13. Melikyan, G. B., Lin, S. S., Roth, M. G., and Cohen, F. S. (1999) Amino acid sequence requirements of the transmembrane and cytoplasmic domains of influenza virus hemagglutinin for viable membrane fusion, *Mol. Biol. Cell* 10, 1821-1836.
14. Leikina, E., LeDuc, D. L., Macosko, J. C., Epand, R., Shin, Y. K., and Chernomordik, L. V. (2001) The 1-127 HA2 construct of influenza virus hemagglutinin induces cell-cell hemifusion, *Biochemistry* 40, 8378-8386.
15. Kim, C. S., Epand, R. F., Leikina, E., Epand, R. M., and Chernomordik, L. V. (2011) The final conformation of the complete ectodomain of the HA2 subunit of Influenza Hemagglutinin can by itself drive low pH-dependent fusion, *J. Biol. Chem.* 286, 13226-13234.
16. Cho, K. J., Lee, J.-H., Hong, K. W., Kim, S.-H., Park, Y., Lee, J. Y., Kang, S., Kim, S., Yang, J. H., Kim, E.-K., Seok, J. H., Unzai, S., Park, S. Y., Saelens, X., Kim, C.-J., Lee, J.-Y., Kang, C., Oh, H.-B., Chung, M. S., and Kim, K. H. (2013) Insight into structural diversity of influenza virus haemagglutinin, *J. Gen. Virol.* 94, 1712-1722.
17. Xiong, X., Corti, D., Liu, J., Pinna, D., Foglierini, M., Calder, L. J., Martin, S. R., Lin, Y. P., Walker, P. A., Collins, P. J., Monne, I., Suguitan, A. L., Jr., Santos, C., Temperton, N. J., Subbarao, K., Lanzavecchia, A., Gamblin, S. J., and Skehel, J. J. (2015) Structures of complexes formed by H5 influenza hemagglutinin with a potent broadly neutralizing human monoclonal antibody, *Proc. Natl. Acad. Sci. U.S.A.* 112, 9430-9435.
18. Ratnayake, P. U., Ekanayaka, E. A. P., Komanduru, S. S., and Weliky, D. P. (2016) Full-length trimeric influenza virus hemagglutinin II membrane fusion protein and shorter constructs lacking the fusion peptide or transmembrane domain: Hyperthermostability of the full-length protein and the soluble ectodomain and fusion peptide make significant contributions to fusion of membrane vesicles, *Protein Expression Purif.* 117, 6-16.
19. Ranaweera, A., Ratnayake, P. U., and Weliky, D. P. (2018) The stabilities of the soluble ectodomain and fusion peptide hairpins of the Influenza virus hemagglutinin subunit II protein are positively correlated with membrane fusion, *Biochemistry* 57, 5480-5493.
20. Chlanda, P., Mekhedov, E., Waters, H., Schwartz, C. L., Fischer, E. R., Ryham, R. J., Cohen, F. S., Blank, P. S., and Zimmerberg, J. (2016) The hemifusion structure induced by influenza virus haemagglutinin is determined by physical properties of the target membranes, *Nature Microbiology* 1, Art. No. 16050.
21. Gui, L., Ebner, J. L., Mileant, A., Williams, J. A., and Lee, K. K. (2016) Visualization and sequencing of membrane remodeling leading to influenza virus fusion, *J. Virol.* 90, 6948-6962.

22. Das, D. K., Govindan, R., Nikic-Spiegel, I., Krammer, F., Lemke, E. A., and Munro, J. B. (2018) Direct visualization of the conformational dynamics of single influenza hemagglutinin trimers, *Cell* 174, 926-937.
23. Eddy, N. R., and Onuchic, J. N. (2018) Rotation-activated and cooperative zipping characterize class I viral fusion protein dynamics, *Biophys. J.* 114, 1878-1888.
24. Carr, C. M., and Kim, P. S. (1993) A spring-loaded mechanism for the conformational change of influenza hemagglutinin, *Cell* 73, 823-832.
25. Durrer, P., Galli, C., Hoenke, S., Corti, C., Gluck, R., Vorherr, T., and Brunner, J. (1996) H<sup>+</sup>-induced membrane insertion of influenza virus hemagglutinin involves the HA2 amino-terminal fusion peptide but not the coiled coil region, *J. Biol. Chem.* 271, 13417-13421.
26. Lorieau, J. L., Louis, J. M., and Bax, A. (2010) The complete influenza hemagglutinin fusion domain adopts a tight helical hairpin arrangement at the lipid:water interface, *Proc. Natl. Acad. Sci. U.S.A.* 107, 11341-11346.
27. Ghosh, U., Xie, L., and Weliky, D. P. (2013) Detection of closed influenza virus hemagglutinin fusion peptide structures in membranes by backbone <sup>13</sup>CO-<sup>15</sup>N rotational-echo double-resonance solid-state NMR, *J. Biomol. NMR* 55, 139-146.
28. Du, T. P., Jiang, L., and Liu, M. L. (2014) NMR structures of fusion peptide from influenza hemagglutinin H3 subtype and its mutants, *J. Peptide Sci.* 20, 292-297.
29. Ghosh, U., Xie, L., Jia, L. H., Liang, S., and Weliky, D. P. (2015) Closed and semiclosed interhelical structures in membrane vs closed and open structures in detergent for the Influenza Virus hemagglutinin fusion peptide and correlation of hydrophobic surface area with fusion catalysis, *J. Am. Chem. Soc.* 137, 7548-7551.
30. Macosko, J. C., Kim, C. H., and Shin, Y. K. (1997) The membrane topology of the fusion peptide region of influenza hemagglutinin determined by spin-labeling EPR, *J. Mol. Biol.* 267, 1139-1148.
31. Marcsisin, S. R., and Engen, J. R. (2010) Hydrogen exchange mass spectrometry: what is it and what can it tell us?, *Anal. Bioanal. Chem.* 397, 967-972.
32. Pirrone, G. F., Iacob, R. E., and Engen, J. R. (2015) Applications of hydrogen/deuterium exchange MS from 2012 to 2014, *Anal. Chem.* 87, 99-118.
33. Koshy, S. S., Eyles, S. J., Weis, R. M., and Thompson, L. K. (2013) Hydrogen exchange mass spectrometry of functional membrane-bound chemotaxis receptor complexes, *Biochemistry* 52, 8833-8842.
34. Koshy, S. S., Li, X., Eyles, S. J., Weis, R. M., and Thompson, L. K. (2014) Hydrogen exchange differences between chemoreceptor signaling complexes localize to functionally important subdomains, *Biochemistry* 53, 7755-7764.

35. Chodavarapu, S., Jones, A. D., Feig, M., and Kaguni, J. M. (2016) DnaC traps DnaB as an open ring and remodels the domain that binds primase, *Nucl. Acids Res.* *44*, 210-220.
36. Garcia, N. K., Guttman, M., Ebner, J. L., and Lee, K. K. (2015) Dynamic changes during acid-Induced activation of influenza hemagglutinin, *Structure* *23*, 665-676.
37. Tamm, L. K. (2003) Hypothesis: spring-loaded boomerang mechanism of influenza hemagglutinin-mediated membrane fusion, *Biochim. Biophys. Acta* *1614*, 14-23.
38. Benton, D. J., Nans, A., Calder, L. J., Turner, J., Neu, U., Lin, Y. P., Ketelaars, E., Kallewaard, N. L., Corti, D., Lanzavecchia, A., Gamblin, S. J., Rosenthal, P. B., and Skehel, J. J. (2018) Influenza hemagglutinin membrane anchor, *Proc. Natl. Acad. Sci. U.S.A.* *115*, 10112-10117.
39. Chang, D. K., Cheng, S. F., Kantchev, E. A. B., Lin, C. H., and Liu, Y. T. (2008) Membrane interaction and structure of the transmembrane domain of influenza hemagglutinin and its fusion peptide complex, *BMC Biology* *6*:2.
40. Borrego-Diaz, E., Peeples, M. E., Markosyan, R. M., Melikyan, G. B., and Cohen, F. S. (2003) Completion of trimeric hairpin formation of influenza virus hemagglutinin promotes fusion pore opening and enlargement, *Virology* *316*, 234-244.
41. Park, H. E., Gruenke, J. A., and White, J. M. (2003) Leash in the groove mechanism of membrane fusion, *Nature Struct. Biol.* *10*, 1048-1053.
42. Liang, S., Ratnayake, P. U., Keinath, C., Jia, L., Wolfe, R., Ranaweera, A., and Weliky, D. P. (2018) Efficient fusion at neutral pH by Human Immunodeficiency Virus gp41 trimers containing the fusion peptide and transmembrane domains, *Biochemistry* *57*, 1219-1235.
43. Weis, D. D., Engen, J. R., and Kass, I. J. (2006) Semi-automated data processing of hydrogen exchange mass spectra using HX-Express, *J. Am. Soc. Mass Spectr.* *17*, 1700-1703.
44. Lorieau, J. L., Louis, J. M., and Bax, A. (2013) Impact of Influenza hemagglutinin fusion peptide length and viral subtype on its structure and dynamics, *Biopolymers* *99*, 189-195.
45. Tatulian, S. A., and Tamm, L. K. (2000) Secondary structure, orientation, oligomerization, and lipid interactions of the transmembrane domain of influenza hemagglutinin, *Biochemistry* *39*, 496-507.
46. Kuzmin, P. I., Zimmerberg, J., Chizmadzhev, Y. A., and Cohen, F. S. (2001) A quantitative model for membrane fusion based on low-energy intermediates, *Proc. Natl. Acad. Sci. U.S.A.* *98*, 7235-7240.
47. Epand, R. M., Epand, R. F., Martin, I., and Ruyschaert, J. M. (2001) Membrane interactions of mutated forms of the influenza fusion peptide, *Biochemistry* *40*, 8800-8807.

48. Gabrys, C. M., Yang, R., Wasniewski, C. M., Yang, J., Canlas, C. G., Qiang, W., Sun, Y., and Weliky, D. P. (2010) Nuclear magnetic resonance evidence for retention of a lamellar membrane phase with curvature in the presence of large quantities of the HIV fusion peptide, *Biochim. Biophys. Acta* 1798, 194-201.
49. Tristram-Nagle, S., Chan, R., Kooijman, E., Uppamoochikkal, P., Qiang, W., Weliky, D. P., and Nagle, J. F. (2010) HIV fusion peptide penetrates, disorders, and softens T-cell membrane mimics, *J. Mol. Biol.* 402, 139-153.
50. Yao, H. W., and Hong, M. (2013) Membrane-dependent conformation, dynamics, and lipid Interactions of the fusion peptide of the paramyxovirus PIV5 from solid-state NMR, *J. Mol. Biol.* 425, 563-576.
51. Lai, A. L., and Freed, J. H. (2015) The interaction between Influenza HA fusion peptide and transmembrane domain affects membrane structure, *Biophys. J.* 109, 2523-2536.
52. Lakomek, N. A., Kaufman, J. D., Stahl, S. J., Louis, J. M., Grishaev, A., Wingfield, P. T., and Bax, A. (2013) Internal dynamics of the homotrimeric HIV-1 viral coat protein gp41 on multiple time scales, *Angew. Chem. Int. Ed.* 52, 3911-3915.
53. Caffrey, M., Cai, M., Kaufman, J., Stahl, S. J., Wingfield, P. T., Covell, D. G., Gronenborn, A. M., and Clore, G. M. (1998) Three-dimensional solution structure of the 44 kDa ectodomain of SIV gp41, *EMBO J.* 17, 4572-4584.
54. Yang, Z. N., Mueser, T. C., Kaufman, J., Stahl, S. J., Wingfield, P. T., and Hyde, C. C. (1999) The crystal structure of the SIV gp41 ectodomain at 1.47 Å resolution, *J. Struct. Biol.* 126, 131-144.
55. Sackett, K., Nethercott, M. J., Epand, R. F., Epand, R. M., Kindra, D. R., Shai, Y., and Weliky, D. P. (2010) Comparative analysis of membrane-associated fusion peptide secondary structure and lipid mixing function of HIV gp41 constructs that model the early pre-hairpin intermediate and final hairpin conformations, *J. Mol. Biol.* 397, 301-315.
56. Panahi, A., and Feig, M. (2010) Conformational sampling of Influenza fusion peptide in membrane bilayers as a function of termini and protonation states, *J. Phys. Chem. B* 114, 1407-1416.
57. Baylon, J. L., and Tajkhorshid, E. (2015) Capturing spontaneous membrane insertion of the Influenza virus Hemagglutinin fusion peptide, *J. Phys. Chem. B* 119, 7882-7893.
58. Jia, L. H., Liang, S., Sackett, K., Xie, L., Ghosh, U., and Weliky, D. P. (2015) REDOR solid-state NMR as a probe of the membrane locations of membrane-associated peptides and proteins, *J. Magn. Reson.* 253, 154-165.
59. Victor, B. L., Lousa, D., Antunes, J. M., and Soares, C. M. (2015) Self-assembly molecular dynamics simulations shed light into the interaction of the influenza fusion peptide with a membrane bilayer, *J. Chem. Inform. Model.* 55, 795-805.



60. Worch, R., Dudek, A., Krupa, J., Szymaniec, A., and Setny, P. (2018) Charged N-terminus of influenza fusion peptide facilitates membrane fusion, *Int. J. Mol. Sci.* 19.
61. Lorieau, J. L., Louis, J. M., and Bax, A. (2011) Helical hairpin structure of Influenza Hemagglutinin fusion peptide stabilized by charge-dipole interactions between the N-terminal amino group and the second helix, *J. Am. Chem. Soc.* 133, 2824-2827.
62. Markosyan, R. M., Cohen, F. S., and Melikyan, G. B. (2000) The lipid-anchored ectodomain of influenza virus hemagglutinin (GPI-HA) is capable of inducing nonenlarging fusion pores, *Mol. Biol. Cell* 11, 1143-1152.

## **Chapter 5**

### **HA2 structure by protein crystallography**

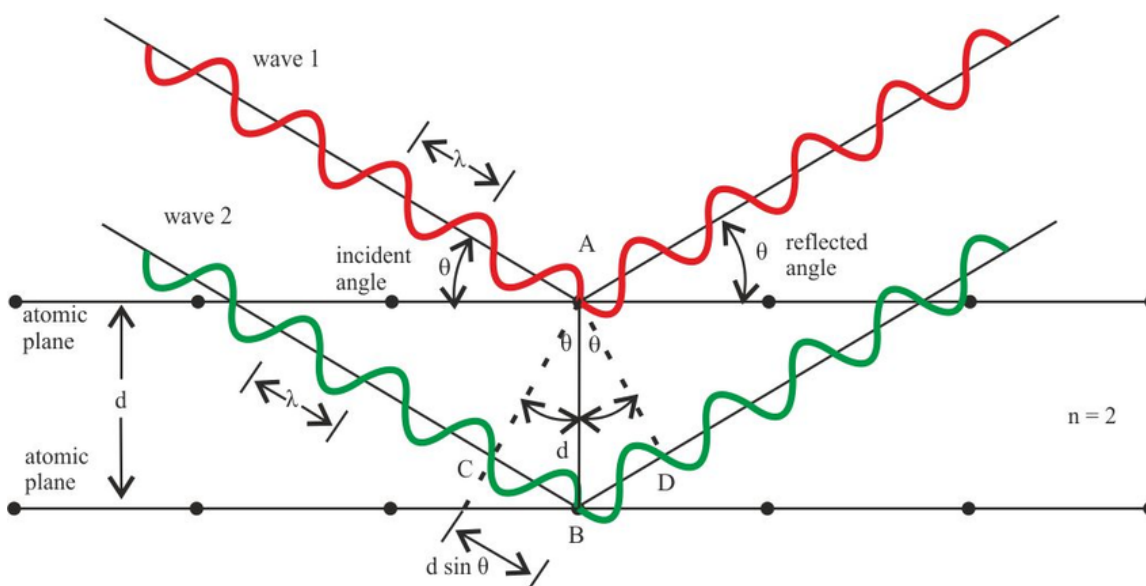
## 5.1 Introduction

High resolution structures of proteins are highly demanded in the field of therapeutic drug designing. Several techniques are available for this purpose including X-ray crystallography, NMR spectroscopy, and electron microscopy (EM). Each of these techniques has advantages and limitations. X-ray crystallography provides three-dimensional structures with atomic level resolution for wide range of proteins that include small peptides to large complexes with other proteins, nucleic acids etc.<sup>1</sup> However, to use this method in structure determination, the protein needs to be in the crystalline form, and only the proteins that can be crystallized are examinable. The NMR spectroscopy can be applied to proteins in the solution form, and can be used to obtain solution dynamics of the protein. However, the size of the molecule is a limiting factor with this method (<50 kDa).<sup>2</sup> EM requires smaller amount of sample compared to the NMR and crystallography. Until recently, the resolution obtained from EM was limited (compared to the other two methods), however with the development of Cryo-EM the resolution has now improved to ~4.0 Å.<sup>3</sup>

In X-ray crystallography, the diffraction pattern from a molecule in a crystalline form is used for the structure determination. In a crystal, the molecules, atoms or ions are arranged three dimensionally in a regular manner with a common repeating unit. When an X-ray beam encounters the crystal, the beam is diffracted into many specific directions by the electron cloud of the atoms in the crystal. The X-ray diffraction by the crystal can result in both constructive and destructive interferences, which give rise to a diffraction pattern. The constructive interference occurs when the X-ray diffraction follows the Bragg's law ( $2d \sin \theta = n\lambda$ , where,  $d$  = distance between diffracting planes,  $\theta$  = incident angle,  $\lambda$  = wavelength of X-ray beam and  $n$  = an integer, Fig. 5.1), whereas other diffractions cause destructive interference. The diffraction pattern provides details

on the arrangement of atoms in a crystal. The intensities of the diffracted peaks are determined by the electron density of individual molecules within the crystal, and the positions of the diffraction peaks are determined by the distance between parallel planes of the repeating lattice of molecules.<sup>4</sup>

Protein X-ray crystallography is widely used to obtain high resolution structure of proteins in the crystallized form. Diffraction by a single molecule is too weak to be measured, and therefore acquiring a protein crystal is the first step in protein crystallography.<sup>5</sup>



**Figure 5.1** Schematic representation of the Bragg's law

### 5.1.1 Membrane protein crystallography

The knowledge of protein structural details with atomic resolution are highly important. They provide details on the proteins' mode of action as well as the physiological function at molecular level, which in turn can be useful for the target drug designing.  $\sim 1/3$  of the genome is consisted of membrane proteins, however, there are very few unique atomically resolved membrane protein

structures compared to those of soluble proteins (e.g. 136709 of soluble protein structures vs. 3456 of transmembrane protein structures by 2017 - PDBTM data<sup>6</sup>). This implies that the obtaining membrane protein structures is a significant challenge. The primary difficulty is the obtaining reasonable amount of protein for X-ray trials because of the low abundance of these proteins in biological membranes. Secondly, because of the amphiphilic character of these molecules they are not soluble in aqueous medium, and detergents are required for solubilization. Selecting a suitable detergent is also challenging, as different membrane proteins require different detergents for proper folding and oligomerization.<sup>7</sup>

### 5.1.2 Protein crystallization methods

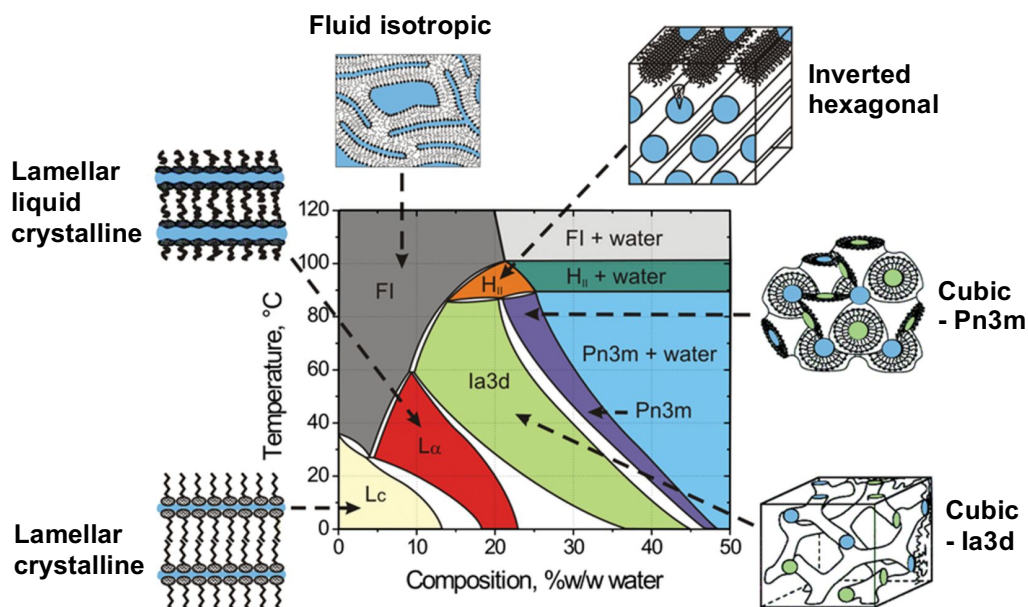
Membrane proteins can be crystallized by various methods, which can be divided into two major categories: the *in surfo* methods and bilayer methods.

***In surfo* methods** - Detergents are extensively used in the membrane protein isolation and solubilization. The *in surfo* methods directly use the detergent-solubilized proteins in the crystallization experiments. This is the first method used for membrane protein crystallization. The vapor diffusion, counter diffusion and micro batch method are among the most common *in surfo* methods. The direct use of detergent-solubilized protein is a major advantage of this method, because of easy handling. Protein crystallogenesis is typically a slow process and it is important to keep the protein without aggregating for considerable amount of time, ranging from days to months. Therefore, the detergent should be carefully chosen to keep the protein solubilized during the incubation period., There are wide range of proteins crystallized from this method, and some

examples include: cytochrome *c* oxidase, cytochrome *bc*<sub>1</sub> complex, ion channels, transporters and G-protein coupled receptors.<sup>8</sup>

**Bilayer methods** - In bilayer methods, the protein is initially confined in a lipid bilayer which act as a hosting medium, from which the protein molecules are introduced into the crystal. These bilayers well resemble the cellular membranes and therefore the protein is in a more familiar environment during the crystallization experiment, which is a potential advantage over the *in surf*o methods. There are various methods under this category including the bicelle method, vesicle method, and the most common *in meso* (lipid cubic phase - LCP) method. The LCP method has been used for structure determination of considerable number of membrane proteins. The LCP method uses a lipidic mesophase (a liquid crystalline phase) as the hosting medium or a protein reservoir from which the crystallogensis initiates.<sup>9</sup>

The mesophase is a highly viscous state of matter intermediate between liquid and solid. The lipid and protein are the main constituents in the bicontinuous lipidic mesophase used in the LCP method. Other than the lipid and protein, it may contain buffers, detergents, salts, polymers, and hydrogen ions.<sup>9</sup> The outcome of the crystallization experiment highly depends on these constituents, because they can affect the lipid phase behavior. The mesophase can have different phases depending on the lipid:water ratio, most commonly cubic, lamellar and hexagonal phases. Monoolein is a common lipid used in the LCP method and Fig. 5.2 shows its temperature-composition phase diagram. In the LCP method, the cubic phase (Pn3m) and the lamellar phase play major roles (discussed in section 5.1.3).

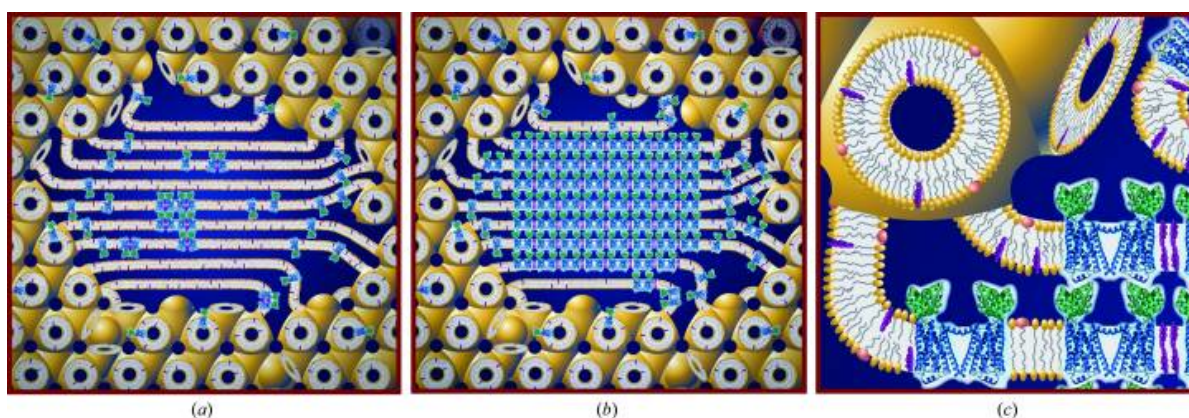


**Figure 5.2** Temperature-composition phase diagram of monoolein.<sup>10</sup>

### 5.1.3 The proposed mechanism of crystallization by LCP method

A detailed mechanism has been proposed to explain the LCP crystallogenesis occurs at a molecular level (Fig. 5.3).<sup>11</sup> First, the membrane protein of interest is solubilized using a detergent, and it is purified using common purification methods such as affinity chromatography and gel filtration chromatography. The purified protein is then combined with a hosting lipid in a proper ratio to get a cubic mesophase. When combine with the lipid (typically a monoacylglycerol), the protein may reorganize into the cubic mesophase bilayer. A precipitant solution is then added to the mesophase. The precipitant solution exists as a separate liquid, and its components diffuse into the mesophase without dissolving the cubic mesophase. This results in phase separation. There are several possible phases depending on the lipid used. If one of the phases is a lamellar phase, the protein molecules can diffuse into the lamellar phase favorably as its environment is closer to a

cell membrane. The high concentration of protein in the lamellar phase can lead to nucleation and crystal growth if bathing solution has proper composition that facilitates protein-protein interactions. Under crystallography condition, the medium contains different components including: the protein, detergent, buffer, salt, hydrogen ions, polymers and small molecular additives, which can influence the phase behavior of the lipid.



**Figure 5.3** Cartoon representation of the proposed mechanism of crystallization of a membrane protein from lipidic cubic mesophase.<sup>8</sup>

There are several experimental evidences that support this mechanism. Some experiments provide visual observations for protein movements in the bicontinuous mesophase using colored proteins such as bacteriorhodopsin. In addition, small angle X-ray scattering (SAXS) experiments have proved the existence of the lamellar phase.<sup>11</sup>



## **5.2 HA2 crystallization - part I (A collaboration with Dr. Ferguson Miller in the Department of Biochemistry and Molecular Biology at Michigan State University)**

In the present work we applied the LCP method to obtain a high-resolution structure for full-length HA2. This work is a collaboration with Dr. Ferguson Miller, and the crystallization set up was performed with Dr. Jian Liu, who is a postdoc in the Ferguson-Miller group. The preliminary tests were performed by Dr. Punsisi U. Ratnayake in the Weliky group.

### **5.2.1 Materials**

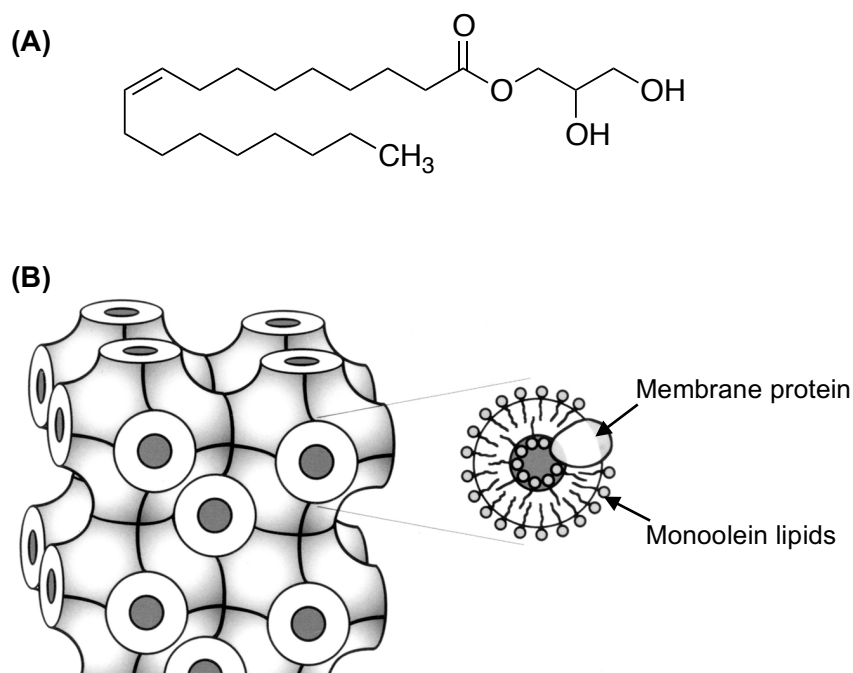
Commercial sources of materials include: Jeffamine M-600, LCP Sandwich set, Gas tight micro syringes and precipitant solution kits, Hampton Research, Aliso Viejo, CA; Monoolein, Sigma-Aldrich, St. Louis, MO; Diffraction™ Starter Kit plastic sandwich plates and precipitant solution kits, Molecular Dimensions, Maumee, OH.

### **5.2.2 Mixing the protein with lipid**

The most commonly use hosting lipid in the LCP method is monoolein which is a monoacylglycerol with 9-cis-octadecenoic acid at the sn-1 position of glycerol (Fig. 5.4). Monoolein can self-assemble into different liquid crystalline structures. Liquid crystal is a type of matter having properties of both liquid state and crystalline solid state. They can flow like a liquid but they have molecules oriented as a crystal. Monoolein can have typically cubic, hexagonal and lamellar liquid crystals depending on the monoolein:water ratio and the temperature of the medium (Fig. 5.2).

In the LCP method, conditions are controlled to obtain a cubic structure with monoolein. The cubic structure creates a similar environment of a cell membrane with a lipid bilayer, and thereby

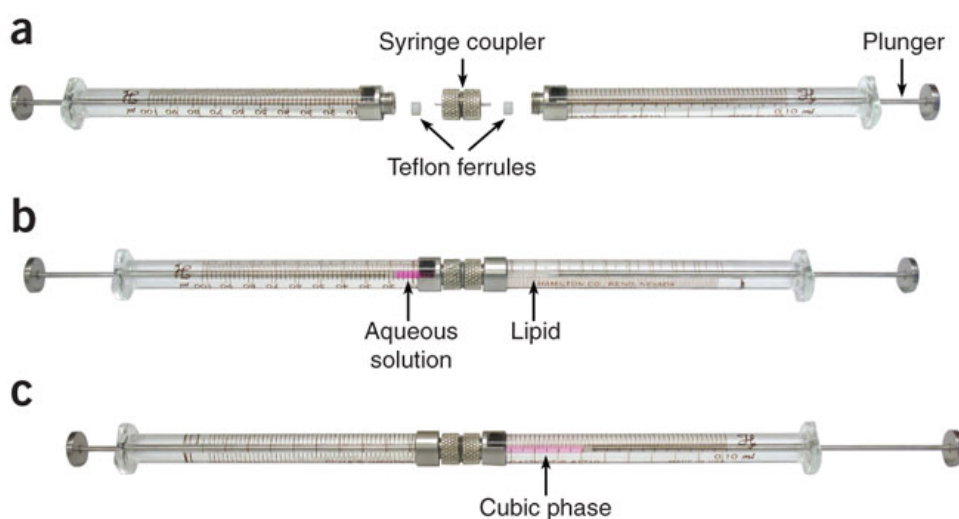
making a familiar place to membrane proteins. Also, the membrane proteins can easily diffuse in the lipid bilayer in a three dimensionally continuous cubic phase, thus the proteins can be concentrated to form crystals.



**Figure 5.4** (A) The chemical structure of monoolein; and (B) A schematic representation of bicontinuous cubic phase of monoolein.

For setting up the crystallization trials, HA2 was purified from ~70 g of wet cells. The SEC peak eluting at ~12.5 ml retention volume (corresponds to the protein trimer - discussed in Chapter 3) was collected, and re-concentrated to 10-20 mg/mL depending on the trial. The SEC followed by re-concentration gave ~70  $\mu$ L of 20 mg/ml trimeric protein from 70 g of wet cells. Then crystallization plates were set up in Dr. Ferguson Miller's lab.

To combine the lipid with the concentrated protein, gas tight micro syringes were used. For the Pn3m cubic phase of monoolein at 20 °C, the full hydration with water occurs at ~40% (w/w) water (Fig. 5.2), which means the required lipid:water weight ratio is 3:2. In the LCP method the protein solution is the water fraction. As the density of monoolein (0.942 g/ml) is approximately equal to the density of water, a 3:2 lipid:water weight ratio  $\approx$  3:2 lipid:water volume ratio. If 30 mg ( $\sim$ 30  $\mu$ l) of the lipid is used for one crystallization set up,  $\sim$ 20  $\mu$ l of protein solution is required to obtain the LCP phase.



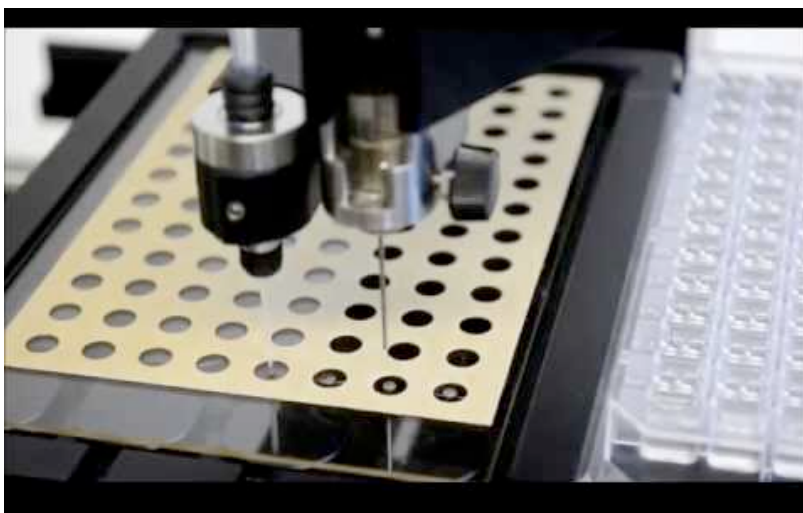
**Figure 5.5** The lipid mixing device: (a) disassembled coupled syringe mixture; and (b,c) assembled coupled syringe mixture.

Once the lipid and the protein in required volumes were placed in the syringes, the two syringes were connected through a syringe coupler (Fig. 5.5), and the protein solution was slowly inserted into the lipid containing syringe. Then the two solutions were mixed well by moving the syringe plungers back and forth. The process was repeated many times (about 10 min) until an optically clear, homogeneous mesophase is produced. At this stage the protein is reconstituted into

the lipid bilayer, and the cubic mesophase has been formed. The protein-loaded mesophase can then be dispensed into individual wells of the crystallization plates.<sup>9</sup>

### 5.2.3 Mixing the mesophase with the precipitant solution

The precipitant solution used in the LCP method typically includes a detergent, buffer, salt, polymers and small-molecular additives. A robot was used to mix the precipitant solution with the mesophase which can measure nano liter level quantities. In-meso robot has two arms programmed to move simultaneously: one dispenses the mesophase (30-50 nl) and the other dispenses the precipitant solution (600-800 nl).



**Figure 5.6** Adding the protein-lipid mesophase to the crystallization plates using the robot tip.

The mixing of mesophase with the precipitant solution was done in a 96-well crystallization plate (Fig. 5.6). First, the wells were loaded with the mesophase and then the precipitant solution was placed on top of each mesophase bolus. A cover slip was placed on the filled wells to cover

them uniformly. The air gaps were eliminated by repeatedly working a spatula or a rod over the surface of the cover slip with moderate applied pressure.

#### 5.2.4 Incubation and tracking crystallogenesis

The plates were incubated in a temperature controlled chamber, at 20 °C. The wells in the crystallization plates were observed using a polarized light microscope (PLM) after 7, 14, 21, 30, and 60 days from the setup date. The presence of the crystals was verified by viewing the bolus in the well between crossed polarizers. The non- birefringent cubic phase appears dark similar to the surrounding precipitant solution, whereas nonisotropic crystals usually “light up” and appears as sharp spots of light in the dark background.

#### 5.2.5 Results and Discussion

##### 5.2.5.1 Initial screening

From the first crystallography trial with 20 mg/mL protein “crystal like” particles were observed in several precipitant solution conditions (Table 5.1).

**Table 5.1** The precipitant solution conditions that gave “crystal-like” particles in trial 1.

	Salt	Buffer	pH	Precipitant 1	Precipitant 2
MemGold					
A3	-	0.015 M Tricine	8.5	24% w/v PEG 4000	-
D1	0.2 M CaCl <sub>2</sub> ·2H <sub>2</sub> O	0.1 M HEPES	7.5	53 % v/v PEG 400	-
D4	0.2 M CaCl <sub>2</sub> ·2H <sub>2</sub> O	0.1 M Tris-HCl	8.0	44% PEG 400	-
F6	0.02 M Sodium citrate tribasic dihydrate	0.08 M Na <sub>3</sub> PO <sub>4</sub>	6.2	18% w/v PEG 2000	-

**Table 5.1 (cont'd)**

G8	0.1 M NaCl	0.1 M MES	6.5	33% v/v PEG 400	4% v/v ethylene glycol
MemGold 2					
G5	-	0.1 M MES	6.0	11% w/v PEG 20,000	-
H11	3 M (NH <sub>4</sub> ) <sub>2</sub> SO <sub>4</sub>	0.1 M MES	5.5	-	-
Index					
B7	-	0.056 M NaH <sub>2</sub> PO <sub>4</sub> .H <sub>2</sub> O	8.2	1.344 M K <sub>2</sub> HPO <sub>4</sub>	-
G10	0.2 M MgCl <sub>2</sub> .6H <sub>2</sub> O	0.1 M Bis-Tris	5.5	25% w/v PEG 3350	-
H7	-	-	7.0	20% w/v PEG 200	0.15 M DL-Malic acid
JB screen membrane					
A11	-	-		30 % w/v PEG 1500	-
B9	300 mM MgCl <sub>2</sub>	100 mM BICINE	9.0	25% w/v PEG 2000	15% w/v glycerol
C6	100 mM NaCl	50 mM MES	6.5	5% w/v PEG 4000	10% w/v glycerol
C10	50 mM Na <sub>2</sub> HPO <sub>4</sub>	-	6.8	12% w/v PEG 4000	-
C11	500 mM KCl	50 mM MOPS	7.0	12% w/v PEG 4000	20% w/v glycerol
D4	-	50 mM Tricine	8.0	22% w/v PEG 4000	-
D5	500 mM NaCl	50 mM Tris	8.5	22% w/v PEG 4000	-
Hampton					
C5	-	0.1 M HEPES Sodium	7.5	0.8 M Potassium Sodium tartarate.4H <sub>2</sub> O	-
D8	-	-		0.2 M Magnesium formate.2H <sub>2</sub> O	-
Cryo I/II					
B10	-	0.1 M Na/K phosphate	6.2	40% v/v ethylene glycol	-
C10	0.2 M MgCl <sub>2</sub>	0.1 M Imidazole	8.0	40% v/v 2-methyl-2,4-pentanediol	-

**Table 5.1 (cont'd)**

D3	-	0.1 M Phosphate-citrate	4.2	40% v/v PEG 600	-
MIDAS					
A2	-	0.1 M MES	5.5	12% w/v polyvinylpyrrolidone	-
B10	-	-	-	30% v/v Jeffamine® M-600	10% v/v Dimethyl sulfoxide
G12	-	0.1 M Bis-Tris	6.0	15% w/v SOKALAN® CP 5	-
H11	0.2 M Ammonium formate	-	-	20% w/v PEG 4000	10% w/v polyvinylpyrrolidone

To check the reproducibility of the crystals, the whole process was repeated twice using commercially available precipitant solution kits/screen kits (MemGold I/ MemGold II/ Index/ JB Screen/ Hampton/ MIDAS/ Cryo I,II/ Wizard I,II/ Morpheus). From these studies, three reproducible conditions were identified.

1. JB Screen A11
2. MIDAS B10
3. MIDAS G12

#### 5.2.5.2 Condition optimization

The size and the perfection of the crystals determine the quality of the X-ray structure analysis. Therefore, condition refinement is a general requirement in the crystallization trials. The improvement of the crystallization conditions is known as the optimization. This includes the optimization of physicochemical parameters like pH, ionic strength, precipitant concentration temperature and in some cases overall methodology.<sup>12</sup>

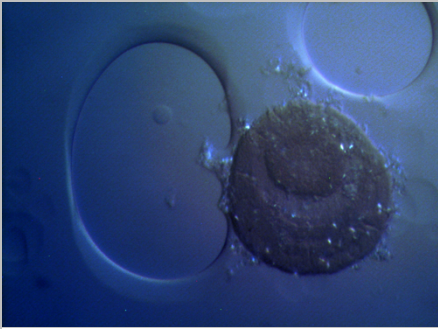
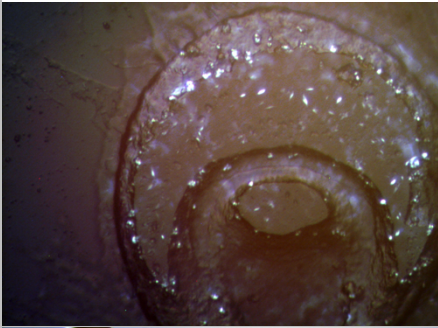

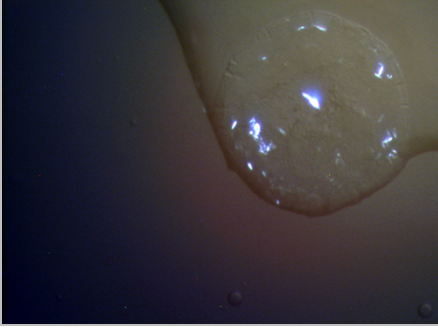
Two screen kits (96-well kits containing precipitant solutions) were prepared by varying the pH and the concentrations of the salt and the precipitants of the conditions that gave positive results. pH optimization is one of the most effective ways of inducing crystallization. In our project as we got only three reproducible conditions from the first 3 trials, new trials were set up over a range with fine pH increments of 0.2 pH units. For example, as pH of the condition MIDAS G12 is pH 6.0 the pH of the medium was adjusted from pH 5.4 - pH 6.4 with 0.2 pH increments. Similarly, the precipitant concentration was also adjusted with 2% (w/v) increments in a range centered upon the current best value. It was difficult to make a homogeneous solution by mixing these salt and precipitant solutions. When the kits were kept aside the screening solutions were separated into distinct layers. We did not observe any positive result from the crystallization trials with these hand-made screening kits.

As the next step of optimization, instead of making new precipitant solutions, the protein concentration was varied. Several trials were performed using 10 and 15 mg/mL protein concentration. Crystal-like particles were obtained from MIDAS B10 and G12 conditions in all the crystallization trials. In addition, MemGold C11 MIDAS E2 and E5 also gave crystal-like particles in at least two trials (Table 5.2).

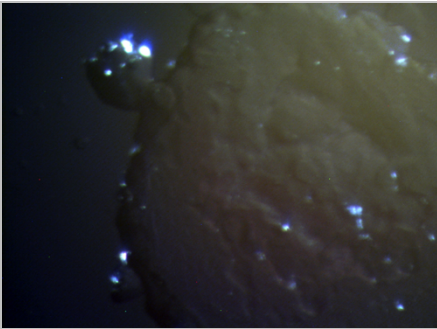
One major challenge with the LCP method is the highly viscous nature of the LCP itself, which result in difficult handling. This has highly limited the number of protein structures that could have been solved by this method.<sup>13</sup> We also encountered this problem in the present experiments. Any of the observed particles couldn't be picked and mounted onto a loop for X-ray diffraction because of the highly viscous nature of the medium.



**Table 5.2** Condition that gave reproducible crystal-like particles (the polarized microscope used for these images does not have a ruler/size information. The size of these crystals can be roughly estimated to ~5-10  $\mu\text{m}$ ).

MIDAS B10	30 % v/v Jeffamine® M-600	
MIDAS G12	0.1 M Bis-Tris 15 % w/v SOKALAN® CP 5 pH 6.0	
MemGold C11	0.02 M Nickel(II) sulfate hexahydrate 0.01 M HEPES 33 % v/v Jeffamine® M-600 pH 7.0	
MIDAS E2	0.2 M Potassium acetate 35 % v/v Pentaerythritol propoxylate (5/4 PO/OH)	

**Table 5.2 (cont'd)**

MIDAS E5	0.2 M Potassium chloride 0.1 M HEPES 15 % v/v Jeffamine® T-403 pH 6.5	
----------	--	--

### 5.2.5.3 Other modifications

Other than the protein concentration optimization, several other changes were made to the crystallization trials as given below.

1. ***Changing the HA2 solubilizing detergent*** - HA2 is soluble in 0.10 % SRC and gives a predominant trimer peak in SEC (discussed in Chapter 3). Crystal set up with SRC gave similar results as in DM with crystal-like particles in JB Screen A11, MIDAS B10 and MIDAS G12 precipitant conditions.
2. ***Salt effect*** - The mesophase in the gas tight syringe was combined with 2  $\mu$ l of 4 M NaCl before mounting onto the crystallization plates through the robot. Similar results were observed in the presence vs. absence of extra salt.
3. ***Use of sandwich plate made from plastic film*** - Small crystals or crystals that are in a sticky, viscous phase such as LCP are typically challenging to pick and mount on to a loop for subsequent X-ray diffraction analysis. As a solution to this, plastic crystalline plates are

available so that each wells in the plate can be easily be cut out, and enable loop-free crystal mounting. This is an emerging technique in the crystallography. Some difficulties we faced were: (1) the mesophase and the precipitant solution on the wells of the plastic plate was mobile, and it was challenging to keep the mixture in one well; and (2) the plastic cover of the sandwich set was sticky, and proper handling required patience and experience. No crystals were obtained with this method.

4. ***Lipidic sponge phase*** - The sticky nature of the LCP phase results in difficulties in the crystal handling. One solution to this difficulty would be lipid sponge phase (LSP) bilayer crystallization method. In the conventional cubic phase method there is a limited size of aqueous pores, hence the hydrophilic domains of the membrane proteins can interfere with the cubic phase. However, the sponge phase has much larger aqueous pores which can stabilize the soluble part of the membrane protein. As HA2 has a much larger soluble ectodomain (~100 Å long), this is another potential method one can think of.<sup>14</sup> The LSP procedure includes<sup>15</sup>: preparation of a LCP by mixing monoolein:water = 3:2 (v/v); mixing the LCP with a LSP initiating solution (20% Jeffamine M- 600, 1 M HEPES at pH 8.0, 0.7 M (NH)<sub>4</sub>SO<sub>4</sub>, and 2.5% 1,2,3-heptanetriol) in a ratio of 1:4 (v/v); allow the mixture to phase separate (will take several hours); Separating the upper LSP layer; combining 1 µL of LSP with 1 µL of protein solution (10 mg/ml) and 1 µL of additive 1M trisodium citrate; and using the resultant mixture as the precipitant solution in a hanging-drop, vapor-diffusion experiment. We could not obtain crystals with the LSP method, probably because of the lack of initial knowledge and experience.

5. *Use of HA2 mutants* - When monoolein is combined with WT-HA2, there was a cloudy mixture formed in the gas-tight syringe before the formation of the transparent LCP. We thought that this is because of the fusogenic properties of the WT-HA2. As this might have a negative effect on the crystallogenesis, a fusion-impaired HA2 G1E mutant was used for the crystallization trials. G1E-HA2 did not form a cloudy mixture during LCP preparation. However, the crystallization results were similar to WT-HA2 with some crystal-like particles, which are difficult to mount on to a loop for subsequent analysis.

### **5.3 HA2 crystallization - part II (collaboration with Creative Biostructure - Shirley, NY 11967)**

As the next step of the HA2 crystallization project, the purified HA2 was sent to “Creative Biostructure”, who are specialized in providing contract services in the field of structural biology. The summary of the project report from Creative Biostructure is given in the following subsections.

#### **5.3.1 Crystallization screening with LCP method**

##### **5.3.1.1 Experimental methods**

- Protein concentration: 30 mg/ml
- LCP with Protein:lipid = 2:3
- Random plates with NT8 robot
- 96×5 and 48×12 crystallization trials
- Droplet size - 1040 nl (40 nl LCP + 1 µl precipitant solution)
- 20 °C incubation

- X-ray diffraction data collection at 100 K at BL17U1 Macromolecular crystallography beamline

**Table 5.3** The screen kits (precipitant solution kits).

Screening Formulations	Company
1 Cubic formulations (96)	Emerald Bio
2 Morpheus kit (96)	Molecular Dimensions (MD)
3 MemGold Box1of2(48)	MD
3 MemGold Box2of2(48)	MD
4 Memplus (48)	MD
4 Memstart (48)	MD
5 MembFac (48)	HAMPTON
5 Mem Sys (48)	MD
6 Crystal Screen 1 (48)	HAMPTON
6 Crystal Screen 2 (48)	HAMPTON
7 ThePGAScreenBox1of2(48)	MD
7 ThePGAScreenBox2of2(48)	MD
8 NeXtal Tubes Anions Suite (96)	Qiagen
9 SG1ScreenBox2of2(48)	MD
9 SG1ScreenBox1of2(48)	MD
10 NeXtal Tubes MBClass I Suite (96)	Qiagen
11 NeXtal Tubes MBClass II Suite (96)	Qiagen

### 5.3.1.2 Results

**Table 5.4** Results from the LCP method.

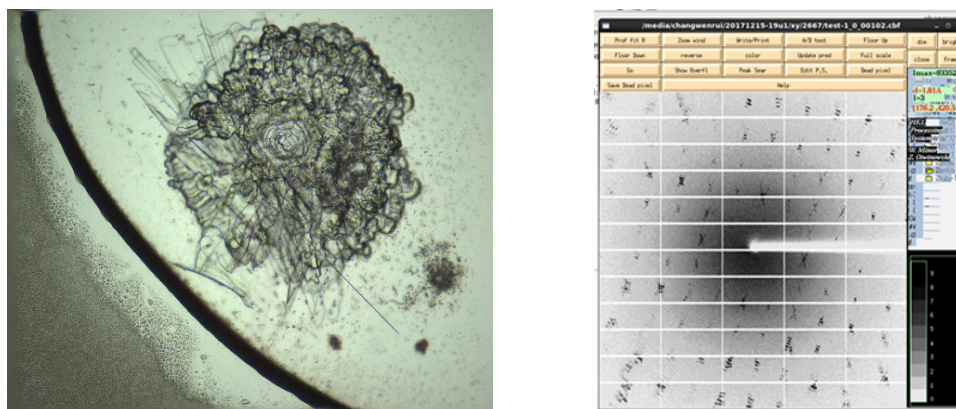
Picture #	Conditions	Crystal morphology	X-ray	
01 1-D10	LCP-1-D10	Large strip	No patterns	01
02 1-E6	LCP-1-E6	Small strip	No patterns	02
03 1-F8	LCP-1-F8	Long flake	No patterns	03
04 3-B9	LCP-3-B9	Diamond	Salt crystal	04
05 3-F12	LCP-3-F12	Small pieces	May be fat crystal	05
06 4-D8	LCP-4-D8	Strips	Salt crystal	06
07 4-F2	LCP-4-F2	Acicular	No patterns	07
08 5-G7	LCP-5-G7	Long flake	Salt crystal	08
09 5-H2	LCP-5-H2	Long flake	Salt crystal	09
10 6-A1	LCP-6-A1	Small pieces	Salt crystal	10
11 6-C3	LCP-6-C3	Long strip	No patterns	11
12 7-C8	LCP-7-C8	Diamond-shaped sheet	No patterns	12
13 7-F9	LCP-7-F9	Triangle piece	No patterns	13
14 7-G4	LCP-7-G4	Flake	<b>Protein crystal 1.5 Å.</b>	<b>14</b>
15 7-H4	LCP-7-H4	Fusiform	Salt crystal	15
16 9-B2	LCP-9-B2	Strips	No patterns	16
17 9-G3	LCP-9-G3	Diamond-shaped sheet	No patterns	17

**Table 5.5** The precipitant solution conditions that gave crystals.

Picture #	Growth condition	Buffer
01 1-D10	LCP-1-D10	0.1 M HEPES pH 7.0, 30 % v/v PEG 500 MME
02 1-E6	LCP-1-E6	0.1 M Sodium chloride, 0.1 M Tris pH 8.0, 0.2 30 % v/v PEG 400
03 1-F8	LCP-1-F8	0.1 M Sodium chloride, 0.1 M Tris pH 8.0, 0.2 30 % v/v PEG500 DME
04 3-B9	LCP-3-B9	0.2 M Sodium chloride, 0.1 M HEPES pH 7.0, 22 % v/v PEG 550 MME

**Table 5.5 (cont'd)**

05 3-F12	LCP-3-F12	0.002 M zinc sulfate, 0.08 M HEPES pH 7.0, 25 % v/v Jeffamine ED2001
06 4-D8	LCP-4-D8	0.1 M tri-sodium citrate pH 5.6, 1 M ammonium dihydrogen phosphate
07 4-F2	LCP-4-F2	0.02 M Tris pH 7.5, 33 % w/v PEG 1500
08 5-G7	LCP-5-G7	0.1 M Sodium acetate trihydrate pH 4.6, 1.0 M Magnesium sulfate heptahydrate
09 5-H2	LCP-5-H2	0.1 M Sodium citrate tribasic dihydrate pH 5.6, 1.0 M Magnesium sulfate heptahydrate
10 6-A1	LCP-6-A1	0.02 M Calcium chloride dehydrate, 0.1 M Sodium acetate, trihydrate, pH 4.6, 30% v/v (+/-)-2- Methyl-2,4-pentanediol
11 6-C3	LCP-6-C3	0.2 M Ammonium acetate, 0.1 M tris hydrochloride pH 8.5, 30% v/v 2-Propanol
12 7-C8	LCP-7-C8	0.1 M Ammonium sulfate, 0.3 M Sodium formate, 0.1 M Sodium cacodylate pH 6.5, 22 % v/v MPD
13 7-F9	LCP-7-F9	0.2 M Potassium thiocyanate, 0.1 M Tris pH 7.8
<b>14 7-G4</b>	<b>LCP-7-G4</b>	<b>0.1 M Ammonium sulfate, 0.3 M Sodium formate, 0.1 M Sodium acetate pH 5.0, 8 % w/v PEG8000</b>
15 7-H4	LCP-7-H4	0.1 M Ammonium sulfate, 0.3 M Sodium formate, 0.1 M Sodium acetate pH 5.0, 8 % w/v PEG20000
16 9-B2	LCP-9-B2	0.2 M Lithium sulfate, 0.1 M Tris pH 8.5, 30% w/v PEG4000
17 9-G3	LCP-9-G3	0.2 M Ammonium chloride, 20 % w/v PEG3350



**Figure 5.7** The crystallization screening image (left) and the diffraction pattern for the growth condition LCP-7-G4 (right).

### **5.3.1.3 Conclusion**

- High-quality protein crystals are obtained under conditions of 0.1 M Ammonium sulfate, 0.3 M Sodium formate, 0.1 M Sodium acetate pH 5.0, 8% w/v PEG 8000 by LCP method.
- the diffraction limit of the crystal is at 1.5 Å.

## **5.3.2 Crystallization screening with sitting-drop method**

### **5.3.2.1 Experimental methods**

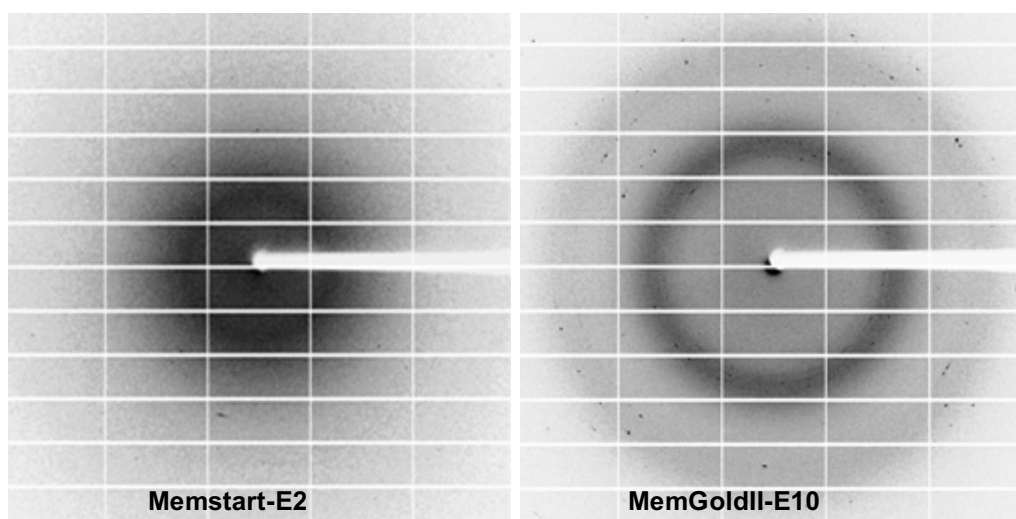
- Protein concentration: 16mg/ml
- 96×5 crystallization trials
- 20 °C incubation for two months
- X-ray diffraction data collection at Synchrotron Radiation Facility

### **5.3.2.2 Results**

Crystals are obtained under conditions of:

- Memstart-E2 (Small strip) - 0.1 M Tris pH 8.5, 2 M ammonium dihydrogen phosphate
- MemgoldII-E10 (Smaller strip) - 0.2 M magnesium chloride, 0.1 M potassium chloride, 0.25 M sodium citrate pH 4.0, 11 % w/v PEG 8000





**Figure 5.8** The diffraction patterns for the growth conditions: Memstart-E2 (left); and MemGoldII-E10 (right).

### 5.3.2.3 Conclusion

- One may be a protein crystal without X- diffraction (Memstart-E2), and the other one may be a salt crystal (MemgoldII-E10).

### 5.3.3 Future directions

The Fig. 5.7 protein crystals obtained from the LCP method are more promising, however the spots in the diffraction pattern are split and unclear. This may be due to the presence of more than one crystal lattice, and further optimization is required to obtain better crystals.

## REFERENCES

## REFERENCES

1. Patrick, C. (2014) A tale of chromatin and transcription in 100 structures *Cell* 159, 985-994.
2. Krishnan, V. V., and Rupp, B. (2012) *Macromolecular Structure Determination: Comparison of X-ray Crystallography and NMR Spectroscopy*, , eLS. John Wiley & Sons, Ltd: Chichester.
3. Hanske, J., Sadian, Y., and Müller, C. W. (2018) The cryo-EM resolution revolution and transcription complexes, *Curr Opin Struct Biol.* 52, 8-25.
4. Noyan, I. C., and Cohen, J. B. (1987) *Fundamental Concepts in X-ray Diffraction. In: Residual Stress*, Materials Research and Engineering. Springer, New York, NY
5. Ilari, A., and Savino, C. (2008) Protein structure determination by x-ray crystallography, In *Bioinformatics*, pp 63-87, Springer.
6. PDBTM: Protein Data Bank of Transmembrane Proteins. [http://pdbtm.enzim.hu/?\\_=/statistics/growth](http://pdbtm.enzim.hu/?_=/statistics/growth) (access date: 12/10/2018)
7. Seddon, A. M., Curnow, P., and Booth, P. J. (2004) Membrane proteins, lipids and detergents: not just a soap opera, *Biochimica et Biophysica Acta (BBA)-Biomembranes* 1666, 105-117.
8. Caffrey, M. (2015) A comprehensive review of the lipid cubic phase or in meso method for crystallizing membrane and soluble proteins and complexes, *Acta Crystallographica Section F: Structural Biology Communications* 71, 3-18.
9. Caffrey, M. (2009) Crystallizing membrane proteins for structure determination: use of lipidic mesophases, *Annual review of biophysics* 38, 29-51.
10. Caffrey, M. (2011) Crystallizing membrane proteins for structure-function studies using lipidic mesophases *Biochemical Society Transactions* 39, 725-732.
11. Caffrey, M. (2008) On the mechanism of membrane protein crystallization in lipidic mesophases, *Crystal Growth and Design* 8, 4244-4254.
12. McPherson, A., and Cudney, B. (2014) Optimization of crystallization conditions for biological macromolecules, *Acta Crystallographica Section F* 70, 1445-1467.
13. Caffrey, M., and Cherezov, V. (2009) Crystallizing Membrane Proteins Using Lipidic Mesophases, *Nat Protoc.* 4, 706-731.

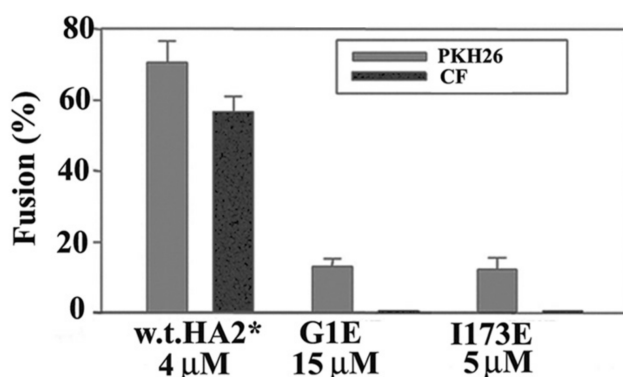
14. Wadsten, P., Wöhri, A. B., Snijder, A., Katona, G., Gardiner, A. T., Cogdell, R. J., Neutze, R., and Engström, S. (2006) Lipidic sponge phase crystallization of membrane proteins, *Journal of molecular biology* 364, 44-53.
15. Wöhri, A. B., Wahlgren, W. Y., Malmerberg, E., Johansson, L. C., Neutze, R., and Katona, G. (2009) Lipidic sponge phase crystal structure of a photosynthetic reaction center reveals lipids on the protein surface, *Biochemistry* 48, 9831-9838.

## **Chapter 6**

### **Cell-cell fusion studies with WT, G1E and I173E HA2**

## 6.1 Introduction

The two truncated HA2 constructs, HA2 1-127 and HA2 1-185 (FHA2), induce lipid mixing between HA0-cells and RBCs.<sup>1, 2</sup> HA0-cells expressed uncleaved HA's that do not induce fusion, but they mediate a physiologically relevant interaction of receptor binding between HA1 and the sialic acid receptors present on RBCs. This interaction brings the HA0-cells and RBCs close to each other without inducing fusion. Other than the lipid mixing, FHA2 also induces the mixing of a small content dye carboxyfluorescein (CF), indicating the opening of a small fusion pore (Fig. 6.1).<sup>1</sup> However, the pores are not large enough to allow the mixing of a large aqueous probe, 10 kDa dextran. The FHA2 G1E and I173E mutants reduce both the lipid mixing and content mixing of HA0-cell/RBC fusion.<sup>1</sup> The vesicle fusion activity by full-length WT-, G1E- and I173E- HA2 was discussed in Chapter 3. Cell fusion studies with full-length HA2's have not been previously studied. In this project, we collaborated with Dr. Chernomordik at NIH to study the cell-cell fusion experiments with full-length HA2.



**Figure 6.1** Fusogenic activity of FHA2  $\equiv$  HA2\* is greatly reduced by G1E and I173E mutations that are known to reduce the HA-mediated fusion. Fusion between HA0 cells and bound RBCs is induced by the medium acidification (pH  $\sim$ 5) in the presence of FHA2, G1E-FHA2, or I173E-

FHA2. Fusion is assayed as redistribution of lipid probe PKH26 (gray bars) or the small aqueous probe CF (black bars).<sup>1</sup>

## **6.2 Methods, Results and Discussion**

### **6.2.1 Cell fusion assay**

WT-, G1E-, and I173E- FHA2 and HA2 were expressed and purified as described in Chapter 3. The FHA2 and HA2 proteins were dialyzed into two different detergents: HA2 into 0.17% DM vs. FHA2 into 0.10% SRC, so as to maximize the amount of trimer relative to aggregated protein. Other than the detergents, the dialysis buffer also included 10 mM Tris at pH 7.4 and 150 mM NaCl.

The cell fusion experimental set up performed by the Chernomordik group includes fusion between HA0-cells and RBCs.<sup>1</sup> In the first few trials of the fusion assay, HA0-cells with bound RBCs were incubated with exogenously added FHA2 or HA2 proteins, and then the medium was acidified. The main challenge with this experiment was that the low pH dependent leakage induced by the exogenously added proteins. This leakage limits the concentration of protein that could be used. The same fusion assay has been performed earlier with FHA2 without any leakage when the protein is dialyzed into a buffer containing 0.1 % (v/v) Triton X-100, 150 mM NaF and 10 mM NaH<sub>2</sub>PO<sub>4</sub> at pH 7.0.<sup>1</sup> As the next step of the project, the purified proteins were dialyzed into the above buffer containing the Triton X-100 detergent. The change in buffer dramatically decreased the FHA2 induced leakage, and the previous cell fusion experiments with FHA2 (= HA2\*)<sup>1</sup> could be reproduced. To figure out why there is such a strong dependence of leakage on the detergent used, the secondary structure and the oligomeric state were compared in the original buffer and the new buffer (discussed in section 6.2.2).

Even in Triton X-100 detergent, the HA2 lyses cells and the leakage continued. In these trials, HA2 was applied to pre-bound HA0-cells and RBCs. In later trials, HA2 was applied only to HA0-cells, and then RBCs were applied followed by medium acidification. In this new setup, full size HA2 induced lipid and content mixing at concentrations lower than the FHA2 concentrations required to induce lipid and content mixing. G1E-HA2 did not induce fusion at the same concentration. Interestingly, the WT-HA2 formed fusion pores large enough to pass 10 kDa dextran dye (Unpublished data). In some contrast, FHA2 formed only the pores that allow passage of the smaller CF dye (~0.5 kDa). The difference between the setups in which proteins were applied to HA0-cells only vs. to HA0 expressing cells with RBCs is likely related to fusion inhibition by HA2 bound to RBC in the latter setup.

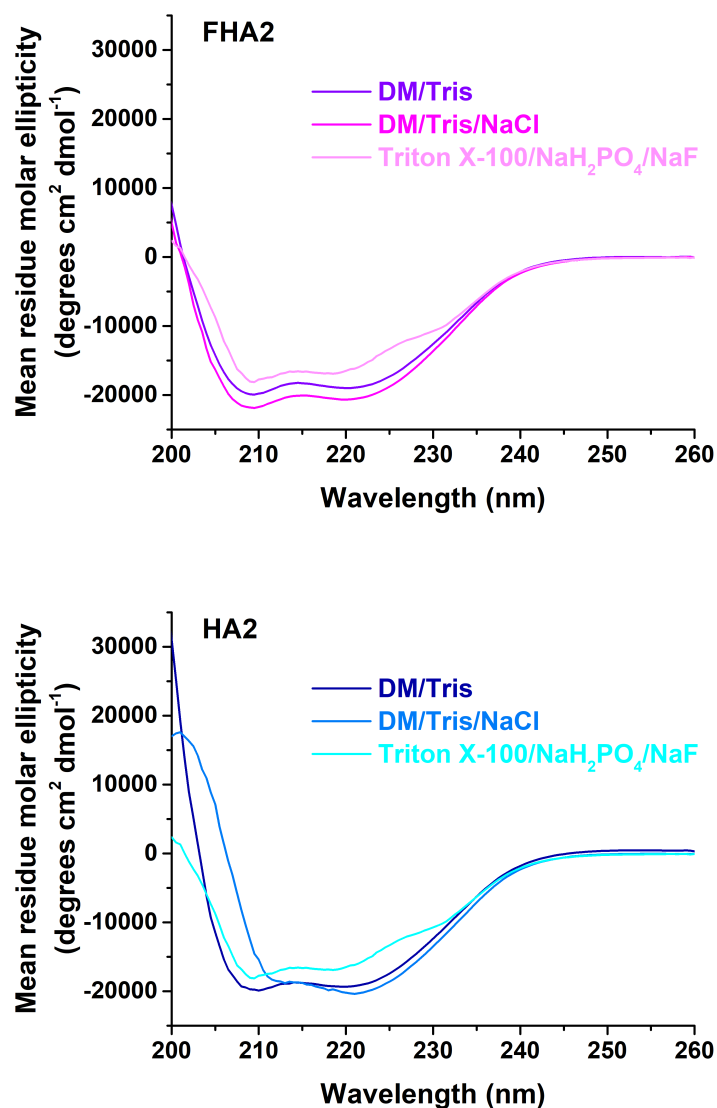
### **6.2.2 Comparative CD and SEC for FHA2 and HA2 in different detergents**

The secondary structures of WT- and G1E- FHA2 have been previously compared in 10 mM NaH<sub>2</sub>PO<sub>4</sub>, 0.14 M NaF at pH 7.4 with 0.5 μM Triton X-100, and the WT shows a 24% α-helical and 25% β-sheet structure, whereas the mutant shows 40% α-helical and 18% β-sheet structure.<sup>1</sup> Our CD spectra in 0.17% DM support ~65% helicity for WT- and G1E- FHA2 and ~58% helicity for WT- and G1E- HA2 (discussed in Chapter 3). In the previous study, there is much less than one detergent molecule per protein, based on the reported 0.5 μM Triton X-100 concentration. The lower helicity may be due to the unfolding associated with lack of enough detergent and/or presence of the F<sup>-</sup> ions, which can form strong hydrogen bonds with amide hydrogens of the protein and disrupt intramolecular hydrogen bonds.<sup>3</sup>

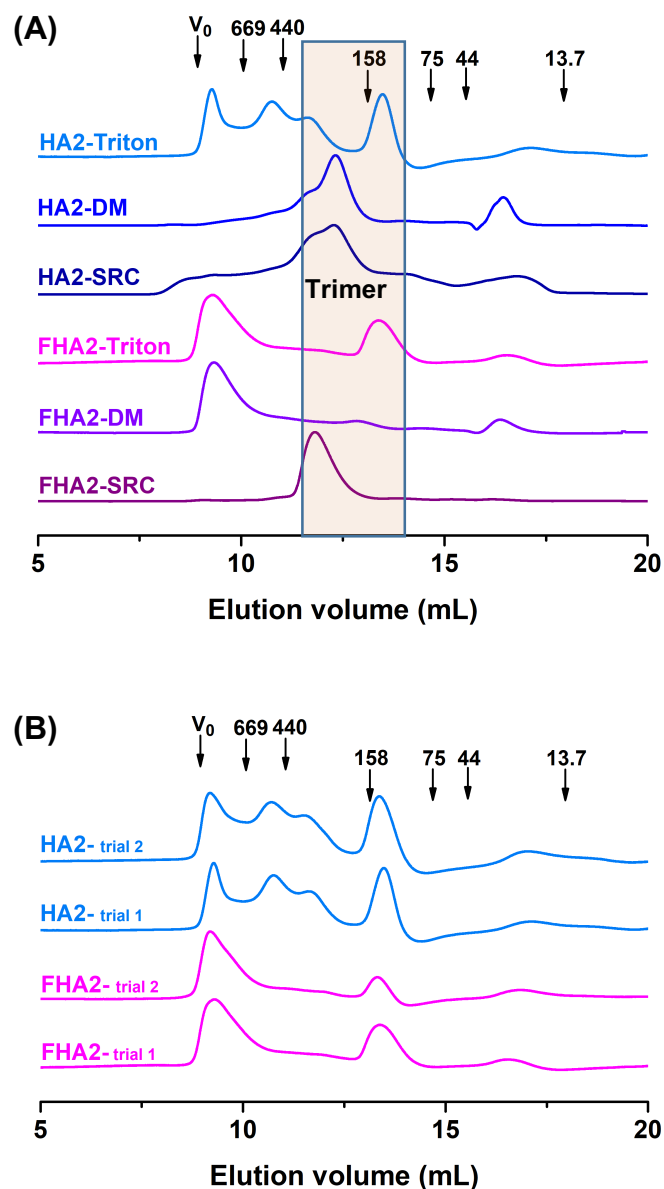
In this part of the project, the secondary structure and oligomeric state of the WT- HA2 and FHA2 proteins were analyzed and compared to that in the original buffer (10 mM Tris at pH



7.4, 150 mM NaCl, and 0.17% DM (HA2) or 0.10% SRC (FHA2)) and in the new buffer (10 mM NaH<sub>2</sub>PO<sub>4</sub> at pH 7.0, 150 mM NaF, and 0.1 % (v/v) Triton X-100) (Figs. 6.2 and 6.3).

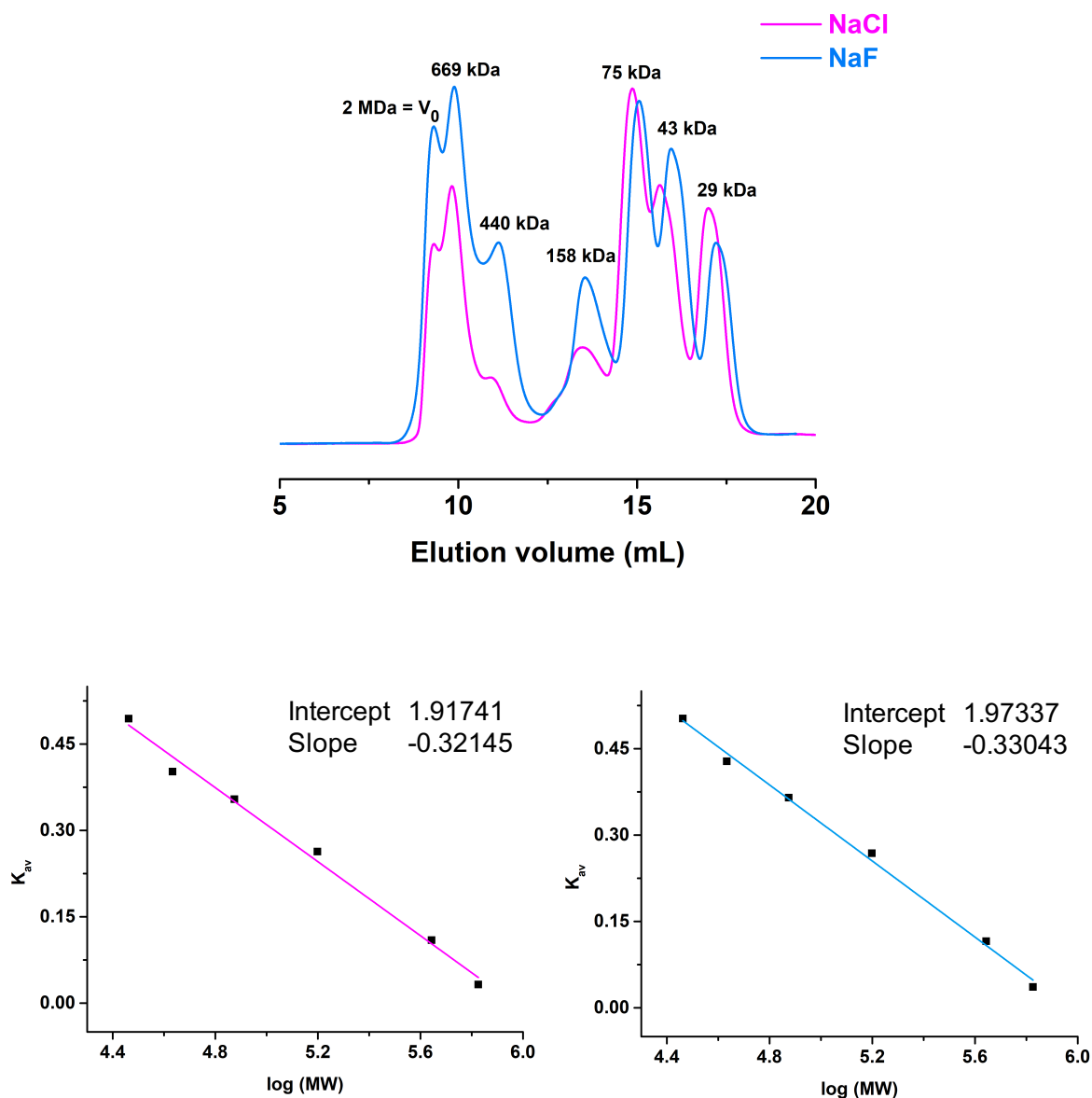


**Figure 6.2** CD spectra of FHA2 and HA2 proteins at ambient temperature. Samples contained 16  $\mu$ M protein in (1) 0.17% DM, 10 mM Tris buffer at pH 7.4; (2) 0.17% DM, 150 mM NaCl, 10 mM Tris buffer at pH 7.4; and (3) 0.1 % (v/v) Triton X-100, 150 mM NaF, 10 mM NaH<sub>2</sub>PO<sub>4</sub> buffer at pH 7.0.



**Figure 6.3** SEC of FHA2 and HA2 proteins: (A) in (1) 0.1% SRC, 150 mM NaCl, 10 mM Tris buffer at pH 7.4; (2) 0.17% DM, 150 mM NaCl, 10 mM Tris buffer at pH 7.4; and (3) 0.1 % (v/v) Triton X-100, 150 mM NaF, 10 mM  $\text{NaH}_2\text{PO}_4$  buffer at pH 7.0; and (B) in buffer (3) for replicate samples. SEC parameters include [loading protein] = 1 mg/mL, [protein running]  $\approx$  0.1 mg/mL, and  $A_{280}$  detection. Arrows mark the elution volumes of the void ( $V_0$ ) and MW standards in kDa.

To see whether the type of salt used (NaCl/NaF) has an effect on the SEC elution volume, the SEC column was calibrated with these two salts. The spectra show no considerable difference as shown in Fig. 6.4.



**Figure 6.4** Size exclusion chromatogram for the  $A_{280}$  detection of MW standards (top), and calibration curves (bottom) using either 150 mM NaCl (pink) or NaF (blue) in the buffer solution.

The CD spectra are similar in both DM and Triton X-100 detergents (Fig 6.2). Considering  $\theta_{222} = -33000$  as 100% helicity,<sup>4</sup> there is ~50% helicity in Triton X-100, and ~60% helicity in DM. The ~150 kDa peak in SEC with Triton X-100 can be considered as trimers with ~75 kDa protein and ~75 kDa detergent contributions.

Successful cell-cell fusion (fusion without protein induced leakage) was obtained when the FHA2 and HA2 proteins are dialyzed into Triton X-100 detergent. Our CD and SEC data support well folded proteins that are mixtures of trimers and larger oligomers in this detergent.

## REFERENCES

## REFERENCES

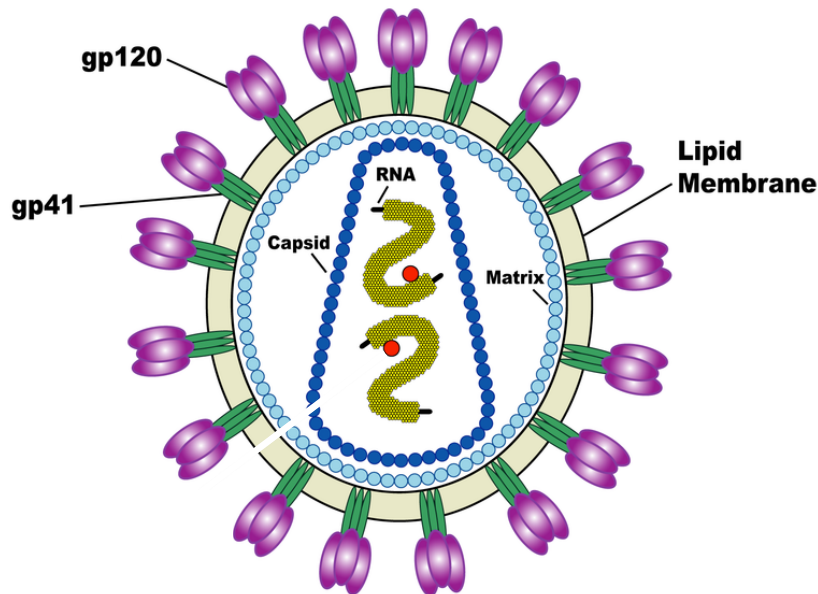
1. Kim, C. S., Epand, R. F., Leikina, E., Epand, R. M., and Chernomordik, L. V. (2011) The final conformation of the complete ectodomain of the HA2 subunit of influenza hemagglutinin can by itself drive low pH-dependent fusion, *Journal of Biological Chemistry* 286, 13226-13234.
2. Leikina, E., LeDuc, D. L., Macosko, J. C., Epand, R., Epand, R., Shin, Y.-K., and Chernomordik, L. V. (2001) The 1– 127 HA2 construct of influenza virus hemagglutinin induces cell– cell hemifusion, *Biochemistry* 40, 8378-8386.
3. Wu, C.-W., Cheng, S.-F., Huang, W.-N., Trivedi, V. D., Veeramuthu, B., Wu, W.-G., and Chang, D.-K. (2003) Effects of alterations of the amino-terminal glycine of influenza hemagglutinin fusion peptide on its structure, organization and membrane interactions, *Biochimica et Biophysica Acta (BBA)-Biomembranes* 1612, 41-51.
4. Chen, Y.-H., Yang, J. T., and Chau, K. H. (1974) Determination of the helix and  $\beta$  form of proteins in aqueous solution by circular dichroism, *Biochemistry* 13, 3350-3359.

## **Chapter 7**

### **Studies on HIV gp41 membrane fusion protein**

## 7.1 Introduction

The human immunodeficiency virus (HIV) affects the human immune system, and impairs the defense mechanisms against infections and certain forms of cancer. The acquired immune deficiency syndrome (AIDS) is the most advanced stage of HIV infection.<sup>1</sup> WHO reported that at the end of 2017, there were ~36.9 million people with HIV globally, and this continues to be a principal public health issue. Although there are antiretroviral drugs that control the virus and prevent transmission, no cure or vaccine is available for the HIV infection.



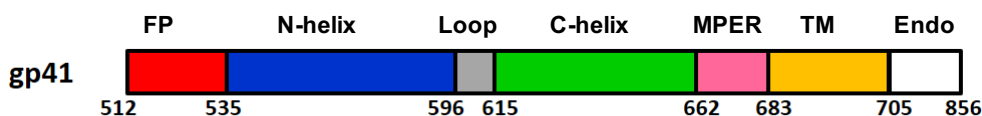
**Figure 7.1** Schematic diagram of an HIV virion. There are ~15 gp41/gp120 trimers per virion.<sup>2-4</sup>

This is much lower compared to the ~400 HA2/HA1 trimers per influenza virion.

HIV is also a class I enveloped virus as is influenza. The viral infection begins by the fusion of viral and cellular membranes. The viral membrane protein gp160 is responsible for the HIV binding to the host cell and subsequent membrane fusion. The precursor gp160 is proteolytically

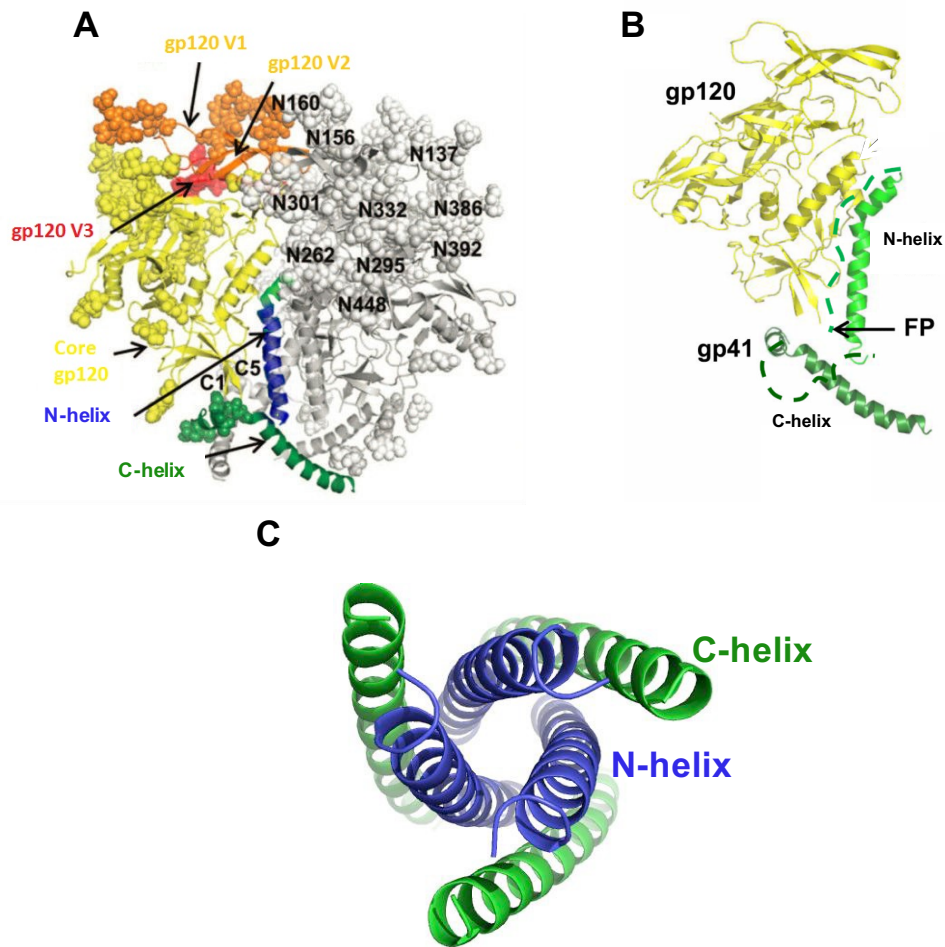


cleaved into two noncovalent linked subunits: the receptor binding subunit gp120, and the fusion subunit gp41.<sup>5</sup> The gp41 is a single pass integral membrane protein with ~170-residue ectodomain and ~150-residue endodomain (Fig. 7.2).



**Figure 7.2** The schematic diagram of full-length gp41 in the final state with domain colored: fusion peptide (FP), red; N-terminal helix, (N-helix), blue; loop, grey; C-terminal helix (C-helix), green; membrane-proximal external region (MPER), pink; transmembrane domain (TM), yellow; and endodomain (Endo), white. The residue numbering is according to gp160 sequence numbering for the HXB2 laboratory strain of HIV.

High resolution structures are available for the initial complex of gp41 ectodomain with gp120 with typical 3-5 Å resolution,<sup>6-11</sup> and soluble regions of gp41 ectodomain without gp120 with ~5Å resolution.<sup>12-14</sup> In the initial complex, gp41 has a set of  $\alpha$ -helices connected by loops, and in the latter structure, it has trimers-of-hairpins with antiparallel N- and C-helices. The hairpin trimer is hyperthermostable ( $T_m \approx 110$  °C), and is probably the final state of fusion.<sup>15, 16</sup> The structures of gp41 in the initial complex of gp120 and the final state structure without gp120 show a large-scale rearrangement of helical segments (Fig. 7.3).

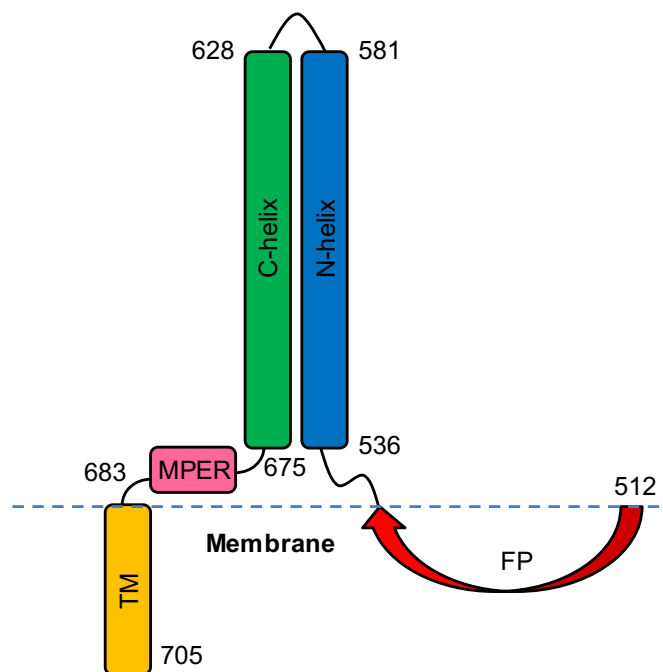


**Figure 7.3** The ribbon diagrams of the: (A) initial complex of gp120 and gp41 (PDB 4NCO). N-linked glycans are shown as gray spheres; (B) close view of the hetero-dimer of gp120 and gp41 in the trimer in (A); and (C) gp41 ectodomain without gp120 in the final state (PDB 2X7R). The FP is absent in this structure.

The gp41 membrane interacting domains are the MPER, TM and FP domains. Structural details for these domains are available based on studies with truncated gp41 constructs or individual peptides of these domains. For example, for the gp41 FP, NMR data support an  $\alpha$ -helical structure in SDS or DPC detergents,<sup>17-19</sup> and an antiparallel  $\beta$ -sheet structure in

membranes.<sup>20-23</sup> For the MPER, NMR and EPR data support two discrete helical segments with a central hinge.<sup>24</sup> A crystal structure containing N-helix, C-helix, and MPER regions shows a trimeric, SHB with an extended C-helix/MPER long helix with a slight bend.<sup>25</sup> The MD stimulation data of the TM in membranes support a tilted  $\alpha$ -helical structure, which has a tendency to form a trimer.<sup>26</sup> In addition, liquid state NMR shows a long uninterrupted  $\alpha$ -helix for the MPER/TM segment.<sup>27</sup>

Fig. 7.4 shows a structural model of gp41 construct containing the FP and TM domains based on CD spectra of truncated gp41 constructs. This was a study by Dr. Shuang Liang in the Weliky group.<sup>28</sup>

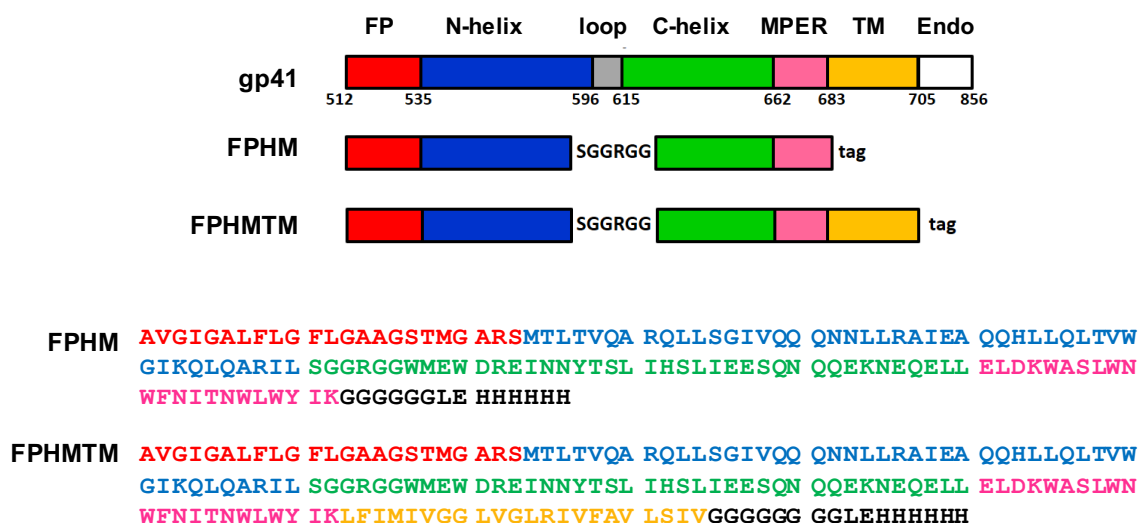


**Figure 7.4** Structural model of a gp41 construct containing the FP and TM (FPHMTM). A monomer is shown for clarity.<sup>28</sup>

## 7.2 Structure-Function relationship of HIV gp41 by site-specific mutagenesis

### 7.2.1 Introduction

In this project V2E mutants of two gp41 constructs, FPHM and FPHMTM, were studied for the purpose of better understanding the gp41-catalyzed fusion. The schematic diagrams of the WT-constructs and their amino acid sequences are given in Fig. 7.5. This mutation in the FP is known to eliminate the gp41-mediated fusion.<sup>29-33</sup> The WT constructs were initially studied by Dr. Shuang Liang, and FPHM V2E mutant was studied by Dr. Lihui Jia in the Weliky group. In the present work, both FPHM and FPHMTM V2E mutants were studied and compared with the respective WT. In these constructs the loop region from residue 582 to 627 in gp41 is replaced with SGGRGG group which does not affect the assembly of the SHB.<sup>34-36</sup>



**Figure 7.5** The schematic diagrams and amino acid sequences of gp41 constructs studied. The residues 582-627 is replaced with a non-native SGGRGG loop. FPHM  $\equiv$  gp41<sub>512-581/SGGRGG/628-683</sub> and FPHMTM  $\equiv$  gp41<sub>512-581/SGGRGG/628-705</sub>. The non-native C-terminal regions are in black.

## 7.2.2 Site-specific mutation

### (A) DNA sequence of WT-FPHMTM

GCG**GTT**GGCATTGGTGCGCTGTTCTGGGCTTTCTGGGTGCGGCGGGTAGCACTATGGGTGCGCGTAGCATGACC  
CTGACCGTTCAAGCGCGTCAACTGCTGAGCGGTATCGTGCAGCAACAGAACAACCTGCTGCGTGCGATTGAGGCG  
CAACAGCACCTGCTGCAGCTGACCGTTTGGGGCATCAAGCAACTGCAGGCGCGTATTCTGAGCGGTGGCCGTGGT  
GGCTGGATGGAGTGGGACCGTGAAATCAACAACCTACACCAGCCTGATCCACAGCCTGATTGAGGAAAGCCAAAAC  
CAACAGGAGAAGAACGAGCAGGAACCTGCTGGAACCTGGATAAATGGGCGAGCCTGTGGAACCTGGTTCAACATCACC  
AACTGGCTGTGGTACATCAAACCTGTTTCATCATGATCGTGGGTGGCCTGGTTGGTCTGCGTATCGTGTTCGCGTT  
CTGAGCATTGTTGGCGGTGGTGGTGGCGGTGGTGGT

### (B) Primer sequence for the V2E mutation

The forward primer - 5' - CAT ATG GCG GAG GGC ATT GGT G - 3'

The reverse primer - 5' - CAC CAA TGC CCT CCG CCA TAT G - 3'

### (C) Amino acid sequence of V2E-FPHMTM after the site-specific mutation

**A**EGIGALFLG FLGAAGSTMG ARSM**T**LT**V**QA RQLLSGIVQQ QNNLLRAIEA QQHLLQLTVW  
GIKQLQARIL SGGRGWMEW DREINNYTSL IHS**L**IEESQN QQEKNEQELL ELDKWASLWN  
WFNITNWLWY IK**L**FIMIVGG LVGLRIVFAV LSIVGGGGGG GGLEHHHHHH

**Figure 7.6** (A) The DNA sequence of WT-FPHMTM; (B) Primer sequences for the V2E mutation; and (C) Amino acid sequence after the mutation based on DNA sequence data.

The plasmid containing the DNA sequence of FPHMTM was extracted from previously sub-cloned *E-coli* cell cultures.<sup>28</sup> The V2E mutant of FPHMTM was generated by PCR using DNA primers and confirmed by DNA sequencing. Each pET24a(+) plasmid was transformed into BL21(DE3) competent cells that were then streaked on an agar plate. After overnight growth in agar, a single colony was transferred to 50 mL Luria-Bertani (LB) medium and grown overnight. All media contained 50 mg/L kanamycin antibiotic, and growth cultures were shaken at 180 rpm and 37 °C. Stock aliquots for future growths were made by mixing 1 mL culture and 0.5 mL 50% glycerol, and were stored at -80 °C.

### 7.2.3 Protein expression and purification

The same expression and purification procedure was used for all four protein constructs, WT- and V2E- FPHM and FPHMTM, and is based on a previous study.<sup>28</sup> An expression culture was prepared by adding 50  $\mu$ L stock to 50 ml LB, followed by overnight growth, and then addition of 1 L fresh LB and growth until  $OD_{600} \approx 0.8$ . Protein expression was induced by addition of 2 mM IPTG and then continued for 5 hours. The incubator settings are 37 °C and 180 rpm. The culture was centrifuged at 9000g and 4 °C for 10 min, followed by harvesting the cell pellet, and storage at -20 °C.

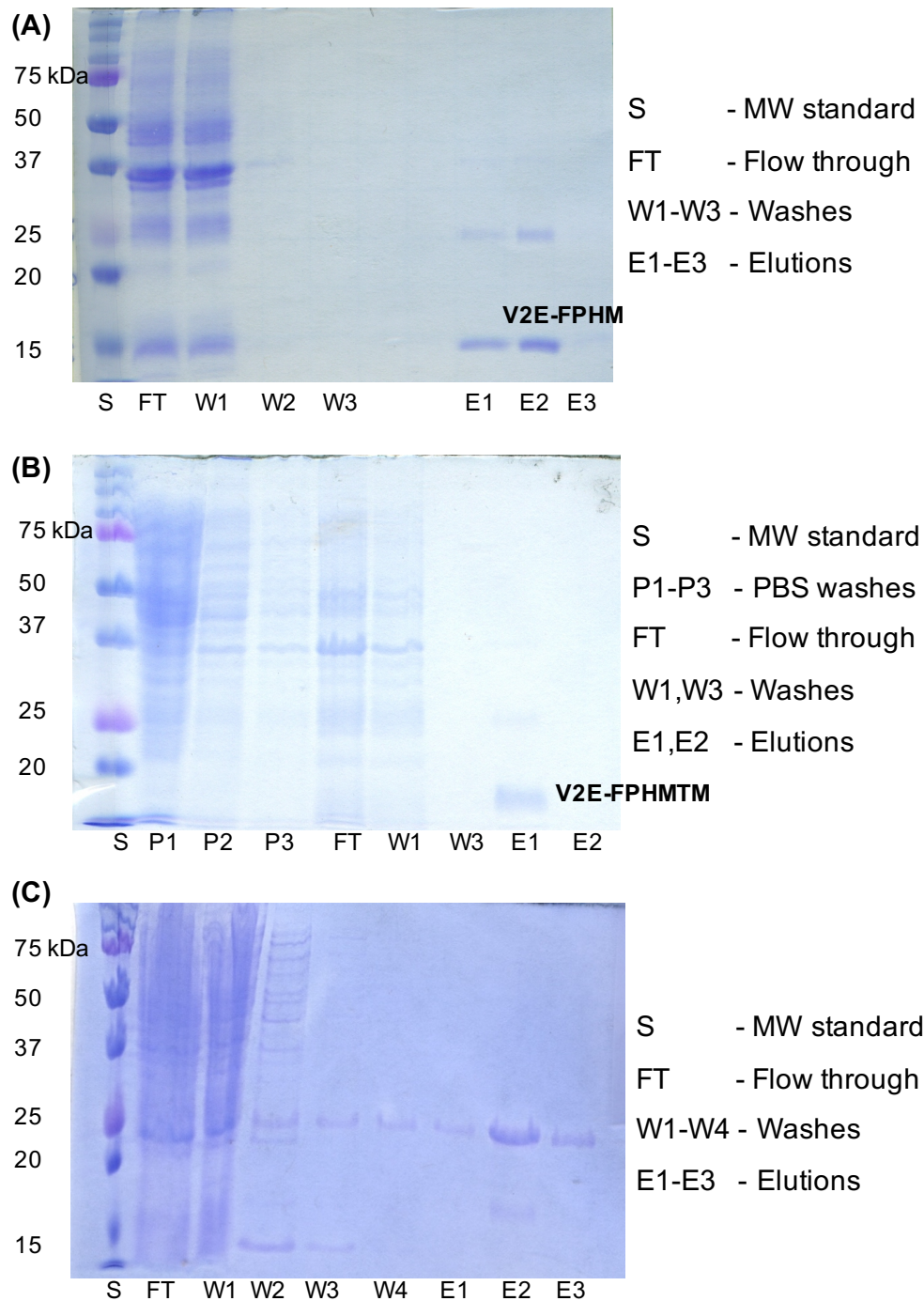
Wet cells (10 g) were subjected to 3 rounds of sonication in PBS (90 ml) followed by centrifugation. Each round lasted for 1.5 min, with 80 % amplitude intervals for 0.8 s followed by 0.2 s off. The pellet was subjected to the same procedure, and the new pellet was also subjected to this procedure. The final pellet was solubilized by sonication (4 rounds) in the 90 mL of solubilizing buffer (8 M urea, 0.5 % SDS, 0.8 % SRC and 10 mM imidazole in PBS at pH 7.4). Each round lasted for 1.5 min, with 80 % amplitude intervals for 0.8 s followed by 0.2 s off. Subsequent affinity purification was done with the supernatant obtained from centrifugation of the lysate at 48000g and 4 °C for 20 min.

This supernatant was mixed with 2 mL of  $Co^{2+}$  affinity resin, and protein binding to resin was achieved during 2 hour stirring at ambient temperature. The protein-coated resin was collected by pouring the suspension through a fritted column, and weakly-bound proteins were removed by washes (4 $\times$  2 ml) with the solubilizing buffer above. The protein was eluted with 4  $\times$  1 mL aliquots of elution buffer (solubilization buffer with 250mM imidazole). The elutions were individually characterized by SDS-PAGE, and those with highest purities and quantities of protein of interest were pooled. The elutions were then mixed with equal volume of ice-cold refolding buffer (10 mM

Tris at pH 8.0, 0.17% DM, 2 mM EDTA, and 1 M *L*-arginine), followed by mixing overnight at 4 °C, and then dialysis into the buffer without arginine for subsequent characterization.

Fig. 7.7 shows the SDS gels of V2E-FPHM and V2E-FPHMTM after the affinity purification. The purified yields of WT- and V2E- FPHM were ~0.4 and 0.8 mg/L culture respectively, whereas the yields of corresponding FPHMTM variants were ~0.3 and 0.5 mg/L culture respectively.

V2E-FPHMTM purification was also tried without the PBS washing, i.e. direct solubilization of the wet cells in above solubilizing buffer. This gave the protein + a high concentration of an impurity protein (Fig. 7.7C). The impurity could be the *E. coli* protein SlyD (MW = 21 kDa), which is removed during the PBS washing.



**Figure 7.7** SDS gels of (A) V2E-FPHM (MW = 16.5 kDa); and (B) V2E-FPHMTM (MW = 18.9 kDa) after the affinity purification with initial PBS washing; and (C) V2E-FPHMTM without PBS washing, which also gave an impurity band at ~22 kDa.



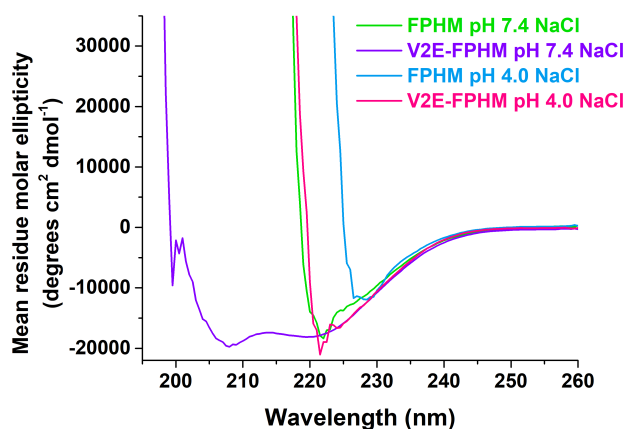
## 7.2.4 CD and SEC

**CD** - Protein was prepared by dialysis of protein in refolding buffer into a SDS containing buffer. Dialysis was initially run for 5 days with 4 buffer changes, however, the proteins were not well folded upon 5-day dialysis. Therefore, increased dialysis times were tested. Protein concentration was adjusted to ~10  $\mu$ M. CD spectra were acquired using a cuvette with 1 mm path. Other parameters include 260-190 nm wavelength range with 0.5 nm increments and 1.5 s signal average per point. Each reported spectrum is the difference between the protein + buffer and buffer- only spectra.

**SEC** - Protein was prepared by dialysis of the protein in refolding buffer against a SDS containing buffer. Dialysis times were varied. SEC was performed using a Tricorn Superdex 200 semi- preparative column, 0.3 mL/min flow rate, and  $A_{280}$  detection at ambient temperature. The column was equilibrated with an initial run with dialysis buffer without protein. Subsequent SEC was done with [loading protein]  $\approx$  1 mg/mL unless given otherwise.

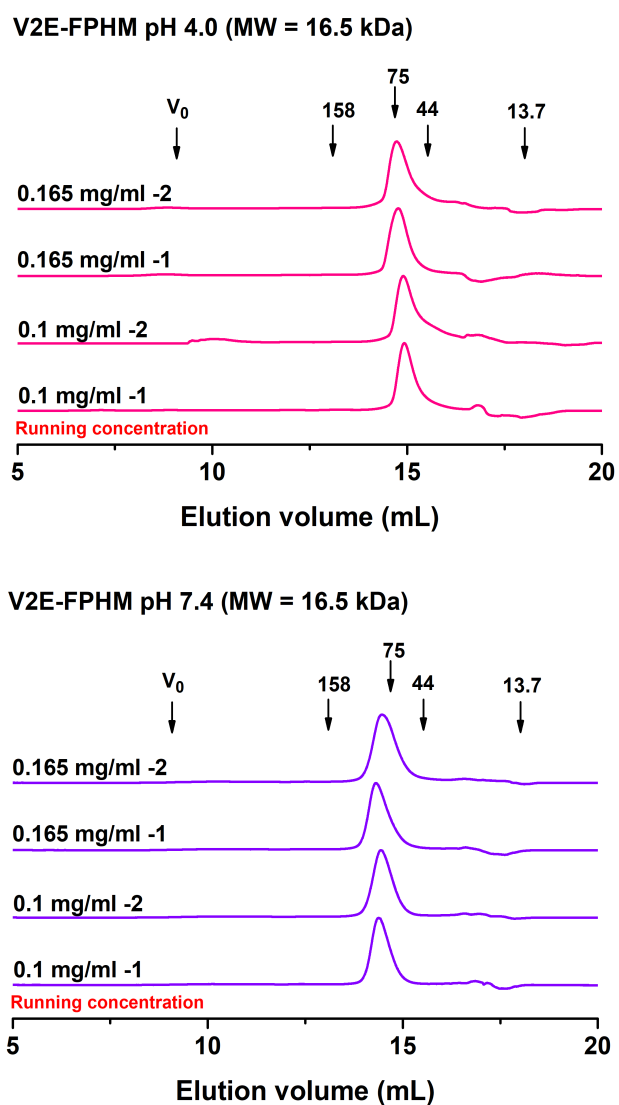
### 7.2.4.1 WT- and V2E- FPHM

Only V2E-FPHM in Fig. 7.8 shows CD spectra characteristics to  $\alpha$ -helical structure. Assuming the other spectra reflect some light scattering during the experiment, the samples were centrifuged for 10 minutes to precipitate any particulate material. However, no precipitate was observed upon centrifugation and the spectra after centrifugation were same as in Fig. 7.8.

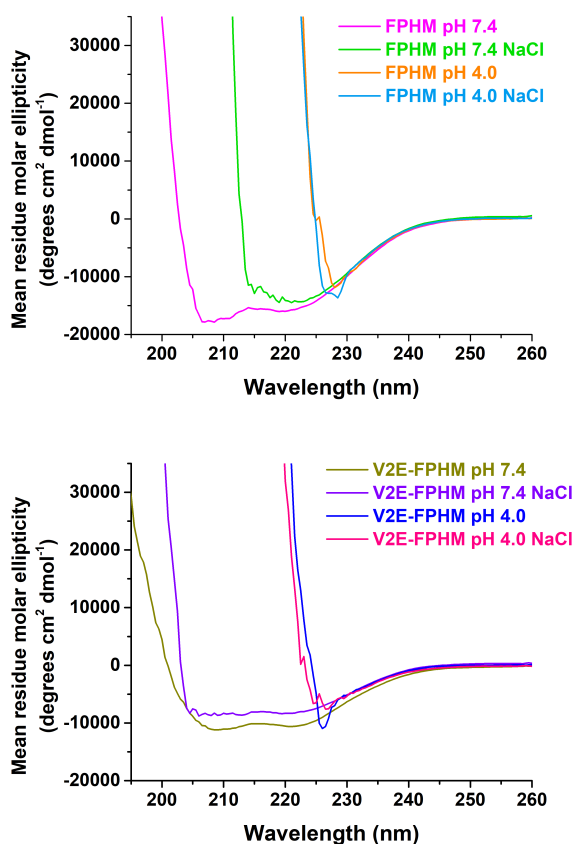


**Figure 7.8** CD spectra of the FPHM proteins at ambient temperature for 7-day dialyzed sample. Sample contained 10  $\mu\text{M}$  protein in either 0.2% SDS/ 150 mM NaCl/ 10 mM Tris at pH 7.4 or 0.2% SDS/ 150 mM NaCl/ 20 mM  $\text{CH}_3\text{COONa}$  at pH 4.0.

In SEC, no prominent peaks were obtained for the WT-FPHM in 7-day dialyzed sample. The protein became highly viscous upon concentrating (0.3 mg/ml - 1 mg/ml loading concentration). However, former data from our group show a similar peak as the V2E mutant for WT-FPHM at pH 7.4 in SEC,<sup>28</sup> and is probably a trimer. This latter study shows CD spectra characteristic to  $\alpha$ -helical structure for WT-FPHM at neutral pH. One difference in the CD experiment in the present vs. former study was the presence vs. absence of 150 mM NaCl in the protein sample. Therefore, as the next step, the effect of NaCl on CD spectra was tested. Fig. 7.10 CD spectra support  $\alpha$ -helical structures for WT- and V2E- FPHM at neutral pH. However further experimentation is required to check the reproducibility.

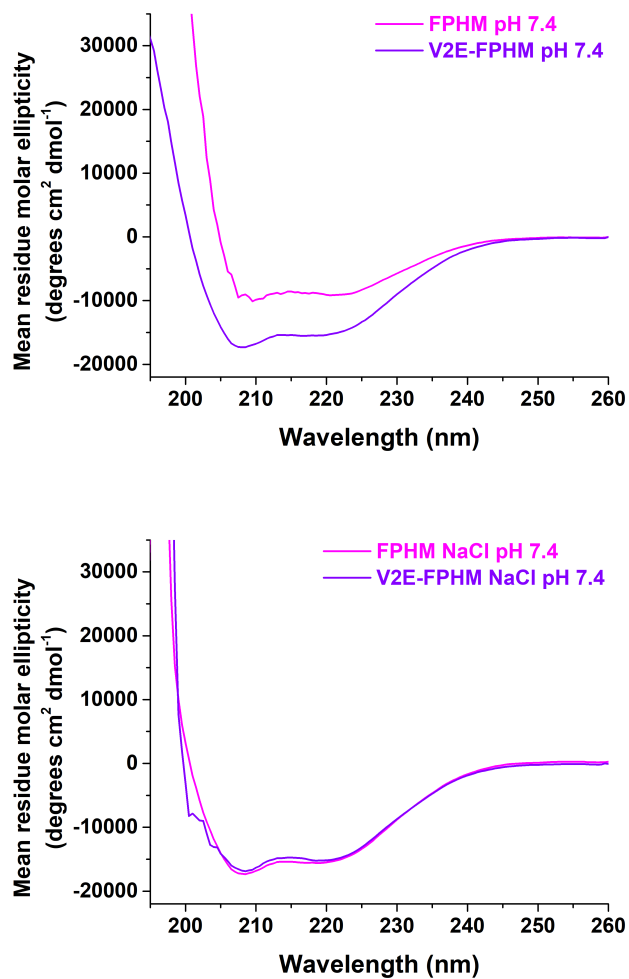


**Figure 7.9** SEC traces of V2E-FPHM at ambient temperature for 7-day dialyzed sample in 0.2% SDS/ 150 mM NaCl/ 20 mM CH<sub>3</sub>COONa at pH 4.0 (top), or 0.2% SDS/ 150 mM NaCl/ 10 mM Tris at pH 7.4 (bottom).



**Figure 7.10** CD spectra of the FPHM proteins at ambient temperature for 7-day dialyzed sample. Sample contained 10  $\mu$ M protein in either 0.2% SDS/ 10 mM Tris at pH 7.4 with and without 150 mM NaCl; or 0.2% SDS/ 20 mM CH<sub>3</sub>COONa at pH 4.0 with and without 150 mM NaCl.

The CD and SEC in Figs. 7.8, 7.9 and 7.10 were obtained in a buffer containing 0.2% SDS. As we couldn't obtain a proper SEC for the WT-FPHM in 0.2% SDS (due to the increase in viscosity upon concentration), we considered testing these proteins in 1% SDS. Fig. 7.11 shows the CD spectra obtained in 1% SDS, and they support  $\alpha$ -helical structures for both WT- and mutant FPHM.

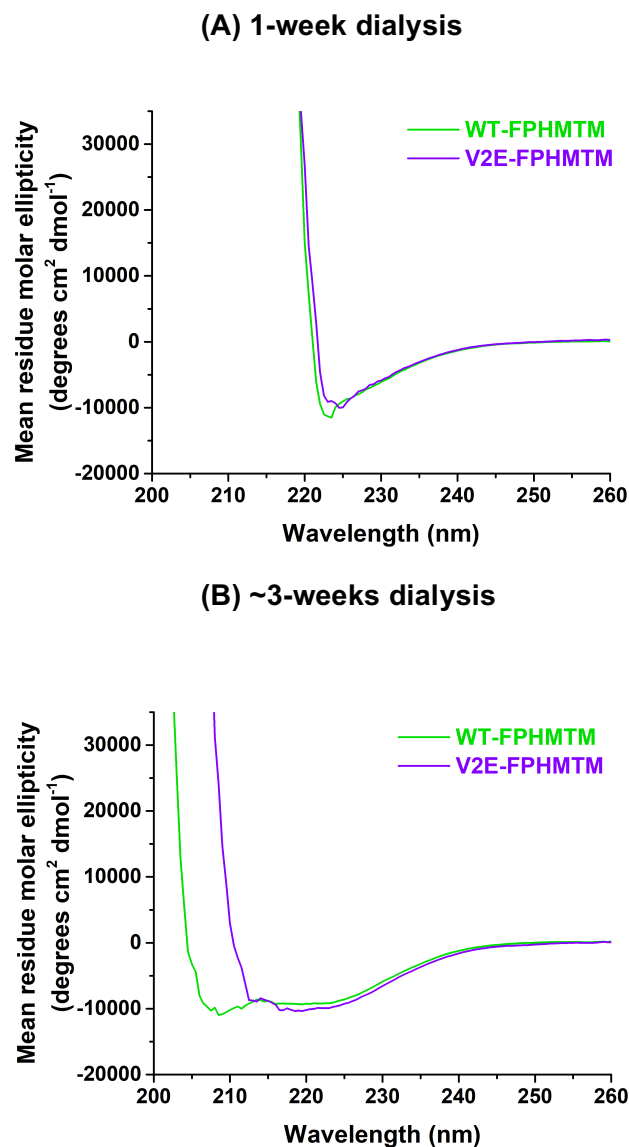


**Figure 7.11** CD spectra of the FPHM proteins at ambient temperature for 7-day dialyzed sample. Sample contained 10  $\mu$ M protein in 1% SDS/ 10 mM Tris at pH 7.4/ without (top) or with (bottom) 150 mM NaCl.

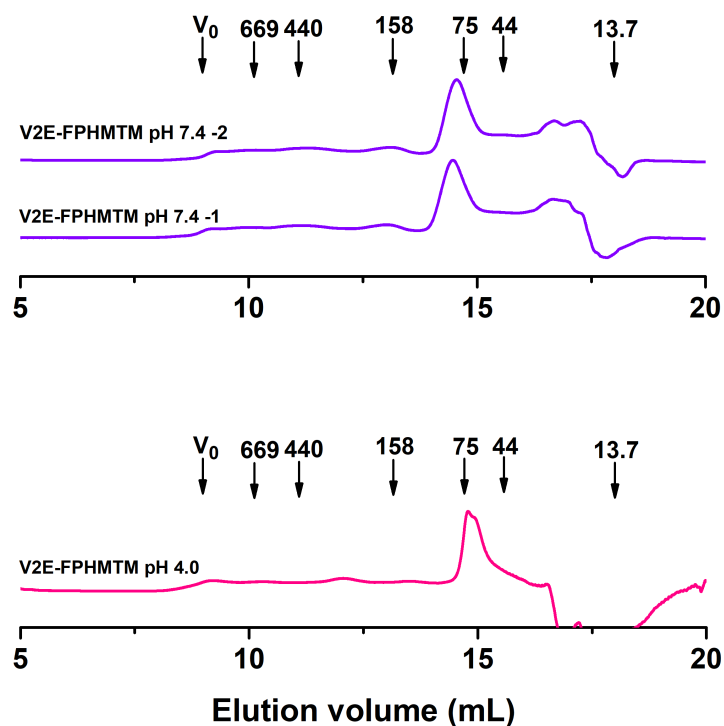
#### 7.2.4.2 WT- and V2E- FPHMTM

The CD spectra after 7-days dialysis in 0.2% SDS with NaCl do not support  $\alpha$ -helical

structures, and the FPHMTM proteins are probably aggregated (Fig. 7.12A). Similar spectra were obtained in the absence of NaCl as well. Some  $\alpha$ -helicity could be observed after ~3 weeks dialysis (Fig. 7.12B).

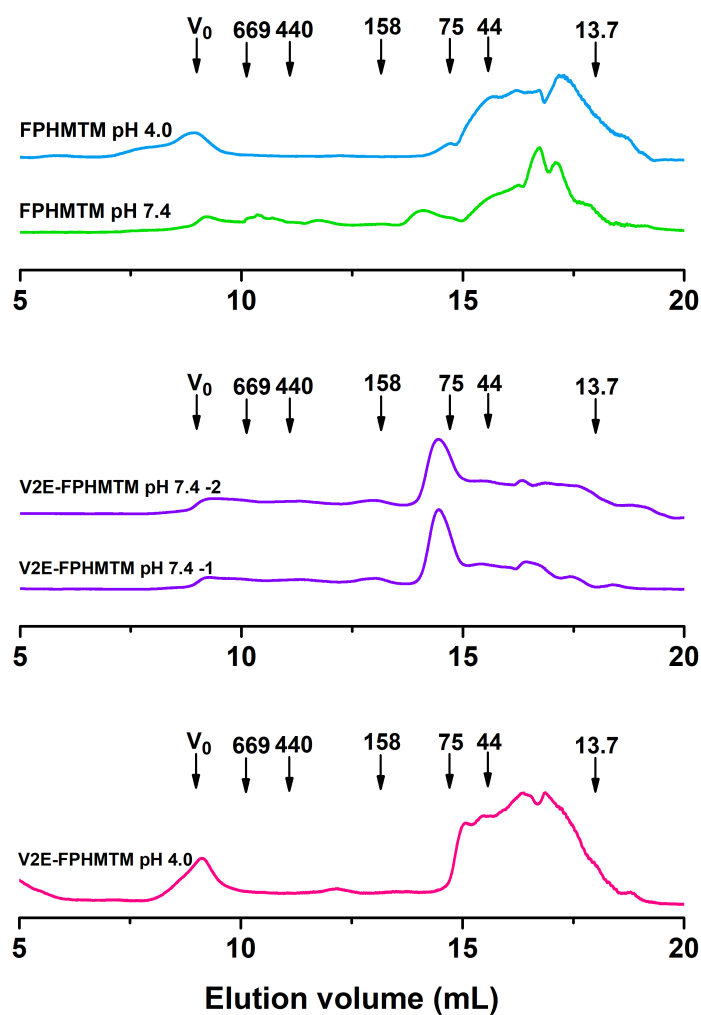


**Figure 7.12** CD spectra of the FPHMTM proteins at ambient temperature for (A) 1-week and (B) 3-weeks dialyzed samples. Sample contained 10  $\mu$ M protein in 0.2% SDS/ 150 mM NaCl/ 10 mM Tris at pH 7.4.



**Figure 7.13** SEC traces of V2E-FPHMTM at ambient temperature for 7-day dialyzed sample in 0.2% SDS/ 150 mM NaCl/ 10 mM Tris at pH 7.4 (purple), or 0.2% SDS/ 150 mM NaCl/ 20 mM  $\text{CH}_3\text{COONa}$  at pH 4.0 (pink).

Fig. 7.13 SEC traces show a peak at ~80 kDa for V2E-FPHMTM similar to V2E-FPHM, and is probably the protein trimer. WT-FPHMTM became highly viscous upon concentration and didn't give a peak in SEC for the 7-day dialyzed sample. A weak peak was observed at ~35 kDa after dialysis for ~4 weeks (Fig. 7.14), and may be the protein monomer with ~20 kDa protein and ~15 kDa detergent contribution.



**Figure 7.14** SEC traces of FPHMTM proteins at ambient temperature for 4-week dialyzed sample in 0.2% SDS/ 150 mM NaCl/ 10 mM Tris at pH 7.4 or 0.2% SDS/ 150 mM NaCl/ 20 mM CH<sub>3</sub>COONa at pH 4.0.

### 7.2.5 Future directions

Obtaining well folded proteins was a challenge with gp41 constructs specially for the FPHMTM proteins, and further experimentation is required. It would be interesting to perform both CD and SEC experiments in 1% SDS for both WT and mutant FPHM and FPHMTM

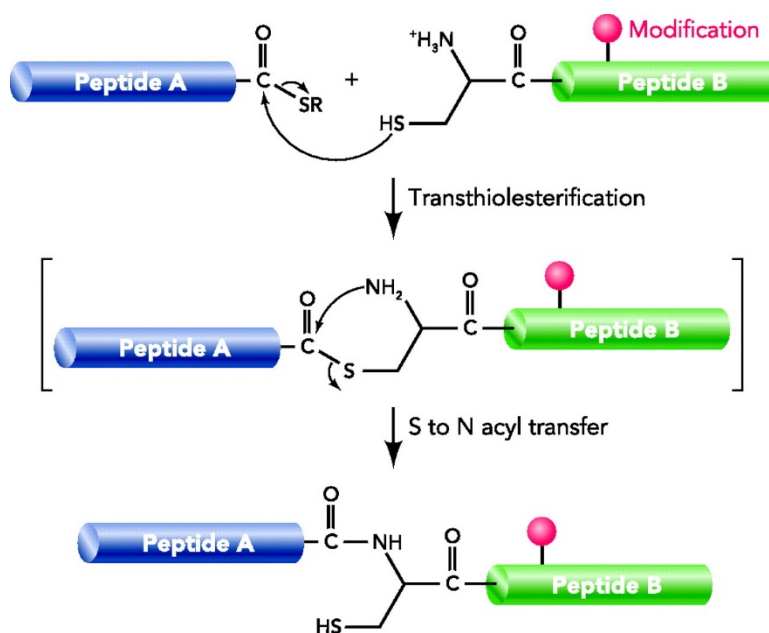


constructs. This project can be further developed by performing CD melting experiments and vesicle fusion experiments after obtaining well-folded proteins.

### 7.3 Native Chemical Ligation of FPHP

#### 7.3.1 Introduction

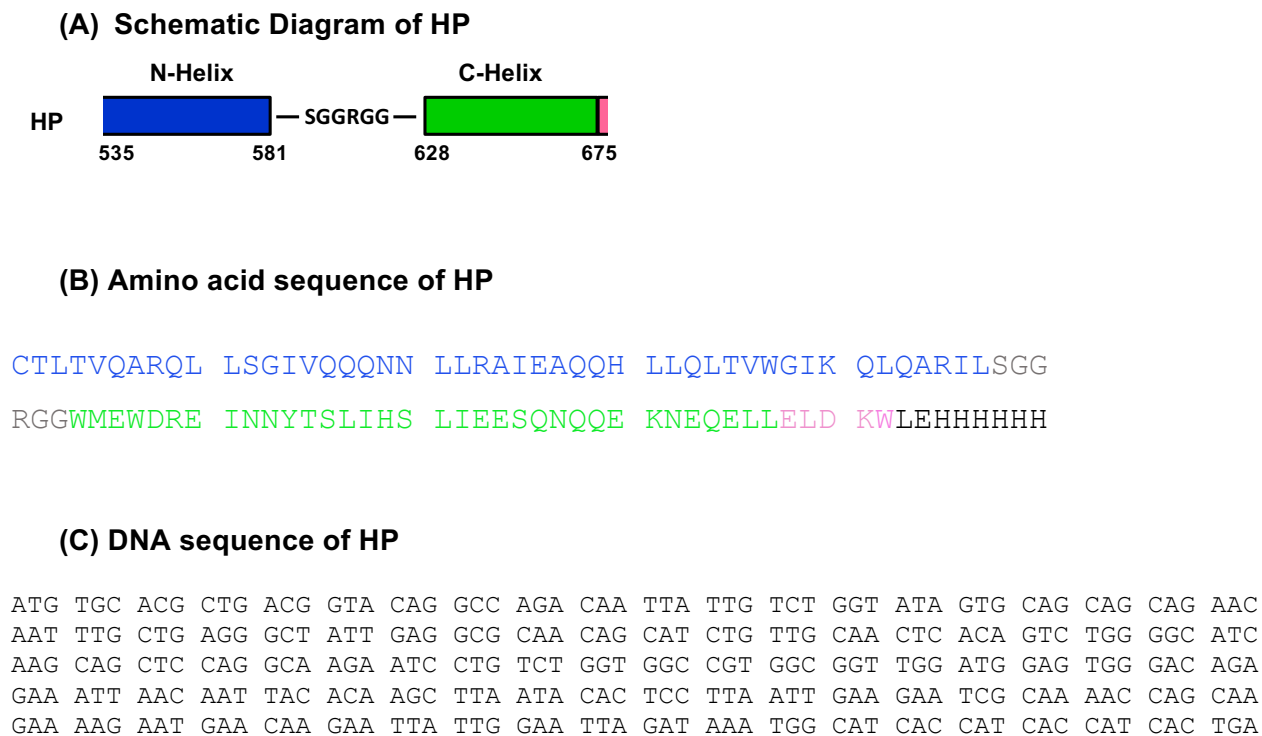
The native chemical ligation (NCL) is a method of constructing longer polypeptide chains by the assembly of two unprotected peptides. The proposed mechanism of NCL is given in Fig. 7.15.



**Figure 7.15** The mechanism of native chemical ligation

NCL of gp41 hairpin (HP), residues 535-581/SGGRGG/628-675 (Fig. 7.16), with the synthetically prepared FP has been previously studied by Dr. Matthew Nethercott in the Weliky group. In this previous study, the FP was isotopically labelled with <sup>13</sup>C and <sup>15</sup>N and the HP was

unlabeled. The present ligation reaction was performed to combine labelled HP (expression in labelled minimal media) and synthetically labelled FP. The resultant, differently labelled, FPHP construct can then be used for future solid state NMR experiments. First, the effectiveness of NCL reaction was tested with labelled FP and unlabeled HP.

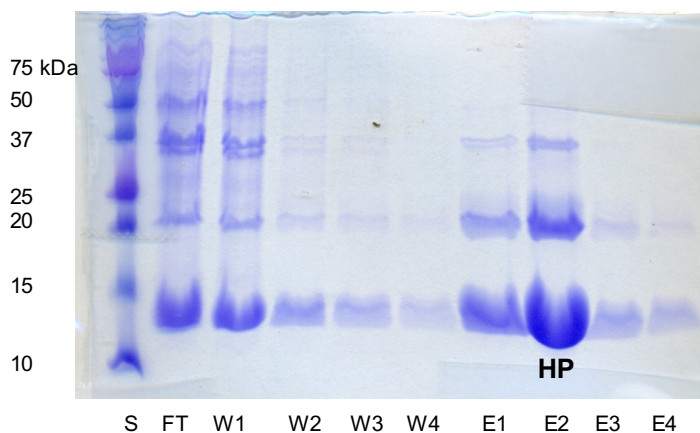


**Figure 7.16** (A) Schematic diagram, (B) amino acid sequence, and (C) DNA sequence of gp41 hairpin (HP).

### 7.3.2 Methods, Preliminary Results and Discussion

The pet24a(+) plasmid containing HP DNA sequence was purchased from GenScript, Piscataway, NJ. The plasmid was transformed into *E. coli* BL21(DE3) strain, followed by preparation of HP glycerol stocks, which were then stored at -80 °C. The HP was expressed and

purified using the same procedure as other gp41 constructs in section 7.2.3, except the refolding step. The refolding was not required as unfolded protein is used for the NCL. The HP yield was 13 mg/ml.



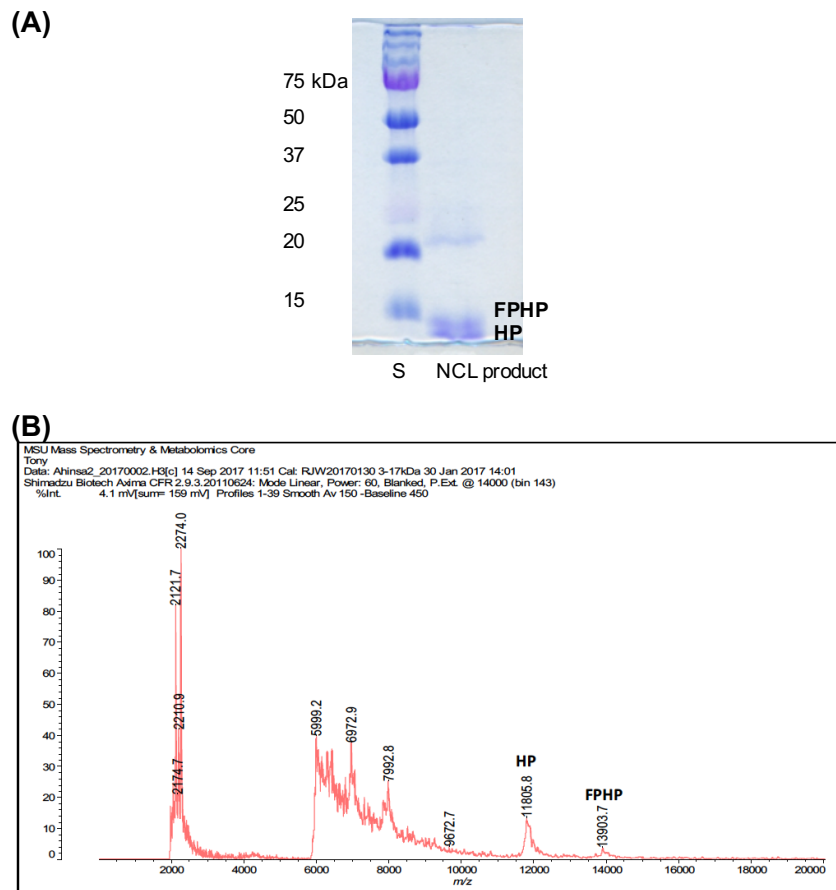
**Figure 7.17** The SDS gel of HP (MW = 11.8 kDa) after the affinity purification.

The eluent containing HP after the affinity purification was dialyzed against water (acidified with few drops of HCl) overnight, until HP is precipitated as a white powder. The precipitated HP was separated as a pellet by centrifugation and subsequent decanting the supernatant. The pellet was washed with water several times. HP was dissolved in the ligation buffer (8M GnCl, 2.5 M Imidazole, 1 M Na<sub>2</sub>HPO<sub>4</sub> and 1 M NaH<sub>2</sub>PO<sub>4</sub> at pH 6) by shaking the mixture in the labquake overnight to prepare an ~5 mg/ml HP solution.

The FP was a kind gift from Robert Wolfe in the Weliky group, and it has been synthesized using the solid phase peptide synthesis. The NCL procedure used here was developed by Robert Wolfe for the NCL of another gp41 construct, FPHM, which contain the complete MPER domain. For NCL, 600 µl of the ligation buffer was shaken in a lab quake for 1 hr at ambient temperature,

with the FP (5 molar eq to HP), 4-Mercaptophenylacetic acid (250 molar eq to HP) and 400  $\mu$ l of 1 M Tris(2-carboxyethyl)phosphine (in ligation buffer). The mixture was then combined with 200  $\mu$ l of 5 mg/ml HP in ligation buffer. The mixture was shaken for 2 days in the labquake at ambient temperature, and the product was analyzed with SDS-PAGE and MALDI-TOF mass spectrometry (Fig. 7.18).

To run the SDS-PAGE the NCL product was dialyzed against water to remove GnCl and the precipitate was washed with water several times. The precipitate was then solubilized in the solubilization buffer (8 M urea, 0.5 % SDS, and 0.8 % SRC in PBS at pH 7.4), and the SDS-PAGE was run. The unreacted HP and the product FPHP bands were very close in the SDS-gel, and was difficult to differentiate (Fig. 7.18A). The precipitated NCL product was then dissolved in formic acid to analyze with MALDI-MS. Fig. 7.18B shows the MALDI-MS of NCL product. HP peak can be clearly seen at 11.8 kDa, and the peak at 13.9 kDa was assigned to FPHP (11.8 kDa (HP) + 2.1 kDa (FP)).



**Figure 7.18** (A) The SDS-PAGE; and (B) the MALDI-MS spectrum of the NCL product.

This project is now conducted by Weliky group undergraduate Abdullah under the supervision of Rob Wolfe.

## REFERENCES

## REFERENCES

1. WHO/ HIV AIDS. <http://www.who.int/mediacentre/factsheets/fs360/en/> (accessed Dec 11, 2018).
2. Zhu, P., Liu, J., Bess, J., Chertova, E., Lifson, J. D., Grise, H., Ofek, G. A., Taylor, K. A., and Roux, K. H. (2006) Distribution and three-dimensional structure of AIDS virus envelope spikes, *Nature* **441**, 847-852.
3. Smith, D. L., Deng, Y., and Zhang, Z. (1997) Probing the non-covalent structure of proteins by amide hydrogen exchange and mass spectrometry, *Journal of mass spectrometry* **32**, 135-146.
4. Chojnacki, J., Staudt, T., Glass, B., Bingen, P., , Engelhardt, J., Anders, M., Schneider, J., Muller, B., Hell, S. W., and Krausslich, H. G. (2012) Maturation-dependent HIV-1 surface protein redistribution revealed by fluorescence nanoscopy., *Science* **338**, 524-528.
5. Blumenthal, R., Durell, S., and Viard, M. (2012) HIV entry and envelope glycoprotein-mediated fusion *Biol. Chem.* **287**, 40841-40849.
6. Julien, J. P., Cupo, A., Sok, D., Lyumkis, D., Deller, M. C., Klasse, P. J., Burton, D. R., Sanders, R. W., Moore, J. P., Ward, A. B., and Wilson, I. A. (2013) Crystal structure of a soluble cleaved HIV-1 envelope trimer, *Science* **342**, 1477-1483.
7. Lyumkis, D., Julien, J.-P., de Val, N., Cupo, A., Potter, C. S., Klasse, P.-J., Burton, D. R., Sanders, R. W., Moore, J. P., Carragher, B., Wilson, I. A., and Ward, A. B. (2013) Cryo-EM structure of a fully glycosylated soluble cleaved HIV-1 envelope trimer, *Science* **342**, 1484-1490.
8. Pancera, M., Zhou, T. Q., Druz, A., Georgiev, I. S., Soto, C., Gorman, J., Huang, J. H., Acharya, P., Chuang, G. Y., Ofek, G., Stewart-Jones, G. B. E., Stuckey, J., Bailer, R. T., Joyce, M. G., Louder, M. K., Tumba, N., Yang, Y. P., Zhang, B. S., Cohen, M. S., Haynes, B. F., Mascola, J. R., Morris, L., Munro, J. B., Blanchard, S. C., Mothes, W., Connors, M., and Kwong, P. D. (2014) Structure and immune recognition of trimeric pre-fusion HIV-1 Env., *Nature* **514**, 455-461.
9. Garces, F., Lee, J. H., de Val, N., Torrents de la Pena, A., Kong, L., Puchades, C., Hua, Y., Stanfield, R. L., Burton, D. R., Moore, J. P., Sanders, R. W., Ward, A. B., and Wilson, I. A. (2015) Affinity maturation of a potent family of HIV antibodies is primarily focused on accommodating or avoiding glycans, *Immunity* **43**, 1053-1063.
10. Lee, J. H., Ozorowski, G., and Ward, A. B. (2016) Cryo-EM structure of a native, fully glycosylated, cleaved HIV-1 envelope trimer, *Science* **351**, 1043-1048.
11. Ozorowski, G., Pallesen, J., de Val, N., Lyumkis, D., Cottrell, C. A., Torres, J. L., Copps, J., Stanfield, R. L., Cupo, A., Pugach, P., Moore, J. P., Wilson, I. A., and Ward, A. B.

- (2017) Open and closed structures reveal allostery and pliability in the HIV-1 envelope spike, *Nature* 547, 360-363.
12. Caffrey, M., Cai, M., Kaufman, J., Stahl, S. J., Wingfield, P. T., Covell, D. G., Gronenborn, A. M., and Clore, G. M. (1998) Three dimensional solution structure of the 44 kDa ectodomain of SIV gp41. , *EMBO J.* 17, 4572-4584.
  13. Zhang, W., Qi, J., Shi, Y., Li, Q., Gao, F., Sun, Y., Lu, X., Lu, Q., Vavricka, C. J., and Liu, D. (2010) Crystal structure of the swine-origin A (H1N1)-2009 influenza A virus hemagglutinin (HA) reveals similar antigenicity to that of the 1918 pandemic virus, *Protein & cell* 1, 459-467.
  14. Buzon, V., Natrajan, G., Schibli, D., Campelo, F., Kozlov, M. M., and Weissenhorn, W. (2010) Crystal structure of HIV-1 gp41 including both fusion peptide and membrane proximal external regions, *PLoS Pathog* 6, e1000880.
  15. Lev, N., Fridmann-Sirkis, Y., Blank, L., Bitler, A., Epand, R. F., Epand, R. M., and Shai, Y. (2009) Conformational stability and membrane interaction of the full-length ectodomain of HIV-1 gp41: Implication for mode of action, *Biochemistry* 48, 3166-3175.
  16. Sackett, K., Nethercott, M. J., Epand, R. F., Epand, R. M., Kindra, D. R., Shai, Y., and Weliky, D. P. (2010) Comparative analysis of membrane-associated fusion peptide secondary structure and lipid mixing function of HIV gp41 constructs that model the early prehairpin intermediate and final hairpin conformations, *J. Mol. Biol* 397, 301-315.
  17. Li, Y. L., and Tamm, L. K. (2007) Structure and plasticity of the human immunodeficiency virus gp41 fusion domain in lipid micelles and bilayers, *Biophysical Journal* 93, 876-885.
  18. CP, J., Kaufman JD, Stahl SJ, Viard M, Blumenthal R, Wingfield PT, and A., B. (2005) Structure and dynamics of micelle-associated human immunodeficiency virus gp41 fusion domain., *Biochemistry* 44, 16167-16180.
  19. Gabrys, C. M., and Weliky, D. P. (2007) Chemical shift assignment and structural plasticity of a HIV fusion peptide derivative in dodecylphosphocholine micelles, *Biochimica Et Biophysica Acta- Biomembranes* 1768, 3225-3234.
  20. Schmick, S. D., and Weliky, D. P. (2010) Major Antiparallel and Minor Parallel beta Sheet Populations Detected in the Membrane-Associated Human Immunodeficiency Virus Fusion Peptide, *biochemistry* 49, 10623-10635.
  21. Qiang, W., Sun, Y., and Weliky, D. P. (2009) A strong correlation between fusogenicity and membrane insertion depth of the HIV fusion peptide, *Proceedings of the National Academy of Sciences of the United States of America* 106, 15314-15319.
  22. Lihui Jia, Shuang Liang, Kelly Sackett, Li Xie, Ujjayini Ghosh, and Weliky, D. P. (2015) REDOR solid-state NMR as a probe of the membrane locations of membrane-associated peptides and proteins, *Journal of Magnetic Resonance* 253, 154-165.



23. Kelly Sackett, M. J. N., Zheng, Z., and Weliky, D. P. (2014) Solid-State NMR Spectroscopy of the HIV gp41 Membrane Fusion Protein Supports Intermolecular Antiparallel  $\beta$  Sheet Fusion Peptide Structure in the Final Six- Helix Bundle State, *Journal of molecular biology* 426, 1077-1094.
24. Sun, Z. Y. J., Oh, K. J., Kim M. Y., Yu, J., Brusic, V., Song, L. K., Qiao, Z. S., Wang, J. H., Wagner, G., and Reinherz, E. L. (2008) HIV-1 broadly neutralizing antibody extracts its epitope from a kinked gp41 ectodomain region on the viral membrane, *Immunity* 28, 52-63.
25. Shi, W. X., Bohon, J., Han, D. P., Habte, H., Qin, Y. L., Cho, M. W., and Chance, M. R. (2010) Structural characterization of HIV gp41 with the membrane-proximal external region., *J. Biol. Chem.* 285, 24290-24298.
26. Kim JH, Hartley TL, Curran AR, and DM., E. (2009) Molecular Dynamics Studies of the Transmembrane Domain of Gp41 from HIV-1., *Biochimica et biophysica acta.* 1788, 1804-1812.
27. Apellaniz, B., Rujas, E., Serrano, S., Morante, K., Tsumoto, K., Caaveiro, J. M. M., Angeles Jimenez, M., and Nieva, J. L. (2015) The atomic structure of the HIV-1 gp41 transmembrane domain and its connection to the immunogenic membrane-proximal external region, *J. Biol. Chem.* 290, 12999-13015.
28. Liang, S., Ratnayake, P., Keinath, C., Jia, L., Wolfe, R., Ranaweera, A., and Weliky, D. P. (2018) Efficient Fusion at Neutral pH by Human Immunodeficiency Virus gp41 Trimers containing the Fusion Peptide and Transmembrane Domain, *Biochemistry*.
29. Freed EO, Delwart EL, Buchschacher GL, and AT, P. (1992) A mutation in the human immunodeficiency virus type 1 transmembrane glycoprotein gp41 dominantly interferes with fusion and infectivity *Proc. Natl. Acad. Sci. U. S. A* 89, 70-74.
30. Martha D. Delahunty, Ina Rhee, Eric O. Freed, Juan S. Bonifacin (1996) Mutational analysis of the fusion peptide of the human immunodeficiency virus type 1: Identification of critical glycine residues *Virology* 218, 94-102.
31. Mobley PW, Waring AJ, Sherman MA, Gordon LM (1999) Membrane interactions of the synthetic N-terminal peptide of HIV-1 gp41 and its structural analogues *Biochim. Biophys. Acta*, 1418, 1-18.
32. Pereira FB, Goñi FM, Nieva JL (1995) Liposome destabilization induced by the HIV-1 fusion peptide: effect of a single amino acid substitution *FEBS letters* 362, 243-246.
33. Pereira FB, Goñi FM, Nieva JL (1997) Permeabilization and fusion of uncharged lipid vesicles induced by the HIV-1 fusion peptide adopting an extended conformation: dose and sequence effects, *Biophysical Journal* 73, 1977-1986.
34. Banerjee, K., and Weliky, D. P. (2014) Folded monomers and hexamers of the ectodomain of the HIV gp41 membrane fusion protein: potential roles in fusion and synergy between

- the fusion peptide, hairpin, and membrane-proximal external region, *Biochemistry* 53, 7184-7198.
35. Sackett K, Nethercott MJ, Shai Y, and David P, Weliky. (2009) Hairpin folding of HIV gp41 abrogates lipid mixing function at physiologic pH and inhibits lipid mixing by exposed gp41 constructs, *biochemistry* 48, 2714-2722.
  36. Kemin Tan, Jin-huan Liu, Jia-huai Wang, Steven Shen, and Lu, M. (1997) Atomic structure of a thermostable subdomain of HIV-1 gp41, *Proc. Natl. Acad. Sci. U. S. A* 94, 12303-12308.

## **Chapter 8**

### **Summary and Future Directions**

## 8.1 Summary

Influenza virus is a membrane-enveloped virus and is responsible for seasonal flu. The virus enters the host respiratory epithelial cell through endocytosis. The influenza viral entry into the host cell is mediated by the viral membrane protein hemagglutinin (HA), which forms an initial complex that contains three HA1 subunits in association with a helix bundle containing ectodomains of three HA2 subunits. The HA2 subunit catalyzes the fusion of viral and endosomal membranes, which is an initial step of viral infection. HA2 spans the viral membrane through an ~25-residue transmembrane domain (TM), and contains an ectodomain outside the virus that includes the ~25-residue fusion peptide (FP) and ~160-residue soluble ectodomain (SE), and an ~10-residue endodomain inside the virus. A structure of the SE without HA1 shows a trimeric hairpin structure, where both N- and C-termini are at one end opposite to the hairpin turn. Residues 38-105 from the SE hairpins form an interior parallel helical bundle, while 154-176 are strands within the grooves of the bundle exterior. This structure is likely the final HA2 state in fusion based on its hyperthermostability with the melting temperature >90 °C. The FP and TM are the only HA2 segments that are deeply inserted in the fused membrane. The FP forms a separate 1-12 helix/turn/14-22 helix hairpin structure with antiparallel and close packing of the two FP helices. CD spectra of the TM peptide in detergent and membrane correlate with the expected  $\alpha$  helical structure. The research discussed in this dissertation focus on better understanding the full-length HA2 structure and membrane fusion mechanism.

Chapter 3 describes structural and functional comparisons between WT-HA2 and its G1E and I173E mutants, where both mutants result in highly-impaired HA-mediated fusion. For all the studies discussed in this dissertation, there is only HA2 (without HA1), and the HA2 SE is always in its final hairpin state. The WT-HA2 still induces fusion of neutral and anionic vesicles at pH 5.0, which reflect conditions for influenza virus fusion. Vesicle fusion is greatly reduced for G1E-

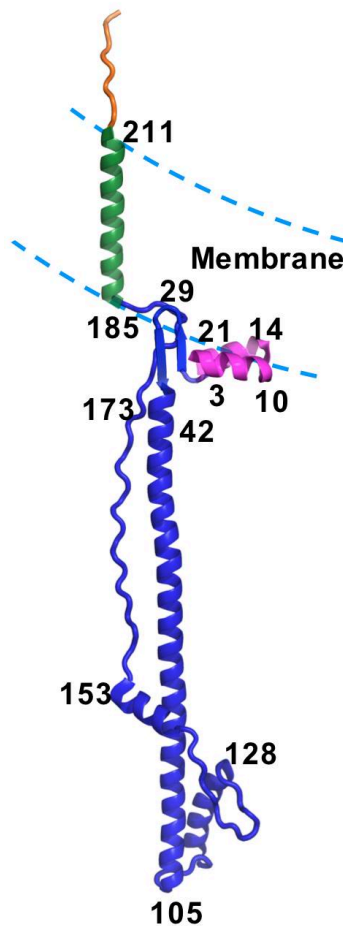
and I173E- HA2, which correlates with similarly reduced HA-mediated fusion. CD vs. temperature data for the WT- and mutant- HA2 and FHA2  $\equiv$  FP+SE constructs show dramatic losses in stability for the mutants, including a 40 °C reduction in melting temperature for I173E- vs. WT- FHA2. We propose that these observations reflect destabilization of the SE hairpins due to dissociation of the C-terminal strands from the N-terminal helical bundle. Such destabilization correlates with larger monomer fractions for mutant vs. WT proteins. SE destabilization by the G1E mutant is unexpected because the mutation is in the FP rather than SE, and we propose that the interhelical packing of the FP is disrupted for the G1E mutant, and results in FP binding to SE strands which then dissociate from the interior helical bundle. Although it is commonly proposed that free energy released by the large-scale HA2 structural rearrangement catalyzes HA-mediated fusion, the present study as well as earlier cell-cell fusion studies support an alternate mechanistic model. In this model, fusion is preceded by FP insertion in the target membrane and formation of the final SE hairpin, which apposes the virus and target membranes. Fusion then proceeds in a ~1 min timescale via stochastic motions of the lipids that are increased by interactions with the FP. Reduced fusion by the G1E and I173E mutants is due to loss of hairpin stability and therefore membrane apposition, and for G1E, also loss of FP contacts with the lipids.

Chapter 4 describes structural comparisons between WT-HA2 and its G1E and I173E mutants by hydrogen-deuterium exchange-mass spectrometry (HDX-MS). Fusion models often suggest a formation of a transmembrane FP/TM complex, and some studies report that this complex play a major role in fusion pore expansion. Our HDX-MS data do not support the existence of this complex in HA2, and show high aqueous exposure for the FP and very low aqueous exposure for the TM. This supports non-interacting FP and TM domains with membrane interfacial location of FP and membrane traversal by TM. The HDX data support a well-structured

and protected 22-38 segment. Although electron density of much of this segment is not reported in the HA2 final hairpin conformation, the 23-35 region is a well-structured  $\beta$ -sheet hairpin in the initial HA2/HA1 complex. Our data support the retention of this  $\beta$ -sheet hairpin in the final HA2 structure. The  $\beta$ -sheet may add a semi rigid connection between the FP and SE that aids maintaining the membrane apposition during fusion. I173E shows much greater HDX for the 22-69 and 150-191 regions compared to the WT, which support the I173E model proposed in Chapter 3, which is the dissociation of the C-terminal strand from the internal helical bundle. G1E data support two distinct SE states with either the WT trimer-of-hairpins structure or the I173E partially -dissociated and -unfolded structure. The latter state also has unfolding of the FP helical hairpin and FP binding to the SE, which is supported by the smaller HDX for G1E FP vs. WT. Both mutants exhibit highly-impaired fusion that may be due to increased distance between the apposed membranes and for G1E, FP binding to SE rather than target membrane.

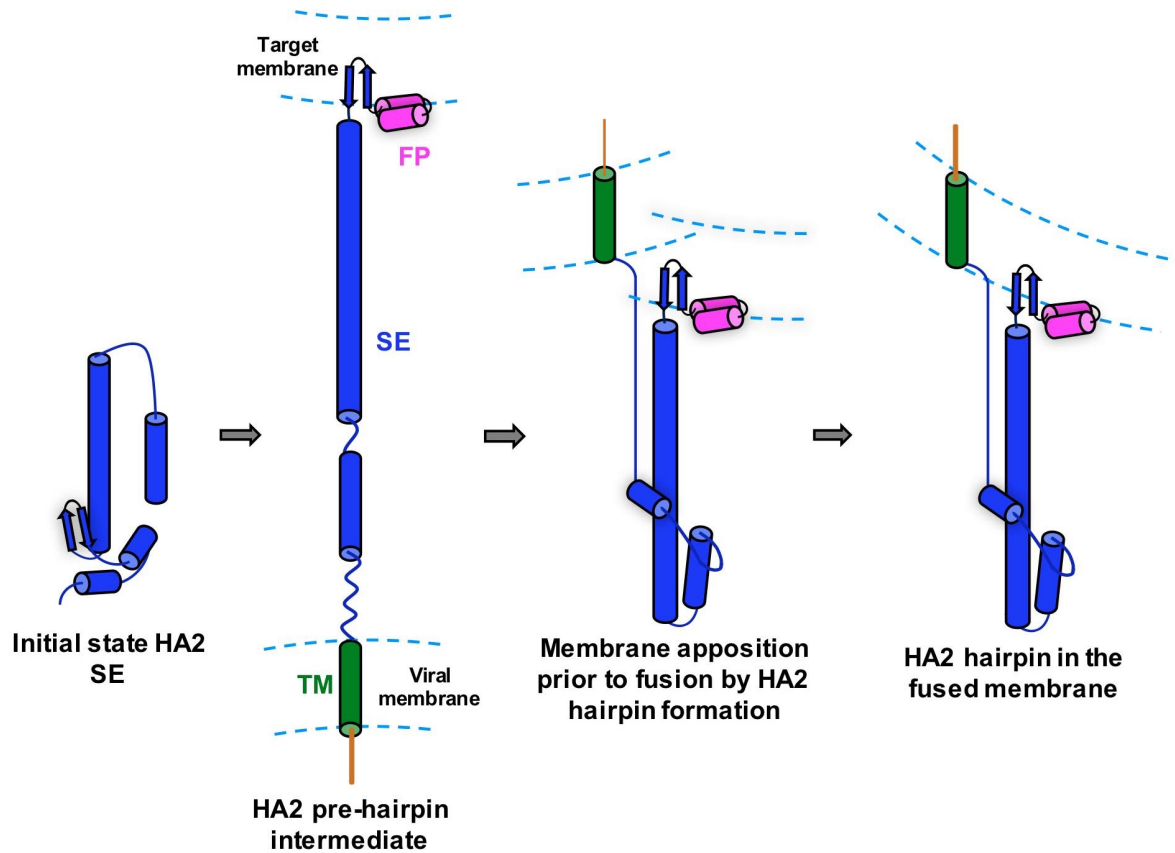
The present and previous data are integrated to obtain models for full-length HA2 structure (Fig. 8.1) and HA2-mediated membrane fusion (Fig. 8.2). The fusion model is characterized by: (1) FP binds to the target membrane during the HA2 structural transformation from the initial state to the final hairpin state; (2) The HA2 hairpin formation brings the two membranes in close apposition prior to fusion. The close apposition is maintained by the hyperthermostable trimeric SE hairpin structure and the 23-35 antiparallel  $\beta$  sheet that provides a semi-rigid connection between the FP and SE; (3) The close inter-membrane distance compensates for  $\sim 60$  kT (theoretically calculated) activation barrier between the initial two-membranes state and the hemifusion stalk intermediate. Subsequent hemifusion and pore formation steps have much lower calculated activation barriers ( $\sim 20$  kT), and occur with the SE remaining in the hairpin state.<sup>1, 2</sup> The FP-induced membrane perturbation may also contribute to lower these energy barriers<sup>3-5</sup>; and

(4) The TM segments facilitate pore expansion by shielding the lipid acyl chains from water during their transition from the hemifusion diaphragm separating the two bodies to the final continuous bilayer enclosing the fused body. The positive membrane curvature associated with the HA2 structure in Fig. 8.1 may also aid pore expansion, which has an associated decrease in negative membrane curvature.



**Figure 8.1** The structural model for full length HA2 in the fused membrane. HA2 exhibits trimer-of-hairpins structure for the SE, but only a monomer is shown for clarity. The model combines residue 1-22 closed FP helical hairpin (PDB 2KXA), 23-35 antiparallel  $\beta$  sheet of the initial HA2/HA1 complex (PDB 2HMG), 36-178 hairpin SE structure (PDB 1QU1), 186-211 continuous

$\alpha$  helix for TM, and 212-221 unstructured endodomain. The 179-185 extended structure is found in one monomer in PDB 1QU1. The small positive membrane curvature accommodates the FP and TM.



**Figure 8.2** A mechanistic model for the HA2-mediated membrane fusion. The FP binds to the target membrane during the HA2 structural transformation from the initial state to the final hairpin state. The final trimer-of-hairpins SE structure holds the two membranes in close apposition prior to fusion. The clustering of hairpin trimers and therefore FP domains in the target membrane perturbs this membrane and lowers the activation barrier to achieve hemifusion and pore formation. For this model, the SE hairpin is a fusogenic state rather than a post-fusion state.



## 8.2 Future directions

In Chapter 3, the WT and mutant HA2's were analyzed with simple biophysical instruments like CD, SEC and fluorimeter, which are typically easy to handle, but the data provide valuable insight into the HA2 mediated fusion. Application of similar experiments to the gp41 constructs discussed in Chapter 7 (WT- and V2E- FPHM and FPHMTM) will aid better understanding the gp41 mediated membrane fusion.

The HDX experiments discussed in Chapter 4 were performed in detergent rich media. However, it would be more relevant to perform these experiments in lipid membranes as the fusion occurs in membranes rather than in detergent micelles. The use of phospholipid bilayer nanodiscs will allow us to study the protein in a more native-like environment. However, we have to face few challenges when using nanodiscs in HDX-MS set up. First, the nanodiscs need to be disassembled after the quench step and this can be accomplished by adding a cholate to the quenched reaction mixture. Second, the nanodiscs lipids, which may interfere with the MS analysis, need to be removed, and  $\text{ZrO}_2$  beads are commonly used to capture the lipids. The HDX-MS set up of protein inserted lipid nanodiscs is not yet compatible with the online column digestion, instead immobilized pepsin beads are used. We need to get rid of both the  $\text{ZrO}_2$  and immobilized pepsin beads before the LC-MS analysis by quick filtration.

In our HDX experiments, the protein +  $\text{D}_2\text{O}$  buffer solution was incubated at pH 7 and 25 °C. It would be interesting to see the HDX data in pH ~5.5 and 37 °C, which are closer to the conditions at which the fusion takes place. One practical challenge with these conditions will be the tendency of protein aggregation at low pH.

Our HDX-MS data provide new insight into the HA2 structure and function. It will be a meaningful study to apply this method to WT and mutant HIV gp41 constructs.

## REFERENCES

## REFERENCES

1. Cohen FS, and GB, M. (2004) The energetics of membrane fusion from binding, through hemifusion, pore formation, and pore enlargement, *J. Membr. Biol* 199, 1-14.
2. Boonstra, S., Blijleven, J. S., Roos, W. H., Onck, P. R., van der Giessen, E., and van Oijen, A. M. (2018) Hemagglutinin-mediated membrane fusion: A biophysical perspective, *Annual review of biophysics* 47, 153-173.
3. Ratnayake, P. U., Ekanayaka, E. P., Komanduru, S. S., and Weliky, D. P. (2016) Full-length trimeric influenza virus hemagglutinin II membrane fusion protein and shorter constructs lacking the fusion peptide or transmembrane domain: Hyperthermostability of the full-length protein and the soluble ectodomain and fusion peptide make significant contributions to fusion of membrane vesicles, *Protein expression and purification* 117, 6-16.
4. Ghosh, U., Xie, L., Jia, L., Liang, S., and Weliky, D. P. (2015) Closed and semiclosed interhelical structures in membrane vs closed and open structures in detergent for the influenza virus hemagglutinin fusion peptide and correlation of hydrophobic surface area with fusion catalysis, *Journal of the American Chemical Society* 137, 7548-7551.
5. Lai, A. L., and Freed, J. H. (2015) The interaction between influenza HA fusion peptide and transmembrane domain affects membrane structure, *Biophysical journal* 109, 2523-2536.