THESIS
4
2009

LIBRARY
Michigan State
University

This is to certify that the
dissertation entitled

STRUCTURAL STUDIES OF THE INFLUENZA AND HIV VIRAL FUSION PROTEINS AND BACTERIAL INCLUSION BODIES

presented by

JAIME LYN CURTIS-FISK

has been accepted towards fulfillment
of the requirements for the

Doctoral degree in Chemistry

David P. Weliky
Major Professor's Signature

March 27, 2009
Date

MSU is an Affirmative Action/Equal Opportunity Employer

PLACE IN RETURN BOX to remove this checkout from your record.
TO AVOID FINES return on or before date due.
MAY BE RECALLED with earlier due date if requested.

DATE DUE	DATE DUE	DATE DUE

5/08 K /Proj/Acc&Pres/CIRC/DateDue.indd

STRUCTURAL S
PROT

STRUCTURAL STUDIES OF THE INFLUENZA AND HIV VIRAL FUSION PROTEINS AND BACTERIAL INCLUSION BODIES

By

Jaime Lyn Curtis-Fisk

A DISSERTATION

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

DOCTOR OF PHILOSOPHY

Chemistry

2009

ABSTRACT

STRUCTURAL STUDIES OF THE INFLUENZA AND HIV VIRAL FUSION PROTEINS AND BACTERIAL INCLUSION BODIES

by

Jaime Lyn Curtis-Fisk

The infection of a cell by enveloped viruses, such as influenza and HIV, begins with the fusion of the viral and cellular membranes, which is mediated by proteins in the membrane of the virus referred to as fusion proteins. For influenza this is the HA2 protein and for HIV the gp41 protein. The structure of several constructs of these proteins have been studied, focusing on the ectodomain regions. Expression in E. Coli has been successfully optimized to produce large quantities of fusion protein with amino acid specific ^{13}C and ^{15}N labeling. Analysis of the protein conformation at specific residues was done by solid state NMR. Detection of specific ^{13}C in the membrane associated and inclusion body protein was achieved by use of a filtering REDOR pulse sequence. The measured ^{13}C chemical shifts are correlated with local conformation. Initial studies focused on determining the conformation of specific residues in the membrane associated ectodomain of the HA2 protein.

After method development regarding purified membrane associated protein, the focus shifted to the structural study of protein while still within the bacterial cell. Recombinant protein expression is typically plagued by the production of insoluble aggregates of the expressed protein, known as inclusion bodies. Little it known about the structure of proteins in this form, but theories range from an amyloid-like beta sheet structure to aggregates of fully folded, functional protein. Previous structural studies of inclusion bodies have mostly used IR, which only gives overall structural information.

We were able to develop

information can be obtai

possible fraction of the c

ching and the REDOX

ained for multiple p

ase proteins. Compari

ained on the membr

ssist bodies maintain

ormation. This is a simple

ve active form, while st

ssist body producing p

We were able to develop a solid-state NMR method in which site-specific structural information can be obtained of the inclusion bodies both in the whole-cell and in the insoluble fraction of the cell without dehydrating the samples. Using amino acid specific labeling and the REDOR pulse sequence, detection of specific carbonyl signal was achieved for multiple positions throughout the constructs of both the HA2 and gp41 fusion proteins. Comparing the inclusion body results to structural studies we previously conducted on the membrane associated protein, our results indicate that the FHA2 inclusion bodies maintain their native structure and there is no indication of beta sheet formation. This is a simple, efficient method to study the structure of inclusion bodies in their native form, while still in the cell, and that should be easily transferable to any other inclusion body producing protein.

Dedicated

Dedicated to Julia Mae, my motivation for everything I do.

The most valuable

ssible if you surround

ve made this long journ

Through the past

y family. I never would

avorites Erica, Matt, S

me of my favorite me

ks Erica I truly enjoy

ge great results and

e so long and still

Sunship than we rea

ce far for you as p

e not. All of the t

hers seemed appre

Professure. Charles,

Charles of the lab

ks for your succe

Gail Working w

ve the opportun

Cheryl countless

Ask not what you ne

ACKNOWLEDGEMENTS

The most valuable lesson that I have learned in graduate school is that anything is possible if you surround yourself with the right people. My friends, family, and advisor have made this long journey not only successful, but enjoyable as well.

Through the past few years I have spent more time with my labmates than with my family. I never would have imagined that I would become such close friends with my co-workers. Erica, Matt, Scott, Charles, Ryan, and Jesse, our Friday trips to Thai Kitchen are one of my favorite memories of graduate school, and one of the things I will miss the most. Erica, I truly enjoyed teaching you my project and I am sure that you will continue to get great results and take the projects to new levels. Scott, sitting across from someone for so long and still wanting to talk them means that we have developed a greater friendship than we realized. Hopefully planning Erica's wedding with her will be as much fun for you as planning mine. Matt, you are one of the most generous people I have ever met. All of the times that you went out of your way to help in lab may not have always seemed appreciated at the time, but I can only hope to have co-workers like you in the future. Charles, the group would not be the same without you. Your insight into the dynamics of the lab and comic relief allowed me to survive many tough days. Ryan, I consider your success in research to be one of the most gratifying aspects of graduate school. Working with you helped me to realize what I want my career to be and I hope to have the opportunity to mentor many more students like you. Yan and Qiang, you answered countless questions about the NMR that probably seemed like dumb questions to you, but you never made me feel that way. You both have a level of patience that I am

sons of Jesse, as an "

e through the tough time

As an advisor, I

pected. All advisors w

well. You allowed me

ny advisors would not

ce. You make your stud

ages.

I am so grateful fo

ve always understood w

something important to m

graduate school if it w

ving up. To my pare

ments you share in this

The person that I

ost amazing days of our

s impossible to expre

ars, if I would have m

e this journey that we

envious of. Jesse, as an "auxiliary member" of the Weliky group your friendship helped me through the tough times, and I so appreciative of that.

As an advisor, Dr. Weliky has gone above and beyond what I would have expected. All advisors want their research to go well, but you also want your students to do well. You allowed me to develop my skills as a teacher as well as a researcher, and many advisors would not give their students this opportunity, let alone be so supportive of it. You make your students goals a priority, and I hope that is something that never changes.

I am so grateful for the support I have received from my family. They may not have always understood what graduate school was about, but my family knew that it was something important to me, and that made it important to them. I would not have made it to graduate school if it weren't for the values of hard work and dedication that I learned growing up. To my parents, grandparents, my brother, and many aunts, uncles, and cousins, you share in this success.

The person that I owe the most to is my husband. We have experienced both the most amazing days of our lives and the most challenging of times over the last five years. It is impossible to express how much your love and support has meant to me. I don't know if I would have made it through those tough times without you by my side. Sharing May this journey that we have began continue through many more happy years.

LIST OF TABLES.....

LIST OF FIGURES.....

LIST OF ABBREVIATIO

Structural Analysis
Inclusion Bodies ...
I. Influenza V
a. Antivira
b. Hemagg
c. HA2 Str
II. Human Imm
a. Antivira
b. Glycopro
c. gp41 Str
III. Membrane P
a. Expressi
b. Inclusion
c. Structura
IV. Protein NMR
a. Liquid S
b. Solid Sta
c. RFDOR

Expression and Puri
I. Introduction
a. Protein E
b. Growth
c. Protein P
d. Membra
II. Materials an
a. Materia
b. Amino
c. Gel Elec
d. Culture
e. Isotopic
f. Purifica
g. Purifica
h. Circular
i. Lipid M
j. Membran
III. Fusion Pr
a. Optimiza
b. Expressi
c. Expressi

TABLE OF CONTENTS

LIST OF TABLES..... xi

LIST OF FIGURES..... xii

LIST OF ABBREVIATIONS..... xx

1. Structural Analysis of the Influenza and HIV Viral Fusion Proteins and Inclusion Bodies..... 1
   - I. Influenza Virus..... 1
     - a. Antiviral Drugs..... 3
     - b. Hemagglutinin Protein..... 5
     - c. HA2 Structural Studies..... 7
   - II. Human Immunodeficiency Virus..... 11
     - a. Antiviral Drugs..... 12
     - b. Glycoprotein..... 13
     - c. gp41 Structural Studies..... 14
   - III. Membrane Proteins..... 16
     - a. Expression..... 16
     - b. Inclusion Bodies..... 17
     - c. Structural Studies..... 19
   - IV. Protein NMR..... 20
     - a. Liquid State NMR..... 21
     - b. Solid State NMR..... 21
     - c. REDOR..... 24

2. Expression and Purification of Natively Folded Fusion Proteins..... 38
   - I. Introduction..... 38
     - a. Protein Expression..... 39
     - b. Growth Conditions..... 42
     - c. Protein Purification..... 45
     - d. Membrane Reconstitution..... 48
   - II. Materials and Methods..... 52
     - a. Materials..... 52
     - b. Amino Acid and DNA Sequences..... 52
     - c. Gel Electrophoresis..... 57
     - d. Culture Growth..... 58
     - e. Isotopic Labeling..... 59
     - f. Purification of Native FHA2 and Fgp41..... 59
     - g. Purification of Native MHA2..... 60
     - h. Circular Dichroism Spectroscopy..... 61
     - i. Lipid Mixing Assay..... 61
     - j. Membrane Reconstitution of FHA2..... 63
   - III. Fusion Protein Expression..... 65
     - a. Optimization of FHA2 expression in E. coli..... 65
     - b. Expression of MHA2..... 70
     - c. Expression of gp41 and Hgp41..... 70

IV. Native Pur
a. Purifica
b. Purifica
c. Protec
d. Egg4
V. Structural a
a. Circula
b. Lipid n
c. Recons
VI. Conclusion
VII. References

Solid State NMR A
I. Introduction
a. The FH
b. Solid St
II. Materials and
a. Materials
b. Isotopica
c. Protein P
d. Membra
e. Solid-St
III. Structural A
a. Single S
b. Multiple
c. Fusion P
d. Missing
e. Soluble
f. Hinge re
g. C-Term
IV. Effect of Ter
Protein
V. Cholesterol
VI. Conclusions
VII. References

Expanding the Poten
I. Introduction
a. DNA Pu
b. Polymer
c. Site Dire
d. DPN: D
e. Trans
f. DNA Se
II. Materials an
a. Materials
b. Molecula
III. Investigation
IV. Investigation
V. Conclusions a

IV. Native Purification of Fusion Proteins..... 71
a. Purification of native FHA2 from soluble cell lystate..... 71
b. Purification of MHA2..... 73
c. Proteolytic cleavage of the maltose binding protein..... 74
d. Fgp41 purification..... 76
V. Structural and Functional Assays..... 77
a. Circular dichroism of FHA2..... 77
b. Lipid mixing of FHA2 and Fgp41..... 77
c. Reconstitution of FHA2..... 82
VI. Conclusions and Future Work..... 84
VII. References..... 88

3. Solid State NMR Analysis of Membrane Associated Fusion Protein..... 92
I. Introduction..... 92
a. The FHA2 Protein..... 92
b. Solid State NMR Analysis..... 94
II. Materials and Methods..... 97
a. Materials..... 97
b. Isotopically Labeled Protein Expression..... 97
c. Protein Purification..... 98
d. Membrane Reconstitution..... 99
e. Solid-State NMR Experiments..... 101
III. Structural Analysis..... 104
a. Single Site Analysis..... 105
b. Multiple site structural analysis..... 107
c. Fusion Peptide..... 109
d. Missing Link..... 112
e. Soluble ectodomain..... 113
f. Hinge region..... 114
g. C-Terminal region..... 116
IV. Effect of Temperature on Analysis of the Membrane Associated Protein..... 119
V. Cholesterol Containing Membranes and pH 7.4 Samples..... 120
VI. Conclusions and Future Work..... 123
VII. References..... 126

4. Expanding the Potential of REDOR Through Site Directed Mutagenesis..... 128
I. Introduction..... 128
a. DNA Purification..... 129
b. Polymerase Chain Reaction..... 130
c. Site Directed Mutagenesis..... 134
d. DPN1 Digestion..... 137
e. Transformation..... 137
f. DNA Sequencing..... 139
II. Materials and Methods..... 142
a. Materials..... 142
b. Molecular Biology..... 142
III. Investigation of the Fusion Peptide Kink..... 145
IV. Investigation of the Missing Link Region..... 148
V. Conclusions and Future Work..... 153

5 Expression and Re
I Introductio
a. Inclusi
b. Protein
c. Additive
II. Materials a
a. Materia
b. Culture
c. Isotopic
d. Solub
Inclusion
e. Circular
f. Membra
g. Solid St
III. Electron Mi
IV. Expression
V. Solubilizatio
Bodies
VI. Membrane R
VII. Solid State N
VIII. Conclusions
IX. References

6 Solid State NMR St
I. Introduction
a. Targets
b. Previous
II. Materials an
a. Express
b. Solid St
III. FHA2 Inclu
IV. Hpa2: Inclu
V. Conclusions
VI. References

7 Implementing an O
Chemistry Lecture C
I. Introduction
a. Chemist
b. Online
c. Online
d. Online
II. Study Des
III. General Che
IV. Organic Che
a. Quiz 1
b. Quiz 2

5. Expression and Refolding of Inclusion Body Proteins.....156
I. Introduction.....156
a. Inclusion Body Purification.....157
b. Protein Refolding.....158
c. Additives to Refolding.....160
II. Materials and Methods.....161
a. Materials.....161
b. Culture Growth.....161
c. Isotopically Labeled Protein Production.....162
d. Solubilization, Purificatiton, and Refolding of FHA2 from Inclusion Bodies.....163
e. Circular Dichroism Spectroscopy.....164
f. Membrane Reconstitution.....164
g. Solid State NMR Spectroscopy.....165
III. Electron Microscopy.....167
IV. Expression.....173
V. Solubilization, Purification, and Refolding of FHA2 in Inclusion Bodies.....175
VI. Membrane Reconstitution.....180
VII. Solid State NMR Spectroscopy.....181
VIII. Conclusions and Future Work.....184
IX. References.....187

6. Solid State NMR Structural Analysis of Bacterial Inclusion Bodies.....190
I. Introduction.....190
a. Targets of Study.....190
b. Previous Structural Studies.....192
II. Materials and Methods.....199
a. Expression and Isotopic Labeling of Inclusion Body Proteins.....199
b. Solid State NMR Analysis.....200
III. FHA2 Inclusion Body Analysis.....202
IV. Hgp41 Inclusion Body Analysis.....209
V. Conclusions and Future Work.....215
VI. References.....224

7. Implementing an Online Homework Program in a Large Organic Chemistry Lecture Course.....227
I. Introduction.....227
a. Chemistry Homework.....228
b. Online Component to Chemistry Courses.....228
c. Online Homework.....230
d. Online Homework Administration.....240
II. Study Design.....242
III. General Chemistry Students.....245
IV. Organic Chemistry.....247
a. Quiz 1.....247
b. Quiz 2.....250

c. Exam
d. Quiz 3
e. Quiz 4
f. Exam 2
g. Quiz 5
h. Final E
i. Course
j. Survey
V. Conclusions
VI. Future Work
VII. References

Appendix 1: FHA2

Appendix 2: HGP4

c. Exam 1 253
d. Quiz 3 262
e. Quiz 4 264
f. Exam 2 266
g. Quiz 5 281
h. Final Exam 283
i. Course Analysis 286
j. Surveys 290
V. Conclusions 297
VI. Future Work 305
VII. References 353

Appendix 1: FHA2 NMR Data Files 307

Appendix 2: HGP41 NMR Data Files 309

Figure 5

Figure 5-1 Secondary Str...

Figure 5-2 ^{13}CO chemical ...

Figure 5-3 ^{13}CO chemical ...
-11 Cvs -50 °C.....

Figure 5-4 ^{13}CO chemical ...
-10 °C.....

Figure 6

Figure 6-1 Chemical shift ...
...ular bodies of the H...
...mplex reconstituted ...

Figure 6-2 Chemical shift ...
...ular bodies of the H...

Figure 7

Figure 7-1 General chem...

Figure 7-2 OHW comp...

Figure 7-3 Recommend...

Figure 7-4 OHW compl...

Figure 7-5 Recommend...

Figure 7-6 OHW comp...

Figure 7-7 Recommend...

Figure 7-8 Oxidative cl...

Figure 7-9 OHW attem...

LIST OF TABLES

Chapter 3

Table 3-1. Secondary Structure CO Chemical Shifts..... 107

Table 3-2. ^{13}CO chemical shifts PC/PG-associated FHA2 at –10 °C..... 118

Table 3-3. ^{13}CO chemical shifts PC/PG-associated FHA2 at –10 °C vs. –50 °C..... 120

Table 3-4. ^{13}CO chemical shifts PC/PG- or PC/PG/cholesterol associated FHA2 at –10 °C..... 122

Chapter 6

Table 6-1. Chemical shift information from the bacterial cells containing inclusion bodies of the FHA2 protein and previous analysis of the membrane reconstituted native protein..... 211

Table 6-2. Chemical shift information from the bacterial cells containing inclusion bodies of the Hgp41 protein..... 213

Chapter 7

Table 7-1. General chemistry OHW performance.....246

Table 7-2. OHW completion for quiz 1.....247

Table 7-3. Recommended OHW completion for quiz 1.....249

Table 7-4. OHW completion for quiz 2.....251

Table 7-5. Recommended OHW completion for quiz 2 252

Table 7-6. OHW completion for exam 1 254

Table 7-7. Recommended OHW completion for exam 1 256

Table 7-8. Oxidative cleavage exam performance vs. OHW completion 260

Table 7-9. OHW attempts vs. exam performance 261

Page 10. OHW compl
Page 11. Recommend
Page 12. OHW compl
Page 13. Recommend
Page 14. OHW compl
Page 15. Recommend
Page 16. Change in ex
Page 17. Stereochemis
Page 18. Alkyne OHW
Page 19. Synthesis OH
Page 20. OHW compl
Page 21. Recommend
Page 22. Total OHW c
Page 23. Recommend
Page 24. OHW in gene
Page 25. Tutorial part

Table 7-10. OHW completion for quiz 3.....263

Table 7-11. Recommended OHW completion for quiz 3 264

Table 7-12. OHW completion for quiz 4.....265

Table 7-13. Recommended OHW completion for quiz 4 266

Table 7-14. OHW completion for exam 2 267

Table 7-15. Recommended OHW completion for exam 2 268

Table 7-16. Change in exam performance 269

Table 7-17. Stereochemistry OHW questions 272

Table 7-18. Alkyne OHW completion vs. exam 2 performance 276

Table 7-19. Synthesis OHW completion vs. exam 2 performance 279

Table 7-20. OHW completion for quiz 5 281

Table 7-21. Recommended OHW completion for quiz 5 282

Table 7-22. Total OHW completion vs. final exam performance 284

Table 7-23. Recommended OHW completion vs. final exam performance 285

Table 7-24. OHW in general vs. organic chemistry 287

Table 7-25. Tutorial participation 289

Chapter 1

Figure 1-1. Influenza virus

Figure 1-2. Membrane fu

Figure 1-3. The HA2 solu

Figure 1-4. The HA2 fusi

Figure 1-5. The HA2 fusi

Figure 1-6. The gp41 hairp

Figure 1-7. Structural and

Figure 1-8. The REDOR p

Figure 1-9. REDOR active

Chapter 2

Figure 2-1. Membrane re

Figure 2-2. Amino acid

Figure 2-3. DNA seque

Figure 2-4. Amino aci

Figure 2-5. DNA seq

Figure 2-6. Amino a

Figure 2-7. DNA se

Figure 2-8. Amino a

Figure 2-9. DNA seq

Figure 2-10. Amino acid

LIST OF FIGURES

Chapter 1

Figure 1-1. Influenza viral lifecycle.....2

Figure 1-2. Membrane fusion by HA2..... 6

Figure 1-3. The HA2 soluble ectodomain crystal structure..... 8

Figure 1-4. The HA2 fusion peptide NMR structure at pH 5..... 9

Figure 1-5. The HA2 fusion peptide NMR structure at pH 7.4..... 9

Figure 1-6. The gp41 hairpin region crystal structure..... 15

Figure 1-7. Structural analysis by REDOR..... 26

Figure 1-8. The REDOR pulse sequence..... 28

Figure 1-9. REDOR active positions..... 29

Chapter 2

Figure 2-1. Membrane reconstitution..... 49

Figure 2-2. Amino acid sequence of FHA2..... 52

Figure 2-3. DNA sequence of FHA2..... 53

Figure 2-4. Amino acid sequence of MHA.....53

Figure 2-5. DNA sequence of MHA2.....54

Figure 2-6. Amino acid sequence of gp41..... 55

Figure 2-7. DNA sequence of gp41..... 55

Figure 2-8. Amino acid sequence of Fgp41.....56

Figure 2-9. DNA sequence of Fgp41..... 56

Figure 2-10. Amino acid sequence of Hgp41..... 56

Figure 2-11. DNA sequ
Figure 2-12. Cell growth
Figure 2-13. Effect of ox
Figure 2-14. Drop in pH
Figure 2-15. Cell growth.
Figure 2-16. Cell growth
Figure 2-17. MHA2 Expr
Figure 2-18. Purification
Figure 2-19. Purification
Figure 2-20. Proteolysis
Figure 2-21. Use of dena
Figure 2-22. Purification
Figure 2-23. Circular di
Figure 2-24. Fluorescen
Figure 2-25. Lipid mixi
Figure 2-26. Lipid mixi
Figure 2-27. Electron m
[illegible]
Figure 2-28. Reconstr
[illegible]
Figure 3-1. Residues o
Figure 3-2. Solid-state
Figure 3-3. Multi-posi
Figure 3-4. REDOR an

Figure 2-11. DNA sequence of Hgp41 57

Figure 2-12. Cell growth in minimal media, LB, and enriched LB 66

Figure 2-13. Effect of oxygenation on cell growth 67

Figure 2-14. Drop in pH with cell growth 68

Figure 2-15. Cell growth as an effect of initial pH 68

Figure 2-16. Cell growth as an effect of carbon source and concentration 69

Figure 2-17. MHA2 Expression 70

Figure 2-18. Purification of FHA2 using nickel and cobalt resins 73

Figure 2-19. Purification of MHA2 73

Figure 2-20. Proteolysis activity of MHA2 purification 74

Figure 2-21. Use of denaturing conditions to increase proteolytic activity 75

Figure 2-22. Purification of Fgp41 76

Figure 2-23. Circular dichroism of the FHA2 protein 77

Figure 2-24. Fluorescent lipid mixing assays 78

Figure 2-25. Lipid mixing activity of FHA2 80

Figure 2-26. Lipid mixing activity of Fgp41 81

Figure 2-27. Electron microscopy of vesicles with and without the addition of FHA2 82

Figure 2-28. Reconstitution of FHA2 83

Chapter 3

Figure 3-1. Residues of membrane associate FHA2 studied by SS-NMR 104

Figure 3-2. Solid-state NMR spectra of PC/PG-associated FHA2 samples 106

Figure 3-3. Multi-position REDOR analysis 109

Figure 3-4. REDOR analysis of the fusion peptide region of FHA2 126

igure 3-5. REDOR anal

igure 3-6. REDOR anal
ctodomain region of H1

igure 3-7. REDOR anal

igure 3-8. REDOR anal

igure 3-9. Temperature

igure 3-10. Membrane e

hapter 4

igure 4-1. Polymerase ch

igure 4-2. Site directed

igure 4-3. DNA sequenc

igure 4-4. Protein seque
nfluenza virus.....

igure 4-5. FHA2 DNA

igure 4-6. DNA resultin

igure 4-7. DNA resultin
S.....

igure 4-8. DNA result

igure 4-9. DNA result

igure 4-10. DNA resul

igure 4-11. DNA resu

hapter 5

igure 5-1. Electron mi

igure 5-2. Electron mi

igure 5-3. Electron mi

Figure 3-5. REDOR analysis of the missing link region of FHA2..... 128

Figure 3-6. REDOR analysis of the N-terminal portion of the soluble ectodomain region of FHA2..... 129

Figure 3-7. REDOR analysis of the hinge region of FHA2..... 131

Figure 3-8. REDOR analysis of the C-Terminal region of FHA2..... 133

Figure 3-9. Temperature effects of the REDOR experiment..... 136

Figure 3-10. Membrane effect of REDOR experiment..... 138

Chapter 4

Figure 4-1. Polymerase chain reaction..... 133

Figure 4-2. Site directed mutagenesis reaction..... 136

Figure 4-3. DNA sequencing..... 141

Figure 4-4. Protein sequence of the FHA2 construct from the X31 strain of the influenza virus..... 143

Figure 4-5. FHA2 DNA sequence in the plasmid pET24(+) with HA(1-185)..... 144

Figure 4-6. DNA resulting form the mutagenesis reaction N12A..... 146

Figure 4-7. DNA resulting from mutagenesis reaction to create the unique position EN..... 147

Figure 4-8. DNA resulting from the mutagenesis reaction I18V..... 149

Figure 4-9. DNA resulting from the mutagenesis reaction N28A..... 150

Figure 4-10. DNA resulting from mutagenesis reaction W21Y..... 151

Figure 4-11. DNA resulting from mutagenesis reaction G156A..... 152

Chapter 5

Figure 5-1. Electron microscopy of un-induced bacterial cells..... 167

Figure 5-2. Electron microscopy of un-induced bacterial cells..... 168

Figure 5-3. Electron microscopy of un-induced bacterial cells..... 168

Figure 5-4. Electron micr
ment.....
Figure 5-5. Electron micr
ment.....
Figure 5-6. Electron micr
ment.....
Figure 5-7. Electron micr
ment.....
Figure 5-8. Electron micr
ment.....
Figure 5-9. Electron micr
ment.....
Figure 5-10. Electron mic
ment.....
Figure 5-11. Electron mic
ment.....
Figure 5-12. Electron mic
ment.....
Figure 5-13. Electron mi
ment.....
Figure 5-14. Electron mi
ment.....
Figure 5-15. Gel electro
Figure 5-16. Gel electro
Figure 5-17. Gel electro
nections.....
Figure 5-18. Circular di
Figure 5-19. Membrane
Figure 5-20. ^{13}C solid-sta

Figure 5-4. Electron microscopy of bacterial cells induced to produce Hgp41 protein..... 168

Figure 5-5. Electron microscopy of bacterial cells induced to produce Hgp41 protein..... 169

Figure 5-6. Electron microscopy of bacterial cells induced to produce Hgp41 protein..... 169

Figure 5-7. Electron microscopy of bacterial cells induced to produce Hgp41 protein..... 169

Figure 5-8. Electron microscopy of bacterial cells induced to produce Hgp41 protein..... 170

Figure 5-9. Electron microscopy of bacterial cells induced to produce Hgp41 protein..... 170

Figure 5-10. Electron microscopy of bacterial cells induced to produce Hgp41 protein..... 170

Figure 5-11. Electron microscopy of bacterial cells induced to produce FHA2 protein..... 171

Figure 5-12. Electron microscopy of bacterial cells induced to produce FHA2 protein..... 171

Figure 5-13. Electron microscopy of bacterial cells induced to produce FHA2 protein..... 171

Figure 5-14. Electron microscopy of bacterial cells induced to produce FHA2 protein..... 172

Figure 5-15. Gel electrophoresis of FHA2 expression cell fractions.....174

Figure 5-16. Gel electrophoresis of Hgp41 expression cell fractions..... 175

Figure 5-17. Gel electrophoresis of the soluble lysis and inclusion body purifications..... 177

Figure 5-18. Circular dichroism structural analysis of FHA2..... 179

Figure 5-19. Membrane reconstitution of refolded FHA2.....181

Figure 5-20. ^{13}C solid-state NMR spectra of membrane-associated FHA2..... 183

Chapter 6
Figure 6-1. REDOR anal
ble fraction of the
Figure 6-2. REDOR anal
ural cells.....
Figure 6-3. REDOR anal
ble insoluble cell fra
Figure 6-4. REDOR anal
ural cells.....
Figure 6-5. REDOR anal
ble insoluble cell fra
Chapter 7
Figure 7-1. General chem
Figure 7-2. OHW comple
Figure 7-3. Recommende
Figure 7-4. OHW comple
Figure 7-5. Recommende
Figure 7-6. OHW comp
Figure 7-7. Recommen
Figure 7-8. Oxidative
Figure 7-9. Oxidative
Figure 7-10. Oxidati
Figure 7-11. OHW a
Figure 7-12. OHW c
Figure 7-13. Recomm
Figure 7-14. OHW com
Figure 7-15. Recommen

Chapter 6
Figure 6-1. REDOR analysis of FHA2 inclusion bodies studied in the insoluble fraction of the bacterial cells..... 203

Figure 6-2. REDOR analysis of FHA2 inclusion bodies while still within bacterial cells..... 204

Figure 6-3. REDOR analysis of FHA2 inclusion bodies for positions in both the insoluble cell fraction and still within bacterial cells..... 205

Figure 6-4. REDOR analysis of Hgp41 inclusion bodies while still within bacterial cells..... 211

Figure 6-5. REDOR analysis of Hgp41 inclusion bodies for positions in both the insoluble cell fraction and still within bacterial cells..... 212

Chapter 7
Figure 7-1. General chemistry OHW performance 246

Figure 7-2. OHW completion for quiz 1 248

Figure 7-3. Recommended OHW completion for quiz 1..... 250

Figure 7-4. OHW completion for quiz 2 251

Figure 7-5. Recommended OHW completion for quiz 2..... 253

Figure 7-6. OHW completion for exam 1 255

Figure 7-7. Recommended OHW completion for exam 1..... 256

Figure 7-8. Oxidative cleavage OHW questions..... 259

Figure 7-9. Oxidative cleavage exam 1 questions..... 260

Figure 7-10. Oxidative cleavage exam performance vs. OHW completion 261

Figure 7-11. OHW attempts vs. exam performance 262

Figure 7-12. OHW completion for quiz 3.....263

Figure 7-13. Recommended OHW completion for quiz 3..... 264

Figure 7-14. OHW completion for quiz 4..... 265

Figure 7-15. Recommended OHW completion for quiz 4 266

Figure 7-16. OHW completion for exam 2..... 267

Figure 7-17. Recommended OHW completion for exam 2 268

Figure 7-18. Change in exam performance..... 270

Figure 7-19. Stereochemistry OHW questions 271

Figure 7-20. Stereochemistry exam 2 questions 272

Figure 7-21. Stereochemistry OHW completion vs. exam performance 273

Figure 7-22. Alkyne OHW questions..... 274

Figure 7-23. Alkyne exam 2 questions..... 275

Figure 7-24. Alkyne OHW completion vs. exam 2 performance 276

Figure 7-25. Synthesis OHW questions..... 278

Figure 7-26. Synthesis exam 2 questions..... 279

Figure 7-27. Synthesis OHW completion vs. exam 2 performance..... 280

Figure 7-28. OHW completion for quiz 5..... 282

Figure 7-29. Recommended OHW completion for quiz 5 283

Figure 7-30. Total OHW completion vs. final exam performance..... 284

Figure 7-31. Recommended OHW completion vs. final exam performance 285

Figure 7-32. OHW in general vs. organic chemistry 287

Figure 7-33. Tutorial participation.....290

Figure 7-34. Survey question results regarding the ability to score higher in chemistry courses by completing OHW..... 291

Figure 7-35. Results of the survey question regarding the students likelihood to complete OHW questions in the future 292

Figure 7-36. Survey question results regarding the students comfort level with completing chemistry problems on a computer..... 293

igure 7-37. Results of s
ility to learn material H

igure 7-38. Survey resu
uizzes felt for quizzes

Figure 7-37. Results of the survey questions regarding the student's ability to learn material by working through OHW problems 294

Figure 7-38. Survey results regarding the questions about how prepared students felt for quizzes and exams after completing OHW 296

A. Absorbance

B. n-Octyl-β-thiog

B. n-Octyl-β-glucop

B. Octyl pentaethyle

C. Circular dichroism

C. Critical Micelle C

D. Deuton

D. Di-O-tetradecyl-p

D. Di-O-tetradecyl-p

E. Fusion peptide

E. Green fluorescen

G. Glycoprotein

G. Glutathione S-tra

H. Hemagglutinin

H. 4-(2-hydroxy

H. Histidine

H. Human Immuno

K. Kelvin

L. Immobilized m

N. Isopropyl-β-D-

N. Normal

N. Nuclear Bernard bra

KEY TO SYMBOLS AND ABBREVIATIONS

A, Absorbance

BTOG, n-Octyl-β-thioglucopyranoside

BOG, n-Octyl-β-glucopyranoside

C8E5, Octyl pentaethylene glycol ether

CD, Circular dichroism

CMC, Critical Micelle Concentration

Da, Dalton

DTPC, Di-O-tetradecylphosphatidylcholine

DTPG, Di-O-tetradecylphosphatidylglycerol

FP, Fusion peptide

GFP, Green fluorescent protein

GP, Glycoprotein

GST, Glutathione S-transferase

HA, Hemagglutinin

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

His, Histidine

HIV, Human Immunodeficiency Virus

Hz, Hertz

IMAC, Immobilized metal affinity chromatography

IPTG, Isopropyl-β-D-1-thiogalactopyranoside

IR, Infrared

LB, Luria-Bertani broth

LM: Lipid mixture thr

LUV: Large unilamellar

MM: Minimal media mi

REDOR: rotational ech

MAS: Magic angle spin

MBP: Maltose binding p

MWCO: molecular weig

MES: 2-N-morpholino

NA: Neuraminidase

NBD-PE: N-(7-nitro-1,

NMR: Nuclear magneti

OPM: Online homew

PAGE: polyacrylamide

PBS: Phosphate buffer

PCR: Polymerase cha

PE: Phosphatidylethan

PEG: Polyethylene gl

PI: Phosphatidylinosi

POPC: 1-palmitoyl-2

POPE: 1-palmitoyl-2

POPG: 1-palmitoyl-2

PPM: Parts per millio

PSM: N-lissamine

LM3, Lipid mixture three

LUV, Large unilamellar vesicles

M9, Minimal media mix

REDOR, rotational echo double resonance

MAS, Magic angle spinning

MBP, Maltose binding protein

MWCO, molecular weight cutoff

MES, 2-(N-morpholino)ethanesulfonic acid

NA, Neuramidase

NBD-PE, N-(7-nitro-1,2,3-benzoxadiazol-4-yl)-phosphatidylethanolamine

NMR, Nuclear magnetic resonance

OHW, Online homework

PAGE, polyacrylamide gel electrophoresis

PBS, Phosphate buffered saline

PCR, Polymerase chain reaction

PE, Phosphatidylethanolamine

PEG, Polyethylene glycol

PI, Phosphatidylinositol

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

POPE, 1-palmitoyl-2-oleoyl-*sn*-glycero-3-phosphoethanolamine

POPS, 1-palmitoyl-2-oleoyl-*sn*-glycero-3-[phospho-L-serine]

PPM, Parts per million

Rh-PE, N-(lissamine rhodamine B sulfonyl)-phosphatidylethanolamine

RPM. Rotations per mi

SDM. Site directed mut

SDS. Sodium dodecyl s

SS. Solid state

RPM, Rotations per minute

SDM, Site directed mutagenesis

SDS, Sodium dodecyl sulfate

SS, Solid state

tructural Analysis of

The influenza v

reats to human health

ar with many more a

ster the effects if in

g go. In order to dev

ver it is that we are

ective drugs has

veloping a better

reat it. This unde

olved in viral in

wards this goal.

luenza Virus

The influen

0 deaths (1).

with limited ar

a new strain

ecently being

ecting clear th

ing the disease

Chapter 1:

Structural Analysis of the Influenza and HIV Viral Fusion Proteins and Inclusion Bodies

The influenza virus and the human immunodeficiency virus (HIV) are significant threats to human health. Thousands of people die each year as a result of these infections, and with many more acquiring the diseases, the battle is on to prevent the infections or to lessen the effects if infection has occurred. Progress has been made, but we still have far to go. In order to develop defenses against these viruses, we need to better understand what it is that we are fighting against. The method of high throughput testing to discover effective drugs has not provided the "magic-bullet" cure. The pursuit now involves developing a better understanding of the target, and the rational design of drugs to combat it. This understanding relies on a better knowledge of the structure of proteins involved in viral infections, and the work in this thesis aims to provide information towards this goal.

I. Influenza Virus

The influenza virus strikes countless individuals each year, and results in over 20,000 deaths (*1*). Vaccines have been developed in an attempt to prevent the infection, but with limited amounts of success. The virus quickly mutates and each season is struck by a new strain for which the previous years vaccine is of no use. Advances are constantly being made in the ability of scientists to predict the likely strains, but it is becoming clear that a small molecule based treatment would be extremely useful in treating the disease, and saving human life. In order to design these small molecules, a

ter understanding

ocess is required.

The virus initi

ns within the cell

nformation change i

gers the process of fu

sion process the con

ir details of this fu

fection occurs, and

develop antiviral

derstanding of the

ure 1-1. Influ

ndrosis. The

viral fusio

membranes

ecting has occ

better understanding of the infection process, and the proteins that take part in this process, is required.

The virus initially enters the cell through endocytosis, as shown in figure 1-1. Once within the cellular endosome, the pH is lowered from 7.4 to 5, causing a conformation change in the viral fusion proteins. These proteins are then activated and begin the process of fusing the viral membrane to the endosomal membrane. Through this fusion process the contents of the virus enter the cell and infection has occurred. Clearly the details of this fusion process are critical in understanding the process by which infection occurs, and therefore developing methods to prevent infection. Current efforts to develop antiviral therapies against the influenza virus have indicated that a better understanding of the process through which the infection occurs would aid in this pursuit.

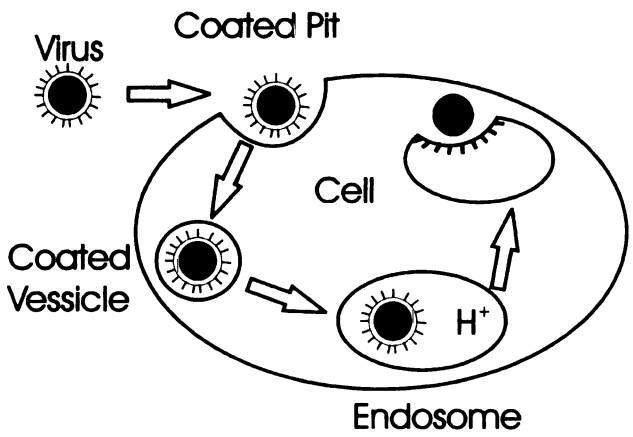


Figure 1-1. Influenza Viral Lifecycle. The virus initially enters the host cell through endocytosis. The pH in the endosome is then lowered, activating a conformation change in the viral fusion proteins and initiating membrane fusion. Fusion of the host and viral cell membranes allows for the entry of the viral capsid contents into the host cell, and infection has occurred.

4 Antiviral Drugs

Currently the n

vaccinations against th

ffective at preventin

determining which str

ho are vaccinated s

imaging effects of t

nfections. These peop

usceptible to the viru

not developing an in

ases in recent deca

hat target the neu

emagglutinin (HA).

Neuraminidase Inh

Influenza virus. Th

ucleocapsid and catal

ecessary binding the

ebris from the c

ells and is therefo

Several dru

ategies for neu

A. Antiviral Drugs

Currently the most common method for preventing influenza infections is annual vaccinations against the strains that are predicted to be the most common, which is effective at preventing up to 90% of infections (*2*). Unfortunately, the process of determining which strains to protect against is not an exact science, and many people who are vaccinated still acquire the infection. The most susceptible group to the damaging effects of the virus are the elderly and those with pre-existing medical conditions. These people typically have weakened immune systems that make them more susceptible to the virus, and also make this approach less effective as they rely on the body developing an immune response to the attenuated viruses in the vaccine. Several studies in recent decades have focused on the affects of small molecule anti-viral drugs that target the neuramidase protein (NA), M2, or the fusion catalyzing protein hemagglutinin (HA).

1. Neuramidase Inhibitors. The protein neuramidase (NA) is a surface protein of the influenza virus. The other surface protein, HA, attaches to receptors of the host cell membrane and catalyzes fusion between the host and viral membranes. NA then cleaves the bond linking the terminal neuraminic acid from the HA receptor, allowing the virus to release from the cellular surface. This process is essential for successful infection to occur, and is therefore an attractive target for drug development.

Several drugs that have proven effective at inhibiting the activity of NA are analogues for neuramidic acid, the target of the enzymes activity. Zanamivir and

oseltamivir are two such molecules that interact differently with the NA active site, but both act as competitive inhibitors (*3*).

2. M2 Inhibitors. The membrane protein M2 is an ion channel that is activated by a drop in pH. The H^+ ions flow into the virus, initiate the uncoating of the virus, and the release of the viral contents into the cytoplasm of the host cell. Two drugs known to inhibit the M2 protein are amantadine and rimantadine. These molecules are effective at blocking infection by binding to the transmembrane region of the protein and preventing the pH induced structural change. A particularly difficult challenge regarding the design of the therapies that target M2 is that the virus quickly develops resistance by mutating amino acids of the transmembrane domain, leaving the drugs useless.

3. Fusion Inhibitors. The HA protein is the first piece of the virus to make contact with the host cell, making it an attractive target for drug design. If the initial viral fusion is blocked, the infection could not progress. Research in this area have led to the discovery of several salicylamide compounds that act as fusion inhibitors (*1*). Studies indicate that this class of drugs is effective at preventing fusion by inhibiting the formation of the low pH conformation, the form that is active in viral fusion. Further confirming the site of action was the analysis of viruses resistant to this drug that arose after exposure. A single mutation at position 110 from a Phe to a Ser appeared to cause the new resistance (*4*). This also indicates that the drug likely binds in that region of the protein.

A separate study confirmed that the mechanism of the drug was the inhibition of the low pH form through the use of ELISA assays with a low-pH-specific monoclonal

antibody (5). This st
prevent the formation
fusion inhibition, wh
that this could be due
HA trimers, and simp
of trimers needed for
more information rega
of environmental chan
will shift the focus of
towards rational desi
compounds.

Hemagglutinin Pro

The hemagglu
of the influenza viru
tightly linked by a
is outside the virus,
endodomain that is
specific cell by re
lower the pH of the
HA structural chan
fusion. The F
shown in figure 1.2

antibody (*5*). This study also found a difference between the concentration needed to prevent the formation of the low pH form and the concentration necessary for membrane fusion inhibition, which required about four fold lower doses. The investigators propose that this could be due to the common theory that fusion requires the presence of many HA2 trimers, and simply blocking a fraction of the protein could prevent the aggregation of trimers needed for effective membrane fusion (*5*). These studies illustrate the need for more information regarding not only the structure of the HA2 protein, but also the effect of environmental changes such as pH. The more knowledge that is gained on the subject will shift the focus of drug development away from random high throughput screening towards rational design, which will lessen the time required to obtaine successful compounds.

B. Hemagglutinin Protein

The hemagglutinin (HA) protein is responsible for the membrane fusion activity of the influenza virus (*6*). The HA protein is composed of HA1 and HA2 subunits initially linked by a disulfide bond. HA2 has a ~185 residue N-terminal ectodomain that lies outside the virus, a ~25 residue transmembrane domain, and a ~10 residue C-terminal endodomain that is inside the virus (*7*). The virus is taken into the host respiratory epithelial cell by receptor-mediated endocytosis and the cell physiological processes lower the pH of the endosome to ~5. The HA1 and HA2 subunits dissociate and a large HA2 structural change results in exposure of the ~20 residue N-terminal "fusion peptide (FP)" region. The FP binds to endosomal membranes and membrane fusion occurs, as outlined in figure 1.2.

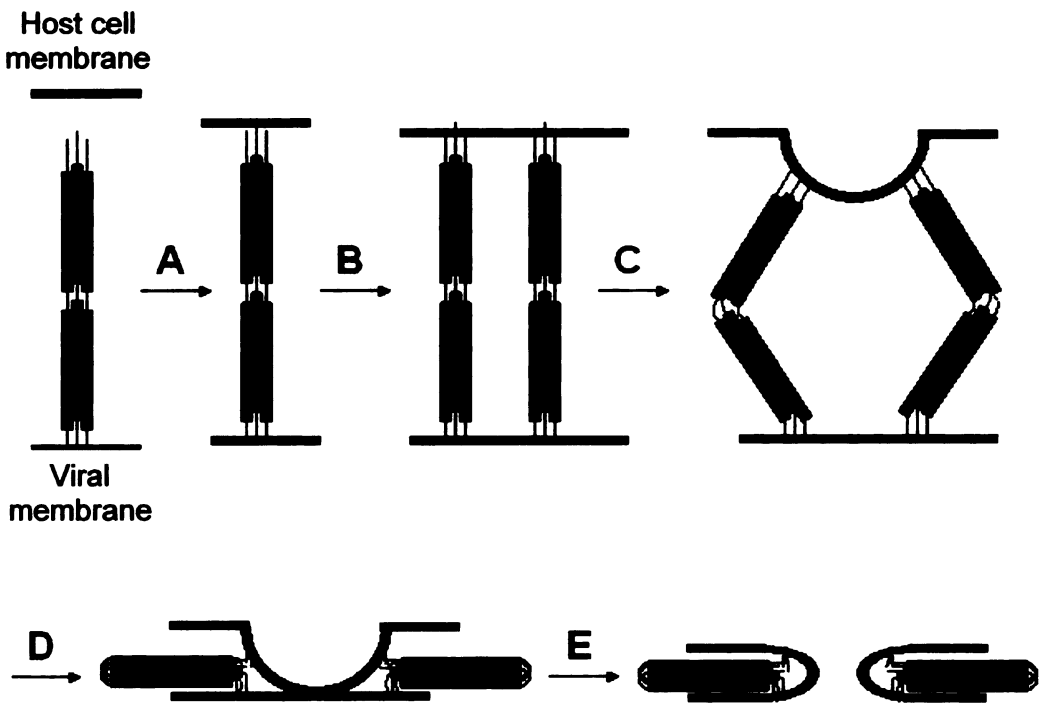


Figure 1-2. Membrane fusion by HA2. (A) Following a drop in pH, the fusion protein changes conformation to expose the fusion peptide (red), which inserts into the host cell membrane. (B) Multiple trimer units aggregate to more efficiently fuse the viral and host cell membranes. (C) The hinge region of the protein begins to fold, bringing the N- and C-terminal coil regions together. (D) Formation of the six-helix bundle brings the host and viral cell membranes in close proximity and initiates the mixing of lipids, known as hemifusion. (E) A fusion pore forms through which the viral contents enter the host cell.

HA2 mediated membrane fusion, along with fusion induced by many other fusion proteins, is thought to occur through a "spring loaded" mechanism. At this point the discussion will focus on the HA protein of influenza and the fusion subunit HA2, but is also applicable to the HIV fusion protein gp120, and the fusion subunit gp41. Upon activation, the HA2 protein undergoes a conformational change to expose the HA2 trimer, which is initially in a metastable conformation (*8*). The HA2 protein changes conformation and propels the fusion peptide region into the host cell membrane. The

HA protein contains

a six-helix bundle ha

ral viral cell memb

mbranes (9). The d

ary structure and fu

understanding of the f

HA activation

pH of the endosome

environment triggers t

portant in understan

The structure of the p

understanding the proc

HA Structural Stud

There has be

crystallized from aqu

from the soluble ec

sion (7, 14). Th

structural studies. T

isolated protein, i

HA2 protein contains two coiled-coiled regions, which collapse onto each other forming a six-helix bundle hairpin structure. The formation of the six-helix bundle brings the host and viral cell membrane in close proximity and leads to the mixing of the lipid membranes (*9*). The details of the formation of the hairpin structure has been the focus of many structure and function studies as it appears to be one of the most critical steps in the understanding of the fusion process (*7, 10-12*).

HA activation occurs after the virus has been taken into a cellular endosome. The pH of the endosome drops from the physiological pH of 7.4 to 5 and this change in environment triggers the fusion process (*11, 13*). This conformational change is then very important in understanding, and eventually developing methods to prevent, viral fusion. The structure of the protein in both the high and low pH form would greatly aid in understanding the process.

C. HA2 Structural Studies

There has been a pH 7.5 structure of the HA1/HA2 ectodomain complex crystallized from aqueous solution and a pH 4.4 structure of residues 34-178 of HA2 that form the "soluble ectodomain" (SHA2) and which was also crystallized from aqueous solution (*7, 14*). The pH 4.4 crystal structure is the model for which we compare our structural studies. This structure, with all positions that were analyzed in the membrane-associated protein, is shown in figure 1-3.

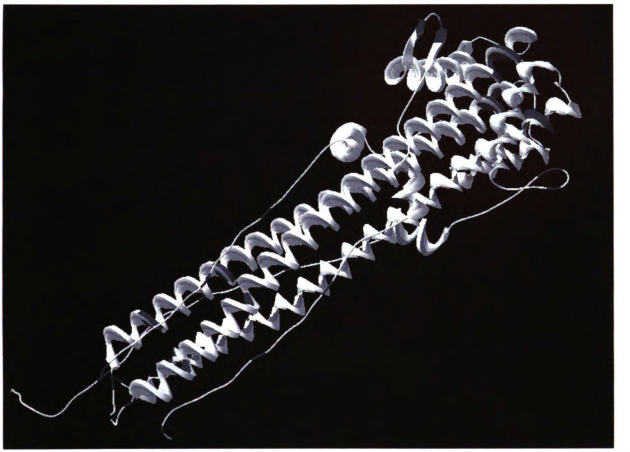


Figure 1-3. The HA2 soluble ectodomain crystal structure (*15*). This figure, created from crystal structure data of the protein crystallized from aqueous solution (PBD ID: 1QU1), indicates in red each residue studied by solid-state NMR. These results are discussed in detail in chapter 4.

In addition, there have been liquid-state NMR structures of the fusion peptide in detergent micelles. Also electron spin resonance measurements of motion and membrane insertion of specific residues of the fusion peptide and of a HA2 construct composed of residues 1-127 (*16, 17*).

The detergent solubilized NMR studies of the fusion peptide fragment have been conducted at both the physiological pH of 7.4, and at the active pH of 5 (*18*). The resulting structures did reveal a difference, consistent with the theory that a conformational change occurs as a result of the pH drop. This study revealed a helix that extends from Leu-2 to Asn-12, followed by a break, and then at the lower pH a second

ein that extends to

defined extended str

CD analysis of the

the resulting spectra

results most likely al

results were the mode

region of the membra

shown in red in figu

structure is shown in f

igure 1-4. The HA

ce solution-state

lution (PBD ID:

ese results are dis

igure 1-5. The HA

ce solution-state

PBD ID:

helix that extends to Asp-19. In the pH 7.4 structure, this break is followed by a well defined extended structure. This study was done in conjunction with circular dichroism (CD) analysis of the peptide in both detergent and lipid environments. The similarity in the resulting spectra led the investigators to conclude that the detergent solubilized NMR results most likely also apply to the structure of the peptides in a lipid membrane. These results were the model to which we compared our structural studies of the fusion peptide region of the membrane associated HA2 protein. The positions studied in this manner are shown in red in figure 1-4. The pH 7.4 detergent solubilized fusion peptide NMR structure is shown in figure 1-5.



Figure 1-4. The HA2 fusion peptide NMR structure at pH 5 (*18*). This figure, created from solution-state NMR structural data of the fusion peptide solubilized in detergent solution (PBD ID: 1IBN), indicates in red each residue studied by solid-state NMR. These results are discussed in detail in chapter 4.



Figure 1-5. The HA2 fusion peptide NMR structure at pH 7.4 (*18*). This figure, created from solution-state NMR structural data of the fusion peptide solubilized in detergent solution (PBD ID: 1IBO).



In addition to
the use of CD, EPR
containing residues
spin label to allow fo
depths of these lab
determined that resid
concluded that the pe
of the membrane at p
Similar membrane as
peptide (16). These st
on the longer construc
pH 5 and 23 at pH
indicated that the pept
compared to the high
endodomain region ma

The segment o
of many structural and
construct of the protei
model to investigate
produced to probe th
had indicated that th
proximity to residue 6
the coils exposed to so

In addition to the comparison of peptide structure in detergent and lipids through the use of CD, EPR studies have been conducted on a lipid associated construct of HA2 containing residues 1-127 (*17, 18*). Residues in the peptide were mutated to contain a spin label to allow for EPR study. EPR measurements can then be used to determine the depths of these labels into the lipid membrane. From these measurements it was determined that residues 2 and 3 were deepest into the lipid membrane. It was also concluded that the peptide enters the membrane at a 25° angle from the horizontal plane of the membrane at pH 7 and a slight change to a 28° angle was observed at pH 5. Similar membrane associated EPR studies have also been conducted on the fusion peptide (*16*). These studies gave a different result than the previously discussed studies on the longer construct of HA2, indicating that the peptide inserted at an angle of 38 at pH 5 and 23 at pH 7.4. Analysis of the immersion depths into the lipid membrane indicated that the peptide inserted 3-6 angstroms deeper in the lower pH structure when compared to the higher pH. This indicates that the presence of a portion of the ectodomain region may affect the binding of the fusion peptide region to the lipid bilayer.

The segment of the HA2 protein containing amino acids 1-127 has been the focus of many structural and functional studies (*19-21*). Lipid mixing assays confirmed that this construct of the protein was active at mixing lipid vesicles, and therefore an appropriate model to investigate protein structure. Cysteine cross linking experiments were conducted to probe the formation of the coiled-coil structure. Previous crystal structures had indicated that the coiled-coil structure places residue 63 of one coil in close proximity to residue 66 of an adjacent coil, and residues 64 and 65 lie on the exterior of the coils exposed to solvent. In these assays, residues 63-66 were selectively mutated to

cysteines and the formation of the disulfide bond monitored. These studies revealed that at pH of 7.4, the structure is not as rigid as previously thought. Disulfide bonds were formed between residues 63 and 66 as expected, but also involved in the bond formation were 64 and 65. These outer residues were far enough apart in the crystals structure that the bond formation would not be allowed, but were still observed to link. This indicates that while the coiled-coil structure does form, the structure possessed flexibility in this region.

The most common target of structure studies is the fusion peptide region of the protein, but the transmembrane domain has been targeted as well. Circular dichroism and IR studies focusing on a peptide fragment of the transmembrane region of the HA2 protein have indicated helical structure that forms stable oligomers (*22*).

These studies all provided pieces to the larger picture, structures of smaller units of the whole fusion protein. Our studies build upon this previous work by focusing on large constructs of the fusion protein, in hopes of building a better understanding of the whole protein structure and how this structure affects membrane fusion.

2. Human Immunodeficiency Virus

The Human Immunodeficiency Virus (HIV) ravages the immune systems of infected individuals and ultimately results in death. The viral fusion proteins allow for entry of the virus into the host cell through a similar mechanism as for the Influenza virus, but the initiation of viral fusion begins at the host cell membrane as opposed to after endocyctosis. The life cycle of the virus involves recognition of host cells through viral surface proteins, cell entry through lipid mixing of the host and viral cell

embranes, transc

integration of the vir

formed viruses that a

Anti-viral the

e progression of

Unfortunately, these

ciated and altere

patients with the fir

e majority of HI

reatments. Resear

ses that allow the

The HIV

ceptor on the c

fection of huma

CR5 and CXCR

Antiviral Dru

The mos

Possesses other

Offers a more

ass of drugs th

as a smaller

Preven

ges of the pro

membranes, transcription of viral DNA by the viral reverse transcriptase enzyme, integration of the viral genetic code into the host cell DNA, and finally budding of newly formed viruses that are activated by protease enzymes.

Anti-viral therapies have been significantly improved in recent years and can slow the progression of the virus for years or even decades when carefully monitored. Unfortunately, these therapies often include several drugs and need to be continuously re-evaluated and altered to combat the ability of the virus to develop resistance. Dedicated patients with the financial resources to stick to such regimens often see great results, but the majority of HIV infected patients are not able to afford the cost of these expensive treatments. Research continues into methods of treatment that are long lasting and at costs that allow their wide use in developing countries.

The HIV virus begins the interaction with host cells through interaction with a receptor on the cell surface, CD4, and a co-receptor from the chemokine family. The infection of human lymphocytes requires that the virus binds to the specific chemokines CCR5 and CXCR4 (4).

A. Antiviral Drugs

The most activity in HIV anti-viral drug development has focused on viral processes other than membrane fusion. While this is the initial step in viral infection, it remains a more elusive target for pharmaceutical therapies. Progress has been made in the class of drugs that aim to prevent the initial entry of the virus into the cell, although they exist in smaller number (*8*).

Preventing the entry of the virus into the host cell has been targeted at several stages of the process. The initial binding of the virus to CD4 and chemokine receptors has

been targeted in an

Antibodies against th

rugs designed to bl

prevent the membra

occurred and has led

inhibitors.

Peptides deriv

gp41 protein have sho

from the N-terminal r

be followed by a C-

concentrations (24).

may be effective th

promising results. It

α-helix bundle by a

ADS before the con

HIV-infected patient

better understandi

including advancing

3 Glycoprotein

The HIV vir

similar in fashion to

gp120 protein) and up

been targeted in an attempt to prevent the initial interaction of virus with the host cell. Antibodies against the receptor needed for binding have shown some success, as well as drugs designed to bind to the recognition proteins of the virus. Another approach is to prevent the membrane fusion of the virus after the initial host cell recognition has occurred, and has led to the development of a class of drugs known as membrane-fusion inhibitors.

Peptides derived from both the C- and N-terminal heptad repeat regions of the gp41 protein have shown promise in preventing membrane fusion (*23*). The first peptide from the N-terminal region, T-21, was effective at micromolar concentrations. This was the followed by a C-terminal peptide, SJ-2176, that was effective down to nanomolar concentrations (*24*). Since the initial results indicated that peptides from these regions may be effective therapy, many new sequences have been studied and also given promising results. It is proposed that these peptides work to prevent the formation of the six helix bundle by associating to the either the C- or N-terminal heptad repeat coiled-coils before the conformation change. The drug Enfuvritide is currently prescribed to HIV infected patients and is proving to be a successful component of anti-viral therapy. A better understanding of the protein structures that these drugs are targeting is needed to continuing advancing in this pursuit.

B. Glycoprotein

The HIV virus initiates infection through the use of an envelope glycoprotein, similar in fashion to the influenza virus. This protein is termed the gp160 protein (glycoprotein) and upon proteolytic cleavage by host cell enzymes yields the subunits

gp41 and gp120. The primary role of the gp120 protein is in receptor binding, while the purpose of gp41 is to catalyze lipid mixing between the viral and host cell membranes. The gp41 protein, like other membrane fusion proteins, also contains an N-terminal region known as the "fusion peptide," which inserts into the host cell membrane. The studies that have been conducted on the HIV and influenza fusion proteins reveal that the nature of their structure and membrane fusion ability is very similar, with studies of the HA2 and gp41 proteins often yielding similar results. The membrane fusion by gp41 is thought to proceed through the same "spring loaded" mechanism as HA2, and also exist as a trimer (*25*).

A significant difference in infection by these two viruses is the method of activation. While HA is activated by the pH drop upon endocytosis, the HIV fusion protein is activated by binding to receptors on the cell surface at neutral pH. The fusion protein binds to the cellular CD4 receptor as well as a chemokine co-receptor, often CCR5 or CXCR4 (*10*). Membrane fusion then occurs between the viral membrane and the host cell membrane.

C. gp41 Structural Studies

The Weliky group has extensively studied the structure of the membrane associated fusion peptide of the gp41 protein. These results have shown that the peptide adopts a β-strand conformation in lipid mixtures similar to that of the host cell (*26-29*). These studies have shown the presence of both parallel and antiparallel β-strands.

While the specific constructs of gp41 studied in this thesis have not been investigated through crystallography, truncated versions of the protein have produced

ystal structures. A

peat regions conne

30). Although this

vel of stability tha

olled hairpin struct

structure was the m

gure 1-6.

gure 1-6. The gp
ystal structure dat
or each residue s
gure 6.

The gp41 pr

dies have focuse

ction 3.1). These

C-terminal reg

crystal structures. A version of the protein that contains the N- and C-terminal heptad repeat regions connected by a non-native linker produced a highly stable hairpin structure (*30*). Although this protein is only a fragment of the total protein, it does exhibit a high level of stability that enforces that current theory that the gp41 protein adopts a coiled-coiled hairpin structure upon activation and binding to the host cell membrane. This structure was the model to which our structural studies are compared, and is shown in figure 1-6.

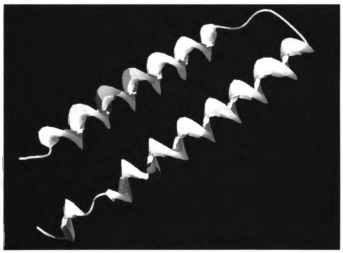


Figure 1-6. The gp41 hairpin region crystal structure (*30*). This figure, created from crystal structure data of the protein in aqueous solution (PBD ID: 1SZT), indicates in color each residue studied by solid-state NMR. These results are discussed in detail in chapter 6.

The gp41 protein is also known to form a hairpin structure, and many structural studies have focused on the two helices of the protein that come together to form the hairpin (*31*). These are commonly referred to as the N- and C-peptides, based on the N- and C- terminal region of the ectodomain from which they arise. Crystal structures of the



gp41 ectodomain have revealed that the N-peptides will form a coiled-coil surrounding by the C-peptides, also in helical formation (*30, 32, 33*).

Lipid mixing studies conducted on the fusion peptide region of gp41 also indicate that the protein natively exists in the trimer formation (*34*). Lipid mixing assays conducted on the monomer, dimer, and trimer version of the fusion peptide reveal that trimerization reduces the activation energy and increases the fusion rate.

3. Membrane Proteins

Membrane proteins are important in many physiological processes. Their roles include the transport of molecules into the cell, regulation of cellular conditions, and membrane fusion (*35*). The work in this thesis focuses on the membrane fusion class of proteins. Membrane fusion is the initial step in viral infection by enveloped viruses and is therefore a process targeted by anti-viral drug design, as discussed earlier in this chapter. While the research presented here is focused on viral fusion proteins, the information that we learn about any fusion proteins can provide insight into the whole class.

A. Expression

Information available on the structure of membrane proteins is lacking in comparison to their soluble counterparts. This is due partly to the difficulties encountered in the expression of membrane proteins. Membrane proteins require a hydrophobic environment to maintain native structure, and typically find this in the cell membrane. The cell membrane has a limited amount of space for these proteins and this results in low expression yields.

Recent adv

oduction of many

odies and expressi

troduction section

B Inclusion Bodies

Expression of

oduction of insolub

ttle practical use. In

about 10% of the

ecessity an attracti

ossible with denat

urification, and re

egarded as a usele

ber of natively fo

rification metho

roteins is to deter

ified form of the

s a difficult task

ecoding previous

There is n

ver that they ar

rotocles not face

Recent advances in the field of membrane protein expression have made the production of many proteins feasible. New methods include expression as inclusion bodies and expression with chaperone proteins, which will be discussed in detail in the introduction sections of chapter 2 and 5, respectively.

B. Inclusion Bodies

Expression of recombinant protein in bacterial cells is notoriously plagued by the production of insoluble protein aggregates, termed inclusion bodies, that typically are of little practical use. Inclusion bodies can compose up to 50% of the total cellular protein, or about 10% of the total cell mass, making improved methods of inclusion body protein chemistry an attractive target (*36*). Obtaining functional protein from inclusion bodies is possible with denaturants, but often includes time consuming steps of solubilization, purification, and refolding to the native, active form. Often inclusion body protein is regarded as a useless byproduct of recombinant expression, but for some proteins the level of natively folded protein expression is so low that these tedious inclusion body purification methods must be used. A logical initial step in the exploration of these proteins is to determine the inclusion body structure and how it relates to the natively folded form of the protein. The structural study of inclusion bodies within bacterial cells is a difficult task, but many new approaches have been presented in recent years, including previous work with solid-state nuclear magnetic resonance (SS-NMR) (*37*).

There is no consensus on the structure of proteins in inclusion bodies, but it is clear that they are non-crystalline solids. The study of solid proteins introduces many obstacles not faced when working with soluble proteins, but methods such as IR and SS-

NMR have given

common structural

inclusion body prot

overall fractions of

there was an increa

compared to the nat

the inclusion bodies

structure similar to

inclusion bodies is s

which hydrogen de

fibrillogenesis and amyl

has been a solid-stat

on the secondary str

fibrillar structure of

While this study in

The samples were

Fourth to form

spectra indicate

39 46. The ov

develop a meth

is opposed to the

NMR have given insight into the secondary structure of these proteins. The most common structural method to date has been the study of dehydrated samples of purified inclusion body protein using IR spectroscopy, which can provide information about the overall fractions of different secondary structure types. The results have suggested that there was an increase in the amount of β sheet structure in the inclusion body protein compared to the native form. These results led to the proposal that individual proteins in the inclusion bodies associated together to form non-native intermolecular β-sheets in a structure similar to amyloid protein (*38-43*). The argument for β sheet structure within inclusion bodies is supported by recent work on three different inclusion body proteins in which hydrogen/deuterium exchange NMR spectroscopy, x-ray diffraction, circular dichroism, and amyloid specific staining which all support this theory (*44*). There has also been a solid-state nuclear magnetic resonance study that was able to provide insight into the secondary structure of several such proteins by observing the presence of partial helical structure of hyperthermophilic protein inclusion bodies expressed in E. coli (*45*). While this study introduced the use of solid state NMR in the study of inclusion bodies, the samples were not of the full bacterial cells, and the labeling scheme used made it difficult to form conclusions about the secondary structure throughout the protein. Some reports indicate that a fraction of inclusion body protein is natively folded and functional (*39, 46*). The overall differing conclusions for the various methods of study have led us to develop a method to study the secondary structure at specific residues within the protein, as opposed to the overall secondary structure.

C. Structural Studie

Due to difficu

ethods of study t

considered success

nalizable due to

yield information re

experiments, as the

inherent draw back

the overall protein st

IR can reveal t

esence of specific

ds reveals secondar

ten indicate the re

rovide information

NMR and cry

ucture of memb

allenges of pro

uccessful at pro

ation experim

ack of crystal

Trials have en

ation. While c

C. Structural Studies

Due to difficulties encountered with the expression of membrane proteins, the methods of study have traditionally been limited. A protein yield of 1 mg/L would be considered successful, but this makes structural studies by crystallography or NMR undesirable due to the large volume of expression that would be required. Methods that yield information regarding the overall structure of the protein are the most reasonable experiments, as they require the least amount of protein for a successful study. The inherent draw back to these methods is that the structural information obtained is about the overall protein structure and not about specific locations in the protein.

IR can reveal the presence of certain secondary structures as indicated by the presence of specific absorptions, but not where in the protein these structures exist. CD also reveals secondary structure, and fitting of the observed curve to known values can even indicate the relative amounts of each secondary structure. Again, CD does not provide information about where in the protein these structures are present.

NMR and crystallography can provide much more detailed information regarding the structure of membrane proteins, but often require larger quantities of protein. Despite the challenges of producing large amounts of membrane proteins, these methods have been successful at providing more specific structural information about membrane proteins in certain experimental conditions. Great advances have been made in recent years in the field of crystallography. The use of detergents, lipids, and antibodies to produce protein crystals have enhanced the ability to study membrane proteins through crystallography (*47-50*). While crystallography and NMR can provide large amounts of complimentary

structural information, ultimately the studies need to be on membranes proteins when in their native membrane environment.

We know that membrane proteins natively exist only in cell membranes. Their function requires them to be at the surface of the cell, a viral fusion protein would not be effective at fusing to a host cell if the protein were on the interior of the virus. For this reason, the most relevant environment to study a membrane protein is in the membrane environment. Crystal structures solved from aqueous solution provide information on stable conformations of the protein, but these conformations may not actually be the same as when the protein is in a hydrophobic environment. NMR studies conducted on detergent solubilized protein is a step closer to a more relevant environment, but structural studies in lipid membrane will be more physiologically relevant. This thesis will present a novel method for the site-specific study of viral fusion proteins while associated to lipid membranes.

4. Protein NMR

NMR has become a key tool in the structural study of proteins. For some proteins, study by crystallography may not be feasible. Well-designed NMR experiments can give similar information regarding structure. The focus can be one-dimensional experiments in which the observed chemical shifts are correlated to structure (*51, 52*). Multidimensional experiments provide information about the proximity of specific atoms within the protein either measured through bonds or through space. As the projects in this thesis illustrate, the study of protein folding is of great interest and NMR can be a useful tool in this

pursuit (53). It is

proteins through N

A Liquid State N

Liquid sta

solubilized memb

purification or re

solubilized form

protein that is fu

structure.

The quest

studies is whethe

NMR has appea

membrane associ

B Solid-State N

Structural

crystallography a

Solid State NMR

for membrane

ated membrane

have been made

NMR, it is first

Section Sailline s

pursuit (*53*). It is even possible to now study protein dynamics or the interactions of proteins through NMR (*54-58*).

A. Liquid State NMR

Liquid state NMR has been used extensively to study the structure of detergent solubilized membrane proteins (*59*). These proteins are obtained from either native purification or refolded from denaturing purifications, but the intent is that the detergent solubilized form of the protein is in the native conformation. When these results reveal a protein that is fully folded, it does appear to be a reasonable initial method of studying structure.

The question that cannot be answered through detergent solubilized structural studies is whether this is the native structure of the protein. This is where solid state NMR has appeared on the scene with the intent of providing information on the membrane associated protein.

B. Solid-State NMR

Structural analysis of large membrane proteins has generally been limited to crystallography and liquid state NMR with the protein solubilized in detergent micelles. Solid State NMR has been used previously to study other large membrane proteins and even membrane associated fusion peptides (*60-64*). Until recently, structural information about membrane associated fusion proteins was difficult to obtain but many advances have been made in recent years (*65*). Before discussing the study of membrane associated NMR, it is first important to understand the previously existing alternatives, such as microcrystalline samples or precipitation to produce solid protein pellets.

1. *Microcrystals.*

pplicable to mem

membrane protein

environment and t

that are desirable f

echnique is usefu

signals that make

2. *Precipitation.*

through dialysis t

which the resultir

ee of lipid ar

been applied to th

acteriophage P2

niformly ^{13}C an

the protein and i

Multidimensiona

completed for a

NG, which coul

An alter

Crystal. Membr

Lipid or deter

ize of the dis

1. *Microcrystals.* Another common type of sample preparation that is not typically applicable to membrane proteins is crystallization (*66, 67*). The hydrophobic nature of membrane proteins often renders them unstable outside of a membrane or detergent environment and therefore many do not maintain native structure in the aqueous solutions that are desirable for crystallization. Despite the difficulties with membrane proteins, this technique is useful in that the highly ordered structure of protein crystals results in sharp signals that make assignment easier.

2. Precipitation. Solid samples of membrane proteins can be prepared by precipitation through dialysis to remove the detergent or with the use of polyethylene glycol (PEG) in which the resulting samples can be associated with lipids, embedded in lipid nanodiscs, or free of lipid and detergent (*68-70*). PEG causes precipitation of the protein. This has been applied to the studies of several proteins. The protein in the coat of the filamentous bacteriophage Pf1 was studied by SS-NMR following precipitation by PEG (*71*). With uniformly ^{13}C and ^{15}N labeling, assignments of specific atoms were successful for 92% of the protein and indicated that the protein was almost completely in a helical structure. Multidimensional SS-NMR studies were conducted and partial chemical shift assignment completed for a uniformly labeled KcsA potassium channel protein precipitated from PEG, which could be a great model for the study of other ion channels (*70*).

An alternative method to membrane protein precipitation is through detergent removal. Membrane proteins are not soluble in aqueous solution, and require the presence of lipid or detergent to remain fully folded and not aggregate. One such example was the study of the disulfide bond forming membrane enzyme DsbB, in which the sample was

prepared by remo

NMR was conduc

Aligned sampl

the alignment of

samples using ma

new structural int

backbone structure

from a mercury tra

conducted on a G

illustrate the feasib

While mos

experiments, the

have also been app

that may be perfo

[5, 7-3].

Membrane a

The fi

proteins. As d

this relevant e

different experi

by Shekhtman w

prepared by removing the detergent through dialysis. Three and four dimensional SS-NMR was conducted on this protein.

3. Aligned samples. The third difference amongst solid-state NMR structural studies is the alignment of the sample. While most studies have been conducted on un-aligned samples using magic angle spinning (MAS), the study of aligned samples has provided new structural information as well. With the use of magnetically aligned bicelles the backbone structure of two transmembrane helices and the 10-residue inter-helical loop from a mercury transport membrane protein has been determined and studies have been conducted on a G protein-coupled receptor to single site resolution, both of which illustrate the feasibility of using this method for the study of membrane proteins (*62, 72*).

While most aligned samples have been studied by $^{1}H/^{15}N$ double resonance experiments, the $^{1}H/^{13}C/^{15}N$ methods commonly used in liquid state and MAS studies have also been applied to these samples. This allows greater versatility in the experiments that may be performed and the information obtained using magnetically aligned bicelles (*65, 73*).

4. Membrane associated.

The final class of solid-state NMR structural studies is membrane associated proteins. As discussed earlier, studying proteins in the membrane-associated form is the most relevant environment. This has been applied to a variety of membrane proteins in different experimental techniques. The structure of the 52-residue peptide phospholamban was studied after reconstitution into lipid bilayers, and the membrane

ssociated results

alidating the SS

rovide unique in

he peptide or pr

epth of the gp41

nformation that c

While the

hen associated t

horter fragments

eptides can be

chemes can eas

he labeling met

ifficult. The w

roduction of m

uch much more

REDOR

The stru

mplification of

ach carbon in th

ctions through

each leucine r

associated results were found to be in agreement with previous biophysical studies, validating the SS-NMR study in membrane (*74*). Membrane associated NMR can also provide unique information not assessable by other methods regarding the interaction of the peptide or protein with membranes. The Weliky lab has conducted research into the depth of the gp41 fusion peptide insertion into membranes, providing another example of information that can be provided only from membrane-associated samples (*75, 76*).

While there are literature examples of the study of fusion peptides and proteins when associated to membranes, most of this work has been conducted on peptides. These shorter fragments avoid some of the experimental complexity of working with proteins. Peptides can be chemically synthesized, and therefore complicated isotopic labeling schemes can easily be produced. When working with recombinantly produced protein, the labeling methods available are much more limited, making these studies more difficult. The work discussed in this thesis describes significant advances in the production of membrane proteins and their isotopic labeling, which makes this type of study much more feasible.

C. REDOR

The structural study of peptides and proteins by NMR has the inherent complication of overlapping signals. If a sample is uniformly ^{13}C labeled, the signal from each carbon in the protein is observed and is of little use in analyzing structure at specific locations throughout the protein. The protein can be labeled as specific amino acids, such as each leucine residue in the protein, but resolving the signal from each specific residue

is often not possi

this problem (

REDOR

combination with

protein. This me

which is discuss

of the carbon

in figure 1-7a. W

NMR active nuc

can be selected

Gly8 form a unic

The pulse

labeled carbonyl

between the labe

protein and the

position. Subtra

position, the onl

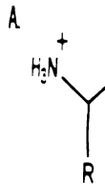
each other. Fig

spectrometer dri

is often not possible. The use of REDOR solid state NMR method can be a solution to this problem (*77-80*).

REDOR (rotational echo double resonance) spectroscopy is a method that in combination with selective isotopic labeling can be used to view only select positions in a protein. This method relies on high yielding, selective isotopic labeling of the protein, which is discussed extensively in this thesis. One such experiment may involve labeling all of the carbonyl carbons of alanine and all of the amide nitrogens of a glycine, shown in figure 1-7a. Whenever there is a Ala-Gly in the protein sequence, there will be two NMR active nuclei next to each other. By analyzing the protein sequence, amino acids can be selected which form unique sequential pairs, figure 1-7b. For example, Ala7 and Gly8 form a unique sequential pair.

The pulse sequence of the REDOR experiment will first acquire a spectrum for all labeled carbonyls in the protein. The second pulse sequence will use the dipolar coupling between the labeled amide and carbonyl to "turn off" the signal at that one position in the protein, and the spectrum acquired will be composed of every position except the labeled position. Subtraction of these two spectra will result in a spectrum of only the desired position, the only position in the protein where the two labeled amino acids are next to each other, figure 1-7c. The two types of spectra are alternately acquired so that spectrometer drifts and related effects will be efficiently subtracted.



C.



Signal f
C in the

igure 1-7. Stru
arbonyls are
he protein whe
he signals for a
eak of only the

The RE

irectly attached

rientation de

agnets. Dipola

between the

ases a distribu

olar coupling

espect to the ex

ach "magic a

olar coupling

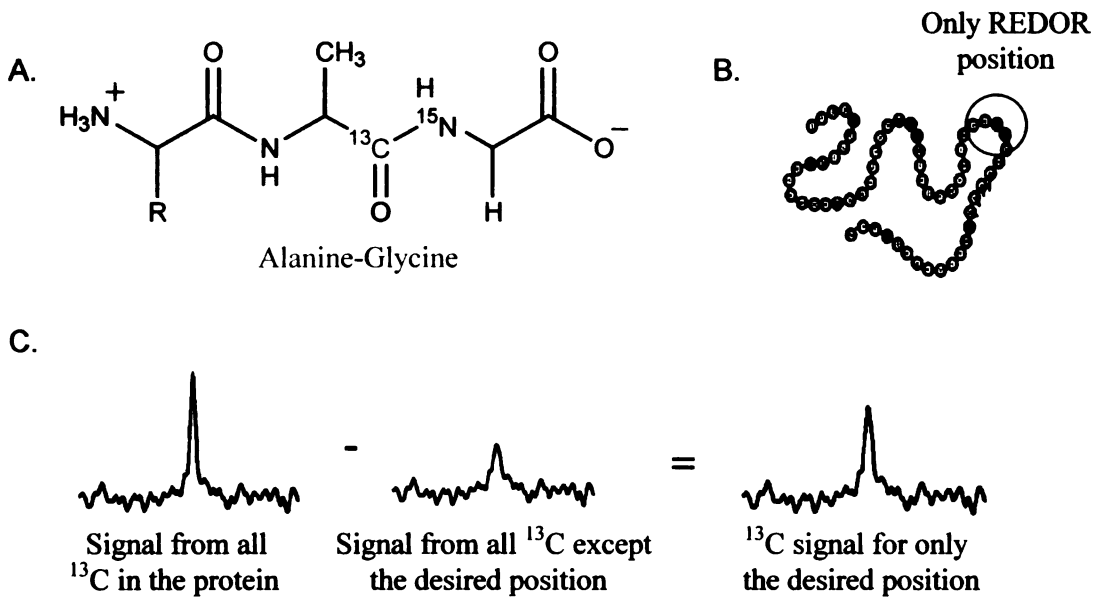


Figure 1-7. Structural Analysis by REDOR. (a) Amino acid sequence in which alanine carbonyls are ^{13}C labeled and glycine amide nitrogens ^{15}N labeled. (b) Only one place in the protein where the two NMR active nuclei are adjacent. (c) A difference spectrum of the signals for all labeled carbonyls and all carbonyls except the desired position yield a peak of only the Ala-Gly position.

The REDOR pulse is a method for exploiting the large dipolar coupling of directly attached ^{13}C-^{15}N unique sequential pairs. Dipole-dipole coupling (figure 1-8a) is an orientation dependent interaction between two nuclei similar to that of two nearby bar magnets. Dipolar coupling has an r^{-3} dependence on distance and is sensitive to the angle θ between the ^{13}C-^{15}N pair axis and the external magnetic field. The dipolar coupling causes a distribution of NMR frequencies and a broadened signal. To obtain sharp lines, dipolar coupling is removed by spinning the sample about an axis tilted at 54.74° with respect to the external magnetic field. Figure 1-8b shows an approximation of the effect of such "magic angle spinning." While the actual dipolar energy does vary as a wave, the dipolar coupling energy will average to zero under MAS over one rotor period. If π

pulses are introdu

energy will chan

pair may either b

dipolar couplin

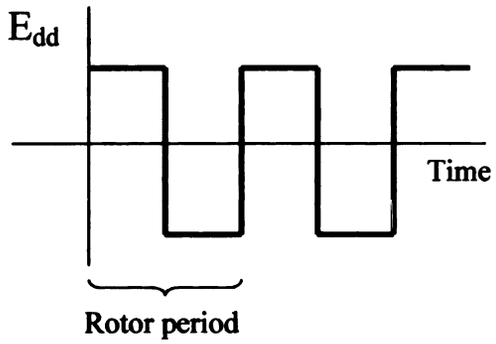
sequential pair.

pulses are introduced on either nucleus twice every rotor period (figure1-8c) the coupling energy will change, as shown in figure 1-8d. Thus, the ^{13}C signal of a unique sequential pair may either be attenuated (π pulses/dipolar coupling) or not attenuated (no π pulse, no dipolar coupling). Difference spectra will yield the filtered ^{13}C signal of the unique sequential pair.

A. Equation for dipolar coupling

$$E_{dd} = \underbrace{\left(\frac{\gamma_C \gamma_N}{r_{CN}^3} 3\cos^2\theta - 1\right)}_{\text{affected by orientation dependent spinning}} \underbrace{\left(I_{CZ} I_{NZ}\right)}_{\text{spin operators affected by rf}}$$

B. Dipolar coupling with time in MAS, averages to 0



C. π pulse twice per rotor period



D. Dipolar coupling reintroduced

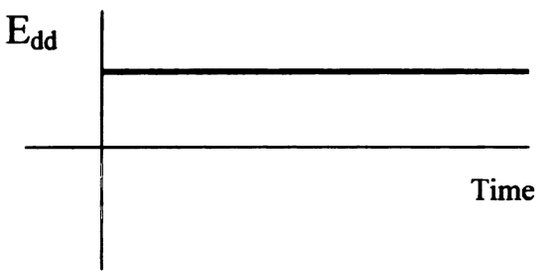


Figure 1-8. The REDOR Pulse Sequence. (a) Energy equation for dipolar coupling where θ is the spinning angle. (b) Approximation of dipolar coupling in time under MAS (averages to zero). (c) π pulses applied twice a rotor period (d) Dipolar coupling is reintroduced.

associated fusion

of this method

change in a mem

isotopic labeling

binding site that

The appl

determination o

membranes thro

measuring the m

nalysis of inter

membrane protei

work in this the

fusion proteins.

The seque

Figure 1-9. Al

underlined, and co

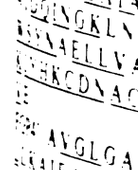


Figure 1-9. REDO

membrane fusion

described.

The REDOR method has been used to study the structure of membrane associated fusion peptides with great success by the Weliky group (*26, 28, 81*). A version of this method has also been applied to the study of a ligand induced conformation change in a membrane associated protein using $^{13}C^{19}F$ labeling (*82*). The complication of isotopic labeling schemes involving two residues was avoided in the study of a ligand binding site that only contained a signal residue of the target amino acid (*83*).

The applicability of the REDOR method has also extended beyond simply the determination of protein structure to the investigation of peptide interaction with membranes through the use of isotopic labeling of lipids and target peptide and measuring the magnitude of the dipolar interaction with this method (*84*). Similarly, the analysis of interaction between the ligand and the protein binding site of a bacterial membrane protein was investigated through distance measurements by REDOR (*63*). The work in this thesis builds upon these previous studies to study membrane associated fusion proteins.

The sequence of both HA2 and gp41 are good for this sort of analysis, as shown in figure 1-9. All positions for which the carbonyl-amide are a unique combination are underlined, and compose greater than half of each protein.

FHA2: GLFGAIAGFIENGWEGMIDGWYGFRHQNSEGTGQAADLKSTQA
AIDQINGKLNRVIEKTNEKEFHQIEKEFSEVEGRIQDLEKYVEDTKIDL
WSYNAELLVALENQHTIDLTDSEMNKLFEKTRRQLRENAEEMGNGSF
KIYHKCDNACIESIRNGTYDHDVYRDEALNNRFQIKGVELKSGYKDW
VE
FGP41: AVGLGAVFLGFLGAAGSTMGAASMTLTVQARQLLSGIVQQQSN
LLKAIEAQQHLLKLTVWGIKQLQARVLAVERYLQDQQLLGIWGCSGK
LICTSFVPWNNSWSNKTYNEIWDNMTWLQWDKEISNYTDTIYRLLED
SQNQQEKNEQDLLALDKLE

Figure 1-9. REDOR active positions. Amino Acid sequence for constructs of the membrane fusion region of HA2 and GP41 with all carbonyl-amide unique positions underlined.

V. References

(1) Combri
M., De
Tiley, I
fusion.

(2) Monto.
Jama-J

(3) Gubarev
neuram.

(4) Luo, G
hemagg

(5) Luo, G.
Malianc
and Kry
fusion i

(6) Chen, J.
C. (199
influenz
417.

(7) Chen, J
combine
cap that
8967-72

(8) LaBranc
Hammer
115.

(9) Bentz,
Biophys

(10) Kielian,
one way

(11) Hernand
Virus-ce

V. References

(1) Combrink, K. D., Gulgeze, H. B., Yu, K. L., Pearce, B. C., Trehan, A. K., Wei, J. M., Deshpande, M., Krystal, M., Torri, A., Luo, G. X., Cianci, C., Danetz, S., Tiley, L., and Meanwell, N. A. (2000) Salicylamide inhibitors of influenza virus fusion. *Bioorganic & Medicinal Chemistry Letters 10*, 1649-1652.

(2) Monto, A. S. (2000) Preventing influenza in healthy adults - The evolving story. *Jama-Journal of the American Medical Association 284*, 1699-1701.

(3) Gubareva, L. V., Kaiser, L., and Hayden, F. G. (2000) Influenza virus neuraminidase inhibitors. *Lancet 355*, 827-835.

(4) Luo, G. X., Colonno, R., and Krystal, M. (1996) Characterization of a hemagglutinin-specific inhibitor of influenza A virus. *Virology 226*, 66-76.

(5) Luo, G. X., Torri, A., Harte, W. E., Danetz, S., Cianci, C., Tiley, L., Day, S., Mullaney, D., Yu, K. L., Ouellet, C., Dextraze, P., Meanwell, N., Colonno, R., and Krystal, M. (1997) Molecular mechanism underlying the action of a novel fusion inhibitor of influenza A virus. *Journal of Virology 71*, 4062-4070.

(6) Chen, J., Lee, K. H., Steinhauer, D. A., Stevens, D. J., Skehel, J. J., and Wiley, D. C. (1998) Structure of the hemagglutinin precursor cleavage site, a determinant of influenza pathogenicity and the origin of the labile conformation. *Cell 95*, 409-417.

(7) Chen, J., Skehel, J. J., and Wiley, D. C. (1999) N- and C-terminal residues combine in the fusion-pH influenza hemagglutinin HA(2) subunit to form an N cap that terminates the triple-stranded coiled coil. *Proc Natl Acad Sci U S A 96*, 8967-72.

(8) LaBranche, C. C., Galasso, G., Moore, J. P., Bolognesi, D. P., Hirsch, M. S., and Hammer, S. M. (2001) HIV Fusion and its inhibition. *Antiviral Research 50*, 95-115.

(9) Bentz, J. (2000) Membrane fusion mediated by coiled coils: a hypothesis. *Biophys. J. 78*, 886-900.

(10) Kielian, M., and Rey, F. A. (2005) Virus membrane-fusion proteins: more than one way to make a hairpin. *Nature Reviews 4*, 67-73.

(11) Hernandez, L. D., Hoffman, L. R., Wolfsberg, T. G., and White, J. M. (1996) Virus-cell and cell-cell fusion. *Annu. Rev. Cell. Dev. Biol. 12*, 627-661.

(12) Tamm
influen
Acta-B

(13) Tamm
connec

(14) Wilson
haema
Nature

(15) Chen.
combin
cap tha
Acadcm

(16) Han. X.
structure
influenz

(17) Macosk
the fusio
EPR. *J.*

(18) Han. X.
structura
influenz

(19) Kim. C
and co
Biochc

(20) Epand
The
mem

(21) Leik
Cher
hema

(22) Tatu
oligom
hemag

(23) *Kin*
2238-224

(12) Tamm, L. K. (2003) Hypothesis: spring-loaded boomerang mechanism of influenza hemagglutinin-mediated membrane fusion. *Biochimica Et Biophysica Acta-Biomembranes 1614*, 14-23.

(13) Tamm, L. K., and Han, X. (2000) Viral fusion peptides: A tool set to disrupt and connect biological membranes. *Bioscience Reports 20*, 501-518.

(14) Wilson, I. A., Skehel, J. J., and Wiley, D. C. (1981) Structure of the haemagglutinin membrane glycoprotein of influenza virus at 3 A resolution. *Nature 289*, 366-73.

(15) Chen, J., Skehel, J. J., and Wiley, D. C. (1999) N- and C-terminal residues combine in the fusion-pH influenza hemagglutinin HA(2) subunit to form an N cap that terminates the triple-stranded coiled coil. *Proceedings of the National Academy of Sciences of the United States of America 96*, 8967-8972.

(16) 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 Biology 8*, 715-720.

(17) 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.

(18) 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. *Nat. Struct. Biol. 8*, 715-720.

(19) 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*, 5359-65.

(20) 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.

(21) 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*, 8378-86.

(22) 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.

(23) Kilby, J. M., and Eron, J. J. (2003) *The New England Journal of Medicine 348*, 2228-2238.

(24) Liu, S
Biolo

(25) Hugh
Curr

(26) Yang
evide
assoc

(27) Yang
and
Bioch

(28) Zhen
flexib
pepti
129

(29) Yang
reson
boun

(30) Tan,
therm
123

(31) Ecke
its in

(32) Chan
from

(33) Weis
(199
430.

(34) Yang
I fusi
which
1422

(35) Schie
mecha

(24) Liu, S., Lu, H., Niu, J., Xu, Y., Wu, S., and Jiang, S. (2005) *The Journal of Biological Chemistry 280*, 11259-11273.

(25) Hughson, F. M. (1997) Enveloped viruses: a common mode of membrane fusion? *Curr Biol 7*, R565-9.

(26) Yang, J., and Weliky, D. P. (2003) Solid state nuclear magnetic resonance evidence for parallel and antiparallel strand arrangements in the membrane-associated HIV-1 fusion peptide. *Biochemistry 42*, 11879-11890.

(27) Yang, R., Yang, J., and Weliky, D. P. (2003) Synthesis, enhanced fusogenicity, and solid state NMR measurements of cross-linked HIV-1 fusion peptides. *Biochemistry 42*, 3527-3535.

(28) Zheng, Z., Yang, R., Bodner, M. L., and Weliky, D. P. (2006) Conformational flexibility and strand arrantements of the membrane-associated HIV fusion peptide trimer probed by solid-state NMR spectroscopy. *Biochemistry 45*, 12960-12975.

(29) 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.

(30) Tan, K., Liu, J., Wang, J., Shen, S., and Lu, M. (1997) Atomic structure of a thermostable subdomain of HIV-1 gp41. *Proc. Natl. Acad. Sci. U.S.A. 94*, 12303-12308.

(31) Eckert, D. M., and Kim, P. S. (2001) Mechansims of viral membrane fusion and its inhibition. *Annual Review of Biochemistry 70*, 777-810.

(32) Chan, D. C., Fass, D., Berger, J. M., and Kim, P. S. (1997) Core structure of gp41 from the HIV envelope glycoprotein. *Cell 89*, 263-273.

(33) Weissenhorn, W., Dessen, A., Harrison, S. C., Skehel, J. J., and Wiley, D. C. (1997) Atomic structure of the ectodomain from HIV-1 gp41. *Nature 387*, 426-430.

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

(35) Sollner, T. H. (2004) Intracellular and viral membrane fusion: a uniting mechanism. *Current Opinion in Cell Biology 16*, 429-435.

(36) Baney
in Esc

(37) Curtis-
specifi
with s
130, 1.

(38) Przyby
Structu
Engine

(39) Oberg,
Second
Reflect

(40) Ami, D
Kinetic
spectro

(41) Petkov
Delagli
amyloi
Natl A

(42) Ritter,
Saupe,
of the F

(43) Carrio,
(2005)
Molecu

(44) Wang,
inclusi

(45) Umets
Kumag
hyperin

(46) Garcia-
polypep
Microb

(47) Caffrey
Biologg

(36) Baneyx, F., and Mujacic, M. (2004) Recombinant protein folding and misfolding in Escherichia coli. *Nature Biotechnology 22*, 1399-1408.

(37) Curtis-Fisk, J., Spencer, R. M., and Weliky, D. P. (2008) Native conformation at specific residues in recombinant inclusion body protein in whole cells determined with solid-state NMR spectroscopy. *Journal of the American Chemical Society 130*, 12568-12569.

(38) Przybycien, T. M., Dunn, J. P., Valax, P., and Georgiou, G. (1994) Secondary Structure Characterization of Beta-Lactamase Inclusion-Bodies. *Protein Engineering 7*, 131-136.

(39) Oberg, K., Chrunyk, B. A., Wetzel, R., and Fink, A. L. (1994) Native-Like Secondary Structure in Interleukin-1-Beta Inclusion-Bodies by Attenuated Total Reflectance Ftir. *Biochemistry 33*, 2628-2634.

(40) Ami, D., Natalello, A., Gatti-Lafranconi, P., Lotti, M., and Doglia, S. M. (2005) Kinetics of inclusion body formation studied in intact cells by FT-IR spectroscopy. *Febs Letters 579*, 3433-3436.

(41) Petkova, A. T., Ishii, Y., Balbach, J. J., Antzutkin, O. N., Leapman, R. D., Delaglio, F., and Tycko, R. (2002) A structural model for Alzheimer's beta-amyloid fibrils based on experimental constraints from solid state NMR. *Proc. Natl. Acad. Sci. U.S.A. 99*, 16742-16747.

(42) Ritter, C., Maddelein, M. L., Siemer, A. B., Luhrs, T., Ernst, M., Meier, B. H., Saupe, S. J., and Riek, R. (2005) Correlation of structural elements and infectivity of the HET-s prion. *Nature 435*, 844-848.

(43) Carrio, M., Gonzalez-Montalban, N., Vera, A., Villaverde, A., and Ventura, S. (2005) Amyloid-like properties of bacterial inclusion bodies. *Journal of Molecular Biology 347*, 1025-1037.

(44) Wang, L., Maji, S., Sawaya, M., Eisenberg, D., and Riek, R. (2008) Bacterial inclusion bodies contain amyloid-like structure. *PLoS Biology 6*, 1791-1801.

(45) Umetsu, M., Tsumoto, K., Ashish, K., Nitta, S., Tanaka, Y., Adschiri, T., and Kumagai, I. (2004) Structural characteristics and refolding of in vivo aggregated hyperthermophilic archaeon proteins. *Febs Letters 557*, 49-56.

(46) Garcia-Fruitos, E., Aris, A., and Villaverde, A. (2007) Localization of functional polypeptides in bacterial inclusion bodies. *Applied and Environmental Microbiology 73*, 289-294.

(47) Caffrey, M. (2003) Membrane Protein Cystrallization. *Journal of Structural Biology 142*, 108-132.

(48) Hunte, C
by antibo

(49) Ai, X., a
Mesopha

(50) Caffrey,
mesophas

(51) Zhang, F
uniformly

(52) Avbelj, F
from a-he
National

(53) Dyson, H
studied by

(54) Zaiderweg
NMR spec

(55) Ishima, R.
Structural

(56) Takahashi,
novel NM
complexes.

(57) Kay, L. E.
517.

(58) Case, D. A
Accounts of

(59) Johansson,
structure of
solvent con

(60) Jaroniec, C.
Griffin, R. G
labeled mem
in the active s

(48) Hunte, C., and Michel, H. (2002) Crystallisation of membrane proteins mediated by antibofdy fragments. *Current Opinion in Structural Biology 12*, 503-508.

(49) Ai, X., and Caffrey, M. (2002) Membrane Protein Crystallization in Lipidic Mesophases: Detergent Effects. *Biophysical Journal 79*, 394-405.

(50) Caffrey, M. (2000) A lipid's eye view of membrane protein crystallization in mesophases. *Current Opinion in Structural Biology 10*, 486-497.

(51) Zhang, H. Y., Neal, S., and Wishart, D. S. (2003) RefDB: A database of uniformly referenced protein chemical shifts. *J. Biomol. NMR 25*, 173-195.

(52) Avbelj, F., Kocjan, D., and Baldwin, R. L. (2004) Protein chemical shifts arising from a-helices and B-sheets depend on solvent exposure. *Proceedings of the National Academy of Sciences of the United States of America 101*, 17394-17397.

(53) Dyson, H. J., and Wright, P. E. (2004) Unfolded proteins and protein folding studied by NMR. *Chemical Reviews 104*, 3607-3622.

(54) Zuiderweg, E. R. P. (2002) Mapping protein-protein interactions in solution by NMR spectroscopy. *Biochemistry 41*, 1-7.

(55) Ishima, R., and Torchia, D. A. (2000) Protein dynamics from NMR. *Nature Structural Biology 7*, 740-743.

(56) Takahashi, H., Nakanishi, T., Kami, K., Arata, Y., and Shimada, I. (2000) A novel NMR method for determining the interfaces of large protein-protein complexes. *Nature Structural Biology 7*, 220-223.

(57) Kay, L. E. (1998) Protein dynamics from NMR. *Nature Structural Biology*, 513-517.

(58) Case, D. A. (2002) Molecular dynamics and NMR spin relaxation in proteins. *Accounts of Chemical Research 35*, 325-331.

(59) Johansson, J., Szyperski, T., Curstedt, T., and Wuthrich, K. (1994) The NMR structure of the pulmonary surfactant-associated polypeptide Sp-C in an apolar solvent contains a valyl-rich α-helix. *Biochemistry 33*, 6015-6023.

(60) Jaroniec, C. P., Lansing, J. C., Tounge, B. A., Belenky, M., Herzfeld, J., and Griffin, R. G. (2001) Measurement of dipolar couplings in a uniformly ^{13}C,^{15}N-labeled membrane protein: Distances between the Schiff base and aspartic acids in the active site of bacteriorhodopsin. *J. Am. Chem. Soc. 123*, 12929-12930.

(61) Ya
me
me

(62) Pa
S.
Jo

(63) Wa
dis
rec

(64) W
So
pe
Ch

(65) Op
Pr

(66) He
of
di
sp

(67) Hi
Ko
of

(68) Li
of
sp

(69) Li
as
er
19

(70) Va
as
Ba

(71) Ga
Fi
an
Ch

(61) Yang, J., Yang, R., and Weliky, D. P. (2002) Structural investigation of the membrane-bound HIV-1 fusion peptide by solid state NMR REDOR measurements. *Biophysical Journal 82*, 2636.

(62) Park, S. H., Prytulla, S., De Angelis, A. A., Brown, J. M., Kiefer, H., and Opella, S. J. (2006) High-resolution NMR spectroscopy of a GPCR in aligned bicelles. *Journal of the American Chemical Society 128*, 7402-7403.

(63) Wang, J., Balazs, Y. S., and Thompson, L. K. (1997) Solid-state REDOR NMR distance measurements at the ligand site of a bacterial chemotaxis membrane receptor. *Biochemistry 36*, 1699-703.

(64) Wasniewski, C. M., Parkanzky, P. D., Bodner, M. L., and Weliky, D. P. (2004) Solid-state nuclear magnetic resonance studies of HIV and influenza fusion peptide orientations in membrane bilayers using stacked glass plate samples. *Chem. Phys. Lipids 132*, 89-100.

(65) Opella, S. J., and Marassi, F. M. (2004) Structure Determination of Membrane Proteins by NMR Spectroscopy. *Chem. Rev. 104*, 3587-3606.

(66) Helmus, J. J., Nadaud, P. S., Hofer, N., and Jaroniec, C. P. (2008) Determination of methyl C-13-N-15 dipolar couplings in peptides and proteins by three-dimensional and four-dimensional magic-angle spinning solid-state NMR spectroscopy. *Journal of Chemical Physics 128*.

(67) Hiller, M., Krabben, L., Vinothkumar, K. R., Castellani, F., van Rossum, B. J., Kuhlbrandt, W., and Oschkinat, H. (2005) Solid-state magic-angle spinning NMR of outer-membrane protein G from Escherichia coli. *Chembiochem 6*, 1679-1684.

(68) Li, Y., Kijac, A. Z., Sligar, S. G., and Rienstra, C. M. (2006) Structural analysis of nanoscale self-assembled discoidal lipid bilayers by solid-state NMR spectroscopy. *Biophysical Journal 91*, 3819-3828.

(69) Li, Y., Berthold, D. A., Gennis, R. B., and Rienstra, C. M. (2008) Chemical shift assignment of the transmembrane helices of DsbB, a 20-kDa integral membrane enzyme, by 3D magic-angle spinning NMR spectroscopy. *Protein Science 17*, 199-204.

(70) Varga, K., Tian, L., and McDermott, A. E. (2007) Solid-state NMR study and assignments of the KcsA potassium ion channel of S. lividans. *Biochimica Et Biophysica Acta-Proteins and Proteomics 1774*, 1604-1613.

(71) Goldbourt, A., Gross, B. J., Day, L. A., and McDermott, A. E. (2007) Filamentous phage studied by magic-angle spinning NMR: Resonance assignment and secondary structure of the coat protein in Pf1. *Journal of the American Chemical Society 129*, 2338-2344.

[72] De Al
Struct
in alig
Ameri

[73] Sinha,
resona
labeled

[74] Andro
(2005)
angle-s
Society

[75] Qiang,
resona
intima
14-gly-

[76] Qiang,
Spectr
host-ce
suppor
regist

[77] Gullion
Magn.

[78] Bodner
Wellik
NMR s
membr

[79] Yang,

[80] Yang,
M., an
and co
Mol. G

[81] Yang,
(2002)
labeled
159, 10

(72) De Angelis, A. A., Howell, S. C., Nevzorov, A. A., and Opella, S. J. (2006) Structure determination of a membrane protein with two trans-membrane helices in aligned phospholipid bicelles by solid-state NMR spectroscopy. *Journal of the American Chemical Society 128*, 12256-12267.

(73) Sinha, N., Grant, C. V., Park, S. H., Brown, J. M., and Opella, S. J. (2007) Triple resonance experiments for aligned sample solid-state NMR of C-13 and N-15 labeled proteins. *Journal of Magnetic Resonance 186*, 51-64.

(74) Andronesi, O. C., Becker, S., Seidel, K., Heise, H., Young, H. S., and Baldus, M. (2005) Determination of membrane protein structure and dynamics by magic-angle-spinning solid-state NMR spectroscopy. *Journal of the American Chemical Society 127*, 12965-12974.

(75) Qiang, W., Yang, J., and Weliky, D. P. (2007) Solid-state nuclear magnetic resonance measurements of HIV fusion peptide to lipid distances reveal the intimate contact of β strand peptide with membranes and the proximity of the ala-14-gly-16 region with lipid headgroups. *Biochemistry 46*, 4997-5008.

(76) Qiang, W., Bodner, M. L., and Weliky, D. P. (2008) Solid-state NMR Spectroscopy of human immunodeficiency virus fusion peptides associated with host-cell-like membranes: 2D correlation spectra and distance measurements support a fully extended conformation and models for specific antiparallel strand registries. *Journal of the American Chemical Society 130*, 5459-5471.

(77) Gullion, T., and Schaefer, J. (1989) Rotational-echo double-resonance NMR. *J. Magn. Reson. 81*, 196-200.

(78) Bodner, M. L., Gabrys, C. M., Parkanzky, P. D., Yang, J., Duskin, C. A., and Weliky, D. P. (2004) Temperature dependence and resonance assignment of ^{13}C NMR spectra of selectively and uniformly labeled fusion peptides associated with membranes. *Magn. Reson. Chem. 42*, 187-194.

(79) Yang, J. (2003), Michigan State University, East Lansing, MI.

(80) Yang, J., Parkanzky, P. D., Khunte, B. A., Canlas, C. G., Yang, R., Gabrys, C. M., and Weliky, D. P. (2001) Solid state NMR measurements of conformation and conformational distributions in the membrane-bound HIV-1 fusion peptide. *J. Mol. Graph. Model. 19*, 129-135.

(81) Yang, J., Parkanzky, P. D., Bodner, M. L., Duskin, C. G., and Weliky, D. P. (2002) Application of REDOR subtraction for filtered MAS observation of labeled backbone carbons of membrane-bound fusion peptides. *J. Magn. Reson. 159*, 101-110.

(82) Mur
Site-
char

(83) Hin
inve
Esch
865

(84) Tok
stru
RE

(82) Murphy, O. J., 3rd, Kovacs, F. A., Sicard, E. L., and Thompson, L. K. (2001) Site-directed solid-state NMR measurement of a ligand-induced conformational change in the serine bacterial chemoreceptor. *Biochemistry 40*, 1358-1366.

(83) Hing, A. W., Tjandra, N., Cottam, P. F., Schaefer, J., and Ho, C. (1994) An investigation of the ligand-binding site of the glutamine-binding protein of Escherichia coli using rotational-echo double-resonance NMR. *Biochemistry 33*, 8651-61.

(84) Toke, O., Maloy, W. L., Kim, S. J., Blazyk, J., and Schaefer, J. (2004) Secondary structure and lipid contact of a peptide antibiotic in phospholipid Bilayers by REDOR. *Biophys. J. 87*, 662-674.

Chapter 2:

Expression and Purification of Natively Folded Fusion Proteins

I. Introduction

While there have been a number of high-resolution structures of bacterial membrane proteins in recent years, there have been fewer structures of viral and eukaryotic membrane proteins in part because of difficulties with production of large quantities of pure and folded protein by heterologous expression in *E. coli* (*1*). This chapter describes methods to address this problem with a particular focus on production of isotopically labeled membrane protein in *E. coli* for nuclear magnetic resonance studies. Isotopic labeling of proteins is usually done in a minimal medium with consequent reduction in cell growth relative to rich medium and lower production of protein. For membrane proteins, the yield may be reduced further because of higher hydrophobicity and limited cell membrane space needed to maintain native structure.

The protein production in this study was accomplished with conventional shake flask fermentation rather than a commercial fermenter. Although cell density can often be increased in a fermenter by controlling growth parameters such as carbon concentration, pH, and dissolved oxygen, the fermenter is expensive and not available in all laboratories. This paper describes the alternate but related approach of controlling growth parameters in shake flask fermentation. The culture was grown to maximum cell density in rich medium and then switched into fresh minimal medium prior to induction of protein expression (*2*). Labeled amino acids were added to the minimal medium at induction and were incorporated into the expressed protein.

A Protein Expre

Bacterial

produce protein

encoding for the

to high cell den

many specific va

are used in most

discussed in the

through the use of

I. The *lac operon*

protein production

xpressed in resp

ode for proteins

epressor binds ti

ranscription of th

ses in the cell

nformation and

esponsible for la

epressor is again

sponse to change

Recombin

combinant DN

A. Protein Expression

Bacterial recombinant protein expression is the most commonly used method to produce protein for structural or functional studies. A plasmid containing the DNA encoding for the desired protein is inserted into the bacterial cell, the cell culture grown to high cell density, and protein production induced followed by purification. While many specific varieties of recombinant expression exist, the same set of key principles are used in most situations. Plasmid DNA is prepared and inserted into the cells as discussed in the chapter 4. Induction of protein production is typically accomplished through the use of the lac operon.

1. The lac operon. The lac operon is natively found in bacteria, and used to induce protein production in recombinant expression. The lac operon consists of genes that are expressed in response to an increase of lactose concentration in the cell. These genes code for proteins that are responsible for the transport and metabolism of lactose. A repressor binds tightly just downstream of the promoter of the lac operon preventing the transcription of the genes when lactose is not present. When the concentration of lactose rises in the cell a metabolite called allolactose binds to the repressor changing the conformation and releasing it from the DNA. This allows for transcription of the genes responsible for lactose metabolism. As the lactose concentration lowers in the cell, the repressor is again free to bind to the DNA. This clever switch is often used in cells to respond to changes in concentrations of specific molecules.

Recombinant expression takes advantage of the lac operon by placing the recombinant DNA for the desired protein following the operon. A non-hydrolyzable

version of lactose is added to the cells that will bind to the repressor in the same manner, the molecule isopropyl-β-D-1-thiogalactopyranoside (IPTG). High concentrations are added to ensure that all repressors are removed and protein production occurs in all possible locations. Since IPTG is not metabolized, the removal of the repressor is essentially irreversible. The result is that IPTG acts as the switch to turn on the production of the recombinant protein.

The general principles of the lac operon and IPTG induction are still the basis of protein expression, but the design of expression systems have become much more complex. The specific vector used for the work in this project is the pET vector. This vector also contains a lac repressor system, and a unique T7 promoter region that does not occur anywhere in the bacterial genome (*3*). It is specific for only the T7 RNA polymerase and not the native bacterial RNA Polymerase (*4*). When the lac operon is not repressed, and the T7 polymerase is present, transcription occurs rapidly. This system relies on the presence of the T7 polymerase, which is not natively produced by bacterial cells. This requires the use of a special strain of bacteria in which the lac operator system and the gene for the polymerase have been incorporated into the bacterial genome, such as the BL21(DE3) strain used in this work. Addition of IPTG will now induce transcription on both the pET vector encoding for the recombinant protein, and the region of the bacterial genome encoding for the necessary polymerase.

2. Chaperone Protein Expression. Membrane proteins typically do not express in high yield, especially if the desired product is the natively folded, active form. Approaches to this situation are to increase the size of fermentations, retrieve protein from inclusion

bodies, or to alt

refolding protein

expressed const

different approac

of chaperone pro

soluble protein.

expressed in the

purification, one

designed to bind

examples of chap

glutathione S-tra

enhanced throug

over 35% of the

Many stu

experiments inv

chaperone often

thrombin. Depe

approach 100%

expression leve

expression and

folded active p

can be superior

bodies, or to alter the construct to optimize expression. Increased fermentations is costly, refolding protein from inclusion bodies can be time consuming, and at times changing the expressed construct will result in a protein that is not useful for the desired studies. A different approach is to increase the amount of soluble protein expressed through the use of chaperone proteins. This process expresses the desired protein fused to a high yielding, soluble protein. The result is often that the chaperone will cause the whole unit to be expressed in the soluble form. Many chaperone proteins are also designed for easy purification, one common example is the maltose binding protein (*5-9*). This protein is designed to bind maltose, and readily binds to amylose resin for easy purification. Other examples of chaperone proteins include SlyD, FKP 506 binding protein A (FkpA), and glutathione S-transferase (GST) (*10, 11*). The amount of expressed protein is greatly enhanced through the use of these chaperone proteins, increasing the amount of protein to over 35% of the total cellular protein and up to 100 mg per liter of expression (*7, 8, 11*).

Many studies would not be possible with the fusion protein still attached, so most experiments involve removal. The linking region connecting the target protein to the chaperone often contains a protease cleavage site for an enzyme such as factor Xa or thrombin. Depending on the accessibility of the linker to the protease, this step can approach 100% completion (*8, 11*). This brings troublesome proteins into the range of expression level that makes most any further experiments feasible. In some cases, the expression and following purification and cleavage steps may not result in properly folded, active protein. In these situations refolding may be necessary but still the process may be superior to expression without the chaperone (*10*).

B. Growth Cond

Effectiv

culture (12). T

increase, so the

increasing the g

medium, temper

may be a good

optimal cell gro

these conditions.

Stating th

protein yield in c

but the two are

within the cell. A

protein production

faster rate will c

greater (15-18). T

containing more r

protein. The grow

growth rate will n

requiring less expe

production.

B. Growth Conditions

Effective protein productive initially relies on the growth of a high density cell culture (*12*). The amount of protein produced by each individual cell is harder to increase, so the easier approach is to increase the number of cells producing protein. Increasing the growth of cells relies on optimizing conditions such as the growth medium, temperature, and oxygenation (*13, 14*). While optimal conditions for one system may be a good starting point for optimization of another, each situation is unique and optimal cell growth, and therefore protein production, involves reconsidering each of these conditions.

Stating that increasing the number of cells is the easier method of increasing protein yield in comparison to increasing the amount of protein per cell is generally true, but the two are actually linked. Recombinant protein production relies on ribosomes within the cell. An increase in the bacterial growth rate also requires an increase in native protein production, therefore relying on more ribosomes. Bacterial culture growing at a faster rate will contain more ribosomes per cell, as the need for protein production is greater (*15-18*). This is a benefit when it comes to recombinant expression as well, cells containing more ribosomes will have more of the necessary machinery for producing the protein. The growth of cells and protein produced per cell are therefore linked. Increased growth rate will not only produce the cell density needed for protein production faster, requiring less experimental time, but will also result in cells better equipped for protein production.

1. Growth

thrive in

conditions

bacterial c

contains n

essential b

cultures, th

se increas

sugars, such

Eac

for protein

cells to pr

nutrients in

decrease th

higher cell

components

2. Acidity. I

Most media

grow well a

acetic acid

acid low of

pH, but the

1. Growth Medium. While E. coli bacteria are surprisingly resilient organisms that can thrive in a wide variety of conditions, optimization of the growth and expression conditions can result in greatly increased protein production. The most critical aspect of bacterial culture growth is the liquid medium in which cells are grown. This medium contains nutrients that the cells need to grow, including amino acids, sugars, and other essential biomolecules. Typically, commercially available mixes are used to grow cell cultures, the most common being Luria Bertini broth (LB). Many other versions exist that use increased amounts of certain ingredients or differing relative amounts of certain sugars, such as Terrific broth.

Each medium is typically associated with a range of cell density that is optimal for protein expression. This is a delicate balance, as a high density of cells means more cells to produce protein, but typically higher densities have depleted much of the nutrients in the media. This combined with increasing levels of cellular waste can decrease the amount of protein produced per cell. The amount of protein produced at higher cell densities can be increased by accounting for the cellular metabolism and the components of the medium.

2. Acidity. Bacteria cells, like most other organisms, are most content at a neutral pH. Most media preparation protocols suggest an initial pH in the range of 7 to 8. The cells grow well at this level, but as the culture grows in density a significant byproduct is acetic acid (*19*). This causes a drastic drop in the pH and if protein production is induced at too low of a pH the yield can suffer. An obvious solution may be to increase the initial pH, but the bacteria will typically not thrive much above pH 8. A second approach is to

increase the bu

increase the ab

maintain a neut

composition can

buffer concentra

As there

production, a m

Glucose is the m

this sugar is resp

E. coli will typic

as a byproduct

optimized at vary

Isotopic Label

enriched medias w

becomes much m

minimal medium

available. If unifo

glucose, or if ^{15}N

ed. These situati

positions rather th

and the form of

increase the buffer in the medium. Increasing the amount of phosphate buffer can increase the ability of the medium to counteract the production of acetic acid and maintain a neutral pH. This may work in some situations, but any change in medium composition can lower overall cell growth and this is observed with simply increasing the buffer concentration.

As there is no simple solution to the pH drop associated with acetic acid production, a more fundamental approach is to limit the amount of acid produced. Glucose is the most common sugar used in bacterial fermentations and the metabolism of this sugar is responsible for the production of the acetic acid. An alternative is glycerol. E. coli will typically grow as well on glycerol as glucose but acetic acid is not produced as a byproduct of glycerol metabolism (*19, 20*). Fermentations should therefore be optimized at varying concentrations of both glucose and glycerol.

3. Isotopic Labeling. If the goal is simply to produce large amount of protein, these enriched medias work well. If the goal is the incorporation of isotopic labels, the situation becomes much more complex (*2, 21-23*). The typical approach is to grow the cells on a minimal medium in which only isotopically labeled versions of certain molecules are available. If uniform ^{13}C labeling is the goal, the only carbon source may be enriched glucose, or if ^{15}N is desired, the only source of nitrogen may be an enriched ammonium salt. These situations are relatively easy compared to incorporating labels into selective positions rather than uniform labeling. This involves considering the cellular metabolism and the form of the label added, as well as timing and concentration. The studies

discussed in

labeling.

C. Protein P

Deve

protein is on

can also be a

procedures

chromatograp

is amenable

these goals is

to a resin wh

protein easily

chromatograph

both methods

works on both

Immobilized

ses of an ins

from a solut

ion, the meta

purpose are de

mechanisms of

discussed in this thesis involve incorporating specific amino acids with either ^{13}C or ^{15}N labeling.

C. Protein Purification

Developing methods for high yielding expression of the desired recombinant protein is only the first challenge in producing large amounts of pure protein. Purification can also be a difficult task that requires much optimization. There are a wide variety of procedures used to purify recombinant proteins, including many versions of chromatography. The ideal method will quickly produce pure, natively folded protein and is amenable to large-scale purifications. The most commonly used method to achieve these goals is to design the protein with a purification tag. This tag will selectively bind to a resin when other cellular contaminants can be easily washed away, and then the protein easily washed from the resin. This project used both immobilized metal-affinity chromatography (IMAC) and the maltose binding protein/amylose resin system, with both methods producing pure, folded protein from bacterial cells in less than 3 hours and works on both small and large scales.

1. Immobilized metal-affinity chromatography. The purification method IMAC makes uses of an insoluble resin with attached metal ions, most commonly nickel or cobalt. When a solution containing a protein designed to bind to this metal is passed over the resin, the metal acts as a lewis acid, binding the protein (*24*). Typically proteins for this purpose are designed with a His-Tag, a sequence of 6 or 8 histidine residues at the C- or N-terminus of the protein, which will ideally be in an accessible position for the metal

affinity resin. T

of an elutant c

concentrations

resin that are w

histidines. Eluti

imidazole, comr

This pro

desired protein is

contaminating pr

bind strongly, or

decrease. A typic

which all protein

rea. This can ev

Most commercial

IMAC is

often be contami

that contains a h

target protein (2

amount of protei

contaminating pr

[3]. A new app

nickel resin. Alth

does not bind to t

affinity resin. The protein will remain bound to the resin until exposed to a concentration of an elutant compound that will bind stronger to the resin, such a imidazole. Low concentrations of this elutant are first used to wash contaminating compounds from the resin that are weakly bound, such as proteins that may contain two or three consecutive histidines. Elution of pure protein is then achieved by eluting with high concentration imidazole, commonly 250 mM.

This process can yield quantitatively pure protein in ideal situations, but if the desired protein is in very low concentrations difficulties can arise. The cell also produces contaminating proteins that posses an amino acid sequence that can also allow them to bind strongly, or if the His-tag is not accessible to the resin, the purity and yield can decrease. A typical approach to remedy these problems is a denaturing purification, in which all proteins in the solutions are unfolded by the presence of a denaturant such as urea. This can expose a hidden His-tag, or lower the affinity of contaminating proteins. Most commercially available IMAC resins are designed for use in these conditions.

IMAC is one of the most commonly used forms of protein purification, but can often be contaminated by a specific protein, SlyD. This protein is a *cis-trans* isomerase that contains a high percentage of histidines and often binds stronger to resins than the target protein (*25, 26*). This issue can be overcome in some cases by increasing the amount of protein solution initially bound to the resin in attempt to out-compete the contaminating protein, using ionic exchange chromatography, or denaturing purifications (*25*). A new approach is the use of cobalt purification resin in place of the traditional nickel resin. Although the explanation for this observation is not clear, the SlyD protein does not bind to the cobalt resin as strongly and can result in simpler purifications.

2 *The Maltose*

to increase the p

simple purificat

natively binds n

except that in pi

is a polymer of

recombinant pr

concentrations o

concentrations f

effective in seve

An intri

size of the fusi

Often removal o

by designing a

ideal situation.

maintain the nat

by again using t

desired protein t

not accessible t

refolding. Even

ack of the maltos

2. The Maltose Binding Protein. Use of the maltose binding protein serves two purposes, to increase the production of soluble protein (as discussed earlier in this chapter) and for simple purification. The maltose binding protein is a periplasmic bacterial protein that natively binds maltose, a dimer of glucose. The general principle is the same as IMAC, except that in place of metal ions attached to the resin, amylose is instead used. Amylose is a polymer of glucose and is tightly bound by the maltose binding protein, but the recombinant protein can be easily released in the presence of maltose. Low concentrations of maltose are used to wash contaminants from the resin, and higher concentrations for elution of purified protein. This purification scheme has proven effective in several published studies (*6, 8, 9, 27*).

An intrinsic difficulty of maltose binding protein purifications is that the large size of the fusion protein can make further studies of the expressed product difficult. Often removal of the maltose binding protein is required, and is typically accomplished by designing a protease cleavage site in the linker connecting the two proteins. In the ideal situation, the protease cleavage reaction would occur rapidly in conditions that maintain the native structure of the expressed protein, and the fusion protein is removed by again using the amylose resin. The released MBP will bind to the resin, allowing the desired protein to elute. Difficulties arise when the linking region between the proteins is not accessible to the protease, requiring denaturation of the complex and subsequent refolding. Even with these potential difficulties, the expression and purification with the aid of the maltose binding protein can be far superior to other methods.

D. Membrane R

The prir

is the environm

in the environ

studying a prote

aqueous or dete

have opened th

associated with

from the bacter

native structure.

and has been ach

Dialysis. The

detergent from th

method for accor

with a second so

through dialysis

through centrifug

ching, but most

the pellet followin

Figure 2-1, wher

ecies in the prote

D. Membrane Reconstitution

The primary deficit of previous structural studies focusing on membrane proteins is the environment of the protein. It is well known that membrane proteins are most stable in the environment of a lipid membrane, but due to the experimental difficulties of studying a protein in this solid form, past studies have focused on proteins solubilized in aqueous or detergent solutions. The development of the new solid state NMR techniques have opened the potential for studying the protein in its more native form, when associated with lipid membranes. The challenge now becomes how to purify the protein from the bacterial cell and return it to a membrane environment while still maintaining native structure. This process is generally referred to as membrane protein reconstitution, and has been achieved through several different approaches.

1. Dialysis. The key step in membrane protein reconstitution is the removal of the detergent from the solution, forcing the protein into association with lipid. A simple method for accomplishing this is to solubilize the protein in a detergent solution, combine with a second solution containing the desired lipids, and to then remove the detergent through dialysis (*28-31*). The result is a lipid/protein suspension that can be retrieved through centrifugation. Typically very little of the detergent remains in the dialysis tubing, but most of what does remain will exist in the soluble supernatant and not be in the pellet following centrifugation. The process is believed to follow the steps illustrated in figure 2-1, where the removal of detergent initiates a reorganization of the mixture and results in the protein embedded in a lipid bilayer.

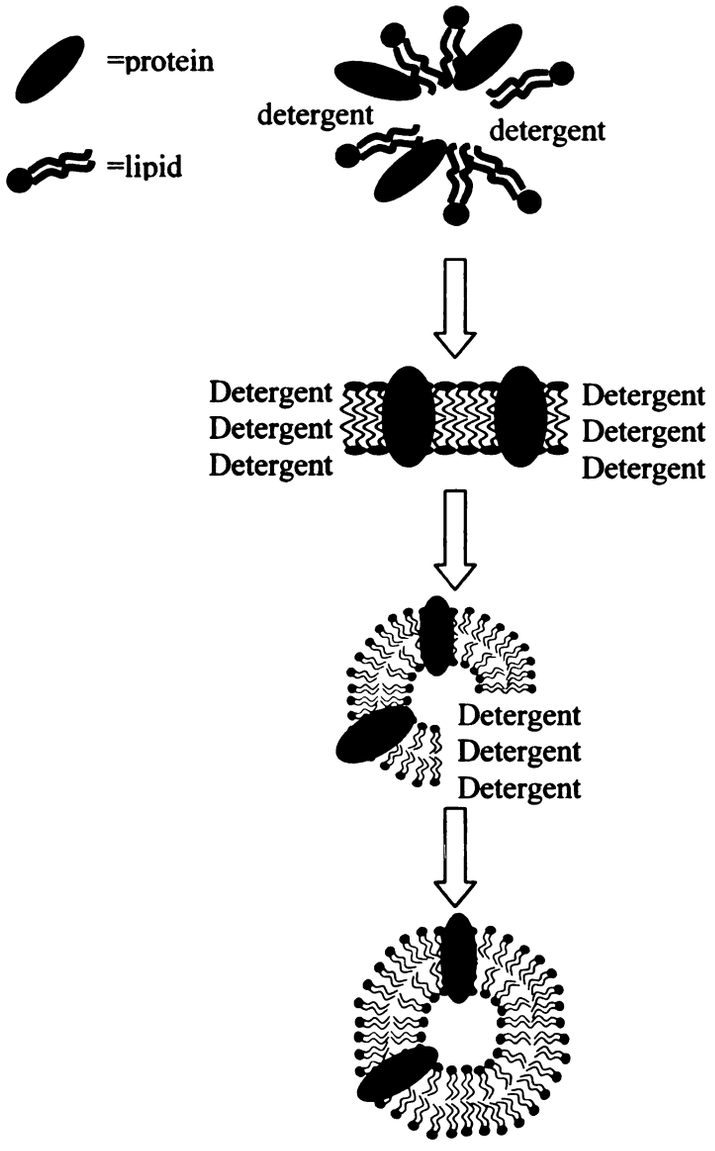


Figure 2-1. Membrane reconstitution. Initially a soluble mixture of lipid, detergent, and protein exists in solution. As the dialysis continues, the concentration of detergent decreases, causing a reorganization of the lipids into a bilayer. As the amount of detergent continues to decrease, the bilayer is finally closed into vesicle, at some point during this process precipitation occurs.

A crucial aspect of this approach is the detergent to be used. It must be one that adequately solubilizes the protein in the native form, while being easily removed through dialysis. Octylglucosides are often successful at this approach (*28, 32*). This class of

detergents ha

monomeric a

detergent wit

detergent is

through the

detergent trit

While

membrane

properties th

has a CMC

character wh

of the deterg

lipid require

times the cor

success (33).

2. *Detergent*

the solution

fraction of th

with this ver

ower and ca

conducted w

the product m

detergents have a high critical micelle concentration (CMC), meaning that they are monomeric at low concentrations and can easily be removed through dialysis. If a detergent with a low CMC is used, then at the working conditions of the experiment the detergent is more likely in the micelle form, with a size that may be too large to flow through the dialysis tubing. This typically prohibits the use of the commonly used detergent triton, which has a relatively low CMC.

While several versions of octyl-glucosides have been effective at reconstituting membrane proteins, specifically B-D-*n*-octyl-thioglucopyranoside (BTOG) possess properties that should allow it to perform better than the non-thiol version (BOG). BTOG has a CMC of 9 mM compared to 22 mM for BOG. This indicates a more hydrophobic character which potentially contributes to the higher level of success with the thio version of the detergent (*32*). Optimization of the relative concentrations of the detergent and lipid required for successful reconstitution revealed that if the detergent was at least four times the concentration of the lipid that reconstitution proceeded with the highest level of success (*33*).

2. Detergent Dilution. A simpler approach to dialysis is to simply dilute the detergent in the solution (*34*). Similar to the dialysis method, the detergent that remains in the soluble fraction of the mixture will largely be separated upon centrifugation. A potential problem with this version of reconstitution is that the concentration of the reaction will be much lower and can slow or prevent the reorganization. Also, if the structural studies to be conducted will be affected by the presence of detergent, the small amount remaining in the product may not be acceptable.

3. Detergent Extraction. A second, yet conceptually similar approach to reconstituting membrane proteins is to selectively remove the detergent from the detergent/lipid/protein mixture. Cyclodextrins posses a much higher affinity for detergents than for lipids, and will selectively bind these molecules in the solution (*35*). The membrane associated protein and cyclodextrin associated detergents can then be separated by column chromatography. A second approach is to use adsorbent beads that will bind the detergent, removing it from solution in a manner similar to the cyclodrextins (*36*). The vesicle containing portion of the mixture can also be removed through gel filtration, where the beads will remain at the top of the column. A difficulty with this approach is removal of the vesicles from the beads, as filtration and centrifugation often resulted in a high loss of the product.

II. Materials a

A. Materials

The F

University, th

the Weismann

Unless noted,

Bernani Brot

detergents n-

ether (C8E5)

were obtaine

purchased fro

B. Amino Ac

1 FHA2. Th

include the fu

GLFC
GKL
VALI
ESIR

Figure 2-2. A

influenza viru

II. Materials and Methods

A. Materials

The FHA2 plasmid was obtained from Dr. Yeon-Kyun Shin at Iowa State University, the MHA2 plasmid from Susanne Swalley, the Hgp41 from Yiechel Shai at the Weismann Institute. All plasmids were transformed into *E. coli* BL21(DE3) cells. Unless noted, all chemicals were purchased from Sigma-Aldrich (St. Louis, MO). Luria-Bertani Broth (LB) medium was purchased from Acumedia (Lansing, MI). The detergents n-octyl-β-D-thioglucopyranoside (BTOG) and octyl pentaethylene glycol ether (C8E5) were purchased from Anatrace (Maumee, OH). All lipids and cholesterol were obtained from Avanti Polar Lipids (Alabaster, AL). Labeled amino acids were purchased from Cambridge Isotope Labs (Andover, MA).

B. Amino Acid and DNA Sequences

1. FHA2. The FHA2 construct contains the residues 1-185 of the HA2 protein, which include the full ectodomain and a C-terminal polyhistidine tag for purification.

GLFGAIAGFIENGWEGMIDGWYGFRHQNSEGTGQAADLKSTQAAIDQIN
GKLNRVIEKTNEKEFHQIEKEFSEVEGRIQDLEKYVEDTKIDLWSYNAELL
VALENQHTIDLTDSEMNKLFEKTRRQLRENAEEMGNGSFKIYHKCDNACI
ESIRNGTYDHDVYRDEALNNRFQIKGVELKSGYKDWVEHHHHHH

Figure 2-2. Amino acid sequence of FHA2. This protein is from the X31 strain of the influenza virus. The last eight residues are non-native.

*CTAGAAATAATTTTGTTTAACTTTAAGAAGGAGATATACATATG*GGCCTATT
CGGCGCAATAGCAGGTTTCATAGAAAATGGTTGGGAGGGAATGATAG
ACGGTTGGTACGGTTTCAGGCATCAAAATTCTGAGGGCACAGGACAA
GCAGCAGATCTTAAAAGCACTCAAGCAGCCATCGACCAAATCAATGG
GAAATTGAACAGGGTAATCGAGAAGACGAACGAGAAATTCCATCAAA
TCGAAAAGGAATTCTCAGAAGTAGAAGGGAGAATTCAGGACCTCGAG
AAATACGTTGAAGACACTAAAATAGATCTCTGGTCTTACAATGCGGAG
CTTCTTGTCGCTCTGGAGAATCAACATACAATTGACCTGACTGACTCG
GAAATGAACAAGCTGTTTGAAAAAACAAGGAGGCAACTGAGGGAAAA
TGCTGAAGAGATGGGCAATGGT*A*GCTTCAAAATATACCACAAATGTGA
CAACGCTTGCATAGAGTCAATCAGAAATGGGACTTATGACCATGATGT
ATACAGAGACGAAGCATTAAACAACCGGTTTCAGATCAAAGGTGTTG
AACTGAAGTCTGGATACAAAGACTGGGTCGAGCACCACCACCACCAC
CACTGA*GATCCGGCTGCTAACAAAGCCCGAAAGGAAGCT*
GAGTTGGCTGCTGC

Figure 2-3. DNA Sequence of FHA2. The DNA encoding for this protein is in the plasmid pET24a(+) with the surrounding DNA shown in italics.

2. MHA2. The MHA2 construct contains the full HA2 protein expressed with the maltose binding protein to enhance expression and aid in purification, connected by a linker region with a thrombin cleavage site.

KDLLPNPPKTWEEIPALDKELKAKGKSALMFNLQEPYFTWPLIAADGGYA
FKYENGKYDIKDVGVDNAGAKAGLTFLVDLIKNKHMNADTDYSIAEAAF
NKGETAMTINGPWAWSNIDTSKVNYGVTVLPTFKGQPSKPFVGVLSAGI
NAASPNKELAKEFLENYLLTDEGLEAVNKDKPLGAVALKSYEEELAKDP
RIAATMENAQKGEIMPNIPQMSAFWYAVRTAVINAASGRQTVDEALKDA
QTNSSSNNNNNNNNNNLGLVPR-end of mbp-GLFGAIAGFIENGWEGM
IDGWYGFRHQNSEGTGQAADLKSTQAAIDQINGKLNRVIEKTNEKFHQIE
KEFSEVEGRIQDLEKYVEDTKIDLWSYNAELLVALENQHTIDLTDSEMNK
LFEKTRRQLRENAEEMGNGSFKIYHKCDNACIESIRNGTYDHDVYRDEAL
NNRFQIKGVELKSGYKDWILWISFAISSFLLAVVLLGFIMWASQRGNIRSN
ISI

Figure 2-4. Amino acid sequence of MHA2. The MHA2 construct contains the full length HA2 sequence attached to the maltose binding protein, with the end of the maltose binding protein noted.

AAAGATCTGCTGCCGAACCCGCCAAAAACCTGGGAAGAGATCCCGGC
GCTGGATAAAGAACTGAAAGCGAAAGGTAAGAGCGCGCTGATGTTCA
ACCTGCAAGAACCGTACTTCACCTGGCCGCTGATTGCTGCTGACGGGG
GTTATGCGTTCAAGTATGAAAACGGCAAGTACGACATTAAAGACGTG
GGCGTGGATAACGCTGGCGCGAAAGCGGGTCTGACCTTCCTGGTTGAC
CTGATTAAAAACAAACACATGAATGCAGACACCGATTACTCCATCGCA
GAAGCTGCCTTTAATAAAGGCGAAACAGCGATGACCATCAACGGCCC
GTGGGCATGGTCCAACATCGACACCAGCAAAGTGAATTATGGTGTAAC
GGTACTGCCGACCTTCAAGGGTCAACCATCCAAACCGTTCGTTGGCGT
GCTGAGCGCAGGTATTAACGCCGCCAGTCCGAACAAAGAGCTGGCAA
AAGAGTTCCTCGAAAACTATCTGCTGACTGATGAAGGTCTGGAAGCGG
TTAATAAAGACAAACCGCTGGGTGCCGTAGCGCTGAAGTCTTACGAGG
AAGAGTTGGCGAAAGATCCACGTATTGCCGCCACTATGGAAAACGCC
CAGAAAGGTGAAATCATGCCGAACATCCCGCAGATGTCCGCTTTCTGG
TATGCCGTGCGTACTGCGGTGATCAACGCCGCCAGCGGTCGTCAGACT
GTCGATGAAGCCCTGAAAGACGCGCAGACTAATTCGAGCTCGAACAA
CAACAACAATAACAATAACAACAACCTCGGG***CTGGTTCCGCGT***GGCCT
ATTCGGCGCAATAGCAGGTTTCATAGAAAATGGTTGGGAGGGAATGA
TAGACGGTTGGTACGGTTTCAGGCATCAAAATTCTGAGGGCACAGGAC
AAGCAGCAGATCTTAAAAGCACTCAAGCAGCCATCGACCAAATCAAT
GGGAAATTGAACAGGGTAATCGAGAAGACGAACGAGAAATTCCATCA
AATCGAAAAGGAATTCTCAGAAGTAGAAGGGAGAATTCAGGACCTCG
AGAAATACGTTGAAGACACTAAAATAGATCTCTGGTCTTACAATGCGG
AGCTTCTTGTCGCTCTGGAGAATCAACATACAATTGACCTGACTGACT
CGGAAATGAACAAGCTGTTTGAAAAAACACGTCGTCAACTGCGTGAA
AATGCTGAAGAGATGGGCAATGGTAGCTTCAAAATATACCACAAATG
TGACAACGCCTGCATAGAGTCAATCAGAAATGGGACTTATGACCATGA
TGTATACAGAGACGAAGCATTAAACAACCGGTTTCAGATCAAAGGTGT
TGAACTGAAGTCTGGATACAAAGACTGGATCCTGTGGATTTCCTTTGC
CATATCATCCTTTTTGCTTGCTGTTGTTTTGCTGGGGTTCATCATGTGG
GCCTCCCAGAGAGGCAACATTAGGTCCAACATTTCCATTTGA*AAGCTT*
GGCTGTTTTGGCGGATGAGAGAAGATTTTC

Figure 2-5. DNA sequence of MHA2. The region encoding for the thrombin cleavage site (leu-val-pro-arg) in bold, this marks the end of the maltose binding protein. Surrounding DNA is in italics.

3. gp41

transmemb

purificatio

AV
AQ
WY
QD

Figure 2-6

TT
GT
TC
CS
TT
GC
TC
GC
A
CT
GC
CT
TC
CT
GT
CT
AC

Figure 2-7

3. gp41. This construct contains the full gp41 protein, including the ectodomain, transmembrane region, and endodomain, with a C-terminal polyhistidine tag for purification.

AVGLGAVFLGFLGAAGSTMGAASMTLTVQARQLLSGIVQQQSNLLKAIE
AQQHLLKLTVWGIKQLQARVLAVERYLQDQQLLGIWGCSGKLICTSFVP
WNNSWSNKTYNEIWDNMTWLQWDKEISNYTDTIYRLLEDSQNQQEKNE
QDLLALDKWANLWNWFSITNWLWYIKLEHHHHHH

Figure 2-6. Amino acid sequence of gp41. This construct is the full gp41 protein.

*TTTTGTTAACTTTAAGAAGGAGATATACATATG*GCAGTTGGACTAGGAGCT
GTCTTCCTTGGGTTCTTGGGAGCAGCAGGGAGCACTATGGGCGCGGCG
TCAATGACGCTGACGGTACAGGCCAGACAATTATTGTCTGGCATAGTG
CACCAGCAAAGCAATTTGCTGAAGGCTATAGAGGCTCAACAGCATCTG
TTGAAACTCACGGTCTGGGGTATTAAACAGCTCCAGGCAAGAGTCCTG
GCTGTGGAAAGATACCTACAGGATCAACAGCTCCTGGGAATTTGGGGC
TGCTCTGGAAAACTCATCTGCACCTCTTTTGTGCCCTGGAACAATAGTT
GGAGTAACAAGACTTATAATGAGATTTGGGACAACATGACCTGGTTGC
AATGGGATAAAGAAATTAGCAATTACACAGACACAATATACAGGCTA
CTTGAAGACTCGCAGAACCAGCAGGAAAAGAATGAACAAGACTTATT
GGCATTAGATAAATGGGCAAATTTGTGGAATTGGTTTAGCATAACAAA
CTGGCTGTGGTATATAAAGCTCGAGCACCACCACCACCACCAC*TGAGA*
TCCGGCTGCTAACAAAGCCCGAAAGGAAGCTGAGTTGGCTGCTGCCACCG
CTGAGCAATAACTAGCATAACCCCTTGGGGCCTCTAAACGGGTCTTGAGGG
GTTTTTTGCTGAAAGGAGGAACTATATCCGGATTGGCGAATGGGACGCGCC
CTGTAGCGGCGCATTAAGCGCGGCGGGTGTGGTGGTTACGCGCAGCGTG
ACCGCTACACTTGCCAGCGCCCTAGCGCCCGCTCCTTTC

Figure 2-7. DNA of gp41. The surrounding regions are shown in italics.

4. Fgp41. The Fgp41 construct contains the ectodomain and a C-terminal polyhistidine tag for purification.

AVGLGAVFLGFLGAAGSTMGAASMTLTVQARQLLSGIVQQQSNLLKAIE
AQQHLLKLTVWGIKQLQARVLAVERYLQDQQLLGIWGCSGKLICTSFVP
WNNSWSNKTYNEIWDNMTWLQWEKEISNYTDTIYRLLEDSQNQQEKNE
QKLLALDKLEHHHHHH

Figure 2-8. Amino acid sequence of Fgp41. This construct contains the full ectodomain of the gp41 protein.

*TTTTGTTAACTTTAAGAAGGAGATATACATATG*GCAGTTGGACTAGGAGCT
GTCTTCCTTGGGTTCTTGGGAGCAGCAGGGAGCACTATGGGCGCGGCG
TCAATGACGCTGACGGTACAGGCCAGACAATTATTGTCTGGCATAGTG
CACCAGCAAAGCAATTTGCTGAAGGCTATAGAGGCTCAACAGCATCTG
TTGAAACTCACGGTCTGGGGTATTAAACAGCTCCAGGCAAGAGTCCTG
GCTGTGGAAAGATACCTACAGGATCAACAGCTCCTGGGAATTTGGGGC
TGCTCTGGAAAACTCATCTGCACCTCTTTTGTGCCCTGGAACAATAGTT
GGAGTAACAAGACTTATAATGAGATTTGGGACAACATGACCTGGTTGC
AATGGGATAAAGAAATTAGCAATTACACAGACACAATATACAGGCTA
CTTGAAGACTCGCAGAACCAGCAGGAAAAGAATGAACAAGACTTATT
GGCATTAGATAAACTCGAGCACCACCACCACCACCAC*TGAGATCCGGC*
TGCTAACAAAGCCCGAAAGGAAGCTGAGTTGGCTGCTGCCACCGCTGAGC
AATAACTAGCATAACCCCTTGGGGCCTCTAAACGGGTCTTGAGGGGTTTTTT
GCTGAAAGGAGGAACTATATCCGGATTGGCGAATGGGACGCGCCCTGTAG
CGGCGCATTAAGCGCGGCGGGTGTGGTGGTTACGCGCAGCGTGACCGCT
ACACTTGCCAGCGCCCTAGCGCCCGCTCCTTTC

Figure 2-9. DNA sequence of Fgp41. Surrounding DNA is shown in italics.

5. Hgp41. This construct contains the hairpin region of the gp41 protein with no added regions for purification.

CTLTVQARQLLSGIVQQQNNLLRAIEAQQHLLQLTVWGIKQLQARILSGG
RGGWMEWDREINNYTSLIHSLIEESQNQQEKNEQELLELDKW

Figure 2-10. Amino acid sequence of Hgp41. This construct contains the hairpin region of gp41.

CTCTAGAAATAATTTTGTTTAACTTTAAGAAGGAGATATACATATG
TGCACGCTGACGGTACAGGCCAGACAATTATTGTCTGGTATAGTGCAG
CAGCAGAACAATTTGCTGAGGGCTATTGAGGCGCAACAGCATCTGTTG
CAACTCACAGTCTGGGGCATCAAGCAGCTCCAGGCAAGAATCCTGTCT
GGTGGCCGTGGCGGTTGGATGGAGTGGGACAGAGAAATTAACAATTA
CACAAGCTTAATACACTCCTTAATTGAAGAATCGCAAAACCAGCAAGA
AAAGAATGAACAAGAATTATTGGAATTAGATAAATGG
TGATAGGGATCCTAATCACTAGTGCGGCCGGCCTGCAGGTCGACCATATG
GGAGAGCT

Figure 2-11. DNA sequence of Hgp41. The surrounding DNA is in italics.

C. Gel Electrophoresis

Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) analysis of protein mixtures was conducted using 12% acrylamide gels of 0.75 mm thickness with 10 sample lanes. The running portion of the gel (the lower 80%) was poured from a solution of containing 2.5 mL 1.5 M Tris pH 8.8, 100 μL 10% w/v SDS, 4 mL acrylamide/Bis 30% stock, 50 μL 10% ammonium persulfate, 3.3 mL water, and 15 μL TEMED added to initiate polymerization. The stacking portion of the gel (the upper 20%) was poured from a solution containing 1.25 mL 0.5 M Tris pH 6.8, 50 μL 10% w/v SDS, 0.65 mL acrylamide/Bis 30% stock, 25 μL 10% ammonium persulfate, 3 mL water, and 10 μL TEMED. Prior to loading samples onto the gel, protein solutions were diluted to approximately 50% in sample buffer containing 1 mL 0.5 M Tris pH 6.8, 0.8 mL glycerol, 1.6 mL 10% (w/v) SDS, 0.4 mL 2-mercaptoethanol, 0.4 mL 0.05% (w/v) bromophenol blue and 5.4 mL water. Following loading, the protein was advanced through the gel by the application of 200 mV current. Protein bands were visualized by overnight staining in coomassie stain containing 2.5 g coomassie blue, 900 mL methanol, 100 mL acetic acid. Repeated steps of destaining removed the background producing a

clear gel w

140% meth

solution 2 (

D. Culture

All

containing

kanamycin

resistance.

from a gly

contained

overnight

shaking at

10 minutes

medium fo

commercia

NaCl 1.0 g

for unlabele

Mar

glucose con

optimizatio

Erlenmeyer

determined a

clear gel with blue protein bands. The destaining was first repeated in destain solution 1 (40% methanol, 10% acetic acid, 50% water) followed by the final destain in destain solution 2 (40% methanol, 10% acetic acid, 48% water, 2% glycerol).

D. Culture Growth

All cell cultures were grown in media containing antibiotic to select for cells containing the desired plasmid. Expression of FHA2 and Fgp41 included 15 mg/L kanamycin, MHA2 and Hgp41 included 50 mg/L ampicillin. Other than the antibiotic resistance, the expression of all constructs was identical. Bacterial growth was initiated from a glycerol stock of the recombinant bacterial cells to 1 L of "enriched LB" which contained LB supplemented with 10 mL glycerol. The cell suspension was grown overnight to maximum cell density in a 2.8 L baffled Fernbach flask with a foam closure, shaking at 37°C and 140 rpm. The cell suspension was then centrifuged at 10,000 x*g* for 10 minutes to produce a solid cell pellet. The pellet was resuspended into 1 L of minimal medium for isotopically labeled expression, whose optimal composition included the commercial M9 minimal medium salts (6.8 g/L Na_2HPO_4, 3.0 g/L NaH_2PO_4, 0.50 g/L NaCl, 1.0 g/L NH_4Cl), 2.5 g/L $MgSO_4$, and 10 g/L glycerol at pH 8.0, or in enriched LB for unlabeled expression. Cell growth was continued by shaking at 37°C and 140 rpm.

Many of the experiments in this chapter examine the effects of parameters such as glucose concentration on cell growth using FHA2 as the model protein. Each growth optimization experiment was done using a 100 mL culture volume in a 250 mL baffled Erlenmeyer flask and all of the growths as a function of a single parameter were performed at the same time in the same shaker. All of the comparative experiments

presented in this chapter were repeated and for each parameter, the trends in cell growth were reproducible. The value of the parameter which yielded the highest growth was also reproducible. Examination of all of the A600 or ΔA600 values between the different comparative growths showed that there was systematic variation by as much as 20%, e.g. for each parameter value, $(A600)_{\text{first growth}}/(A600)_{\text{second growth}}$ would be ~1.2. It was therefore difficult to estimate error bars, which would reflect data uncertainties within a single comparative growth.

E. Isotopic Labeling

After one hour of resuspended cell growth in optimized minimal medium, protein expression was induced by addition of IPTG to a final concentration of 0.2 mM. For isotopic labeling, 100 mg of each labeled amino acid to be incorporated was added per liter of expression at the time of induction. Protein production was continued for three hours at 37°C. The cell pellet was harvested by centrifugation at 10,000 x*g* for 10 minutes, and the pellet was then stored at –80°C until purification.

F. Purification of Native FHA2 and Fgp41

The FHA2 and Fgp41 purifications were based on the C-terminal His-tag. The purification was conducted at room temperature, but all buffers were refrigerated at 4°C immediately prior to use to limit protease activity. The following protocol provided optimal purity and yield from 5 g of FHA2 cells. Purification of larger cell quantities should follow the same procedures but may require optimization of specific values. Cells (5 g) were suspended in 25 mL of wash buffer A (50 mM sodium phosphate at pH 8.0,

300 mM N

detergent and

a 1 minute d

cycles with

centrifugation

bound to 0.5

LabQuake sh

the pellet wa

of wash buf

column vo

piperazinee

(MES), 20 m

wash buffer

at pH 7.4). P

buffer (5 mM

7.4). Protein

M^{-1} cm^{-1} an

tyrosines and

G Purificati

Purif

were lysed in

$MgCl_2$ and

300 mM NaCl, 20 mM imidizole, and 0.5% (w/v) *N*-lauroylsarcosine (Sarkosyl) detergent and cell walls were lysed during 4 sonication periods of 1 minute duration with a 1 minute delay between sonication periods. Each period contained 0.8 s on/0.2 s off cycles with 80% amplitude during the on cycle. Cell debris was removed by centrifugation at 48000 x*g* and 4°C for 20 minutes. Fusion protein in the supernatant was bound to 0.5 mL of chelated cobalt His-Select resin with one hour of mixing on a LabQuake shaker. A resin pellet was formed by centrifugation at 1000*g* for 1 minute and the pellet was transferred to a column and then washed with 1 column volumes (0.5 mL) of wash buffer A. Detergent/buffer exchange was then completed by washing with 1 column volume each of wash buffer B (5 mM 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES), 10 mM 2-(*N*-morpholino)ethanesulfonic acid, (MES), 20 mM imidazole, 0.5% Sarkosyl, 0.5% BTOG, and 0.4% C8E5 at pH 7.4) and wash buffer C (5 mM HEPES, 10 MES, 20 mM imidazole, 0.5% BTOG, and 0.4% C8E5 at pH 7.4). Protein was eluted from the resin with 5 column volumes (2.5 mL) of elution buffer (5 mM HEPES, 10 MES, 250 mM imidazole, 0.5% BTOG, and 0.4% C8E5 at pH 7.4). Protein concentrations were quantified using A280 with the FHA2 $\varepsilon 280 = 34{,}000$ M^{-1} cm^{-1} and Fgp41 $\varepsilon 280 = 46{,}000$ M^{-1} cm^{-1} which was calculated from the numbers of tyrosines and tryptophans in the sequence and their extinction coefficients.

G. Purification of Native MHA2

Purification of the MHA2 protein is based on a published procedure (*6*). The cells were lysed in a lysis buffer containing 0.2 mg/ml lysozyme, 0.1 mg/ml DNAse 1, 10 mM $MgCl_2$, and 1% decyl maltoside. The lysate was centrifuged at 50,000 xg for 20 minutes.

The supernat

The column

with 0.17%

containing 0

H. Circular

Circ

controller to

Kingdom), s

separated b

difference s

circular dic

a quantity

concentrati

Lipid Mi

The

The mixtur

based on th

PC). The

POPC), 1

Cholesteryl-s

PI, and ch

The supernatant was washed over amylose resin at a rate no exceeding 1 ml per minute. The column was washed with 10 column volumes of phosphate buffered saline (PBS) with 0.17% decyl maltoside. The protein was eluted with 5 column volumes of PBS containing 0.17% decyl maltoside and 10 mM maltose.

H. Circular Dichroism Spectroscopy

Circular dichroism spectra were obtained on an instrument with a temperature controller to maintain constant at 4°C (Chirascan, Applied Photophysics, Surrey, United Kingdom), a 1 mm pathlength cuvette, a 200-260 nm spectral window, wavelength points separated by 0.5 nm, and 0.5 seconds signal averaging per point. For each sample, a difference spectrum was obtained by subtracting the sample buffer from the protein. The circular dichroism signal was reported in units of mean residue molar ellipticity, which is a quantity normalized to the concentration of protein residues, and therefore protein concentration.

I. Lipid Mixing Assay

The first step in the lipid mixing assay is the preparation of unilamellar vesicles. The mixture of lipids selected for this assay was termed the "LM3" mixture and was based on the composition of human epithelial cells, the target of the influenza infection (*37*). The LM3 mixture contained 1-palmitoyl-2-oleoyl-*sn*-glycero-3-phosphocholine (POPC), 1-palmitoyl-2-oleoyl-*sn*-glycero-3 phosphoethanolamine (POPE), 1-palmitoyl-2-oleoyl-*sn*-glycero-3-[phospho-L-serine] (POPS), sphingomyelin, phosphatidylinositol (PI) and cholesterol in a 10:5:2:2:1:10 mol ratio. One set of LUVs was made from LM3

(5.4 µmol to

LMB mixtur

the lipid N-

2 mol% of

PE). Lipids

with nitrog

was added

homogeniz

polycarbon

On

this assay

fluorescent

and unlab

quenching

monitored

of each ur

150 µM.

Spectrimete

and emiss

the vesicl

addition o

concentra

the initial

(5.4 μmol total lipid + 2.7 μmol cholesterol) mixture while another set was made from LM3 mixture (0.6 μmol total lipid + 0.3 μmol cholesterol) and an additional 2 mol% of the lipid *N*-(7-nitro-2,1,3-benzoxadiazol-4-yl)-phosphatidylethanolamine (NBD-PE) and 2 mol% of the lipid *N*-(lissamine Rhodamine B sulfonyl) phosphatidylethanolamine (Rh-PE). Lipids and cholesterol were dissolved in chloroform and the solvent then removed with nitrogen gas and overnight vacuum pumping. A 1 ml aliquot of HEPES/MES buffer was added to each of the dry lipid/cholesterol mixture and the dispersion was homogenized with ten freeze-thaw cycles. LUVs were prepared by extrusion through a polycarbonate filter with 100 nm diameter pores (*38*).

One feature of vesicle fusion is mixing of lipids between different vesicles, and in this assay between vesicles containing different lipids. The NBD and Rh groups are fluorescent and quenching functionalities. Lipid mixing between the fluorescently labeled and unlabeled LUVs will increase the average distance between fluorescent and quenching lipids and an increase in fluorescence will be detected (*39*). Fluorescence was monitored in mixtures made from ~1.9 mL of HEPES/MES buffer at pH 7.4 or 5, 50 μL of each unlabeled and labeled LUV solution. The resultant solutions had [total lipid] = 150 μM, [total cholesterol] = 75 μM, and [labeled LUV]/[unlabeled LUV] = 0.11. The fluorimeter (Fluoromax 2, HORIBA Jobin Yvon Inc, Edison, NJ) was set with excitation and emission wavelengths of 465 and 530 nm. For each data set, the "F_0" fluorescence of the vesicle solution was recorded and then the "F_1" fluorescence was detected after addition of an aliquot of FHA2 solution in HEPES/MES buffer pH 7.4. The typical FHA2 concentration was ~100 μM and the volume of buffer was adjusted in each trial so that the total solution volume $V_{FHA2} + V_{vesicles} + V_{buffer} = 2.0$ mL. F_1 was the equilibrium

fluorescence

change was

fluorimeter.

added to the

large averag

fluorescence

[40].

Perc

The typical

different tr

fluorescenc

or contain

be attribute

2. Membran

The

tio of the

tradecyl ph

predominan

influenza vi

tumor fra

ter-linked

tion prob

fluorescence of the FHA2/LUV solution as the time for the FHA2-induced fluorescence change was shorter than the ~2 s between addition of the FHA2 aliquot and closing the fluorimeter. After measurement of F_1, a 10 μL aliquot of a 20% w/v triton solution was added to the FHA2/LUV solution. The triton solubilized the vesicles and resulted in a large average distance between the fluorescent and quenching lipids and a maximum "F_2" fluorescence. The "percent lipid mixing" was calculated using the literature convention (*40*):

$$\text{Percent lipid mixing} = 100 \times (F_1 - F_0)/(F_2 - F_0)$$

The typical variation in percent lipid mixing was ±2% as determined by comparison of different trials with the same vesicle solutions and FHA2 stocks. As a control, fluorescence was also recorded with addition of aliquots of HEPES/MES buffer that did not contain FHA2. There was very little change in fluorescence and $F_1 - F_0$ can therefore be attributed to FHA2, rather than BOG, induced lipid mixing.

J. Membrane Reconstitution of FHA2.

The membrane composition used for solid-state NMR samples was a 4:1 molar ratio of the ether linked lipids di-O-tetradecylphosphatidyl-choline (DTPC) and di-O-tetradecylphosphatidylglycerol (DTPG) and was chosen because: (1) choline is a predominant headgroup of lipids of membranes of respiratory epithelial host cells of the influenza virus; (2) the headgroup of DTPG is negatively charged like the headgroups of a minor fraction of the host cell lipids; and (3) DTPC and DTPG are ether- rather than ester-linked lipids and do not have a natural abundance ^{13}C contribution to the carbonyl region probed in the NMR experiments (*41, 42*). The lipids (~40 mg total) and the

detergent BTOG (~160 mg) were dissolved in chloroform. The solvent was removed by a stream of nitrogen gas and subsequent overnight pumping in a vacuum chamber. The lipid/detergent mixture was then dissolved in ~5 mL of HEPES/MES buffer at pH 7.4. The FHA2 solution was added to the detergent/lipid solution to form a co-micelle solution of ~8 mg FHA2, detergent, and lipid. The solution was transferred to 10,000 MWCO tubing and dialyzed against 2 L of HEPES/MES buffer at pH 5.0. This pH is comparable to the one for fusion between influenza and endosomal membranes. The dialysis was done at 4°C for three days with one buffer change. The FHA2 reconstituted in membranes was then harvested by centrifugation at 50000*g* for 3 hours.

III. Fusion Protein Expression

A. Optimization of FHA2 expression in E. coli.

Expression of amino acid specific isotopically labeled protein requires induction in minimal medium. However, when minimal medium was used for all stages of bacterial growth the yield of FHA2 protein was less than 0.5 mg FHA2 per liter of culture. High-resolution structural methods such as NMR typically require at least 5 mg of protein, so this would require large volumes of bacterial culture as well as large quantities of labeled amino acids. There are at least two general approaches to improving protein yield: increasing cell density and increasing protein expression for each cell. This chapter describes progress with both approaches.

The effort to increase cell density was focused on an initial growth of the cells in rich medium, followed by centrifugation and resuspension of the cells in minimal medium. Fig. 2-12a shows a comparison of the cell densities after growth in different media: minimal medium, LB, and LB supplemented with 10 g/L glycerol. The highest cell density was achieved with enriched LB and these cells also had the best subsequent growth in minimal medium. Fig. 2-12b demonstrates that the final cell density obtained with cell growth in rich medium and then minimal medium is ~5 times greater than the density obtained with growth only in minimal medium.

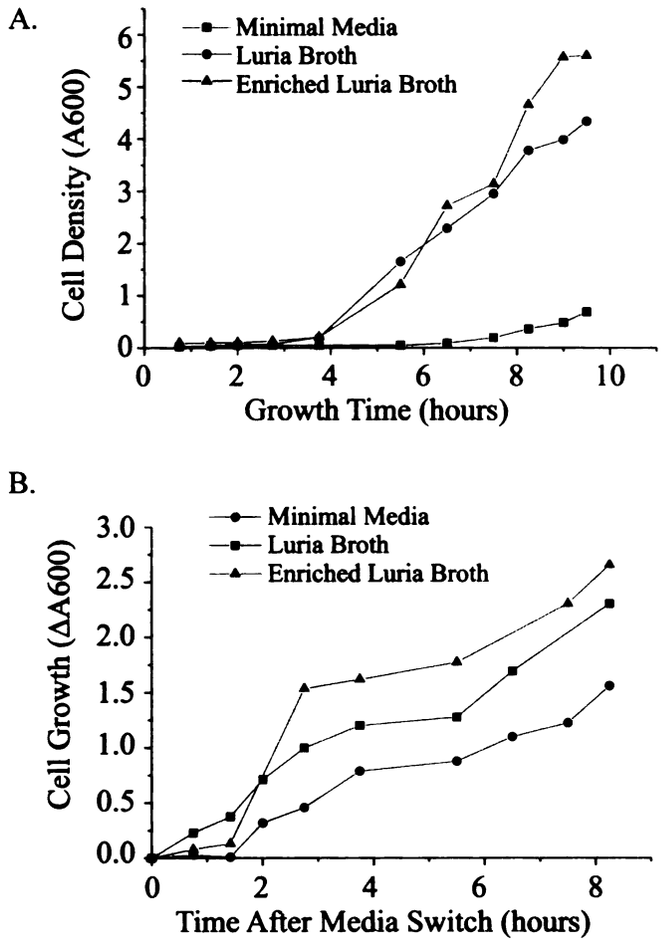


Figure 2-12. Cell growth in minimal media, LB, and enriched LB. (A) The initial cell growth varies with media. Minimal medium shows the slowest cell growth and yields the lowest maximum density, while LB and LB enriched with 10 g/L of glycerol show more rapid initial cell growth, and enriched LB yields the highest final cell density. (B) Cell growth continues after the switch into minimal medium and the greatest growth is from enriched LB.

Because the cell densities for our protein expression are well above the normal range for shake flask fermentation, oxygenation may be a limiting factor on cell growth (*13, 14*). The effect of oxygenation was first investigated by comparing growth with different types of flasks and closures, see Fig. 2-13. Baffled flasks are designed to increase oxygenation of the medium, and in our case significantly improved cell growth.

The ty

(used

better

of usi

from 5

Figure
growth

resusp

growth

Medium

above t

varied

reductio

experim

of initia

The type of closure also had an effect on the oxygenation of the cell culture. Parafilm (used as a control), foil, or a foam plug were compared, and the foam plug allowed for better oxygenation and increased growth. These improvements to the traditional method of using flat bottom Erlenmeyer flasks with foil closures increased the final cell density from 5 to 9.

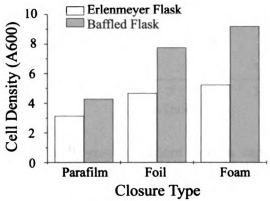


Figure 2-13. Effect of oxygenation on cell growth. The total cell density after overnight growth in enriched LB for different combinations of flask and closure types.

The next set of experiments focused on growth in minimal medium after the resuspension and were based on the hypothesis that there was a correlation between cell growth in this medium and expressed protein yield. The initial choice for the minimal medium was the commercially available M9 mix. Because the cell density was well above the typical density for this mixture, parameters such as pH and carbon source were varied to attain maximum cell growth.

As discussed earlier, growth of bacteria at high cell density may be impaired by a reduction in pH correlated with over-production of acetic acid and this was experimentally observed, see Fig. 2-14. The cell growth was then monitored as a function of initial pH of the minimal medium, see Fig. 2-15, and the highest growth was found for

pH 8.0. Pre

more amena



Figure 2-14.
resuspension

Figure 2-15.
switch was m
represents the
contained 10 g

Glycer

Glycerol into t

nd consequer

pH 8.0. Presumably, the pH reduction from an initial pH of 8.0 occurs in a pH range more amenable to bacterial growth.

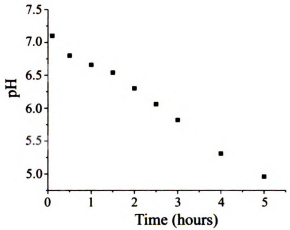


Figure 2-14. Drop in pH with cell growth. There was a rapid drop in pH after resuspension of the cells in minimal medium containing 10 g/L glucose.

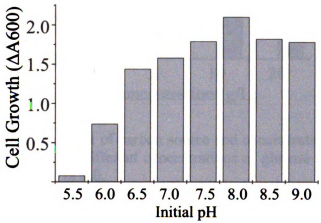


Figure 2-15. Cell growth as an effect of initial pH. The cell growth after the medium switch was monitored as a function of the initial pH of the minimal medium. Each bar represents the change in A600 3 hours after the medium switch. The minimal medium contained 10 g/L glucose.

Glycerol may be a better carbon source than glucose because the uptake of glycerol into the cell occurs at a lower rate with less saturation of metabolic pathways and consequent lower production of acetate and reduction in pH (*19, 20*). Experiments

were therefore carried out to study the effect of glycerol vs. glucose on bacterial growth after the medium switch, see Fig. 2-16. For glycerol, the highest growth was obtained with 10 g/L and significantly less growth was observed with 5 or 20 g/L. For glucose, the highest growth was obtained with 5 g/L although reasonable growth was also obtained with 10 or 20 g/L. For 10 g/L of glucose, the pH of the medium was 5.0 after three hours of growth and for 10 g/L of glycerol, the pH was 6.4. This result correlated with the difference in cellular uptake rates of glucose and glycerol.

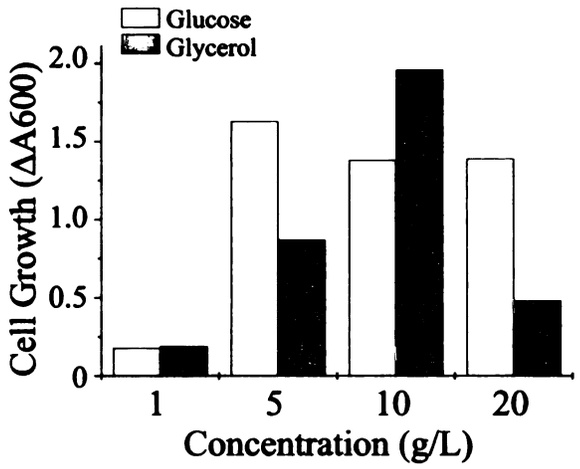


Figure 2-16. Cell growth as an effect of carbon source and concentration. Cell growth 3 hours after the medium switch for different concentrations of glucose or glycerol in the minimal medium with initial pH of 8.0.

In summary, the highest cell density of A600 ≈ 9 was achieved with: (1) overnight growth in LB enriched with 10 g/L glycerol; and (2) a switch to minimal medium containing M9 salts, 2.5 g/L $MgSO_4$, and 10 g/L glycerol at pH 8.0. For these conditions, there were ~10 g of wet cell mass per liter of culture.

B. Expression of MHA2.

The MHA2 plasmid was obtained from Susanne Swalley and transformed into BL21(DE3) cells. The expression of MHA2 was initially attempted using a previously published method, but the procedure optimized for FHA2 expression yielded similar results in a simpler process (*6*). The MHA2 protein expressed in much higher yields than the FHA2 protein, which can be seen in the gel of the whole cells following transformation, shown in figure 2-17.

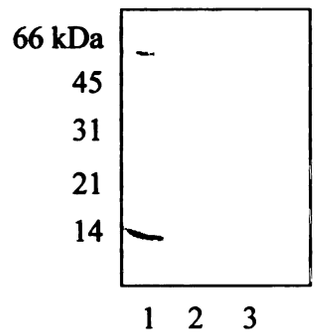


Figure 2-17. MHA2 expression. Gel electrophoresis of the whole cells expressed following transformation. Lane 1-standards, Lane 2- 1 μl DNA used in transformation, Lane 3- 3 μl DNA used in transformation.

C. Expression of gp41 and Hgp41.

The expression of both Fgp41 and gp41 initially appeared to produce much less protein than the FHA2 expression. Attempts to visually monitor the production of protein using gel electrophoresis of the whole cells was unsuccessful, as the amount of expressed protein was not above the background proteins levels of the sample. Conditions such as the IPTG concentration, cell density at induction, and amount of time allowed for protein production were varied, but with no significant increase in expression as measured

through gels and purification yields. As discussed in a later section of this chapter, purification of Fgp41 was successful, but only obtained 0.1 mg per liter of fermentation, much below the level required for solid-state NMR experiments. No detectable level of the full-length construct was ever obtained from purifications.

As an alternative method to monitor the production of recombinant protein, a labeled expression was conducted and the produced protein analyzed by whole cell NMR. The details of this experiment are discussed in more detail in chapter 6, but the results indicated a significant amount of Fgp41 present in the cell, comparable to the amount of FHA2 produced. The low level of protein obtained from purifications indicates that the majority of the produced protein is in the form of inclusion bodies and not natively folded protein. The expression of Hgp41 was successful and high yielding. This protein was not expressed for the purpose of native purification and is discussed in detail in chapter 5.

IV. Native Purification of Fusion Proteins

A. Purification of Native FHA2 from the Soluble Cell Lysate.

Purification of the FHA2 protein from *E. coli* cells takes advantage of the poly-histidine tag attached to the C-terminus of the protein. Initial attempts at purification using nickel resin were hindered by the presence of SlyD in the cell lysate, as shown in figure 2-18. Attempts to load the resin with more cell lysate revealed that the contaminant protein was binding as strongly to the resin as FHA2, the proportion of the two proteins were always near equal. Denaturing purification also resulted in a mixture of protein. The

use of cobalt resin in place of the nickel was an effective solution to the problem, and pure FHA2 was obtained as shown in figure 2-18.

The cells were lysed by sonication, the insoluble components separated by centrifugation, and the soluble cell lysate bound to a chelated cobalt resin. Contaminants were removed by washing with a low concentration of imidazole buffer solution, and high purity protein was eluted with high concentration imidazole. Sarkosyl was chosen to initially solubilize the protein during sonication but this detergent is difficult to remove after purification. Residual sarkosyl can interfere with membrane reconstitution and has a large background signal in circular dichroism spectroscopy. The sarkosyl detergent was therefore exchanged for a more desirable mixture while the protein was bound to the resin. The first wash buffer contained both sarkosyl, the second wash buffer contained both sarkosyl and the target detergent mixture of BTOG and C8E5, and the third wash buffer and the elution buffer contained only the target detergent mixture. Relative to other detergents, FHA2 aggregation appeared smaller in the mixture of BTOG and C8E5 detergents. Evidence for removal of most of the sarkosyl by this procedure included: (1) no obvious precipitation after addition of solutions containing divalent cations; (2) high signal-to-noise in the circular dichroism spectra without interference from strong far ultraviolet absorption; and (3) ability to reconstitute the FHA2 into membranes. Prior to exchange of the sarkosyl, there was respectively: (1) obvious precipitation; (2) low signal-to-noise because of the far ultraviolet absorption of Sarkosyl; and (3) FHA2 did not reconstitute well into membranes. The typical yield of FHA2 from this purification was 8-10 mg/L culture.

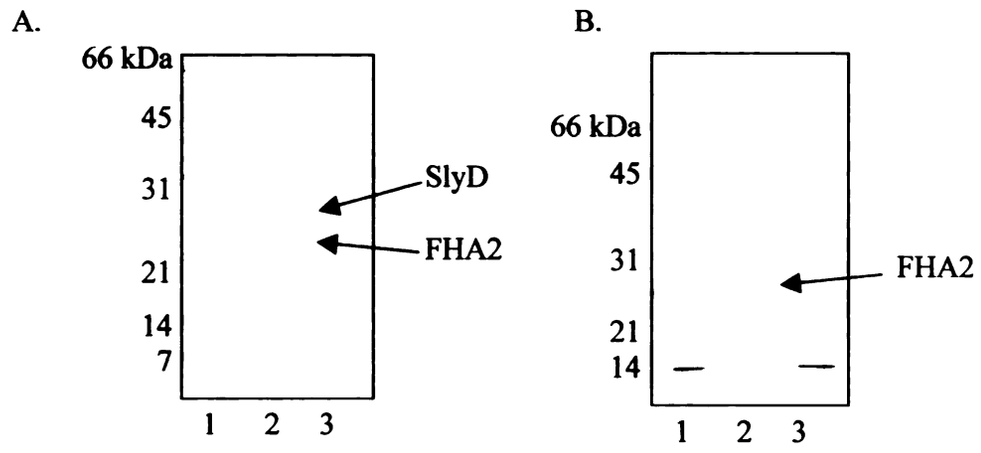


Figure 2-18. Purification of FHA2 using nickel and cobalt resins. (A) Purification with nickel resin results in a large fraction of SlyD in the elution. (B) Purification with cobalt resin results in qualitatively pure FHA2.

B. Purification of MHA2.

The purification of MHA2 using amylose resin was successful at obtaining high purity protein. Figure 2-19 shows the protein solution obtained by eluting the MHA2 from the resin using a maltose solution.

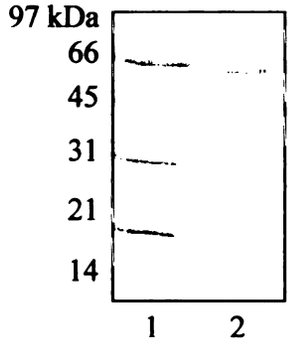


Figure 2-19. Purification of MHA2. Lane 1-standards, Lane 2-purified MHA2.

Previous reports of MHA2 purification included several protease inhibitors. In certain cases, proteases natively found with the bacterial cell can begin to degrade the

protein during the purification steps. Completing the reactions at 4°C to lower the protease activity can typically lessen this, or minimizing the time required to complete the purification can help. Initial purifications of the MHA2 protein without the protease inhibitors did not appear to be degraded. To test for the presence of protease activity, samples of the MHA2 cell lysate were allowed to incubate at varying temperatures and lengths of time. Following each condition, the samples were frozen until gel electrophoresis analysis. The gel in figure 2-20 shows that there is no detectable change in the protein composition. This indicates that for our experiments, we do not need to use protease inhibitors.

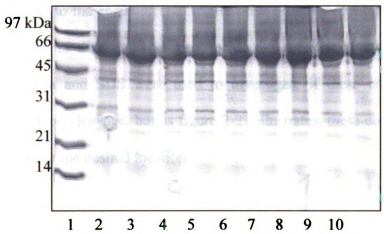


Figure 2-20. Proteolysis activity of MHA2 purification. Lane 1-standards, Lane 2-immediately frozen sample, Lane 3-4°C for 1 hr, Lane 4-23°C for 1 hr, Lane 5-4°C for 2 hr, Lane 6-23°C for 2 hr, Lane 7-4°C for 4 hr, Lane 8-23°C for 4 hr, Lane 9-4°C for 18 hr, Lane 10-23°C for 18 hr.

C. Proteolytic Cleavage of the Maltose Binding Protein.

Expression and purification of the full length HA2 protein bound to the maltose binding protein were successful. The next step in the preparation of protein suitable for solid state NMR structural studies is to remove the maltose binding protein. Not only

could such a large protein affect the structure of HA2, but the intended studies will use the REDOR pulse sequence, which require a large proportion of unique sequential pairs in the amino acid sequence. As the protein lengthens, the number of unique pairs will lessen, making these studies more difficult.

This construct was designed with a thrombin protease recognition sequence in the linker connecting the two proteins. Initial attempts to use thrombin on the protein solution obtained from column elution were not successful.

The next attempt at removing the maltose binding protein involved using slightly denaturing conditions. If the cleavage site is buried within the folded protein complex, it may not be accessible to the enzyme. The inherent risk of using denaturants to improve the protease reaction is that it could also lower the activity of the enzyme. Low concentrations of urea and SDS were used to investigate if these conditions would enhanced the proteolytic cleavage, but as figure 2-21 illustrates, these conditions did not enhance the cleavage at the desired location.

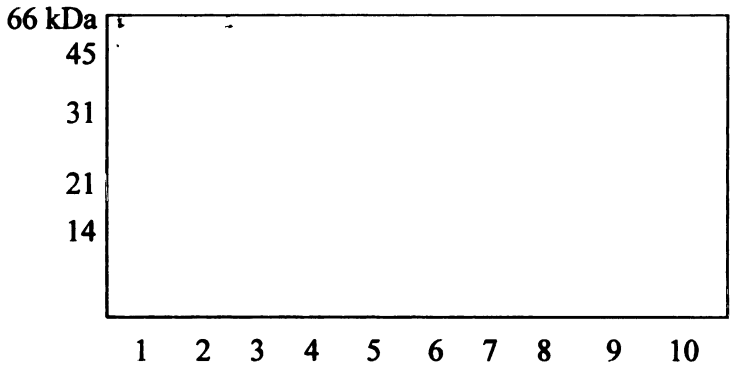


Figure 2-21. Use of denaturing conditions to increase proteolytic activity. Lane 1-standards, Lane 2-.01% SDS, Lane 3-0.025% SDS, Lane 4-0.05% SDS, Lane 5-1% SDS, Lane 6-.1 M urea, Lane 7-.5 M urea, Lane 8-1 M urea, Lane 9-2 M urea, Lane 10-3 M urea

D. Fgp41 purification.

The purification of gp41 was conducted following the same procedures are FHA2, but with much lower yield. Attempts to further optimize the purification were unsuccessful. Despite the difficulties obtaining large quantities of the protein, high purity was accomplished, figure 2-22.

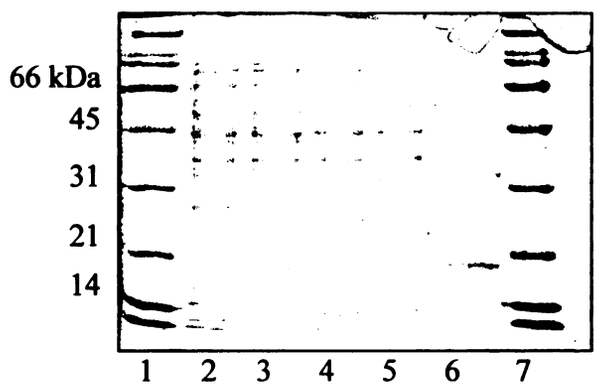


Figure 2-22. Purification of Fgp4. Lanes 1 and 7-standards, Lanes 2, 3, 4, and 5-column washes, Lane 6-column elution, purified Fgp41.

V. Stru

A. Cir

dichro

physio

helical

The sli

as the

differer

Figure 2
Observed
sults of

B. Lipid

Th

hiking he

V. Structural and Functional Assays

A. Circular Dichroism of FHA2.

In order to assess the secondary structure, and therefore the folding, circular dichroism was conducted on the purified FHA2. The protein was analyzed at both the physiological pH of 7.4 and the fusion active pH of 5. Both results indicate a primarily helical structure, with minimum points of the absorption occurring at 208 and 220 nm. The slight difference between the pH indicates that a conformational change may occur as the pH is lowered, which is consistent with the observation that the protein possess different levels of activity at these two pH values, figure 2-23.

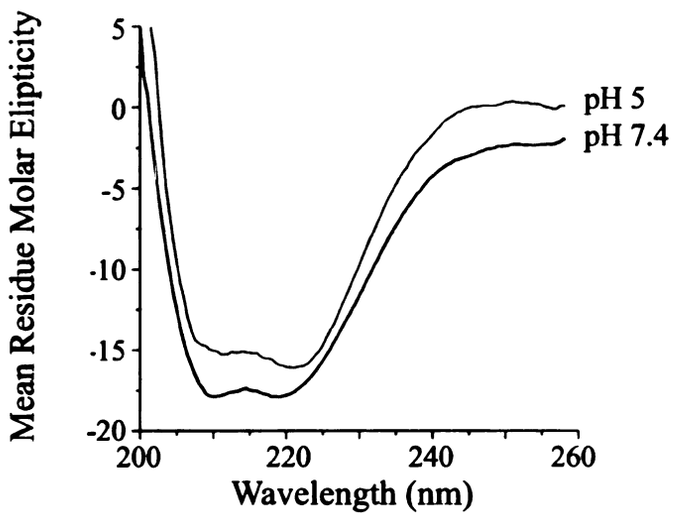


Figure 2-23. Circular dichroism of the FHA2 protein. The secondary structure was observed at both pH 5 (top) and pH 7.4 (bottom). Mean Residue Molar Elipticity is in units of 10^3 deg-cm^2-dmol^{-1}.

B. Lipid Mixing of FHA2 and Fgp41.

The role of the fusion protein in viral infection is to catalyze the initial lipid mixing between the host and viral cell membranes. A simple way to test the activity of

the puri

observe

mixing

speed a

Second

addition

produc

small p

of thes

the soli

the flu

In orde

deterg

resultin

of prot

is [illegible]

the purified product is to add the protein to a solution containing lipid vesicles and observe the interaction. This was accomplished by two complimentary methods. First, mixing assays were conducted using fluorescently labeled lipids that gave insight into the speed and completeness of the process at varying protein:lipid ratios, figures 2-24. Second, electron microscopy was used to visualize lipid vesicles before and after the addition of the fusion protein to investigate the structure and order of the lipid mixing product.

The fluorescence assay was conducted on a mixture of lipid vesicles containing a small proportion with both fluorescent and quenching lipids. Initially, the close proximity of these lipids resulted in a low level of observed fluorescence, but as protein is added to the solution and the vesicles containing the labeled lipids mix with the unlabeled vesicles the fluorescent and quenching lipids are diluted, resulting in an increase of fluorescence. In order to quantify the percent lipid mixing, the final step of the assay is to add the detergent triton, which will break up the vesicles and completely disperse the lipids, resulting in maximum fluorescence. The initial fluorescence observed before the addition of protein is the 0% lipid mixing value, and the observed level after the addition of triton is 100% lipid mixing.

90%

10%

Figure
protei
mixin
mixed
fluore

is muc

in wh

with t

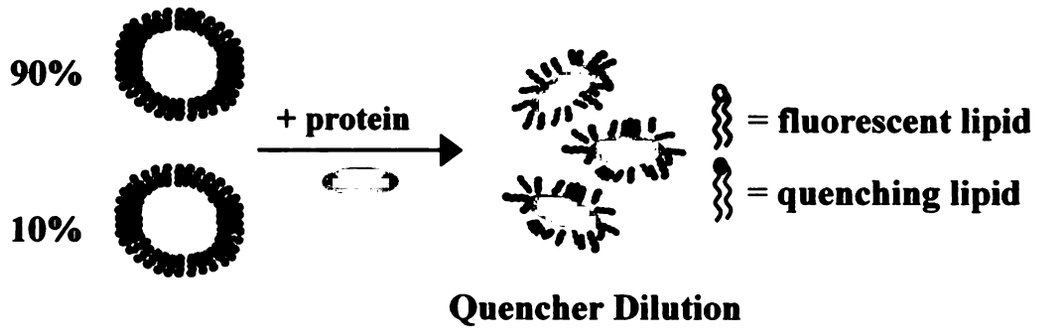
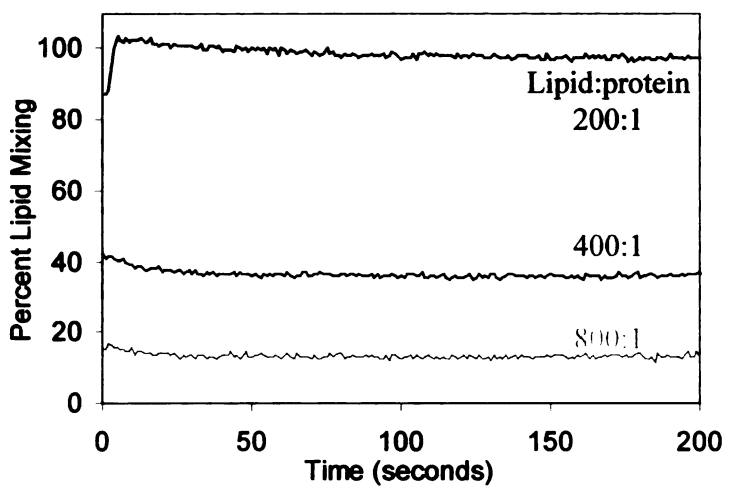


Figure 2-24. Fluorescent lipid mixing assay. This is used to determine the activity of the protein. A small fraction of vesicles containing both quenching and fluorescing lipids is mixing with a unlabeled vesicles. When protein is added to the mixture, the lipids are mixed and the quenching and fluorescing lipids are diluted resulting in an increase in fluorescence.

Figure 2-25 shows the results of the FHA2 fluorimetry assay. The extent of fusion is much greater at pH 5 than at 7.4. This is consistent with the mechanism of viral fusion, in which the HA2 is activated by the endosomal pH drop. This result is also consistent with the circular dichroism results that indicate a different conformation of pH 5 and 7.4.

A.



B.

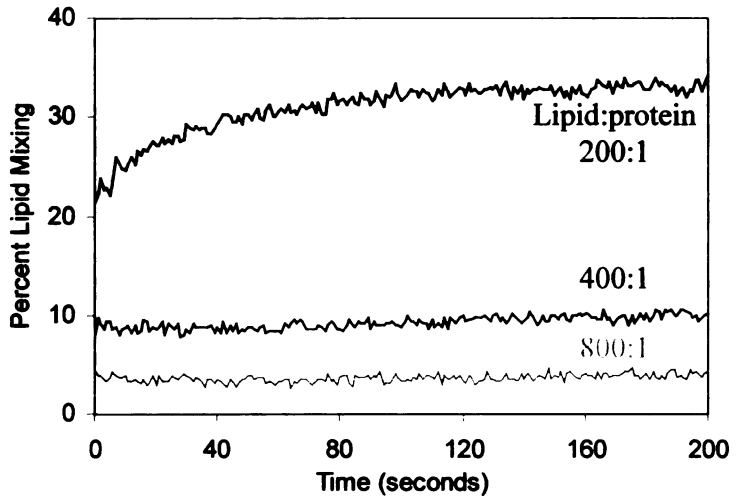


Figure 2-25. Lipid mixing activity of FHA2. At the lower pH of 5, shown in panel A, the activity is much greater than at pH 7.4, shown in panel B. As the lipid : protein ratio increases, the amount of activity observed decreases.

The activity of Fgp41 was also analyzed through the lipid mixing assay, figure 2-26. The activity of the protein was much lower than for FHA2, at similar lipid : protein ratios the mixing ability of the protein was <10%.

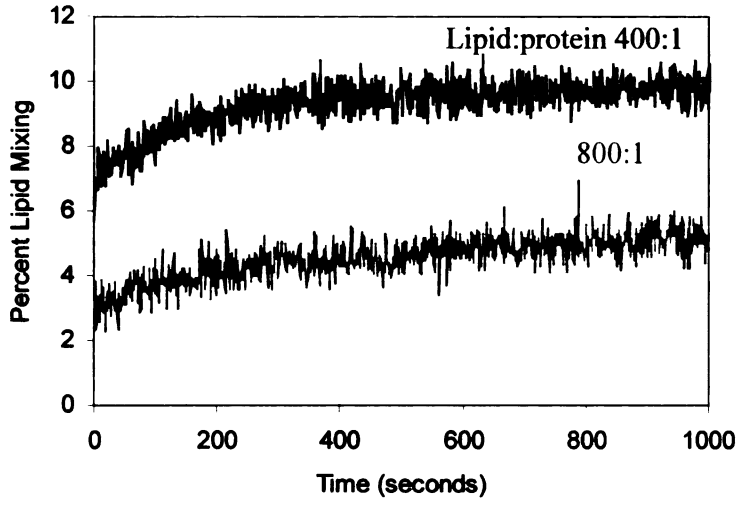


Figure 2-26. Lipid mixing activity of Fgp41. The activity level was much lower than for FHA2, but the observation of activity indicates that the initial purification procedure is a good starting point for further optimization.

The second method of studying the protein fusion activity is through electron microscopy. Figure 2-27 shows vesicles before and after the addition of FHA2 protein. The result is much different from the initial vesicles. It is not clear if any level of ordered structure exists in the product, but the vesicles have been disrupted, indicating activity by the protein.

Figur
Wher
After
highe

C. Re

poten

chang

struct

recons

lipid d

memb

indicat

recons

of 5, re

small fi

uniform

2 to 5

mixture

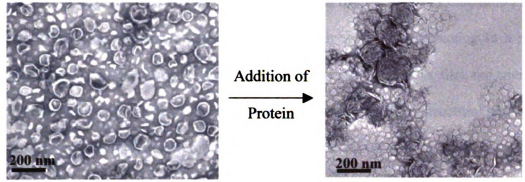


Figure 2-27. Electron microscopy of vesicles with and without the addition of FHA2. When no protein is present, the vesicle appear to be intact and relatively uniform in size. After the addition of protein , individual vesicles are clearly no longer present and a higher level of disorder exists.

C. Reconstitution of FHA2.

Reconstitution of the fusion protein from a detergent solution into lipid vesicles is potentially similar to the viral fusion process. The protein may undergo a conformation change from the pre-fusion state as it embeds into the lipid vesicles, resulting in a structure similar to that of the protein when bound to the host cell. FHA2 was reconstituted in membranes by initially mixing protein solubilized in detergent with a lipid/detergent mixture, followed by dialysis to remove the detergent and produce membrane-associated fusion protein. Previous lipid mixing studies of the FHA2 protein indicate much less activity at physiological pH, and this was also observed with the reconstitution of the protein into lipids. When the process is carried out at the active pH of 5, reconstitution of all protein occurs. If the same process is repeated at pH 7.4, only a small fraction of the protein will be associated with the lipid. This could be due to a conformational change that occurs when the pH of the FHA2 solution is lowered from 7.4 to 5, causing the formation of a structure that more readily reconstituted into the lipid mixture.

While it is not yet clear the exact mechanism by which the reconstitution occurs, these experimental observations are consistent with the protein behaving in a manner similar to that which occurs during viral infection. This indicates that the membrane reconstituted protein is a physiologically relevant form suitable for structural studies. Due to the fact that the endosomal pH at the time of viral fusion is near 5, and the increased rate of reconstitution, this condition was chosen for the initial structural studies of the membrane-associated protein. Membrane reconstitution was also achieved at a pH of 7.4 by instigating the reconstitution process at a pH of 5 for one day, followed by dialysis at pH 7.4 for two days, a process that successfully associates all of the protein with the lipid. This results in a protein/lipid mixture in which the membrane associated protein may be studied at the higher pH.

The reconstitution of FHA2 is effective, as shown in figure 2-28. Following the reconstitution process, the lipid pellet obtained is boiled in gel sample buffer to dissociate the protein/lipid vesicles and analyzed by gel electrophoresis. Lane 2 shows the protein present in the lipid sample, and Lane 3 the contents of the supernatant. Clearly, there is no significant amount of protein present in the supernatant, implying that all is embedded in the lipids.

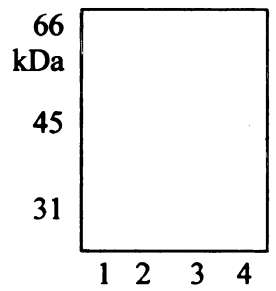


Figure 2-28. Reconstitution of FHA2. Lane 1-standards, Lane 2-lipid/protein pellet, Lane 3-soluble supernatant, Lane 4-standards.

VI. Conclusions and Future Work

This chapter describes methods to express, purify, and refold the 23 kDa FHA2 membrane protein which comprises the full ectodomain of the influenza HA2 fusion protein. This protein was used as a model for fusion protein purification and the optimized methods were applied to the expression and purification of similar proteins. Each protein and its characteristics are different and require a separate process of optimization, but these procedures did provide a good starting point. A yield of ~20 mg isotopically labeled FHA2 per liter of cell culture was achieved by: (1) initially growing cells in rich medium followed by a switch to minimal medium prior to induction of expression; and (2) solubilization, purification, and refolding of FHA2 from inclusion bodies. The total time from initializing cell growth to obtaining pure, folded protein is less than a week and most of that time is devoted to cell growth and dialysis, which require little attention from the scientist. The total yield is at least 40 times higher than the best yield obtained from growth solely in minimal medium. One reason for the larger yield was the increase in cell density A600 from ~2 in minimal medium growth to ~9 in the medium switch method. This increase relied on optimization of oxygenation and carbon source in the rich and minimal medium and on optimization of the pH and carbon source in minimal medium.

The growth in rich medium is also rapid and should lead to a higher concentration of ribosomes within the cell and therefore greater production of expressed protein (*16, 43*). This theory appeared valid for FHA2 as evidenced by the analysis of yields of FHA2 from the soluble lysate. Relative to growth solely in minimal medium, there was a ~20-fold increase in the yield of FHA2 from growth in rich medium/minimal medium. This

increase can be compared to the smaller ~5-fold increase in cell density. This difference suggests that relative to minimal medium growth, there was a ~4-fold increase in FHA2 per cell from growth in rich medium/minimal medium. This finding is consistent with a higher concentration of ribosomes in the cells.

Purification of the FHA2 protein was also highly successful, the protein obtained from the IMAC purification was greater than 95% pure, and following reconstitution into membranes there were no detectable contaminants in the sample. The protein proved to be folded to an extent consistent with previous structural work as observed through circular dichroism. Lipid mixing assays indicate that the protein highly active, although much more so at the physiologically active pH of 5 rather than 7.4.

The optimized methods developed for FHA2 were also applied to the production of the MHA2 construct (containing the full length HA2 protein), Fgp41 (containing the ectodomain of the protein) and the full-length gp41 construct. Unfortunately, these proteins were not as successful. The MHA2 protein is produced in high yield and is easily purified, but removing the maltose binding protein from HA2 has proven difficult. Fgp41 appears to express almost exclusively as inclusion bodies, and the full-length gp41 not at all.

The expression and purification of native FHA2 has been extremely successful, and can now be used as the goal to work towards on the other proteins. The work with MHA2 has shown that the maltose binding protein can be useful in increasing the amount of soluble protein produced, but the main difficulty is still the removal of the maltose binding protein. Expression of natively folded fusion region of gp41 (Fgp41) has been extremely low compared to the similar region of HA2 (FHA2), but the purification

procedure yields high purity protein that has proven to be active by the lipid mixing assays.

The MHA2 project is significant to the study of the HA2 protein because it possess the transmembrane and endodomain regions of the protein which are not in the FHA2 construct. The current difficulty is the proteolytic cleavage of the maltose binding protein from the N-terminus of HA2. An easy solution to this problem may be to change the protease recognition sequence in the linker region to that of another enzyme, such as factor Xa, which may be more successful. It is possible that the difficulty with the cleavage reaction is due to the folding of the protein complex. This could bury the protease recognition site in a way that is inaccessible and no matter which enzyme is used the reaction may not be successful.

A second approach to obtaining the full length HA2 may be to add the additional 40 residues to the current FHA2 construct. This could be done through either cloning or through several site directed mutagenesis additions. The expression level of the FHA2 construct is high enough that it should be a good starting point for the additional residues, but the addition of the transmembrane and endodomain could affect the expression level. Future graduate students of the Weliky group will study these two approaches further.

Expression of gp41 also has several possibilities. As MHA2 showed, the maltose binding protein is extremely effective at solubilizing protein and increasing overall expression. Through cloning, the maltose binding protein could be added to gp41 in an attempt to increase yield. This is a potential solution for either of the fragments. A second option for the Fgp41 protein would be to target the insoluble protein that is produced. Though not evident from gel electrophoresis, solid state NMR of whole bacterial cells

expressing this protein indicates that a significant amount of the protein is present. The low yields obtained from purification indicate that this additional protein is most likely in the form of inclusion bodies. As chapter 5A will discuss in great detail, it is possible to purify and refold proteins from inclusion bodies, and this may be the best option for the Fgp41 protein.

VII. References

(1) Link, A. J., and Georgiou, G. (2007) Advances and chattenges in membrane protein expression. *Aiche Journal 53*, 752-756.

(2) Cai, M. L., Huang, Y., Sakaguchi, K., Clore, G. M., Gronenborn, A. M., and Craigie, R. (1998) An efficient and cost-effective isotope labeling protocol for proteins expressed in Escherichia coli. *Journal of Biomolecular Nmr 11*, 97-102.

(3) Novagen (2006) *pET System Manual*, Vol. 11th Edition.

(4) Voet, D., and Voet, J. G. (1995) *Biochemistry*, John Wiley & Sons, Inc., New York.

(5) Nomine, Y., Ristriani, T., Laurent, C., Lefevre, J., Weiss, E., and Trave, G. (2001) Formation of soluble inclusion bodies by HPV E6 oncoprotein fused to maltose-binding protein. *Protein Expression and Purification 23*, 22-32.

(6) 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(2) refolds into the trimeric low-pH-induced conformation. *Biochemistry 43*, 5902-5911.

(7) Kapust, R. B., and Waugh, D. S. (1999) *Escherichia coli* maltose-binding protein is uncommonly effective at promoting the solubility of polypeptides to which it is fused. *Protein Science 8*, 1668-1674.

(8) Korepanova, A., Moore, J. D., Nguyen, H. B., Hua, Y., Cross, T. A., and Gao, F. (2007) Expression of membrane proteins from Mycobacterium tuberculosis in Escherichia coli as fusions with maltose binding protein. *Protein Expression and Purification 53*, 24-30.

(9) Bach, H., Mazor, Y., Shaky, S., Berdichevsky, A., Gutnick, D. L., and Benhar, I. (2001) *Escherichia coli* maltose binding protein as a molecular chaperone for recombinant intracellular cytoplasmic single-chain antibodies. *Journal of Molecular Biology 312*, 79-93.

(10) Scholz, C., Schaarschmidt, P., Engel, A. M., Andres, H., Schmitt, U., Faatz, E., Balbach, J., and Schmid, F. X. (2005) Functional solubilization of aggregation-prone HIV envelope proteins by covalent fusion with chaperone modules. *Journal of Molecular Biology 345*, 1229-1241.

(11) Feng, X., Wang, J., Shan, A., Teng, D., Yang, Y., Yao, Y., Yang, G., Shao, Y., Liu, S., and Zhang, F. (2006) Fusion expression of bovine lactoferricin in *Escherichia coli. Protein Expression and Purification 47*, 110-117.

(12) Lee, S. Y. (1996) High cell-density culture of Escherichia coli. *IBTech 14*, 98-105.

(13) Nikakhtari, H., and Hill, G. A. (2006) Closure effects on oxygen transfer and aerobic growth in shake flasks. *Biotechnology and Bioengineering 95*, 15-21.

(14) McDaniel, L. E., and Bailey, E. G. (1969) Effect of shaking speed and type of closure on shake flask cultures. *Applied Microbiology 17*, 286-290.

(15) Harvey, R. J. (1973) Fraction of Ribosomes Synthesizing Protein as a Function of Specific Growth-Rate. *Journal of Bacteriology 114*, 287-293.

(16) Harvey, R. J. (1970) Regulation of Ribosomal Protein Synthesis in Escherichia-Coli. *Journal of Bacteriology 101*, 574.

(17) Gourse, R. L., De Boer, H. A., and Nomura, M. (1986) DNA determinants of rRNA synthesis in E. coli: growth rate dependent regulation, feeback inhibition, upstream activation, antitermination. *Cell 44*, 197-205.

(18) Milne, A. N., Mak, W. W. N., and Wong, J. T. F. (1975) Variation of Ribosomal-Proteins with Bacterial-Growth Rate. *Journal of Bacteriology 122*, 89-92.

(19) Luli, G. W., and Strohl, W. R. (1990) Comparison of Growth, Acetate Production, and Acetate Inhibition of Escherichia-Coli Strains in Batch and Fed-Batch Fermentations. *Applied and Environmental Microbiology 56*, 1004-1011.

(20) Eiteman, M. A., and Altman, E. (2006) Overcoming acetate in Escherichia coli recombinant protein fermentations. *Trends in Biotechnology 24*, 530-536.

(21) Lian, L. Y., and Middleton, D. A. (2001) Labelling approaches for protein structural studies by solution-state and solid-state NMR. *Progress in Nuclear Magnetic Resonance Spectroscopy 39*, 171-190.

(22) Goto, N. K., and Kay, L. E. (2000) New developments in isotope labeling strategies for protein solution NMR spectroscopy. *Current Opinion in Structural Biology 10*, 585-592.

(23) Marley, J., Lu, M., and Bracken, C. (2001) A method for efficient isotopic labeling of recombinant proteins. *Journal of Biomolecular Nmr 20*, 71-75.

(24) Porath, J. (1992) Immobilized metal ion affinity chromatography. *Protein Expression and Purification 3*, 263-281.

(25) Mukherjee, S., Shukla, A., and Guptasarma, P. (2003) Single-step purification of a protein-folding catalyst, the SlyD peptidyl prolyl isomerase (PPI), from cytoplasmic extracts of Escherichia coli. *Biotechnology and Applied Biochemistry 37*, 183-186.

(26) Hottenrott, S., Schumann, T., Pluckthun, A., Fischer, G., and Rahfeld, J. U. (1997) The *Escherichia coli* SlyD is a metal ion-regulated peptidyl-prolyl *cis/trans*-isomerase. *The Journal of Biological Chemistry 272*, 15697-15701.

(27) Henning, L., and Schafer, E. (1998) Protein purification with C-terminal fusion of maltose binding protein. *Protein Expression and Purification 14*, 367-370.

(28) Petri, W. A., and Wagner, R. R. (1979) Reconstitution into Liposomes of the Glycoprotein of Vesicular Stomatitis-Virus by Detergent Dialysis. *Journal of Biological Chemistry 254*, 4313-4316.

(29) Helenius, A., Sarvas, M., and Simons, K. (1981) Asymmetric and Symmetric Membrane Reconstitution by Detergent Elimination Studies with Semliki-Forest-Virus Spike Glycoprotein and Penicillinase from the Membrane of Bacillus-Licheniformis. *European Journal of Biochemistry 116*, 27-35.

(30) Surrey, T., and Jahnig, F. (1992) Refolding and oriented insertion of a membrane protein into a lipid bilayer. *Proceedings of the National Academy of Sciences of the United States of America 89*, 7457-7461.

(31) Engelhard, V. H., Guild, B. C., Helenius, A., Terhorst, C., and Strominger, J. L. (1978) Reconstitution of purified detergent-soluble HLA-A and HLA-B antigens into phospholipid vesicles. *Proceedings of the National Academy of Sciences of the United States of America 75*, 3230-3234.

(32) Chami, M., Pehau-Arnaudet, G., Lamber, O., Ranck, J.-L., Levy, D., and Rigaud, J.-L. (2001) Use of Octyl B-Thioglucopyranoside in Two-Dimensional Cystrallizaion of Membrane Proteins. *Journal of Structural Biology 133*, 64-74.

(33) Mimms, L. T., Zampighi, G., Nozaki, Y., Tanford, C., and Reynolds, J. A. (1981) Phospholipid vesicle formation and transmembrane protein incorporation using octyl glucoside. *Biochemical Journal Biochemistry*, 833-840.

(34) Sato, K., and Wickner, W. (1998) Fuctional reconstitution of Ypt7p GTPase and a purified vacuole SNARE complex. *Science 281*, 700-702.

(35) DeGrip, W. J., VanOostrum, J., and Bovee-Geurts, P. H. M. (1998) Selective detergent-extraction from mixed detergent/lipid/protein micelles, using cyclodextrin inclusion compounds: a novel generic approach for the preparation of proteoliposomes. *Biochemical Journal 330*, 667-674.

(36) Ueno, M., Tanford, C., and Reynolds, J. A. (1984) Phospholipid vesicle formation using nonionic detergents with lwo monomer solubility. Kinetic factors determine vesicle size and permeability. *Biochemistry 23*, 3070-3076.

(37) Worman, H. J., Brasitus, T. A., Dudeja, P. K., Fozzard, H. A., and Field, M. (1986) Relationship Between Lipid Fluidity and Water Permeability of Bovine Tracheal Epithelial-Cell Apical Membranes. *Biochemistry 25*, 1549-1555.

(38) Hope, M. J., Bally, M. B., Webb, G., and Cullis, P. R. (1985) Production of large unilamellar vesicles by a rapid extrusion procedure - characterization of size distribution, trapped volume and ability to maintain a membrane-potential. *Biochim. Biophys. Acta-Biomembr. 812*, 55-65.

(39) Struck, D. K., Hoekstra, D., and Pagano, R. E. (1981) Use of resonance energy transfer to monitor membrane fusion. *Biochemistry 20*, 4093-4099.

(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. *Journal of the American Chemical Society 129*, 11320.

(42) Rooney, S. A., Nardone, L. L., Shapiro, D. L., Motoyama, E. K., Gobran, L., and Zaehringer, N. (1977) Phospholipids of Rabbit Type-2 Alveolar Epithelial-Cells - Comparison With Lung Lavage, Lung-Tissue, Alveolar Macrophages, and a Human Alveolar Tumor-Cell Line. *Lipids 12*, 438-442.

(43) Harvey, R. J. (1973) Growth and Initiation of Protein-Synthesis in Escherichia-Coli in Presence of Trimethoprim. *Journal of Bacteriology 114*, 309-322.

Chapter 3:

Solid State NMR Analysis of Membrane Associated Fusion Protein

I. Introduction

Structural analysis of large membrane proteins has generally been limited to crystallography and liquid state NMR with the protein solubilized in detergent micelles. Crystallography and NMR have also complemented each other in the pursuit of structural information, and this has historically been the case for membrane proteins as well (*1*). Viral fusion proteins, the viral component that catalyzes fusion of the viral and host cell membranes, are targets of anti-viral therapies. Until recently, structural information about the membrane associated fusion proteins was difficult to obtain. Recent work by our group has demonstrated a new method that combines amino acid specific isotopic labeling and a solid state NMR filtering sequence to obtain spectra from a single residue in a membrane associated fusion protein (*2*). This paper presents a broader view of the project, including site specific information on 24 sites within the ectodomain of the hemagglutinin protein and analysis in several different membrane environments.

A. The FHA2 Protein

The target of our structural studies is the first 185 residues of the hemagglutinin protein containing the fusion peptide and ectodomain regions, termed the FHA2 protein. Hemagglutinin (HA) is the fusion protein for the influenza virus (IFV), an enveloped virus enclosed by a membrane obtained from the host cell. The HA protein is composed of two subunits, HA1 and HA2. HA2 plays a critical role in the viral infection by initiating the fusion between the viral and host cell membranes. An important component

of the HA2 unit is the N-terminal fusion peptide, the first 20 residues of the FHA2 construct. The Influenza fusion peptide (IFP) is the portion of HA2 that inserts into the host cell membrane at the initiation of the fusion process (*3*).

The virus begins the infection process by entering respiratory epithelial cells via receptor-mediated endocytosis. Once in the endosome, the pH drops to about 5. This increase in acidity causes the dissociation of the HA1 and HA2 subunits, and a large structural change in HA2 that activates the protein for insertion into the host membrane. This change exposes the fusion peptide region, allowing it to bind to the endosomal membrane of the host cell. The pH activation of the protein is a critical part of the infection, as this fusion process does not occur until the drop of the endosomal pH. Previous studies using the FHA2 protein and lipid vesicles have demonstrated this pH activation. The FHA2 protein is much less active at physiological pH, and upon raising the acidity, the activity greatly increases (*2*).

As the fusion protein is the first part of the virus to make contact with the host cell, this has become an attractive target to better understand the viral infection process, and more structural information about the HA2 protein would help in this pursuit. Previous structural studies have yielded a crystal structure at pH 4.4 of the soluble ectodomain, containing residues 34-178, as well as a liquid state NMR structure of the IFP in detergent micelles (*4-6*). Crystallography revealed that the low pH structure of the ectodomain is a trimer with each protein subunit composed of a long helix and a long β-strand connected by a hinge region. Overall, each of the subunits of the trimer were very similar in structure, with the exception of several short segments within that hinge region where slight differences in the dihedral angles were observed. The detergent solubilized

NMR structure of the fusion peptide at pH 5 indicates a helix that extends from residues 2 to 9, is interrupted by a kink at residues 10 and 11, and then continues into a well defined extended structure including a short 3_{10}-helix (*6*). At pH 7.4 a change in structure is observed and the ordered structure of the 3_{10}-helix is no longer present. To date, there have been no structural studies of residues 21 to 37 that lie between these two structures, and our study is the first to give insight into the secondary structure of this region.

B. Solid State NMR Analysis

Solid State Nuclear Magnetic Resonance (SSNMR) has been used previously to study other large membrane proteins and even membrane associated fusion peptides (*7-11*). Three main differences amongst these structural studies are the sample preparation, the method of isotopic labeling, and alignment of the sample in the magnetic field. Solid samples of membrane proteins can be prepared by precipitation through dialysis to remove the detergent or with the use of polyethylene glycol (PEG) in which the resulting samples can be associated with lipids, embedded in lipid Nanodiscs, or free of lipid and detergent (*12-14*). A comparison of several sample preparation methods for two dimensional SS-NMR analysis including lyopholization and precipitation revealed that in the case of the α-spectrin SH3 domain, precipitation with ammonium sulfate yielded the most well resolved spectra (*15*). This is the method of choice for many other current studies in the literature as well. The chemical shifts of the transmembrane helices of an integral membrane enzyme were assigned using multidimensional NMR for a lipid associated protein sample precipitated by the removal of detergent through dialysis (*13*). A potassium ion channel protein has been studied by both 2D and 3D SS-NMR using

protein precipitated by the addition of PEG (*14*). The secondary structure of a filamentous phage was studied by 2D SS-NMR in a hydrated sample also prepared by PEG precipitation (*16*). Precipitation of the protein with PEG is also the method used to produce samples of the membrane scaffold protein 1 in lipid Nanodiscs (*12*).

A second common type of sample preparation that is not typically applicable to membrane proteins is crystallization, although in isolated cases it has proven successful (*17, 18*). This type of sample often results in sharper lines (< 0.5 ppm) than what is traditionally observed in membrane samples. The sharper lines allow for multidimensional studies, such as the study of 2D crystals of the outer membrane protein G by 2D NMR (*18*). The dipolar couplings of proteins and peptides have also been determined using three and four-dimensional NMR on microcrystalline samples (*17*).

A second difference among current protein NMR studies is the level of isotopic labeling. Uniformly labeled protein possesses the greatest potential for providing a large amount of information from a single sample, but the observed linewidths must be sufficiently sharp and the peaks well resolved in order to make assignment of specific residues possible. For proteins that yield larger linewidths, a more practical yet time consuming approach is to selectively label the protein and study individual positions in separate samples. Previously reported studies of the membrane associated viral fusion protein FHA2 resulted in signals much broader (> 2ppm) than what is observed in microcrystalline samples, which can be as low as 0.1 ppm (*2, 18*). For these protein samples the broader linewidths and overlapping of signals would make it very difficult, or even impossible, to assign peaks to the individual amino acids and to make structural conclusions regarding these positions.

The third difference amongst solid-state NMR structural studies is the alignment of the sample. While most studies have been conducted on un-aligned samples using magic angle spinning, the study of aligned samples has provided new structural information as well. With the use of magnetically aligned bicelles the structure of two transmembrane helices and the 10-residue inter-helical loop from a mercury transport membrane protein has been determined and studies have been conducted on a G protein-coupled receptor to single site resolution, both of which illustrate the feasibility of using this method for the study of membrane proteins (*9, 19*). While most aligned samples have been studied by $^{1}H/^{15}N$ double resonance experiments, the $^{1}H/^{13}C/^{15}N$ methods commonly used in liquid state and MAS studies have also been applied to these samples which allows greater versatility in the experiments that may be performed and the information obtained using magnetically aligned bicelles (*20, 21*).

The most difficult aspect of the SS-NMR and the limiting factor in most structural studies is the requirement of a relatively large amount of isotopically labeled protein. Our samples require about 0.2 μmol of protein, in the case of FHA2 about 5 mg, for adequate signal to noise in the 50 μL sample. Membrane proteins typically express in much lower yields than soluble proteins, even in recombinant expression, which can make producing 5 mg expensive and time consuming. For FHA2, a protocol was developed to increase the low yield, and included initial growth of the bacterial cells to high density in rich medium before switching to a minimal medium for isotopically labeled protein expression.

II. Materials and Methods

A. Materials

The FHA2 plasmid was generously provided by Dr. Yeon-Kyun Shin at Iowa State University. The plasmid contained kanamycin resistance and was transformed into E. coli BL21(DE3) cells. Unless noted, all chemicals were purchased from Sigma-Aldrich (St. Louis, MO). Luria-Bertani broth (LB) was obtained from Acumedia (Lansing, MI). The detergents n-octyl-β-D-thioglucopyranoside (BTOG) and octyl pentaethylene glycol ether (C8E5) were obtained from Anatrace (Maumee, OH). The ether linked lipids di-*O*-tetradecylphosphatidylcholine (DTPC) and di-*O*-tetradecylphosphatidylglycerol (DTPG) were obtained from Avanti Polar Lipids (Alabaster, AL). Amino acids with 1-^{13}C and/or ^{15}N labeling were obtained from Cambridge Isotope Labs (Andover, MA).

B. Isotopically Labeled Protein Expression

Cell growth was initiated in 1 L of LB enriched with 10 mL glycerol in a 2.8 L baffled fernbach flask at 37°C. After overnight growth to maximum cell density, the cell suspension was centrifuged at 10000*g* for 10 minutes. The resulting pellet was resuspended in 1 L of minimal medium (6.8 g/L Na_2HPO_4, 3.0 g/L NaH_2PO_4, 0.50 g/L NaCl, 1.0 g/L NH_4Cl, 2.5 g/L $MgSO_4$, and 10 g/L glycerol at pH 8.0). After one hour of cell growth in minimal medium, FHA2 expression was induced with 0.2 mM isopropyl thiogalactoside (IPTG). For production of FHA2 with isotopic labeling, 100 mg/L of each labeled amino acid was added at the time of induction. Protein production was continued for three hours at 23°C. The cell pellet was retrieved by centrifugation at 10000*g* for 10

minutes, and the pellet was then stored at –80°C until purification. All cell cultures were grown in media containing 15 mg/L kanamycin.

C. Protein Purification

FHA2 purification was achieved with the use of a C-terminal polyhistidine tag and cobalt resin. The following method was optimized for purification of 5g of cells. Cells were suspended in 25 mL of wash buffer A (50 mM sodium phosphate at pH 8.0, 300 mM NaCl, 20 mM imidizole, and 0.5% (w/v) N-laurylsarcosine (Sarkosyl) detergent). Cells were lysed by 4 periods of 1 minute sonication with 1 minute delay in between. Each period was 0.8 s on/0.2 s off to avoid overheating and 80% amplitude. The insoluble fraction was removed by centrifuation at 48,000g at 4°C for 20 minutes. The soluble FHA2 remained in the supernatant and was bound to 0.5 mL cobalt His-Select resin (Sigma-Aldrich) for one hour with mixing on a shaker. The binding step was done at room temperature. The resin was then pelleted by centrifugation at 1000g for 1 minute and then transferred to a 10 mL column using wash buffer A. The resin was washed with 1 column volume each of wash buffer A, wash buffer B (5 mM 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES), 10 mM 2-(N-morpholino)ethanesulfonic acid, (MES), 20 mM imidazole, 0.5% Sarkosyl, 0.5% BTOG, and 0.4% C8E5 at pH 7.4), and wash buffer C (5 mM HEPES, 10 MES, 20 mM imidazole, 0.5% BTOG, and 0.4% C8E5 at pH 7.4). This series of wash buffers exchanges the detergent from sarkosyl, best used for initially solubilizing protein, to a BTOG/C8E5 mixture, which is best for keeping the protein dispersed and un-aggregated. FHA2 was eluted with 5 column volumes of elution

buffer (5 mM HEPES, 10 MES, 250 mM imidazole, 0.5% BTOG, and 0.4% C8E5 at pH 7.4). Protein concentrations were quantified using A280 with $\varepsilon 280 = 34{,}000\ M^{-1}\ cm^{-1}$.

D. Membrane Reconstitution

Reconstitution of the fusion protein from a detergent solution into lipid vesicles is potentially similar to the viral fusion process. The protein may undergo a conformation change from the pre-fusion state as it embeds into the lipid vesicles, resulting in a structure similar to that of the protein when bound to the host cell. FHA2 was reconstituted in membranes by initially mixing protein solubilized in detergent with a lipid/detergent mixture, followed by dialysis to remove the detergent and produce membrane-associated fusion protein. Previous lipid mixing studies of the FHA2 protein indicate much less activity at physiological pH, and this was also observed with the reconstitution of the protein into lipids. When the process is carried out at the active pH of 5, reconstitution of all protein occurs. If the same process is repeated at pH 7.4, only a small fraction of the protein will be associated with the lipid. This could be due to a conformational change that occurs when the pH of the FHA2 solution is lowered from 7.4 to 5, causing the formation of a structure that more readily reconstituted into the lipid mixture.

While it is not yet clear the exact mechanism through which the reconstitution occurs, these experimental observations are consistent with the protein behaving in a manner similar to that which occurs during viral infection, indicating that the membrane reconstituted protein is a physiologically relevant form suitable for structural studies. Due to the fact that the endosomal pH at the time of viral fusion is near 5, and the increased

rate of reconstitution, this condition was chosen for the initial structural studies of the membrane-associated protein. Membrane reconstitution was also achieved at a pH of 7.4 by instigating the reconstitution process at a pH of 5 for one day, followed by dialysis at pH 7.4 for two days, a process that successfully associates all of the protein with the lipid. This results in a protein/lipid mixture in which the membrane associated protein may be studied at the higher pH.

The membrane composition for most of the samples was a 4:1 molar ratio of the ether linked lipids DTPC and DTPG, although samples were also prepared with a 2:1 molar ratio of the DTPC/DTPG lipid mixture and cholesterol. These lipids were selected based on: (1) choline is a predominant headgroup of lipids of membranes of respiratory epithelial cells, the host cells of the influenza virus; (2) the headgroup of DTPG is negatively charged like the minor fraction of the host cell lipids; and (3) DTPC and DTPG are ether- rather than ester-linked lipids and do not have a natural abundance ^{13}C contribution to the carbonyl region probed in the NMR experiments (*2, 22*). The lipids (~40 mg total), or the lipid/cholesterol mixtures, and the detergent BTOG (~160 mg) were dissolved in chloroform and a thin film of lipid/deterent produced by removing the solvent with a flow of nitrogen gas followed by overnight drying in a vacuum chamber. The lipid/detergent film was then dissolved in ~5 mL of 5 mM HEPES/10 mM MES buffer at pH 7.4. The FHA2 solution was added to the detergent/lipid solution to form a co-micelle solution of ~8 mg FHA2, detergent, and lipid. The solution was transferred to 10 kDa molecular weight cut off dialysis tubing and dialyzed against 2 L volume of HEPES/MES buffer at pH 5.0, the optimal pH for fusion protein induced lipid mixing. The dialysis was carried out for three days at 4°C with one buffer change. The

membrane-associated FHA2 sample was reclaimed by centrifugation at 50000*g* at 4°C for 3 hours.

E. Solid-State NMR Experiments

Data were obtained with a 9.4 T instrument (Varian Infinity Plus, Palo Alto, CA), a triple resonance magic angle spinning (MAS) probe, and a 4.0 mm diameter rotor with ~40 μL sample volume. It is estimated that the sample volume contained ~4 mg FHA2 and ~20 mg total lipid. Typical parameters of the REDOR pulse sequence were: (1) 8.0 kHz MAS frequency; (2) a 6 μs ^{1}H $\pi/2$ pulse; (3) a 1.6 ms cross-polarization period with 63 kHz ^{1}H Rabi frequency and 80 kHz ^{13}C Rabi frequency; (4) a 2 ms dephasing period with alternating 19 μs ^{15}N π pulses and 8 μs ^{13}C π pulses and 88 kHz two-pulse phase modulation (TPPM) ^{1}H decoupling; (5) ^{13}C detection with 88 kHz TPPM ^{1}H decoupling; and (6) 1 sec delay (*2*). The sample was cooled with nitrogen gas at –10°C to counteract radio frequency heating. Data were collected for approximately three days (~100,000 scans of each S_0 and S_1). Spectra were externally referenced to the methylene carbon of adamantane at 40.5 ppm, which corresponds to the ^{13}C referencing used in liquid-state NMR of soluble proteins (*23, 24*). Spectra were processed with 100 or 200 Hz of line-broadening. Data were acquired without (S_0) and with (S_1) the ^{15}N π pulses during the dephasing period and respectively represented the full ^{13}C signal and the ^{13}C signal minus ^{13}C close to ^{15}N nuclei.

Because the FHA2 contained a ^{13}CO/^{15}N unique sequential pair, the $S_0 - S_1$ difference was predominantly the filtered signal of this pair (*8*). For example, spectra of

labeled FHA2 which targeted the ^{13}CO of Gly-4 or Ala-7 the experimental S_1/S_0 integrated intensity ratio was respectively 0.95 or 0.93 and correlated with the expected ratios of 0.94 and 0.91 for putative 100% labeling. These expected ratios were based on an approximate model in which the S_1/S_0 intensity ratio was: (number of FHA2 residues of the same type as the ^{13}CO labeled amino acid minus one)/(number of FHA2 residues of the same type as the ^{13}CO labeled amino acid), e.g. 10/11 for Ala. Analysis of the $S_0 - S_1$ signal for the sample targeting Leu-98 is more complex due to the potential for intra-residue dephasing because both the ^{15}N and ^{13}C labeling are within the same amino acid. The difference signal observed had some contribution from the other Leus (~10% per Leu) in the sequence because of the 2.5 Å intra-residue $^{13}CO \cdots {}^{15}N$ distance in 1-^{13}C, ^{15}N-Leu (*25*). These contributions are probably responsible for the experimental S_1/S_0 integrated intensity ratio, which is smaller than the ratio predicted by the approximate model. Although some of these intra-residue Leu contributions would likely have helical shifts similar to the shift of Leu-98, the contributions from other Leu residues would be dispersed over a range of non-helical shifts and would be less apparent in the overall appearance of the difference spectrum. This idea is supported by the observation that there is a large apparent difference between the S_0 and S_1 intensities in the 177.5-179.5 ppm helical shift region and much smaller difference in the non-helical lower shift region.

Natural abundance ^{13}COs or ^{15}Ns will also contribute to the $S_0 - S_1$ difference signal. A model has been developed for these contributions in the context of the approximation that only the desired positions of protein are labeled and that this labeling

is close to 100%. These approximations appeared to hold for the FHA2 labeling which targeted the Gly-4 or Ala-7 residues. The recombinant protein is considered to have "*D*" residues that are ^{13}CO labeled and "*E*" residues that are ^{15}N labeled. There is a single unique sequential ^{13}CO-^{15}N labeled pair and this pair contributes 1.0 intensity to the $S_0 - S_1$ difference signal. The approximate contribution to the difference signal intensity by natural abundance ^{15}N dephasing of the $D - 1$ other labeled ^{13}COs will be $(D - 1) \times (1.1) \times 0.0037$ where the 1.1 factor considers the effect at 2 ms dephasing time of ^{15}N separated by one and two bonds from the labeled ^{13}CO and 0.0037 is the ^{15}N fractional natural abundance. Similarly, the approximate contribution to the difference signal intensity from natural abundance ^{13}COs dephased by the $E - 1$ other labeled ^{15}Ns is $(E - 1) \times (1.1) \times 0.0111$. The fractional natural abundance contribution to the total $S_0 - S_1$ signal intensity is therefore:

$$\{[(D-1) \times 0.00407] + [(E-1) \times 0.0122]\}/$$
$$\{1.0 + [(D-1) \times 0.00407] + [(E-1) \times 0.0122]\}$$

For FHA2 with labeling which targeted the Gly-4 residue, $D = 16$, $E = 11$, and the ratio was 0.15. A much larger natural abundance contribution to the $S_0 - S_1$ difference signal will be observed if the recombinant protein is much larger than FHA2 and *D* and *E* are much larger.

In the context of the model, the lineshape of the natural abundance contribution will primarily depend on the amino acid identities and the conformations of the residues forming sequential pairs with the labeled residues. Because there will likely be a wide variety of amino acid types and perhaps conformations for these residues, there will

likely be significant chemical shift dispersion in the lineshape. For large values of D and E, the $S_0 - S_1$ difference signal would therefore appear to contain a sharp signal from the ^{13}CO in the unique sequential pair and a broad signal from the natural abundance ^{13}CO contributions.

III. Structural Analysis

The study of FHA2 by SS-NMR proved to be an effective method for determining secondary structure at specific residues in the membrane associated form. Figure 3-1 shows the previous structures used as models in our structural studies. Shown in red is each residue in which the secondary structure was analyzed by SS-NMR.

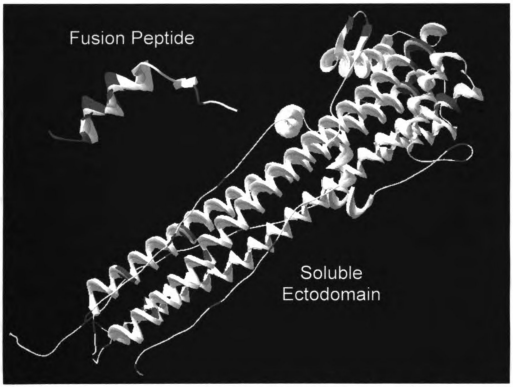


Figure 3-1. Residues of membrane associated FHA2 studied by SS-NMR. This figure, created from previous structural studies of the fusion peptide in detergent solution by liquid state NMR (PBD ID 1IBN), and the soluble ectodomain from aqueous solution by crystallography (PDB ID 1QU1), shows the positions analyzed by SS-NMR in red (*26, 27*).



A. Single Site Structural Analysis.

With the use of REDOR filtering and amino acid specific labeling, it is possible to observe the ^{13}CO signal of a single residue. In general, the scheme involves isotopically labeling the carbonyl carbon of one amino acid type and the amide nitrogen of another. The set of amino acids to be labeled were selected based on the primary sequence of the protein. If the carbonyl and amide labeled amino acids are only adjacent to each other at one position in the protein, with the use of the REDOR filter we can subtract the signal from all other labeled positions to observe only the carbon that is part of the unique pair. These unique sites are termed the REDOR active positions. The dipolar coupling between the adjacent ^{13}CO and ^{15}N results in the observation of decreased ^{13}CO signal at the active position upon the application of the ^{15}N pulses. Alternating signals are collected from the sample, the first is the entire observed carbon signal, mostly comprised of the ^{13}C in the sample (S_0), with no applied nitrogen pulses. The second is an attenuated signal without the REDOR active position (S_1) obtained by applying the nitrogen pulse. Subtraction of the S_1 signal from the S_0 signal yields the spectrum of solely the REDOR active position. Figure 2a and 2d illustrate examples of the S_0 and S_1 spectra. Figs. 3-2a and 3-2d show the S_0 in grey and the S_1 in black for the Leu-98 and the Ala-101/Ala-166 positions, respectively. Fig. 3-2c and 3-2c show the S_0 signal from Met-17 and Try-22, and are marked with black and grey lines to mark the center of the literature distributions of chemical shifts for helical and β strand conformations, respectively. Panel B clearly shows that the majority of the methionine residues in the protein are in a helical conformation, consistent with the strongly helical difference signal observed for residue

17. Panel C indicates that a greater proportion of the tyrosine residues in the protein also lie within the helical conformation, as indicated by the blue line and the shoulder of the CO peak.

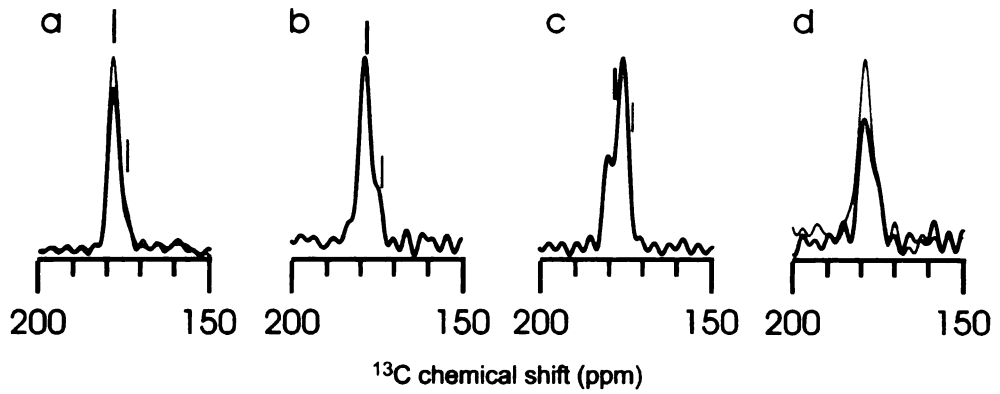


Figure 3-2. Solid-state NMR spectra of PC/PG-associated FHA2 samples. The sample are expressed with labeling that targeted one residue. (a) displays superimposed REDOR S_0 (grey) and S_1 (black) spectra of the sample that targeted Leu-98 and (b) and (c) display the S_0 spectra of the samples that respectively targeted Met-17 and Tyr-22. In (a), the black line marks the ^{13}CO chemical shift which is the center of the literature distribution of Leu residues in helical conformation and the grey line marks the center for β strand conformation. In (b) and (c), the lines mark the centers of the distributions for Met and Tyr residues, respectively. (d) displays the S_0 and S_1 of the sample which targeted Ala-101 and Ala-166.

The chemical shifts observed were correlated to secondary structure based on the data set for ~180 proteins for which both the ^{13}CO assignments and high-resolution structures were known, shown in table 3-1 (*24*). The local environment of the carbonyl in each structure type gives rise to the differences in chemical shift that we can be used to identify the local structure (*28*). Inclusion of a residue in helical conformation was based on its hydrogen bonding and on its values of the backbone dihedral angles φ and ψ with typical helical ranges being $-120° < \varphi < -34°$ and $-80° < \psi < -6°$. These are generally consistent with the experimental ranges of dihedral angles in α helices (*29*). The inclusion of residues in β-strand conformation considered the typical range of dihedral

angles experimentally observed, including $-180^{\circ} < \varphi < -40^{\circ}$ or $160^{\circ} < \varphi < 180^{\circ}$ and $70^{\circ} < \psi < 180^{\circ}$ or $-180^{\circ} < \psi < -170^{\circ}$. If the observed linewidth, (the full width at half maximum), is relatively sharp (~2-4 ppm) this supports the conclusion of a single secondary structure. Broader lines could indicate multiple structures, and the observed signal may be the result of several overlapping peaks.

Table 3-1. Secondary Structure CO Chemical Shifts.

Residue	Coil	Coil SD	Helix	Helix SD	Beta Strand	Beta Strand SD
Ala	177.67	1.57	179.40	1.32	176.09	1.51
Phe	175.59	1.60	177.13	1.38	174.25	1.63
Gly	173.89	1.42	175.51	1.23	172.55	1.58
Ile	175.57	1.67	177.72	1.29	174.86	1.39
Leu	176.89	1.71	178.53	1.30	175.67	1.47
Ser	174.49	1.31	175.94	1.39	173.55	1.50
Val	175.66	1.47	177.65	1.38	174.80	1.39
Tyr	175.39	1.67	177.36	1.40	174.54	1.45
Met	173.35	1.89	177.95	1.12	174.83	1.40

Average observed ^{13}CO chemical shift (ppm) along with standard deviation of the measurements (*24*).

B. Multiple Site Structural Analysis.

Several of the positions studied were amino acid pairs that occur more than once in the protein sequence. This results in difference signal representing two different amino acid locations within the protein. Spectra for the Leu-2 and Leu-119, Ala-5 and Ala-41, and Gly-8 and Gly-23 residues were observed and all gave strong, relatively sharp signal, shown in fig. 3-3. In order for the sum of the individual signals to be sharp, each contributing signal must be as well, because if the individual signals were broad or of different chemical shifts the result would be a broadened signal. This indicates that both of the REDOR active positions in each of these samples are of the same secondary

structure. This allowed us to observe the spectrum of an additional six residues in only three samples, bringing the total number of residues studied to 24. While these examples show the analysis of multiple positions in one sample, this method only provides useful information when the secondary structure of both positions are the same. Other possibilities include a broad signal (>4 ppm) covering multiple structural regions or two distinct signals. Either situation would make it impossible to definitively assign secondary structure to the positions in question, and would be of little usefulness in structural analysis.

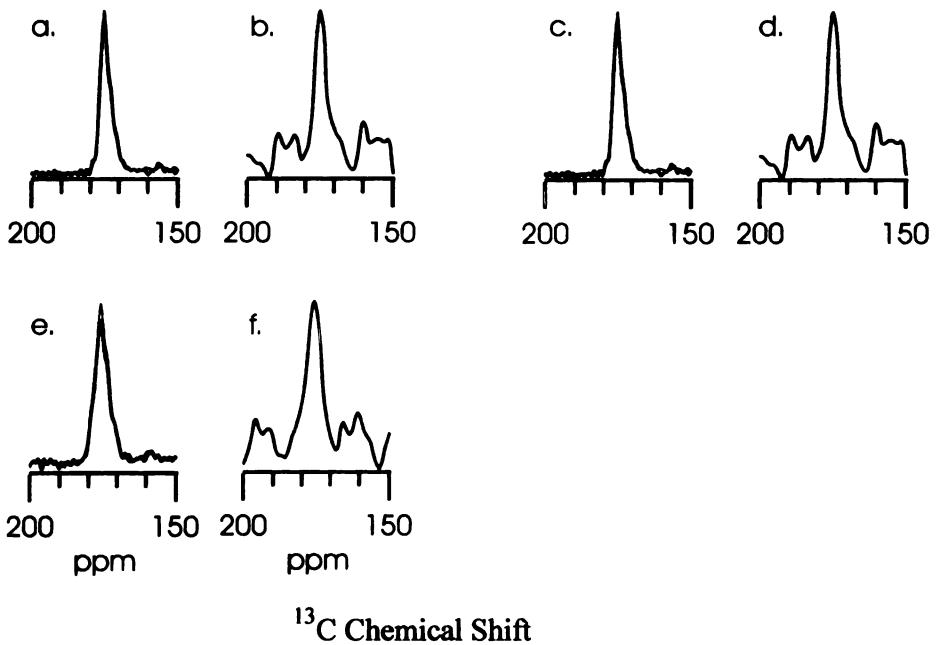


Figure 3-3. Multi-position REDOR analysis. Three of the positions studied by REDOR analysis were not part of a unique pair, but a pair that occurs twice in the protein sequence. In this case, the resulting difference signal is a combination of both positions. If relatively sharp signal is observed, conclusions regarding secondary structure can be reached for both positions in one sample. (a) S_0/S_1 signal from LF labeling and (b) difference signal representing leucines 2 and x. (c) S_0/S_1 signal from AI labeling and (d) resulting difference signal representing alanines 5 and 41. (e) S_0/S_1 signal from GF labeling and (f) resulting difference signal representing glycines 8 and 23.

C. Fusion Peptide.

The carbonyl chemical shift of eight specific positions within the fusion peptide region were studied, Gly-1, Leu-2, Gly-4, Ala-5, Ala-7, Gly-8, Gly-16, and Met-17, shown in figure 3-4. Each of these positions, with the exception of Gly-1, corresponds to helical structure and all are consistent with the liquid state NMR structure. Gly-1, in the N-terminus of the protein, produced a broader signal with a chemical shift that indicates a less ordered structure at this position. This is consistent with previous NMR of the detergent solubilized peptide. In Table 2, the results of the lipid associated protein samples are compared to previous studies conducted by Yan Sun on the lipid associated fusion peptide at pH 5.0.

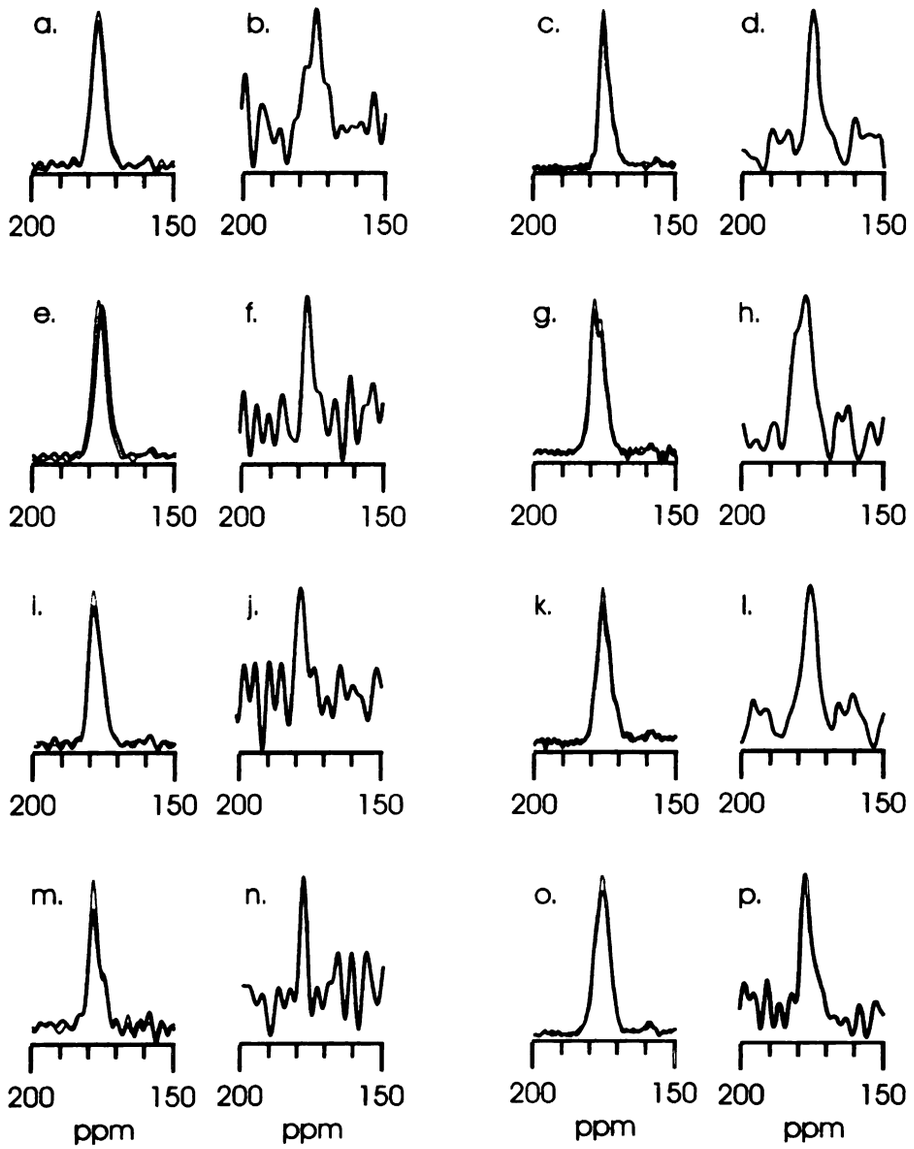


Figure 3-4. REDOR analysis of the fusion peptide region of FHA2. Uniform labeling of selected amino acids in the protein sequence is abbreviated as "AB", where "A" is the ^{13}C-carbonyl labeled amino acid and "B" is the ^{15}N-amide labeled amino acid. This results in an S_0 signal, shown in grey, and a decreased S_1 signal, shown in black. The $S_0 - S_1$ difference represents the signal of the single "A" carbonyl. (A) S_0/S_1 signal from GL labeling and (B) difference signal representing Gly-1. (C) S_0/S_1 signal from LF labeling and (D) resulting Leu-2 signal. (E) S_0/S_1 signal from GA labeling and (F) resulting Gly-4 signal. (G) S_0/S_1 signal from AI labeling and (H) resulting Ala-5 signal. (I) S_0/S_1 signal from AG labeling and (J) resulting Ala-7 signal. (K) S_0/S_1 signal from GF labeling and (L) resulting Gly-8 signal. (M) S_0/S_1 signal from MI labeling and (N) resulting Met-16 signal. (O) S_0/S_1 signal from GL labeling and (P) resulting Gly-17 signal.

D. Missing Link.

The 21-33 residue region of the HA2 protein has not been previously studied by either NMR or crystallography. The residue Tyr-22 was studied by individual site REDOR analysis, and the Gly-23 was studied as a double target sample in combination with Gly-8, shown in figure 3-5. The ^{13}CO chemical shift at both positions is within the range of helical secondary structure, giving the first insight into the structure of this region of the protein. Helical structure has been observed at the end of the fusion peptide construct, residues Gly-16 and Met-17, as well as at the beginning of the ectodomain crystal structure, Ala-36 and Ala-44, as part of a long helix that extend from residues 38 to 105. This new information regarding helical structure in between the regions previously studied presents the possibility of a long helix extending from the fusion peptide domain through part of the soluble ectodomain region to residue 105. Chapter 4 discusses in detail future studies that are planned to further probe the structure of this region.

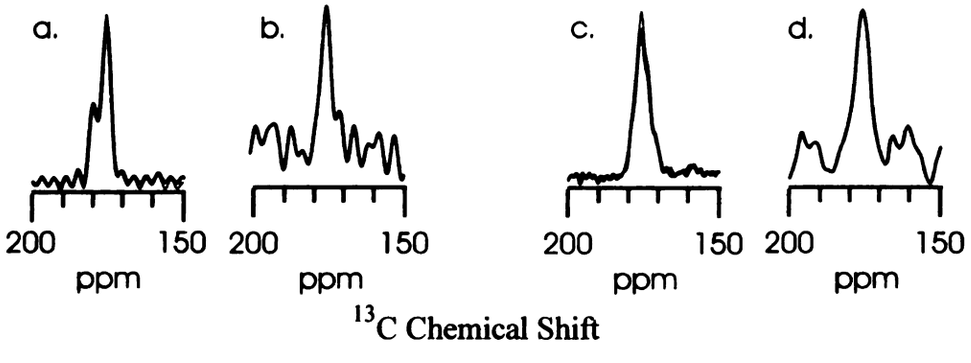


Figure 3-5. REDOR analysis of the missing link region of FHA2. (A) S_0/S_1 signal from YG labeling and (B) difference signal representing Tyr-22. (C) S_0/S_1 signal from GF labeling and (D) resulting Gly-23 signal.

E. Soluble Ectodomain.

A total of 12 positions in the soluble ectodomain previously observed by crystal structures were studied by REDOR analysis. The N-terminal portion of this region shown in figure 3-6, the hinge region in figure 3-7, and the C-terminal region in figure 3-8. Of the positions studied in the N-terminal portion, Ala-36 was of particular interest. This residue is near the N-terminus of the crystal structure, and was determined to lack order in half of the strands and to exist in a beta strand in the others. The observed chemical shift in the membrane-associated protein is 179.3 ppm, solidly within the known helical range of 179.4 +/- 1.3 ppm.

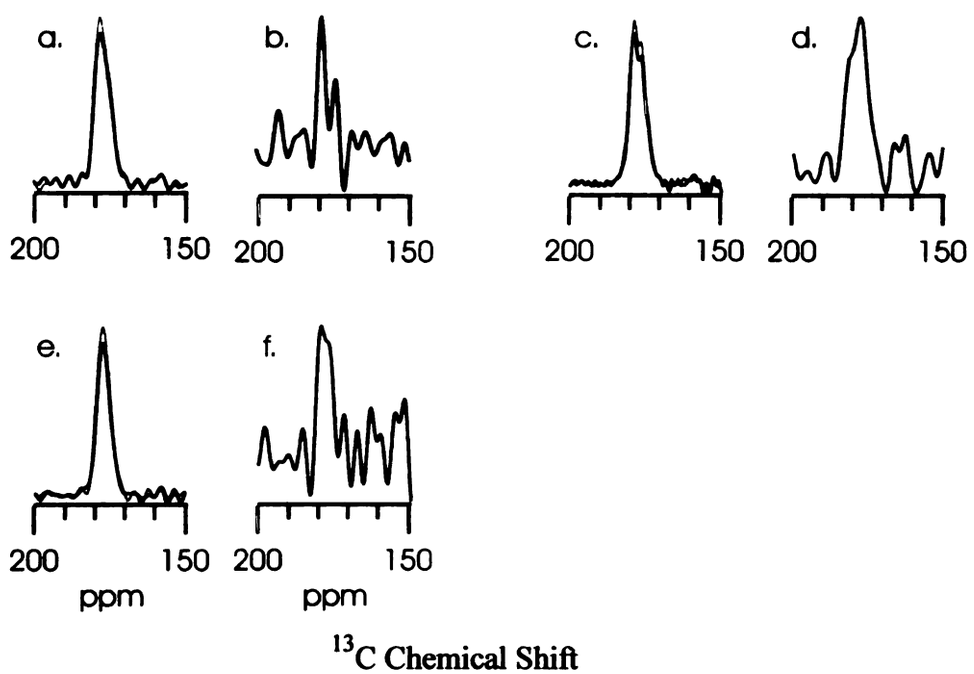


Figure 3-6. REDOR analysis of the N-terminal portion of the soluble ectodomain region of FHA2. (A) S_0/S_1 signal from AD labeling and (B) difference signal representing Ala-36. (C) S_0/S_1 signal from AI labeling and (D) resulting Ala-44 signal. (E) S_0/S_1 signal from VI labeling and (F) resulting Val-55 signal.

F. Hinge Region

The hinge region of the protein, results shown in figure 3-7, was very similar to the previously solved crystal structure. Positions 98, 99, and 100 were observed to be in a helical conformation, the same as the crystal structure. The crystal structure also showed a region around residues 130-140 where the individual proteins of the two trimers in the structure exhibited different secondary structure. For example, Ser-137 exhibited a wide range of dihedral angles, with φ values of 70.7°, -93.7°, -43.0°, 59.0°, -110.8°, and –84.7°. The dihedral angles of ψ were –46.1°, -140.0°, 138.0°, 32.1°, 146.1°, and 161.4°. These values are consistent with a lack of ordered secondary structure at this position. REDOR analysis of positions from this region resulted in a broad signal that overlapped the ranges for multiple secondary structures. A broad signal overlapping the chemical shift range of multiple secondary structures is consistent with observing the protein in multiple conformations at that residue. This confirms the differences seen within the strands of the crystal structure, and the validity of the REDOR method to accurately describe the secondary structure at the position studied. Even if that is a lack of overall consistent secondary structure, this can be indicated by REDOR analysis.

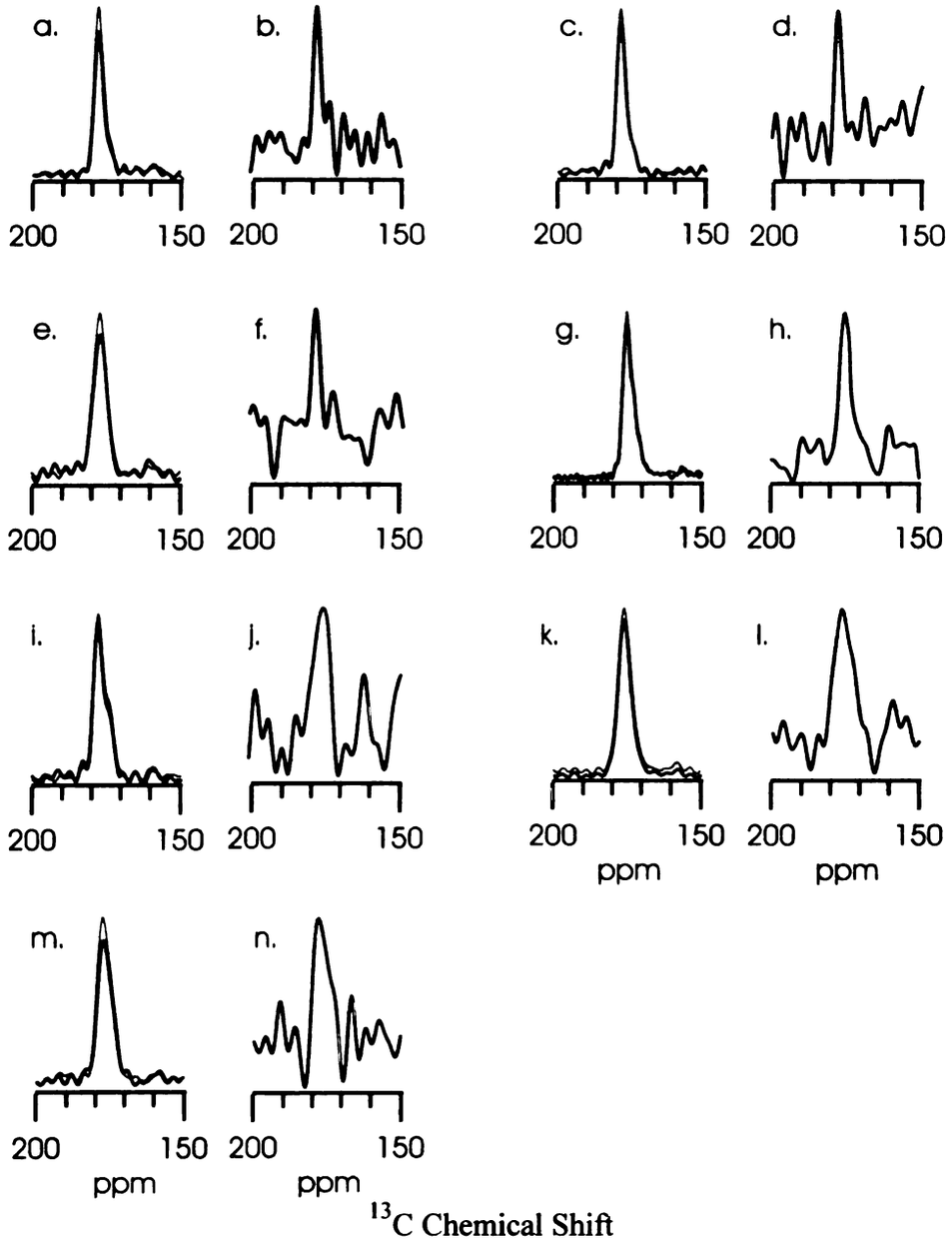


Figure 3-7. REDOR analysis of the hinge region of FHA2. (A) S_0/S_1 signal from LL labeling and (B) difference signal representing Leu-98. (C) S_0/S_1 signal from LV labeling and (D) resulting Leu-99 signal. (E) S_0/S_1 signal from VA labeling and (F) resulting Val-100 signal. (G) S_0/S_1 signal from LF labeling and (H) resulting Leu-119 signal. (I) S_0/S_1 signal from MG labeling and (J) resulting Met-133 signal. (K) S_0/S_1 signal from GS labeling and (L) resulting Gly-134 signal. (M) S_0/S_1 signal from SF labeling and (N) resulting Ser-137 signal.

G. C-Terminal Region

The C-terminal region of the soluble ectodomain also yielded a result different than the crystal structure, at Val-161. A sharp signal at 176.8 ppm was observed, corresponding well to the known helical region of 177.7 +/- 1.4 ppm. The dihedral angles of this region in the crystal structure indicate that this residue is within a long beta strand region (φ values for the six strands of the helical structure ranging from –106.8 to –118.2° and ψ values from 107.0 to 117.8°) The chemical shift range for a valine in β-strand conformation is 174.8 +/- 1.4 ppm, significantly different from the observed chemical shift, shown in figure 3-8. This difference with the crystal structure indicates that more study of the C-terminal domain of the protein is required, but the lack of reasonable target residues in this region makes this difficult. Chapter 4 discusses potential solutions to situations like this where due to cellular metabolism, the cost of the labeled amino acids, or the lack of a unique sequential pair, the original labeling method is impractical.

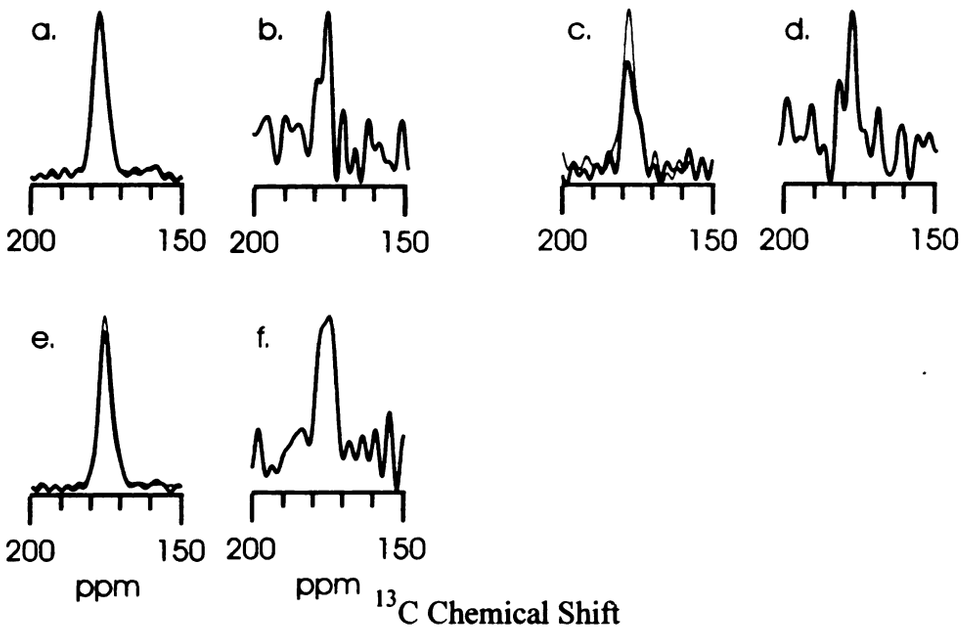


Figure 3-8. REDOR analysis of the C-Terminal region of FHA2. (A) S_0/S_1 signal from VY labeling and (B) difference signal representing Val-161. (C) v signal from AL labeling and (D) resulting Ala-102 and Ala-166 signal. (E) S_0/S_1 signal from GV labeling and (F) resulting Gly-175 signal.

The chemical shift measurements obtained through the REDOR analysis are summarized in table 2, which also shows the chemical shift range of helical and β-strand secondary structure. The sites studied are in both the fusion peptide region previously solved by liquid state NMR and in the soluble ectodomain region solved by crystallography, as well as two positions from the region of residues 21 to 33 in-between the two solved structures, the "missing link." Overall, the membrane-associated protein appears to have similar structure to the previous structural studies by crystallography and liquid state NMR, but there are regions that will require further study to confirm this assertion.

Table 3-2. ^{13}CO chemical shifts PC/PG-associated FHA2 at –10 °C

Residue	Position type	Lineshape	^{13}CO peak shift (ppm)	Helical shift (ppm)	β strand shift (ppm)	FP shift (ppm)[a]
Gly-1	single	broad	174.8	175.51	172.55	171.2
Leu-2	double		177.7	178.53	175.67	177.5
Gly-4	single	narrow	177.2	175.51	172.67	
Ala-5	double		178.1	179.40	176.09	179.5
Ala-7	single	narrow	179.1	179.40	176.09	179.3
Gly-8	double		176.3	175.51	172.55	174.5
Gly-16	single	narrow	177.7	175.51	172.55	175.2
Met-17	single	narrow	178.1	177.95	174.83	
Tyr-22	single	narrow	176.5	177.36	174.54	
Gly-23	double		176.3	175.51	172.55	
Ala-36	single	narrow	179.3	179.40	176.09	
Ala-44	double		178.1	179.40	176.09	
Val-55	single	broad	179.1	177.65	174.80	
Leu-98	single	narrow	178.8	178.53	175.67	
Leu-99	single	narrow	178.5	178.53	175.67	
Val-100	single	narrow	178.2	177.65	174.80	
Ala-101	double		178.9	179.40	176.09	
Leu-119	double		177.7	178.53	175.65	
Met-133	single	broad	177.6	177.95	174.83	
Gly-136	single	broad	176.4	175.51	172.67	
Ser-137	single	broad	178.4	175.94	173.66	
Val-161	single	narrow	176.8	177.65	174.80	
Ala-166	double		178.9	179.40	176.09	
Gly-175	single	broad	176.8	175.51	172.55	

a. Fusion peptide (FP) chemical shift values obtained in DTPC/DTPG lipid.

IV. Effect of Temperature on Analysis of the Membrane Associated Protein.

Most of the analysis presented in this thesis occurred at a sample temperature of –10°C, nominally below the freezing point. Previous solid-state structural studies of fusion peptides were conducted at –50°C due to an increased observed signal relative to higher temperatures (*30*). Several positions were studied at both –10 and –50°C, shown in figure 3-9 and table 3-3. Each of these positions exhibited differences in peak chemical shift of less than 1.0 ppm, but a visible change in the linewidth. The experiments conducted at –10°C resulted in sharper lines than those at –50°C. The broadness of lines observed with these type of samples could be attributed to less overall structure in the protein at the lower temperature, or increased motion at the higher temperatures that results in greater conformational averaging and narrower linewidths. Overall the change in chemical shift was minor and all of the observed positions at each temperature still lie within the helical chemical shift range.

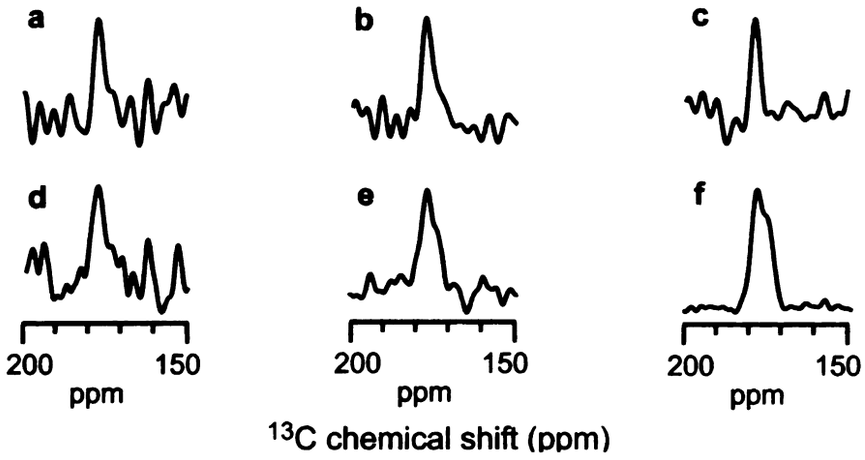


Figure 3-9. Temperature effects of the REDOR experiment. Solid-state NMR $S_0 - S_1$ difference spectra of PC/PG-associated FHA2 samples obtained at (a, b, c) –10 °C or (d, e, f) –50 °C and composed of the filtered ^{13}CO signal of: a, d, Gly-4; b, e, Gly-16; or c, f, Leu-99. The same sample was used at each temperature for the Gly-16 spectra and a different sample was used at each temperature for the Gly-4 and the Leu-99 spectra. Each of the Gly-16 spectra represents approximately the same number of scans.

Table 3-3. ^{13}CO chemical shifts PC/PG-associated FHA2 at –10 °C vs. –50 °C.

Position	^{13}CO peak shift at –10°C (ppm)	^{13}CO peak shift at –50°C (ppm)
Gly-1	174.8	174.7
Gly-4	177.8	177.2
Gly-16	177.7	176.9
Leu-99	178.5	178.0

V. Cholesterol Containing Membranes and pH 7.4 Samples

All of the previously discussed samples were prepared in membranes containing only lipid with a final pH of 5.0, the functional pH of the protein. This is a good start for structural studies, but ideally information should also be obtained for the protein at neutral pH as well. Another variable to be addressed is the composition of the membrane, such as samples associated to membranes containing cholesterol, a composition closer to that of human epithelial cells, which is the target of the influenza virus (*31*). Although we

have not done a complete analysis of the change in protein structure with these alterations to the sample conditions, we have determined that it is possible to prepare the samples and observe REDOR difference signal at an intensity similar to that in our standard samples at pH 5 with no cholesterol.

These changes in sample preparation were tested with the position Leu-98 and Val-100. These positions were chosen for past success with isotopic labeling. Figure 3-10 shows that the signal-to-noise ratio and linewidths for each condition are comparable, and table 3-4 displays the observed chemical shifts. No change in structure in this region is expected with a change in pH or lipid composition and this is consistent with our observed results. Since there have been no structural studies regarding the affect of pH or membrane composition on the membrane associated protein, the future studies will involve first investigating positions scattered throughout the protein to determine which regions are affected by these characteristics. More detailed studies will then be conducted in which several residues are analyzed in these "hot-spots" to determine the changes in secondary structure.

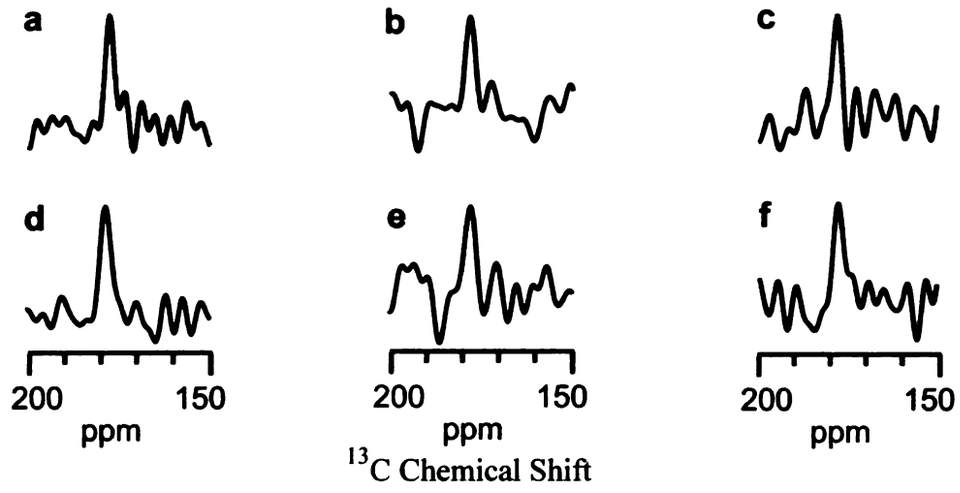


Figure 3-10. Membrane effect of REDOR experiment. Solid-state NMR S_0 – S_1 difference spectra of membrane-associated FHA2 samples as a function of pH and cholesterol content of membranes. Spectra a and d were composed of the filtered ^{13}CO signal of Leu-98 and spectra b, c, e, and f were composed of the filtered ^{13}CO signal of Val-100. Other sample conditions were: a, b, pH = 5.0, PC/PG membranes; c, pH = 7.4, PC/PG membranes; d, e, pH = 5.0, PC/PG/CHOL membranes; f, pH = 7.4, PC/PG/CHOL membranes.

Table 3-4. ^{13}CO chemical shifts PC/PG- or PC/PG/cholesterol-associated FHA2 at –10 °C.

Position	Sample pH	Membrane Composition	^{13}CO peak shift (ppm)
Leu-98	5.0	Lipid	178.8
Leu-99	5.0	Lipid/Cholesterol	178.4
Val-100	5.0	Lipid	178.2
Val-100	5.0	Lipid/Cholesterol	177.6
Val-100	7.4	Lipid	178.1
Val-100	7.4	Lipid/Cholesterol	178.1

VI. Conclusions/Future Work

This chapter presents the use of the REDOR method for secondary structure analysis of 24 positions throughout a large membrane-associated protein, as well as the possible variables such as membrane composition and pH. Structural analysis of membrane proteins, particularly in the lipid environment, is difficult and to our knowledge no studies to date have been able to survey local conformation throughout a membrane protein. The method has proven to be not only effective, but also surprisingly versatile with respect to sample pH and membrane composition. The average amount of researcher hours spent on preparing a sample is less than 10 hours over about a week, which can easily be done in parallel by expressing, purifying, and reconstituting protein with several different labeling schemes at once.

Future work on the structural analysis of the membrane associated influenza fusion protein will focus on investigating the regions where a discrepancy was observed between previous studies and the effect of membrane composition on structure. Changes in the membrane composition such as the presence of cholesterol and pH of the sample preparation will be investigated. This method has the ability to provide information about the missing link region, most likely through the use of molecular biology to introduced REDOR active positions, discussed in more detail in chapter 4. The C-terminal region of the soluble ectodomain appears to have a different structure when associated with membranes than in the crystal structure from aqueous solution. Instead of the previously observed β-strand structure, we have initial information that indicates a helical conformation. Combining this with the initial results from the missing link indicates a possible hairpin structure in which a long helix extends from the fusion peptide region

through the beginning of the soluble ectodomain to the hinge region, after which a second helix extends through the remainder of the ectodomain, possibly to the transmembrane region.

These studies prove the feasibility of studying protein structure when associated with membranes of various composition and pH. This will be further studied to determine the effect of membrane composition, such as the inclusion of cholesterol, and pH on the protein structure. This could give valuable information regarding the proteins specificity for particular host cells, and the observed change in activity at different pH levels.

A third aspect that will be the target of future work is the remainder of the HA2 protein, including the transmembrane and endodomain regions. Currently two approaches are potential options. The first is a construct that includes the full protein attached to the maltose binding protein, for increased expression. The difficulty with this approach is removal of the maltose binding protein following purification. The second approach is to use molecular biology to add the remaining residues to the current FHA2 construct. It is reasonable to attempt the addition of up to 20 amino acids in one site directed mutagenesis experiment, which means that the remainder of the protein could be added by two mutations. The difficulty with this approach may be the addition of so many amino acids in each round of mutagenesis, with the solution being many more rounds adding a fewer number of residues. Also, the longer construct may not express at as high levels as experienced with FHA2. The level of FHA2 expression is high enough that a decrease in expression level of the full protein could still be feasible for structural studies, but as the level decreases the amount of time to prepare each sample will increase making the overall process less efficient.

Overall, all of these areas of future work will continue to be pursued with the intent of general method development. The goal of this project is to not only study the structure of the influenza fusion protein, but to develop methods that can easily be transferred to other fusion proteins.

VII. References

(1) McDaniel, L. E., and Bailey, E. G. (1969) Effect of shaking speed and type of closure on shake flask cultures. *Applied Microbiology 17*, 286-290.

(2) 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. *Journal of the American Chemical Society 129*, 11320-11321.

(3) 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.

(4) Chen, J., Skehel, J. J., and Wiley, D. C. (1999) N- and C-terminal residues combine in the fusion-pH influenza hemagglutinin HA(2) subunit to form an N cap that terminates the triple-stranded coiled coil. *Proceedings of the National Academy of Sciences of the United States of America 96*, 8967-8972.

(5) Wilson, I. A., Skehel, J. J., and Wiley, D. C. (1981) Structure of the haemagglutinin membrane glycoprotein of influenza virus at 3 A resolution. *Nature 289*, 366-73.

(6) 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 Biology 8*, 715-720.

(7) Jaroniec, C. P., Tounge, B. A., Herzfeld, J., and Griffin, R. G. (2001) Frequency selective heteronuclear dipolar recoupling in rotating solids: Accurate ^{13}C-^{15}N distance measurements in uniformly ^{13}C, ^{15}N-labeled peptides. *Journal of the American Chemical Society 123*, 3507-3519.

(8) Yang, J., Parkanzky, P. D., Bodner, M. L., Duskin, C. G., and Weliky, D. P. (2002) Application of REDOR subtraction for filtered MAS observation of labeled backbone carbons of membrane-bound fusion peptides. *J. Magn. Reson. 159*, 101-110.

(9) Park, S. H., Prytulla, S., De Angelis, A. A., Brown, J. M., Kiefer, H., and Opella, S. J. (2006) High-resolution NMR spectroscopy of a GPCR in aligned bicelles. *Journal of the American Chemical Society 128*, 7402-7403.

(10) Wang, J., Balazs, Y. S., and Thompson, L. K. (1997) Solid-state REDOR NMR distance measurements at the ligand site of a bacterial chemotaxis membrane receptor. *Biochemistry 36*, 1699-703.

(11) Wasniewski, C. M., Parkanzky, P. D., Bodner, M. L., and Weliky, D. P. (2004) Solid-state nuclear magnetic resonance studies of HIV and influenza fusion peptide orientations in membrane bilayers using stacked glass plate samples. *Chem. Phys. Lipids 132*, 89-100.

(12) Li, Y., Kijac, A. Z., Sligar, S. G., and Rienstra, C. M. (2006) Structural analysis of nanoscale self-assembled discoidal lipid bilayers by solid-state NMR spectroscopy. *Biophysical Journal 91*, 3819-3828.

(13) Li, Y., Berthold, D. A., Gennis, R. B., and Rienstra, C. M. (2008) Chemical shift assignment of the transmembrane helices of DsbB, a 20-kDa integral membrane enzyme, by 3D magic-angle spinning NMR spectroscopy. *Protein Science 17*, 199-204.

(14) Varga, K., Tian, L., and McDermott, A. E. (2007) Solid-state NMR study and assignments of the KcsA potassium ion channel of S. lividans. *Biochimica Et Biophysica Acta-Proteins and Proteomics 1774*, 1604-1613.

(15) Pauli, J., van Rossum, B., Forster, H., de Groot, H. J. M., and Oschkinat, H. (2000) Sample optimization and identification of signal patterns of amino acid side chains in 2D RFDR spectra of the alpha-spectrin SH3 domain. *Journal of Magnetic Resonance 143*, 411-416.

(16) Goldbourt, A., Gross, B. J., Day, L. A., and McDermott, A. E. (2007) Filamentous phage studied by magic-angle spinning NMR: Resonance assignment and secondary structure of the coat protein in Pf1. *Journal of the American Chemical Society 129*, 2338-2344.

(17) Helmus, J. J., Nadaud, P. S., Hofer, N., and Jaroniec, C. P. (2008) Determination of methyl C-13-N-15 dipolar couplings in peptides and proteins by three-dimensional and four-dimensional magic-angle spinning solid-state NMR spectroscopy. *Journal of Chemical Physics 128*.

(18) Hiller, M., Krabben, L., Vinothkumar, K. R., Castellani, F., van Rossum, B. J., Kuhlbrandt, W., and Oschkinat, H. (2005) Solid-state magic-angle spinning NMR of outer-membrane protein G from Escherichia coli. *Chembiochem 6*, 1679-1684.

(19) De Angelis, A. A., Howell, S. C., Nevzorov, A. A., and Opella, S. J. (2006) Structure determination of a membrane protein with two trans-membrane helices in aligned phospholipid bicelles by solid-state NMR spectroscopy. *Journal of the American Chemical Society 128*, 12256-12267.

(20) Opella, S. J., and Marassi, F. M. (2004) Structure Determination of Membrane Proteins by NMR Spectroscopy. *Chem. Rev. 104*, 3587-3606.

(21) Sinha, N., Grant, C. V., Park, S. H., Brown, J. M., and Opella, S. J. (2007) Triple resonance experiments for aligned sample solid-state NMR of C-13 and N-15 labeled proteins. *Journal of Magnetic Resonance 186*, 51-64.

(22) Rooney, S. A., Nardone, L. L., Shapiro, D. L., Motoyama, E. K., Gobran, L., and Zaehringer, N. (1977) Phospholipids of Rabbit Type-2 Alveolar Epithelial-Cells - Comparison With Lung Lavage, Lung-Tissue, Alveolar Macrophages, and a Human Alveolar Tumor-Cell Line. *Lipids 12*, 438-442.

(23) Morcombe, C. R., and Zilm, K. W. (2003) Chemical shift referencing in MAS solid state NMR. *J. Magn. Reson. 162*, 479-486.

(24) Zhang, H. Y., Neal, S., and Wishart, D. S. (2003) RefDB: A database of uniformly referenced protein chemical shifts. *J. Biomol. NMR 25*, 173-195.

(25) Yang, J., and Weliky, D. P., unpublished experiments.

(26) Tan, K., Liu, J., Wang, J., Shen, S., and Lu, M. (1997) Atomic structure of a thermostable subdomain of HIV-1 gp41. *Proc. Natl. Acad. Sci. U.S.A. 94*, 12303-12308.

(27) 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. *Nat. Struct. Biol. 8*, 715-720.

(28) Avbelj, F., Kocjan, D., and Baldwin, R. L. (2004) Protein chemical shifts arising from a-helices and B-sheets depend on solvent exposure. *Proceedings of the National Academy of Sciences of the United States of America 101*, 17394-17397.

(29) Hovmoller, S., Zhou, T., and Ohlson, T. (2002) Conformations of amino acids in proteins. *Acta Crystallographica Section D-Biological Crystallography 58*, 768-776.

(30) Bodner, M. L., Gabrys, C. M., Parkanzky, P. D., Yang, J., Duskin, C. A., and Weliky, D. P. (2004) Temperature dependence and resonance assignment of ^{13}C NMR spectra of selectively and uniformly labeled fusion peptides associated with membranes. *Magn. Reson. Chem. 42*, 187-194.

(31) Worman, H. J., Brasitus, T. A., Dudeja, P. K., Fozzard, H. A., and Field, M. (1986) Relationship Between Lipid Fluidity and Water Permeability of Bovine Tracheal Epithelial-Cell Apical Membranes. *Biochemistry 25*, 1549-1555.

Chapter 4:

Expanding the Potential of REDOR Through Site Directed Mutagenesis

I. Introduction

The studies conducted on the membrane associated FHA2 fusion protein yielded a great amount of information regarding secondary structure. To this point, the carbonyl chemical shift of 22 residues have been observed, but very few feasible positions exist for further study. Due to the high cost of certain isotopically labeled amino acids or the lack of unique sequential pairs in regions to be studied, the simple labeling methods applied thus far will no longer be useful.

Solutions to these problems use molecular biology. Through the use of site directed mutagenesis, residues of the native protein sequence can be altered. This can be used to insert amino acids which are commercially available for reasonable cost, have high levels of incorporation, or create unique REDOR positions. The creation of unique positions can be accomplished by either switching one of the amino acids of the pair in the region to be studied to create a sequence that did not before exist, or by mutating away the additional pairs in the protein leaving the only REDOR position in the desired location. The main focus of this project is to study the kink region of the fusion peptide or the missing link between the fusion peptide and soluble ectodomain. While the idea of switching amino acids may seem simple, it requires the use of several molecular biology techniques that have not yet been used in this project.

Many projects in protein biochemistry require the use of molecular biology techniques. While a lucky researcher may be provided with a glycerol freeze of bacterial cells ready to express the desired protein, somewhere throughout the project there is

typically a need to make changes to the original plan and this may require alterations to the plasmid used for the expression. Common instances are adding or removing residues from a protein to form a different length of construct, switching certain amino acids to test for the role it plays in structure and function, or to change the amino acid sequence of a protease recognition site. All of these are common when studying proteins, and require an understanding a few basic molecular biology techniques.

A. DNA Purification

Most any experiment involving DNA begins with purifying the DNA from bacteria cells. Cells replicate very rapidly and stocks are easy to replenish, making this an obvious source of plasmid DNA. DNA purification is typically accomplished using commercially available kits, referred to mini-preps for small-scale purifications, medi-preps for medium, and maxi-preps for large-scale purifications. Each company has its own proprietary set of reagents to accomplish the purification, but most use the same basic principles. Initially, the cells are lysed using a high concentration detergent solution, releasing DNA, protein, and other cellular components into the solution. The protein is digested through a general protease reaction, leaving the DNA as the only macromolecules in solution. All other cellular components are then precipitated and removed through centrifugation, producing a clear solution containing the plasmid DNA. This is bound to a resin or membrane, washed with an ethanol solution that will remove contaminants but precipitate the plasmid DNA. The final step is to re-dissolve and elute the plasmid DNA from the membrane using water or a dilute buffer solution. These are simple reactions that can be completed in less than 20 minutes and the typically mini-

prep reaction can yield 5-10 μg of DNA from a 5 mL bacterial culture used for the proteins in this study. The purity of the resulting DNA is measured through the ratio of absorbance at 260 nm to 280 nm. DNA typically dominates absorption at 260 nm, whereas protein (or peptide fragments) dominates at 280. A ratio between 1.80 and 2.0 is considered to be relatively pure DNA, acceptable for use in further experiments.

B. Polymerase Chain Reaction

Polymerase chain reaction (PCR) is used to amplify the quantity of DNA in a sample (*1-3*). Often the DNA we have to start with in any experiment is a very small amount, typically purified from bacterial cells. The process is a simplified version of what occurs *in vivo*, amplifying the genetic material using a DNA synthase enzyme to build the new DNA strand from individual bases using the original as a template.

The PCR reaction is assembled containing the original template DNA, free nucleotide bases, a DNA replicating enzyme, and replication primers in a buffer solution containing an optimized mixture of salts. The free nucleotides, dNTPs, are the building blocks from which the enzyme will synthesize the new strand of DNA. Various enzymes are used for the reaction depending on the specific experiment, but all are specially designed to survive the conditions of the PCR reaction, which includes drastic temperature changes that would inactivate many enzymes. The buffer solution used for the reaction is optimized for maximum performance of the enzyme, as many are sensitive to pH and salt concentrations.

A critical component of the PCR reaction is the primers. Primers are short segments of DNA, typically 12-20 bases in length that are designed to compliment the

strands to be replicated. These are a necessary component of the reaction as the process needs a short segment of DNA to initiate replication. Careful thought must go into designing the primers and involves considering the melting temperature, T_m. This is the temperature at which the primer will separate from the parent DNA, a process referred to as melting. How closely the primer compliments the original DNA and the length of the primer affects this value. As the percent of matching bases increases or the length of the primer increases, so does the temperature required to separate the strands. Another consideration is how specific the primer is to the desired initiation point. A very short primer may match up at several other locations in the parent DNA, but a longer, more specific strand may have a T_m that is above the desired range. The basic formula for calculating the melting temperature of primers is T_m=4°C x (# G and C) + 2°C x (# A and T) (*3*).

Consideration of the T_m is important due to the drastic temperature changes that occur during the PCR reaction. Each cycle is initiated by raising the temperature to over 90°C, causing the original double stranded DNA to melt into two individual strands. The temperature is then lowered below the melting point of the primers, typically around 50 to 60°C, which causes the primers to anneal to the parent DNA. The third step of the cycle is raising the temperature to the optimal point for the replication enzyme, about 72°C. At this point, the enzyme will begin extending from the primers, using the individual DNA bases to assemble the new strand of DNA in complement to the parent strand. The result is double stranded DNA, one being the original parent strand and the other the newly formed copy. The difference between this double stranded unit and the original is that the PCR reaction does not contain a ligase enzyme to seal the ends of the

newly synthesized DNA so the result is a linear strand. If desired, a ligation reaction can be conducted following the PCR reactions, but typically this step is omitted, as the bacterial cells will ligate the linear DNA upon transformation.

This process is then repeated starting from the new parent/daughter strands in which they are separated and each used as a template for a new strand. The primer will associate with the complimentary regions of the DNA, which in the linear DNA will be at the end of the strand, the position where replication was initiated in the first round of PCR. After the second cycle, there is now four times the DNA as in the original solution. The cycles are repeated, which each theoretically doubles the DNA concentration. A general formula for the amount of DNA in the reaction, given that each step is 100% efficient, is $2^{(\#\ cycles)}$. After ten cycles, the amount of DNA in solution is 1024x the original amount. With each cycle only taking about 10 minutes, this is an easy, efficient method to replicate DNA.

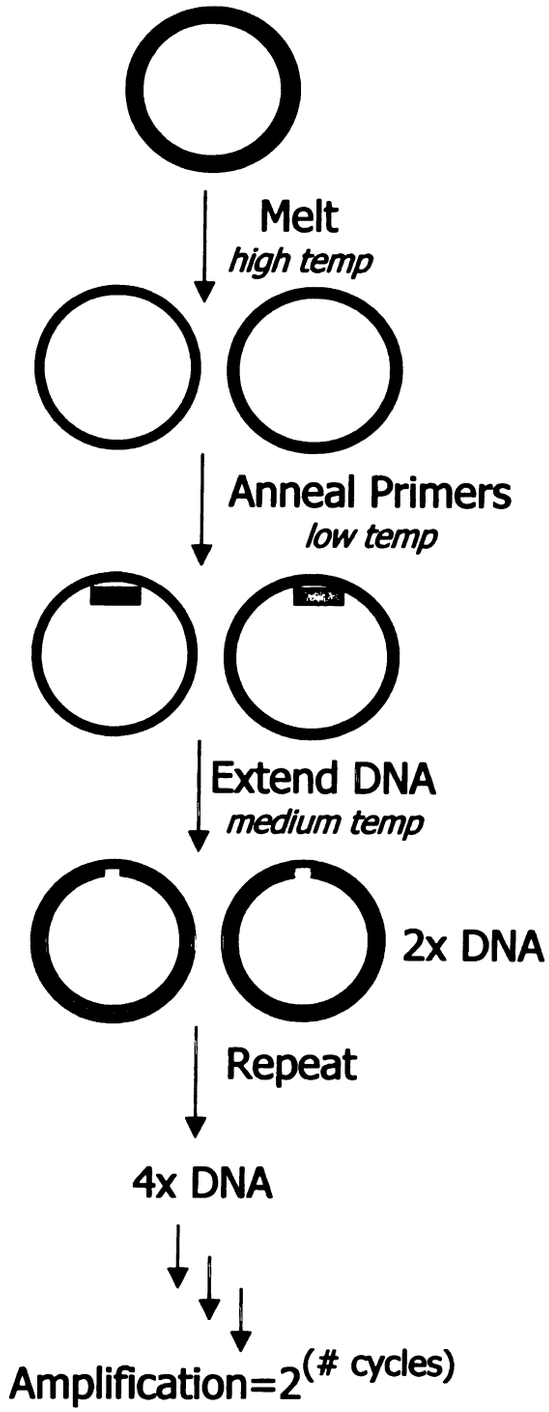


Figure 4-1. Polymerase Chain Reaction. DNA amplification is a repeating cycle of melting the double stranded DNA, annealing oligonucleotide primers, and extending from these primers to produce a copy of the original DNA strand.

C. Site Directed Mutagenesis

Often protein biochemistry requires changing one or several amino acids in the protein sequence. This would be a very difficult task to accomplish after the protein has been produced, and much simpler to achieve by changing the DNA that encodes for that protein. Each amino acid is encoded for by three bases in the DNA sequence, termed codons. Each codon has a complementary tRNA molecule that delivers the appropriate amino acid during protein synthesis. To incorporate a different amino acid, the DNA sequence, and therefore the codon, needs to be changed and this process is referred to as site directed mutagenesis (*3, 4*). For example, the DNA sequence AAA will result in a glycine, but AAT in an alanine. If this is the desired mutation, then the third A simply needs to be switched to a T. This process is referred to as site directed mutagenesis (SDM), and is based on the PCR reaction.

The product of a PCR reaction contains only a very small amount of original DNA and is composed mostly of new DNA originating from the provided primers. If 10 cycles of PCR are completed the original DNA is amplified 1024 times, and the original DNA is only about 0.1% of the product. This makes PCR a good choice for introducing mutations to the original DNA, and takes advantage of the primers used to initiate replication.

For a simple PCR reaction, the primer is designed to exactly match the parent DNA. The length of the primer may be varied in order to obtain the optimal T_m, but typically an exact compliment is used. If the primer is designed that does not perfectly match the parent DNA, if a high enough percentage of the primer still compliments the parent DNA, then it will likely still anneal during the low temperature step of the PCR

reaction. The result is that a bubble is formed at the point where the parent DNA and primer do not match, but if each side of bubble sticks strongly to the DNA, the PCR reaction can still proceed as normal. This is the approach for SDM, to design a primer that contains the desired mutation. Typically the mutated base is surrounded by 12-20 bases that exactly compliment the original DNA, and the PCR reaction proceeds as normal (*3*). Each new strand of DNA comes from the mutated primer and will therefore carry the mutation.

The formula for calculating the melting temperature of SDM primers is: T_m=81.5°C + 0.41 (% GC) – 675/N - % mismatch, where N is the number of bases and the percent mismatch is the fraction of the primer that does not match with a complementary base in the parent DNA (*3*). From this, we can see that as more bases are mutated, the length of the complimentary regions of DNA needs to be extended. It is possible to mutate many amino amino acids in one reaction, with this project succeeding at changing six in one reaction, which requires mutating 18 bases. This large of a change requires a much longer primer in order for proper annealing to occur.

These mutagenesis reactions not only switch one base for another, but also can be used to insert or remove bases, and therefore amino acids in the protein sequence. If bases are to be added, the new bases are inserted into the designed primer and surrounded by complimentary regions. The result is a mutant primer that will form a bubble at the added residues. If residues are to be removed, then the parent strand will then have a bubble at the place in the primer that is missing residues.

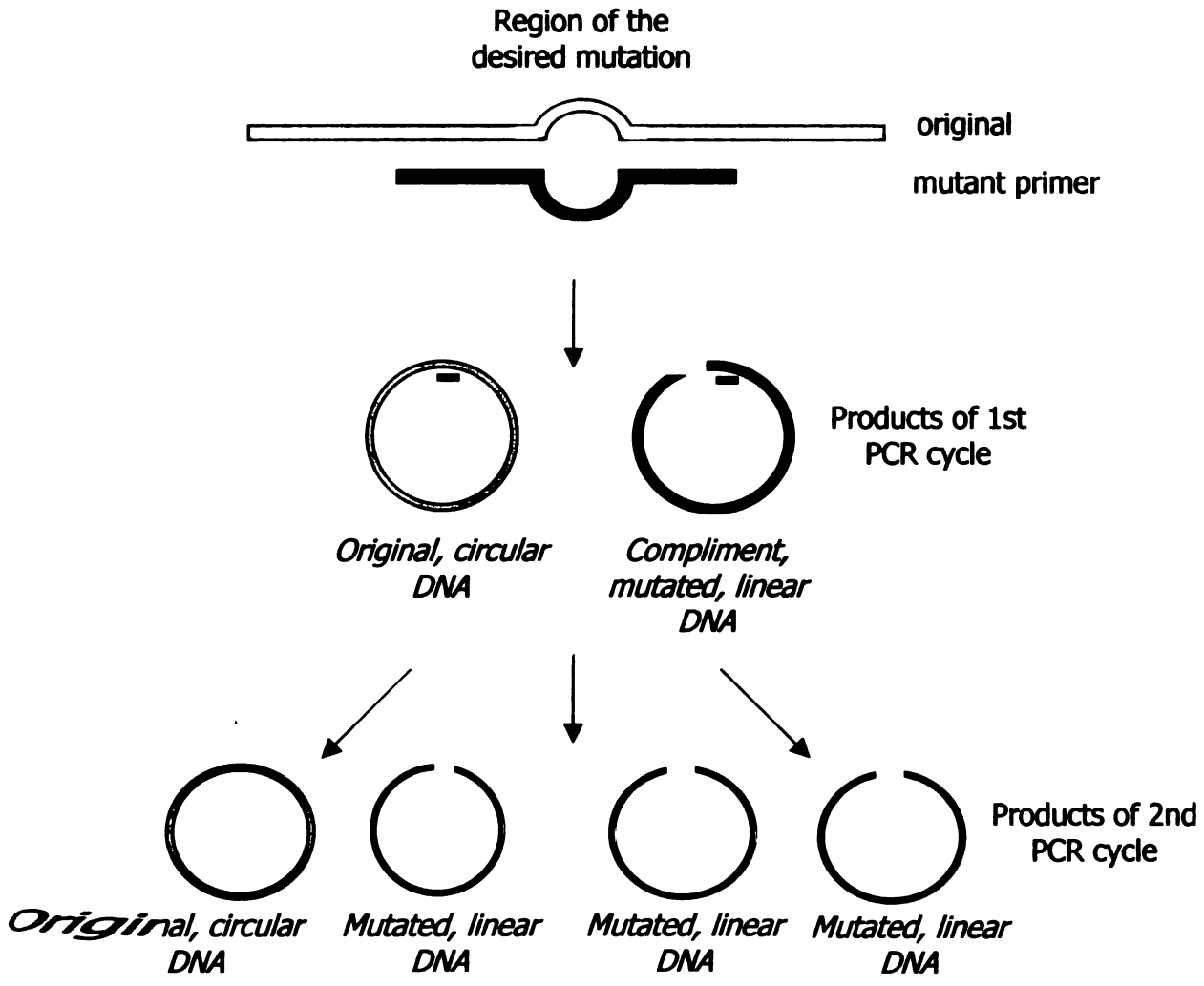


Figure 4-2. Site directed mutagenesis reaction. Mutant primers are used with native DNA, so all newly synthesized DNA contains the mutation and it linear as opposed to the native, circular DNA.

D. DPN1 Digestion

If the SDM reaction is highly efficient and many cycles are completed, then the original DNA will be a very small proportion of the resulting solution, and the mutated DNA will dominate. Any new DNA synthesized must start from the mutated primer, and will then carry the mutation. Unfortunately, the mutagenesis reaction is rarely 100% efficient and the possibility that a noticeable proportion of original DNA may still be in the product. If this DNA were not removed from solution, and the product was carried onto the next step, the resulting cell cultures may contain native DNA. This would result in native protein, and affect the results of future experiments.

To avoid this potential contamination, the un-mutated DNA is commonly removed through a digestion reaction with the DNAse DPN1 (*5, 6*). This enzyme is specific for digesting DNA at the sequence GATC in which the adenine is methylated at the N6 position. Methylated DNA is formed by the DAM enzyme when replication occurs in the bacteria, but in PCR reactions this enzyme is not present and the methylation does not occur. The result is that only the original DNA obtained from the bacteria will contain methylated DNA and be digested by the DPN1 enzyme, and the PCR product DNA will be unmethylated and untouched in the digestion reaction. The DNA fragments can easily be removed in a clean-up reaction that also rids the solution of leftover dNTPS, primers, and enzyme.

E. Transformation

After a PCR or SDM reaction, the next step is often to return the DNA to the bacterial cells for protein expression. This is accomplished through a process called

transformation (*7*). Bacteria are typically very selective about the material that they transport into the cell, but in the right conditions they can take in genetic material from their surrounding solution. Cells treated with calcium chloride are primed for picking up new DNA, and are now referred to as competent cells. Competent cells are often purchased from suppliers and are ready for the addition of DNA. A typical reaction is conducted on a 50 μL sample of competent cells and only requires the addition of 1 ng of DNA, and are often successful with even less. After the addition of the new DNA to the solution, the cells are heat shocked through incubation in a 42°C water bath for 30 seconds. This heat shock step causes the bacteria to transport the DNA into the cell. The component cells are now ready for large-scale growth.

Like any other microbiology reaction, 100% efficiency is typically not achieved. A clever method has been developed to avoid contaminating the new cell culture with cells not containing the new DNA. The new plasmid DNA contains genes for antibiotic resistance, allowing them to grow in media containing that antibiotic when cells without this plasmid would not survive. To select for the newly transformed cells containing the plasmid, the solution is streaked onto a LB plate containing the antibiotic and incubated overnight to grow colonies. All cells that grow will contain the new plasmid.

In the case of SDM, another potential step for contamination of the resulting cell culture is DNA that did not undergo mutation. Original versions of the plasmid that survived the DPN1 digestion may still exist in the newly transformed cells and will also survive the antibiotic resistance selection step. In order to confirm that the cells selected to proceed forward with do contain the mutation, individual colonies of bacterial cells and individually selected from the plate. Each colony grew from a single cell, and will

therefore all contain the same DNA. The DNA from each colony is then sequenced to determine if it contains the mutation, and only successful SDM/transformation colonies are used in future reactions.

F. DNA Sequencing

The final step in a SDM reaction before large-scale expression is to confirm that the obtained cell culture contain the mutated version of the DNA. With the overall dilution of original DNA in the initial SDM reaction, the DPN1 digestion, and the selection for antibiotic resistance, the product is typically the desired DNA. The experiments discussed in this chapter consistently resulted in greater than 90% success in colonies sequenced to confirm the presence of the mutation. Even though the rate of success appears to be high, the consequences of conducting future experiments on proteins that are not actually mutated are so high that DNA sequencing is conducted to confirm the results.

Current DNA sequencing techniques also use the basic principle of PCR. Instead of synthesizing the new DNA strand from native dNTPs, the solution is doped with a small percentage of bases containing a fluorescent label. These bases will add to the growing chain the same as the native base, but the DNA cannot be extended beyond these labeled bases. This caps the new DNA strand at their addition. Since a small amount of these are used, most of the incorporated bases will result in typical DNA synthesis, and only a small fraction will be capped at each position. After the completion of the PCR reaction, there will potentially be a fragment that was capped at each base in the sequence. These fragments can then easily be separated based on size by column

chromatography, and the fluorescence of each eluted species monitored. Each base correlates to a different fluorescence absorportion, and from the order of the observed labeled bases, the DNA sequence can be determined.

This sequencing process is highly efficient and effective, typically sequencing up to 500 bases per analysis. This is usually adequate, but if a larger region of DNA is be sequenced, the process is conducted using multiple primers. Each starts the sequencing at a different position in the plasmid. The results are then pieced together to produce a larger DNA sequence.

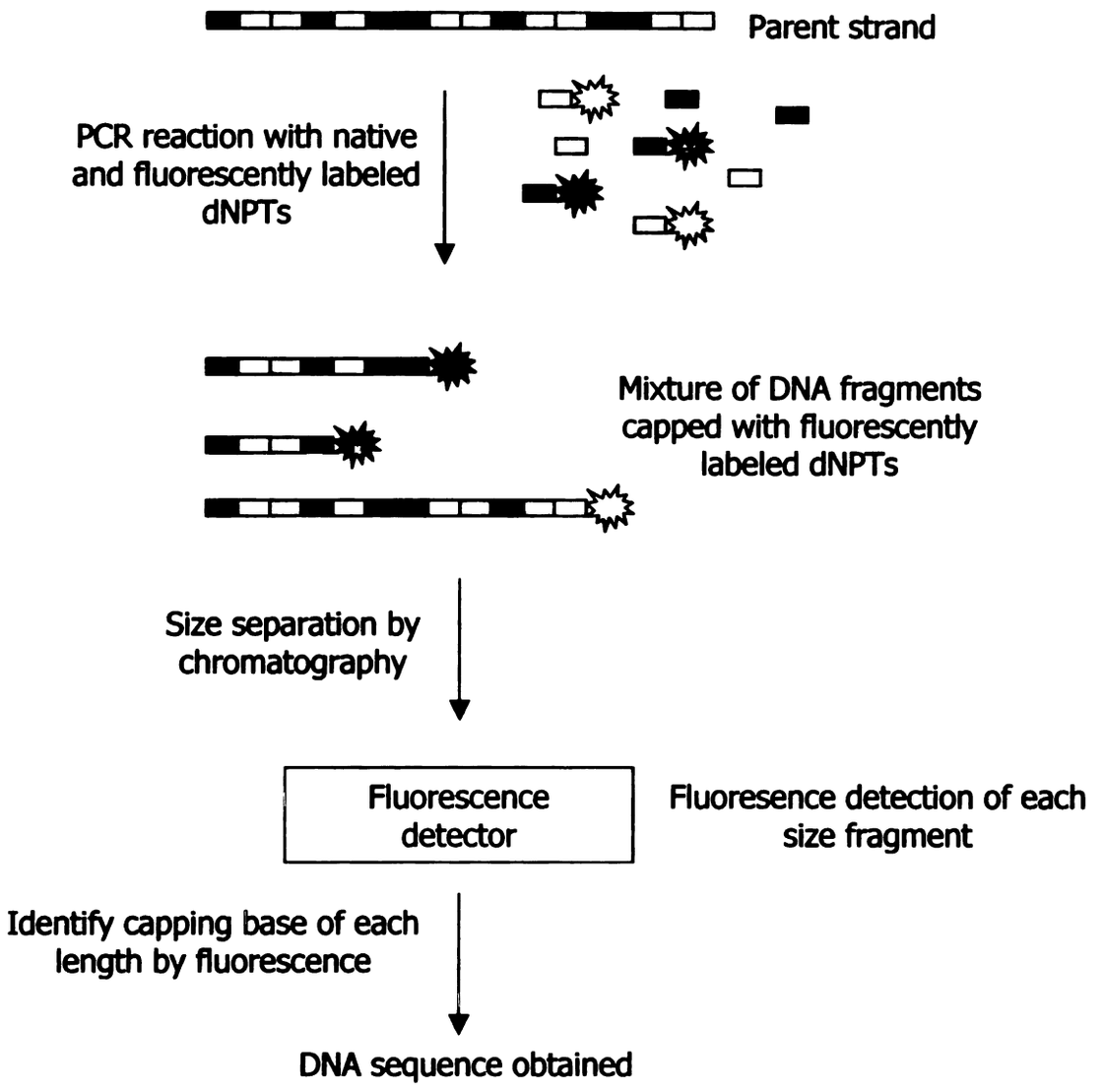


Figure 4-3. DNA Sequencing. The initial step is similar to PCR in which a new DNA strand is synthesized from the original, template DNA. A small portion of the dNTPs used for the replication are tagged with fluorescent markers, and when incorporated terminate the replication. This results in a mixture of fragments that terminate at a variety of lengths. These are separated by chromatography, and the terminal base identified by fluorescence, and from this the sequence is determined.

II. Materials and Methods

A. Materials

The QIAquick PCR purification system was used to purify the DNA following the PCR and site directed mutagenesis reactions. The Wizard Plus SV miniprep kit was used to purify DNA from bacterial cells. PCR and site directed mutagenesis reactions were conducted with the PfuTurbo enzyme from stratagene. Transformations were conducted with BL21(DE3) cells from EMDBiosciences.

B. Molecular Biology

1. DNA Purification. DNA purifications were performed using the Wizard SV Mini-Prep kit following the standard protocol.

2. Polymerase Chain Reaction. All PCR and site directed mutagenesis PCR reactions were conducted using the PfuTurbo polymerase enzyme and following the standard procedure. Clean up reactions were performed using the Qiagen Qiaquick clean-up kit, following the standard procedure.

3. Transformation. All transformation reactions were conducted using BL21(DE3) competent cells following the standard procedure.

4. Selecting Colonies. Following transformation, cells were streaked on LB plates containing the appropriate antibiotic to select for only those colonies containing antibiotic resistance, and therefore the newly introduced plasmid. Plates are made from autoclaved

LB broth containing 15 g/L agar. Antibiotic is added after the media has cooled but has not yet hardened. Immediately prior to streaking cells, the plates are warmed to 37°C but long term storage is in a sealed container at 4°C. Cells are streaked onto the plate by placing several pools of compotent cells (5-20 µl each) and then streaked across the plate using a sterile pipette tip. Plates are incubated overnight at 37°C. Single colonies were selected for further culture growth and sequencing.

5. DNA Sequencing. All site directed mutagenesis reactions were confirmed by high throughput sequencing at the MSU Genomics Facility.

6. DNA and Amino Acid Sequences. All reactions were conducted on the FHA2 construct, which contains the full ectodomain of the HA2 protein.

FHA2 Amino Acid and DNA Sequences. The FHA2 construct contains the residues 1-185 of the HA2 protein, which include the full ectodomain.

GLFGAIAGFIENGWEGMIDGWYGFRHQNSEGTGQAADLKSTQAAIDQIN
GKLNRVIEKTNEKEFHQIEKEFSEVEGRIQDLEKYVEDTKIDLWSYNAELL
VALENQHTIDLTDSEMNKLFEKTRRQLRENAEEMGNGSFKIYHKCDNACI
ESIRNGTYDHDVYRDEALNNRFQIKGVELKSGYKDWVEHHHHHH

Figure 4-4. Protein sequence of the FHA2 construct from the X31 strain of the influenza virus. The last eight residues are non-native.

*CTAGAAATAATTTTGTTTAACTTTAAGAAGGAGATATACAT*ATGGGCCTATT
CGGCGCAATAGCAGGTTTCATAGAAAATGGTTGGGAGGGAATGATAG
ACGGTTGGTACGGTTTCAGGCATCAAAATTCTGAGGGCACAGGACAA
GCAGCAGATCTTAAAAGCACTCAAGCAGCCATCGACCAAATCAATGG
GAAATTGAACAGGGTAATCGAGAAGACGAACGAGAAATTCCATCAAA
TCGAAAAGGAATTCTCAGAAGTAGAAGGGAGAATTCAGGACCTCGAG
AAATACGTTGAAGACACTAAAATAGATCTCTGGTCTTACAATGCGGAG
CTTCTTGTCGCTCTGGAGAATCAACATACAATTGACCTGACTGACTCG
GAAATGAACAAGCTGTTTGAAAAAACAAGGAGGCAACTGAGGGAAAA
TGCTGAAGAGATGGGCAATGGT*A*GCTTCAAAATATACCACAAATGTGA
CAACGCTTGCATAGAGTCAATCAGAAATGGGACTTATGACCATGATGT
ATACAGAGACGAAGCATTAAACAACCGGTTTCAGATCAAAGGTGTTG
AACTGAAGTCTGGATACAAAGACTGGGTCGAGCACCACCACCACCAC
CACTGA*GATCCGGCTGCTAACAAAGCCCGAAAGGAAGCTGAGTTGGCTGC*
TGC

Figure 4-5 FHA2 DNA Sequence in the plasmid pET24a(+) with HA(1-185). The surrounding DNA is in italics.

III. Investigation of the Fusion Peptide Kink.

Previous studies of the HA2 fusion peptide indicate a kink at positions 11 and 12. The goal is to investigate this in the membrane associated fusion protein, but is a difficult task due to the amino acids involved. Asparagine with an ^{15}N amide label is an expensive amino acid, and the functionality of the side chain may make it difficult to achieve a high level of isotopic labeling with minimal scrambling. An attempt to optimize the level of incorporation is impractical due to the cost of the labeled amino acid.

The second REDOR position in this region to be studied is EN, position 11. This position introduced two difficulties in that both amino acids are expensive and due to the functional groups in the side chain pose the potential for problems with incorporation and it is not a unique sequential pair. The first task will be to mutate the other two EN positions, at 104 and 129 by substituting an aspartic acid for the glutamic acid at these positions. This substitution should have a minimal, if any, effect on the structure. Also, both locations are not near the position to be studied, a change at these locations would not necessarily affect our study. If the mutation reactions are successful, the isotopic labeling will be attempted.

A. Site Directed Mutagenesis N12A Primer Design. The codon for the N-12 position is AAT. A change to GCT will encode for an alanine and will require mutating 2 bases.

*CTAGAAATAATTTTGTTTAACTTTAAGAAGGAGATATACAT*ATGGGCCTATT
CGGCGCAATAGCAGGTTTCATAGAA**GC**TGGTTGGGAGGGAATGATAG
ACGGTTGGTACGGTTTCAGGCATCAAAATTCTGAGGGCACAGGACAA
GCAGCAGATCTTAAAAGCACTCAAGCAGCCATCGACCAAATCAATGG
GAAATTGAACAGGGTAATCGAGAAGACGAACGAGAAATTCCATCAAA
TCGAAAAGGAATTCTCAGAAGTAGAAGGGAGAATTCAGGACCTCGAG
AAATACGTTGAAGACACTAAAATAGATCTCTGGTCTTACAATGCGGAG
CTTCTTGTCGCTCTGGAGAATCAACATACAATTGACCTGACTGACTCG
GAAATGAACAAGCTGTTTGAAAAAACAAGGAGGCAACTGAGGGAAAA
TGCTGAAGAGATGGGCAATGGT*A*GCTTCAAAATATACCACAAATGTGA
CAACGCTTGCATAGAGTCAATCAGAAATGGGACTTATGACCATGATGT
ATACAGAGACGAAGCATTAAACAACCGGTTTCAGATCAAAGGTGTTG
AACTGAAGTCTGGATACAAAGACTGGGTCGAGCACCACCACCACCAC
CACTGA*GATCCGGCTGCTAACAAAGCCCGAAAGGAAGCTGAGTTGGCTGC*
TGC

Forward:
GCAATAGCAGGTTTCATAGAAGCTGGTTGGGAGGGAATGATAG
Reverse: CTATCATTCCCTCCCAACCAGCTTCTATGAAACCTGCTATTGC
GC content: 46.51%
Melting temp: 80.1°C
Length: 43 bp
Forward primer MW: 13475.87 Da
Reverse primer MW: 12968.60 Da

Figure 4-6. DNA resulting from the mutagenesis reaction N12A.

B. Site Directed Mutagenesis EN Primer Design. The glutamic acid residues are positions 104 and 129 will be mutated to aspartic acids to leave the EN at position 11 as a unique pair. The codon for glutamic acid 104 is GAG and for 129 is GAA. Both will be mutated to GAC, which for each case will require only a one base substitution. Two sites will need to be mutated, and will require two PCR-SDM reactions, conducted sequentially and separately. After mutation to remove the first EN site and confirmation by DNA sequencing, the second reaction will be conducted.

*CTAGAAATAATTTTGTTTAACTTTAAGAAGGAGATATACAT*ATGGGCCTATT
CGGCGCAATAGCAGGTTTCATAGAAAATGGTTGGGAGGGAATGATAG
ACGGTTGGTACGGTTTCAGGCATCAAAATTCTGAGGGCACAGGACAA
GCAGCAGATCTTAAAAGCACTCAAGCAGCCATCGACCAAATCAATGG
GAAATTGAACAGGGTAATCGAGAAGACGAACGAGAAATTCCATCAAA
TCGAAAAGGAATTCTCAGAAGTAGAAGGGAGAATTCAGGACCTCGAG
AAATACGTTGAAGACACTAAAATAGATCTCTGGTCTTACAATGCGGAG
CTTCTTGTCGCTCTGGA**C**AATCAACATACAATTGACCTGACTGACTCG
GAAATGAACAAGCTGTTTGAAAAAACAAGGAGGCAACTGAGGGA**C**A
ATGCTGAAGAGATGGGCAATGGT*A*GCTTCAAAATATACCACAAATGTG
ACAACGCTTGCATAGAGTCAATCAGAAATGGGACTTATGACCATGATG
TATACAGAGACGAAGCATTAAACAACCGGTTTCAGATCAAAGGTGTTG
AACTGAAGTCTGGATACAAAGACTGGGTCGAGCACCACCACCACCAC
CACTGA*GATCCGGCTGCTAACAAAGCCCGAAAGGAAGCTGAGTTGGCTGC*
TGC

Primers for 104 mutation
Forward: TGTCGCTCTGGA**C**AATCAACATACA
Reverse: TGTATGTTGATT**G**TCCAGAGCGACA
GC content: 44.0%, Melting temp: 68.5°C, Length: 25 bp, Forward primer MW: 7610 Da, Reverse primer MW: 7712 Da

Primers for 129 mutation
Forward: GCAACTGAGGGA**C**AATGCTGAAGAG
Reverse: CTCTTCAGCATT**G**TCCCTCAGTTGC
GC content: 52.0%, Melting temp: 71.8°C, Length: 25, Forward primer MW: 7789 Da, Reverse primer MW: 7534 Da

Figure 4-7. DNA resulting from the mutagenesis reaction to create the unique position EN.

IV. Investigation of the Missing Link Region

Previous structural studies of the HA2 protein have not been able to gain information on the region from positions 18 to 35. The liquid state NMR analysis of the detergent solubilized fusion peptide included region 1 to 20. Crystallography of the soluble ectodomain revealed a structure from positions 34 to 178. This makes the missing region a target for structure studies, as well as the amino acid immediately surrounding this region, as the presence of the full protein as opposed to the truncated version could affect these residues as well. For our purposes, we consider 18 to 35 to be in the "missing link" region we would like to study. The sequence of this region is IDGWYGFRHQNSEGTGQA, which contains many amino acids which are not feasible for REDOR analysis due to high cost or difficulty with isotopic labeling, or pairs which are not unique in the sequence. Several of these positions were targeted for study, including Ile-18, Try-21, Asp-28, Gly-31.

A. Site Directed Mutagenesis I18V Primer Design. The codon for the I-18 position is ATA. A change to GTA will encode for a valine and will require mutating only 1 base. This will create the unique sequential pair VD.

*CTAGAAATAATTTTGTTTAACTTTAAGAAGGAGATATACAT*ATGGGCCTATT
CGGCGCAATAGCAGGTTTCATAGAAGCTGGTTGGGAGGGAATGGTAG
ACGGTTGGTACGGTTTCAGGCATCAAAATTCTGAGGGCACAGGACAA
GCAGCAGATCTTAAAAGCACTCAAGCAGCCATCGACCAAATCAATGG
GAAATTGAACAGGGTAATCGAGAAGACGAACGAGAAATTCCATCAAA
TCGAAAAGGAATTCTCAGAAGTAGAAGGGAGAATTCAGGACCTCGAG
AAATACGTTGAAGACACTAAAATAGATCTCTGGTCTTACAATGCGGAG
CTTCTTGTCGCTCTGGAGAATCAACATACAATTGACCTGACTGACTCG
GAAATGAACAAGCTGTTTGAAAAAACAAGGAGGCAACTGAGGGAAAA
TGCTGAAGAGATGGGCAATGGT*A*GCTTCAAAATATACCACAAATGTGA
CAACGCTTGCATAGAGTCAATCAGAAATGGGACTTATGACCATGATGT
ATACAGAGACGAAGCATTAAACAACCGGTTTCAGATCAAAGGTGTTG
AACTGAAGTCTGGATACAAAGACTGGGTCGAGCACCACCACCACCAC
CACTGA*GATCCGGCTGCTAACAAAGCCCGAAAGGAAGCTGAGTTGGCTGC*
TGC

Forward: GGTTGGGAGGGAATGGTAGACGGTTGGTAC
Reverse: GTACCAACCGTCTACCATTCCCTCCCAACC
GC content: 53.0%
Melting temp: 78.0°C
Length: 30 bp
Forward primer MW: 9400 Da
Reverse primer MW: 9000 Da

Figure 4-8. DNA resulting from the mutagenesis reaction I18V.

B. Site Directed Mutagenesis N28A Primer Design. The codon for the N-28 position is AAU. A change to GCU will encode for an alanine and will require mutating 2 bases. This will create the new unique position AS.

*CTAGAAATAATTTTGTTTAACTTTAAGAAGGAGATATACATATG*GGCCTATT
CGGCGCAATAGCAGGTTTCATAGAAGCTGGTTGGGAGGGAATGATAG
ACGGTTGGTACGGTTTCAGGCATCAA**GC**TTCTGAGGGCACAGGACAA
GCAGCAGATCTTAAAAGCACTCAAGCAGCCATCGACCAAATCAATGG
GAAATTGAACAGGGTAATCGAGAAGACGAACGAGAAATTCCATCAAA
TCGAAAAGGAATTCTCAGAAGTAGAAGGGAGAATTCAGGACCTCGAG
AAATACGTTGAAGACACTAAAATAGATCTCTGGTCTTACAATGCGGAG
CTTCTTGTCGCTCTGGAGAATCAACATACAATTGACCTGACTGACTCG
GAAATGAACAAGCTGTTTGAAAAAACAAGGAGGCAACTGAGGGAAAA
TGCTGAAGAGATGGGCAATGGT*A*GCTTCAAAATATACCACAAATGTGA
CAACGCTTGCATAGAGTCAATCAGAAATGGGACTTATGACCATGATGT
ATACAGAGACGAAGCATTAAACAACCGGTTTCAGATCAAAGGTGTTG
AACTGAAGTCTGGATACAAAGACTGGGTCGAGCACCACCACCACCAC
CACTGA*GATCCGGCTGCTAACAAAGCCCGAAAGGAAGCTGAGTTGGCTGC*
TGC

Forward: GGTTTCAGGCATCAAGCTTCTGAGGGCACAGG
Reverse: CCTGTGCCCTCAGAAGCTTGATGCCTGAAACC
GC content: 50.0%
Melting temp: 75.0°C
Length: 32 bp
Forward primer MW: 9900 Da
Reverse primer MW: 9700 Da

Figure 4-9. DNA resulting from the mutagenesis reaction N28A.

C. Site Directed Mutagenesis W21Y Primer Design. The codon for the W-21 position is UGG. A change to UAC will encode for an tyrosine and will require mutating 2 bases. This will create the new unique position GY.

*CTAGAAATAATTTTGTTTAACTTTAAGAAGGAGATATACAT*ATGGGCCTATT
CGGCGCAATAGCAGGTTTCATAGAAGCTGGTTGGGAGGGAATGATAG
ACGGTT**AC**TACGGTTTCAGGCATCAAAATTCTGAGGGCACAGGACAA
GCAGCAGATCTTAAAAGCACTCAAGCAGCCATCGACCAAATCAATGG
GAAATTGAACAGGGTAATCGAGAAGACGAACGAGAAATTCCATCAAA
TCGAAAAGGAATTCTCAGAAGTAGAAGGGAGAATTCAGGACCTCGAG
AAATACGTTGAAGACACTAAAATAGATCTCTGGTCTTACAATGCGGAG
CTTCTTGTCGCTCTGGAGAATCAACATACAATTGACCTGACTGACTCG
GAAATGAACAAGCTGTTTGAAAAAACAAGGAGGCAACTGAGGGAAAA
TGCTGAAGAGATGGGCAATGGT*A*GCTTCAAAATATACCACAAATGTGA
CAACGCTTGCATAGAGTCAATCAGAAATGGGACTTATGACCATGATGT
ATACAGAGACGAAGCATTAAACAACCGGTTTCAGATCAAAGGTGTTG
AACTGAAGTCTGGATACAAAGACTGGGTCGAGCACCACCACCACCAC
CACTGA*GATCCGGCTGCTAACAAAGCCCGAAAGGAAGCTGAGTTGGCTGC*
TGC

Forward: GGAATGATAGACGGTTACTACGGTTTCAGGCATC
Reverse: GATGCCTGAAACCGTAGTAACCGTCTATCATTCC
GC content: 44.0%
Melting temp: 74.0°C
Length: 34 bp
Forward primer MW: 10500 Da
Reverse primer MW: 10400 Da

Figure 4-10. DNA resulting from the mutagenesis reaction W21Y.

D. Site Directed Mutagenesis G156A Primer Design. The codon for the G-156 position is GGG. A change to GCG will encode for a alanine and will require mutating only 1 base. This will remove the REDOR active position at 156, making the GT at 31 the only active site for this pair.

*CTAGAAATAATTTTGTTTAACTTTAAGAAGGAGATATACAT*ATGGGCCTATT
CGGCGCAATAGCAGGTTTCATAGAAGCTGGTTGGGAGGGAATGATAG
ACGGTTGGTACGGTTTCAGGCATCAAAATTCTGAGGGCACAGGACAA
GCAGCAGATCTTAAAAGCACTCAAGCAGCCATCGACCAAATCAATGG
GAAATTGAACAGGGTAATCGAGAAGACGAACGAGAAATTCCATCAAA
TCGAAAAGGAATTCTCAGAAGTAGAAGGGAGAATTCAGGACCTCGAG
AAATACGTTGAAGACACTAAAATAGATCTCTGGTCTTACAATGCGGAG
CTTCTTGTCGCTCTGGAGAATCAACATACAATTGACCTGACTGACTCG
GAAATGAACAAGCTGTTTGAAAAAACAAGGAGGCAACTGAGGGAAAA
TGCTGAAGAGATGGGCAATGGT*A*GCTTCAAAATATACCACAAATGTGA
CAACGCTTGCATAGAGTCAATCAGAAATG**C**GACTTATGACCATGATGT
ATACAGAGACGAAGCATTAAACAACCGGTTTCAGATCAAAGGTGTTG
AACTGAAGTCTGGATACAAAGACTGGGTCGAGCACCACCACCACCAC
CACTGAGATCCGGCTGCTAACAAAGCCCGAAAGGAAGCTGAGTTGGC
TGCTGC

Forward: GTCAATCAGAAATGCGACTTATGACCATG
Reverse: CATGGTCATAAGTCGCATTTCTGATTGAC
GC content: 38.0%
Melting temp: 71.0°C
Length: 29 bp
Forward primer MW: 8900 Da
Reverse primer MW: 8900 Da

Figure 4-11. DNA resulting from the mutagenesis reaction G156A.

V. Conclusions and Future Work

Several mutants of the FHA2 protein have been produced to allow for the study of the kink in the fusion peptide and missing link region of the protein. DNA sequencing have confirmed successful mutagenesis reactions, and glycerol freezes of each mutant have been produced. Samples produced to investigate the missing link region have yet to studied.

The samples produced to study the kink region of the fusion peptide were unsuccessful. The N12A sample produced quantities of protein similar to the native protein, but did not reconstitute well into lipid membranes. The amino acids isotopically labeled in this sample are ^{13}C-alanine and ^{15}N-Glycine, a pair that has been successfully labeled and studied in the native protein. This study was successful, which indicates that labeling should not be an issue with this sample.

The expression of the protein designed to remove the additional EN positions did not label effectively. Whole cell NMR was used to test for labeling to avoid the use of expensive amino acids in a full scale expression. The result was an extremely weak S_0 signal, indicating a low level of ^{13}C-glutamic acid incorporation. No apparent dephasing was observed, which could indicate a low level of ^{15}N- glutamic acid incorporation but the low level of S_0 signal would make it nearly impossible to observe dephasing even with complete ^{15}N incorporation.

Future work on this aspect of the project will involve producing additional membrane associated samples to further study the missing link region. Based on previous experiments, the isotopic labeling of these positions should be effective and the mutations made should be benign enough to not affect protein expression or membrane

reconstitution. Several other possible REDOR positions that could be created through site directed mutagenesis exist. These include:

- D19L to produce the unique pair LG
- G8A to make 23GF a unique pair
- G23A to produce the unique pair AF
- R25A to create the unique pair FA
- E30A to create the unique position SA
- T32G to create the unique position VG
- Q34G to create the unique position GG

Future work on the study of the kink in the fusion peptide will involve further investigation of the N12A mutant and the isotopic labeling of the EN position. It appears that the N12A mutation may change the structure of the protein, and that the mutant produced is not as functional. This could be confirmed through lipid mixing fluorescence assays. In order for the EN samples to be feasible, the isotopic labeling must be increased. This will involve optimization of variables such as the amount of amino acid added, the time of addition relative to induction, and if additional aliquots of the labeled amino acids are required.

VI. References

(1) Voet, D., and Voet, J. G. (1995) *Biochemistry*, John Wiley & Sons, Inc., New York.

(2) Stryer, L. (1995) *Biochemistry*, 4 ed., W. H. Freeman, New York.

(3) Stratagene (2007) *PfuTurbo DNA Polymerase Product Manual*, Vol. Revision B.

(4) Creighton, T. E. (1993) *Proteins: Structures and Molecular Properties*, 2 ed., W. H. Freeman and Company, New York.

(5) Karttunen, J. T., Lehner, P. J., Sen Gupta, S., Hewitt, E. W., and Cresswell, P. (2001) Distinct functions and cooperative interaction of the subunits of the transporter associated with antigen processing (TAP). *Proceedings of the National Academy of Sciences of the United States of America 98*, 7431-7436.

(6) Hyland, E. M., Cosgrove, M. S., Molina, H., Wang, D. X., Pandey, A., Cotter, R. J., and Boeke, J. D. (2005) Insights into the role of histone H3 and histone H4 core modifiable residues in Saccharomyces cerevisiae (vol 25, pg 10060, 2005). *Molecular and Cellular Biology 25*, 11193-11193.

(7) Novagen (2004) *Compotent Cells Product Manual.*

Chapter 5:

Expression and Refolding of Inclusion Body Proteins

I. Introduction

While there have been a number of high-resolution structures of bacterial membrane proteins in recent years, there have been fewer structures of viral and eukaryotic membrane proteins in part because of difficulties with production of milligram quantities of pure and folded protein by recombinant expression in *E. coli* (*1*). As discussed in earlier chapters, one difficulty with membrane protein expression is the limited cell membrane space. Excess membrane protein is often stored as insoluble aggregates that are termed inclusion bodies (*2-6*). The amount of protein that can be recombinantly produced as inclusion bodies is up to 25% of the total cell mass, making inclusion bodies an attractive target for protein production (*7*).

However, the inclusion body protein has to be solubilized, purified, and properly folded in membranes or detergent in order to be useful for structural studies (*8, 9*). A typical protocol includes denaturation and solubilization using a solution containing a denaturant such as urea or triton, followed by purification and then refolding with solutions containing detergent and possibly lipid vesicles (*10-12*). In this study, refolding was achieved by first mixing the solution of purified denatured fusion protein with a solution containing high concentration detergent and then dialyzing to reduce the denaturant concentration to a negligible level.

While both the influenza and HIV fusion proteins were expressed as inclusion bodies for study by electron microscopy and SS-NMR, the target of the refolding aspects

of this chapter is the FHA2 influenza fusion protein. This protein appears to express into the form of inclusion bodies in a manner common to other membrane proteins, and is therefore a good model for this study.

In the earlier solid-state NMR study, the purified yield of the 23 kDa FHA2 protein expressed in *E. coli* was initially 3 mg/L culture. Through the optimized expression procedures discussed in chapter 2 and the refolding methods of this chapter, the amount of folded protein was significantly increased. Production of 20 mg/L quantities of isotopically labeled FHA2 can be obtained through the combination of the native and inclusion body purifications, which is a very good yield for a non-bacterial membrane protein expressed in *E. coli (13, 14)*.

A. Inclusion Body Purification

The first challenge many researchers face in regards to inclusion bodies is purification. The working definition of an inclusion body is typically a protein that does not solubilize through purification methods that are successful for natively folded protein. Purification for the purpose of studying the inclusion body structure will be discussed in chapter 6, and this chapter will focus on purifying with the intent to refold to active protein.

Inclusion bodies are often observed to be denser than other components of the cell, and this can be taken advantage of for the first purification step. Following cell lysis, the insoluble fraction is separated by centrifugation (*7, 9*). If the centrifugal force is low enough, the inclusion bodies may pellet before other cellular debris and this can act as the

first step in purification. Depending on the level required for future studies, this may be sufficient purity for continuing to the next step of analysis.

After a pellet is obtained containing insoluble material, either by selectively pelleting the inclusion bodies or by collecting all material, two different approaches exist. The first is to suspend the pellet in "gentle" conditions that may act to solubilize other cellular components, leaving the inclusion bodies in the solid form (*8, 9, 15*). In most cases strong denaturants such as triton or urea are harsh enough to solubilize the aggregated protein, but when used a low enough concentrations may only solubilize other cellular components and leave the inclusion bodies in the insoluble form.

The second is to use "harsh" conditions that would solubilize the inclusion bodies, along with many other proteins, producing a solution that can be carried on to other purification steps. Typical reagents used for this process include 8 M urea or 6 M guanadinium chloride, conditions known to denature most any protein (*2, 9*). Once the protein is solubilized, methods used for native purification can often be successful, such as IMAC or gel filtration chromatography (*8*).

B. Protein Refolding

An extensive literature search revealed no current procedures for purifying inclusion body proteins directly to the active, natively folded form. This means that the process of purification is often followed by steps to refold the protein. Typically the purification protocols require the use of denaturing conditions such as urea or harsh detergents. Removal of such reagents is not always an easy task, a delicate balance exists between unfolded or folded monomeric protein and aggregation, typically making this

step the most sensitive (*2, 4, 9*). Quickly changing conditions through dilution or dialysis often results in aggregation. A common method is to remove denaturants through a process of dialysis in which the concentration is lowered stepwise through multiple solutions (*8*).

A second consideration in the refolding process is the concentration of the protein. If the concentration is too high, then the process of aggregation can dominate potentially due to the proximity of the individual protein units in solution (*2, 9, 16*). An answer to this situation can be to add small aliquots of denatured protein to the refolding solution over a long period of time, referred to as pulse renaturation. This method still allows for the refolding of a relatively small amount of protein at a given time, and this approach has proved useful in certain cases (*7*).

A conceptually different approach often used for membrane proteins is to refold by reconstituting into lipid membranes (*11, 17*) . Membrane proteins require the presence of detergent or lipid to maintain their native structure and simply removing the denaturant is not sufficient to achieve total refolding. Transferring from denaturant directly to lipids can achieve both steps in one process.

C. Additives to refolding

The common need to refold proteins has led to many literature examples of successful methods. Many of these rely on the addition of additives such as L-arginine to obtain optimal results (*7, 8, 18-21*). Some protein refolding processes are even benefited by non-denaturing concentrations of urea or guanidinium (*9, 22, 23*). Polyethylene glycol, which at high concentrations often induces precipitation, acts to inhibit aggregation and enhance refolding of certain proteins which proceed through a molten globule intermediate (*24*). For membrane proteins, as well as some soluble proteins, the addition of detergents can increase the success of refolding (*9, 25*).

There appears to be no overall single mechanism by which these additives increase the success of refolding, but it most likely involves the stabilization and increased solubility of the folding intermediates as well as the final product. These additives most likely aid in preventing aggregation, which is typically in competition with protein refolding. The varying possibilities for refolding mechanisms and differing nature of refolding intermediates makes this observation completely reasonable. As studies of protein refolding continue and more information regarding the mechanisms is determined, we may be able to better predict which of these additives would assist in folding a particular protein and limit the amount of "guess and check" currently required.

II. Materials and Methods

A. Materials

The molecular biology aspects of the protein expression are discussed in detail in chapter 2. Luria-Bertani Broth (LB) medium was purchased from Acumedia (Lansing, MI). The detergents n-octyl-β-D-thioglucopyranoside (BTOG) and octyl pentaethylene glycol ether (C8E5) were purchased from Anatrace (Maumee, OH). The ether linked lipids di-*O*-tetradecylphosphatidylcholine (DTPC) and di-*O*-tetradecylphosphatidylglycerol (DTPG) were obtained from Avanti Polar Lipids (Alabaster, AL). Leucine with 1-^{13}C, ^{15}N labeling was purchased from Cambridge Isotope Labs (Andover, MA).

B. Culture Growth

All cell cultures were grown in media containing antibiotic, 15 mg/L kanamycin for FHA2 and 50 mg/L ampicillin for Hgp41. Other than the antibiotic resistance, all other aspects of protein expression were the same for both proteins. Bacterial growth was initiated from a glycerol stock of the recombinant bacterial cells in 1 L of "enriched LB" which contained LB supplemented with 10 mL glycerol. The cell suspension was grown overnight to maximum cell density in a 2.8 L baffled Fernbach flask with a foam closure, shaking at 37°C and 140 rpm. The cell suspension was then centrifuged at 10,000*g* for 10 minutes to produce a solid cell pellet. The pellet was resuspended into 1 L of minimal medium whose optimal composition included the commercial M9 minimal medium salts (6.8 g/L Na_2HPO_4, 3.0 g/L NaH_2PO_4, 0.50 g/L NaCl, 1.0 g/L NH_4Cl), 2.5 g/L $MgSO_4$,

and 10 g/L glycerol at pH 8.0. Cell growth was continued by shaking at 37°C and 140 rpm.

C. Isotopically Labeled Protein Production

After one hour of resuspended cell growth in minimal medium, protein expression was induced by addition of IPTG to a final concentration of 0.2 mM. For production of the FHA2 sample with 1-^{13}C, ^{15}N Leucine isotopic labeling that was studied by SS-NMR discussed later in the this chapter, 100 mg/L of labeled amino acid was added at the time of induction. Protein production was continued for three hours at 37°C. The cell pellet was harvested by centrifugation at 10000*g* for 10 minutes, and the pellet was then stored at –80°C until purification.

1. Gel Electrophoresis. The amount of protein in the soluble vs. insoluble fractions of the cells as well in the membrane-reconstituted samples was monitored using gel electrophoresis, similar to the methods used by other groups studying inclusion bodies and discussed in chapter 2 (*2, 26*). All samples were prepared for analysis by first boiling in 10% SDS for 15 minutes followed by an additional 15 minutes of boiling in Gel Sample Buffer (GSB). The initial SDS solubilization was at an approximate 20x dilution of the sample to a total volume of ~ 0.5 mL, with the addition of 0.5 mL GSB for the second solubilization step. The amount of sample loaded onto the gel was optimized for clear visualization of the protein band intensities, with the intent of comparing the amount of expressed recombinant protein relative to other cellular proteins in each sample. These gels are therefore not adequate for comparing the actual quantity of protein in each sample.

2. Electron Microscopy. Due to the large size of inclusion bodies, microscopy is a common method used to view their presence in bacterial cells (*5, 10, 27, 28*). The control sample was cells that were not induced to produce the recombinant protein. These cells were grown overnight to high density in LB with no addition of IPTG. The samples of inclusion body containing cells were grown in an OD ~ 1 followed by the addition of IPTG to a final concentration of 0.2 mM and overnight protein production in LB. The cells were pelleted, rinsed twice and resuspended with water. Urinyl acetate was used to stain the cells and visualize the dense inclusion body regions.

D. Solubilization, Purification, and Refolding of FHA2 from Inclusion Bodies

The insoluble fraction of the cells was defined as the component which pelleted during the centrifugation of the cell lysate following sonication in lysis buffer containing sarkosyl. The pellet was sonicated in denaturing lysis buffer (8 M urea, 100 mM NaH_2PO_4, 10 mM Tris-Cl), the suspension centrifuged, and the supernatant purified with cobalt His-Select resin using methods similar to those described for native purification in chapter 2. After binding the denatured protein to the resin, the column was washed with 3 column volumes of denaturing wash buffer (8 M urea, 100 mM NaH_2PO_4, 10 mM Tris-Cl, 20 mM imidazole) and the FHA2 was eluted with 5 column volumes of denaturing elution buffer (8 M urea, 100 mM NaH_2PO_4, 10 mM Tris-Cl, 250 mM imidazole). The denatured FHA2 solution was rapidly diluted into twice the volume of ice cold refolding buffer (1 M arginine, 10 mM Tris-Cl, 0.17% decyl-maltoside, 2 mM EDTA at pH 8) and stored at 4°C overnight. Removal of urea and arginine and refolding of FHA2 was

achieved with dialysis at 4°C for two days using 10000 MWCO tubing and a dialysis buffer containing the same concentration buffer and detergent (*12*).

E. Circular Dichroism Spectroscopy

Spectra were obtained at 4°C using a CD instrument (Chirascan, Applied Photophysics, Surrey, United Kingdom), a cuvette with 1 mm pathlength, a 200-260 nm spectral window, wavelength points separated by 0.5 nm, and 0.5 seconds signal averaging per point. For each sample, a difference spectrum was obtained by subtracting the sample buffer from the FHA2 sample. Refolded FHA2 was analyzed in dialysis buffer, and FHA2 from the native purification was analyzed in elution buffer without imidazole. The circular dichroism signal was reported in units of mean residue molar ellipticity, which is a quantity normalized to the concentration of amino acid residues. The samples contained purified FHA2 and the mean residue molar ellipticity was therefore normalized to the FHA2 concentration.

F. Membrane Reconstitution

The membrane composition was a 4:1 molar ratio of the ether linked lipids di-O-tetradecylphosphatidyl-choline (DTPC) and di-O-tetradecylphosphatidylglycerol (DTPG) and was chosen because: (1) choline is a predominant headgroup of lipids of membranes of respiratory epithelial host cells of the influenza virus; (2) the headgroup of DTPG is negatively charged like the headgroups of a minor fraction of the host cell lipids; and (3) DTPC and DTPG are ether- rather than ester-linked lipids and do not have a natural abundance ^{13}C contribution to the carbonyl region probed in the NMR experiments (*29*,

30). The lipids (~40 mg total) and the detergent BTOG (~160 mg) were dissolved in chloroform. The solvent was removed by a stream of nitrogen gas and subsequent overnight pumping in a vacuum chamber. The lipid/detergent mixture was then dissolved in ~5 mL of 5 mM HEPES/10 mM MES buffer at pH 7.4. The FHA2 solution was added to the detergent/lipid solution to form a co-micelle solution of ~8 mg FHA2, detergent, and lipid. The solution was transferred to 10,000 MWCO tubing and dialyzed against 2 L of HEPES/MES buffer at pH 5.0. This pH is comparable to the one for fusion between influenza and endosomal membranes. The dialysis was conducted at 4°C for three days with one buffer change. The FHA2 reconstituted in membranes was then harvested by centrifugation at 50,000*g* for 3 hours.

G. Solid-State NMR Spectroscopy

Data were obtained with a 9.4 T spectrometer (Varian Infinity Plus, Palo Alto, CA), a triple resonance magic angle spinning (MAS) probe, and a 4.0 mm diameter rotor with ~40 μL sample volume. It is estimated that the sample volume contained ~8 mg FHA2 and ~20 mg total lipid. Typical parameters of the rotational-echo double-resonance (REDOR) pulse sequence were: (1) 8.0 kHz MAS frequency; (2) a 6 μs ^{1}H $\pi/2$ pulse; (3) a 1.6 ms cross-polarization period with 63 kHz ^{1}H Rabi frequency and 80 kHz ^{13}C Rabi frequency; (4) a 2 ms dephasing period with alternating 19 μs ^{15}N π pulses and 8 μs ^{13}C π pulses and 88 kHz two-pulse phase modulation (TPPM) ^{1}H decoupling; (5) ^{13}C detection with 88 kHz TPPM ^{1}H decoupling; and (6) 1 sec delay (*29*). Data were acquired without (S_0) and with (S_1) the ^{15}N π pulses during the dephasing period and respectively

represented the full ^{13}C signal and the ^{13}C signal minus ^{13}Cs directly bonded to ^{15}N nuclei. The sample was cooled with nitrogen gas at −10°C to counteract radiofrequency heating. Spectra were externally referenced to the methylene carbon of adamantane at 40.5 ppm, which corresponds to the ^{13}C referencing used in liquid-state NMR of soluble proteins (*31, 32*).

III. Electron Microscopy

The most direct way to observe the presence of inclusion bodies in bacterial cells is through the use of microscopy. As opposed to SS-NMR and gel electrophoresis that give indirect evidence for the presence of the recombinant protein in aggregates, this gives the researcher the ability to directly visualize the contents of the cell. The combination of these two types of results can be used together to make determinations regarding the presence of inclusion bodies in bacterial cells. Transmission electron microscopy was used to visualize inclusion bodies in the induced bacterial cells producing both Hgp41 and FHA2 protein. As a control, cells were also analyzed that were not induced, shown in figures 5-1 through 5-3. These cells exhibit no visible regions of dense, aggregated protein. The cells induced to produce either Hgp41 (figures 5-4 through 5-10) or FHA2 (figures 5-11 through 5-14) all exhibited dense regions of protein attributed to inclusion bodies. The size of the inclusion bodies relative to the cell varied, but in each case appeared to be at least 10% of the cellular volume. The number of inclusion bodies also varied, with some cells possessing one large aggregate and others containing two smaller inclusion bodies.

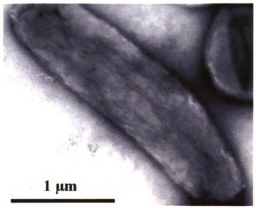


Figure 5-1. Electron microscopy of un-induced bacterial cells.

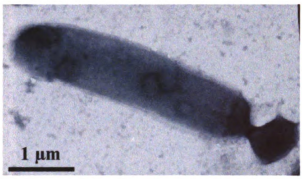


Figure 5-2. Electron microscopy of un-induced bacterial cells.

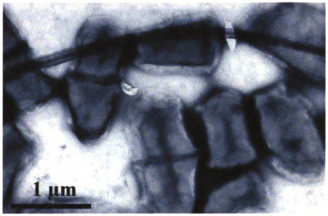


Figure 5-3. Electron microscopy of un-induced bacterial cells.

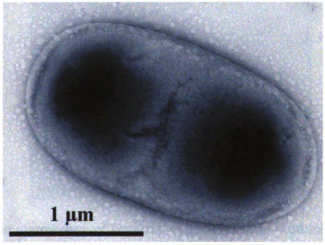


Figure 5-4. Electron microscopy of bacterial cells induced to produce Hgp41 protein.

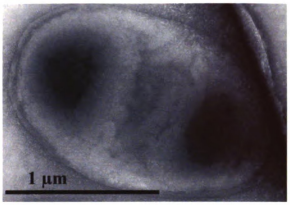


Figure 5-5. Electron microscopy of bacterial cells induced to produce Hgp41 protein.

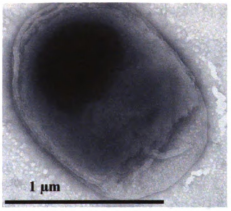


Figure 5-6. Electron microscopy of bacterial cells induced to produce Hgp41 protein.

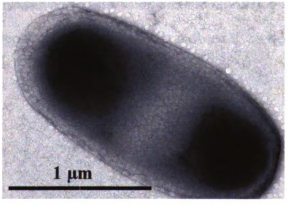


Figure 5-7. Electron microscopy of bacterial cells induced to produce Hgp41 protein.

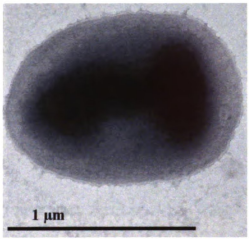


Figure 5-8. Electron microscopy of bacterial cells induced to produce Hgp41 protein.

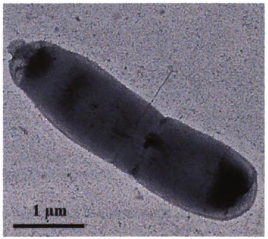


Figure 5-9. Electron microscopy of bacterial cells induced to produce Hgp41 protein.

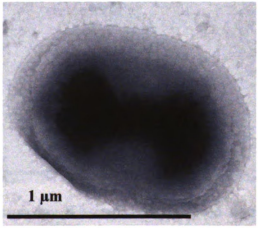


Figure 5-10. Electron microscopy of bacterial cells induced to produce Hgp41 protein.

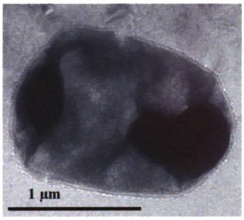


Figure 5-11. Electron microscopy of bacterial cells induced to produce FHA2 protein.

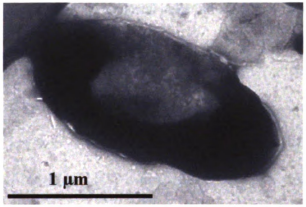


Figure 5-12. Electron microscopy of bacterial cells induced to produce FHA2 protein.

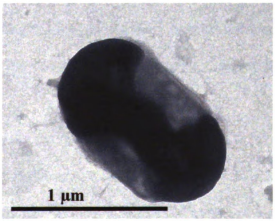


Figure 5-13. Electron microscopy of bacterial cells induced to produce FHA2 protein.

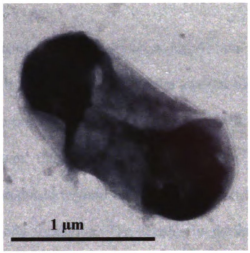


Figure 5-14. Electron microscopy of bacterial cells induced to produce FHA2 protein.

IV. Expression

Overall expression of the fusion proteins Hgp41 and FHA2 was optimized as discussed in chapter 2. All attempts to increase the amount of produced protein did not raise the amount of native protein, but instead the amount of inclusion body protein. For FHA2, a membrane protein, this is most likely due to the limited space in the cell membrane. Any protein produced beyond the level that the membrane can accommodate will result in inclusion body formation. The reason for the low yield of native Hgp41 protein is unclear, as this construct is not in the region of the protein that interacts with the membrane.

The amount of protein in the soluble and insoluble fractions for the expression of each protein was monitored through gel electrophoresis. For the expressed FHA2, the densities of the inclusion bodies was also analyzed. Previous reports have indicated that inclusion bodies are relatively dense in comparison to the other cellular components and may be easily separated by low speed centrifugation (*15*). This was tested by sampling the pellet produced by centrifuging the cell lysate at low (1000x g) vs. high speeds (50,000x g). The high-speed pellet was sampled from both the top and bottom, assuming that more dense components would pellet first and be in higher propitiation at the bottom of the pellet.

Figure 5-15 shows the gel of whole cell as well as these components, and indicates no significant difference in the relative proportion of the expressed FHA2 protein in any of the samples. The whole cell sample exhibits an amount of FHA2 much greater than what is obtained from native purification (~1 mg protein per g of cells, or 0.1% of the total cell mass). The large of amount of protein still present in the insoluble

portions clearly shows that a very small amount of the protein is solubilized through lysis in detergent solution, indicating that the majority of the expressed protein is in the form of inclusion bodies. The similarity of the samples obtained by varying speeds of centrifugation and the location of the sample in the produced pellet shows that for this protein there is no significant difference in the density of the inclusion body protein and the remainder of the cellular components. Future studies could further investigate this by collecting samples at a wider range of centrifugation speeds or in a sucrose gradient.

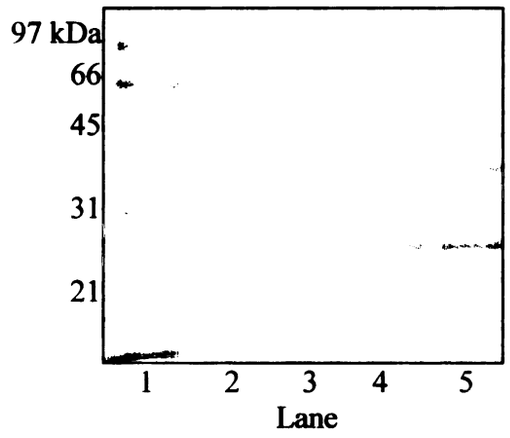


Figure 5-15: Gel electrophoresis of FHA2 expression cell fractions. Lane 1-molecular weight standards, 2-whole cell, 3-pellet after soft spin of 1,000x g, 5-bottom of pellet after hard spin of 50,000x g, 5-top of pellet after hard spin.

The presence of Hgp41 in inclusion bodies was also analyzed through gel electrophoresis. The cells were lysed in sodium phosphate buffer with no detergent followed by high-speed centrifugation to pellet the insoluble components of the cell. The gel in figure 5-16 shows the whole cell, the insoluble fraction, and the soluble fraction. While the relative intensities of other proteins in the samples vary, particularly those between 31 and 45 kDa, the amount of Hgp41 appears to be constant. These results challenge previous reports that this protein is produced solely in an insoluble form that

can only be solubilized in the presence of glacial acetic acid, but do show that a significant portion of the protein is not readily solubilized in aqueous solution as would be expected of natively folded protein.

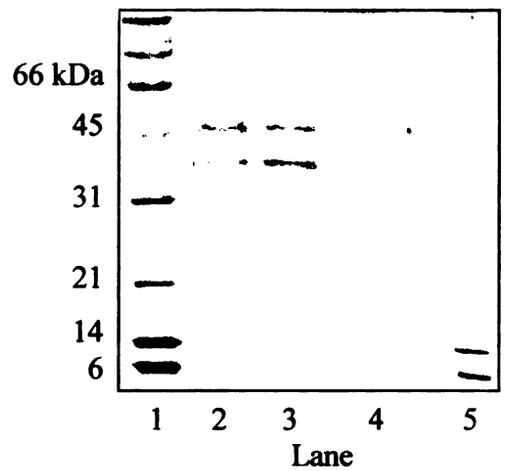


Figure 5-16. Gel electrophoresis of Hgp41 expression cell fractions. Lane 1-standards, Lane 2-whole cell, Lane 3-insoluble fraction, Lane 4-soluble fraction, Lane 5-standards

V. Solubilization, Purification, and Refolding of FHA2 in Inclusion Bodies

The large amount of protein in the form of inclusion bodies that was initially considered to be of no use to our structural studies caused us to reconsider the potential usefulness of this byproduct. In addition to obtaining protein through native purification, methods were also developed to solubilize, purify, and refold the inclusion body protein to a natively folded form that could also be studied through SS-NMR.

The first step in this process was to solubilize the inclusion body protein. Our initial studies showed that this protein could not be easily separated from the remainder of the cellular contents and the purification would require a solubile protein solution. The cell lysis method used to obtain the native protein was considered to be complete because increased lysis time did not increase the purified FHA2 yield. However, comparison of

lanes 2 and 3 in Fig. 5-17 shows that a significant fraction of FHA2 is not solubilized in sarkosyl detergent and is presumably in the form of inclusion bodies. Different amounts of total protein were loaded in the lanes of the gel so the only meaningful comparison between the lanes is the purity of FHA2 relative to the other proteins. Using this metric, the purity of FHA2 is probably somewhat higher in the soluble cell lysate (lane 2) than in the insoluble cell lysate (lane 3).

The insoluble cell lysate pellet was sonicated in 8 M urea denaturing lysis buffer, centrifuged, and the FHA2 in the supernatant was purified using the cobalt resin in the same basic procedure of native purification discussed in chapter 2 with the exception that the purification buffers now all contained 8 M urea. Lanes 4 and 5 in Figure 5-17 show that a large amount of inclusion body FHA2 was soluble in urea and lanes 8 and 9 show that the purification was effective. The FHA2 purity from this protocol is estimated to be ~90% based on relative band intensities in lane 9.

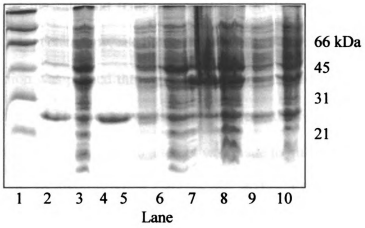


Figure 5-17. Gel electrophoresis of the soluble lysate and inclusion body purifications. A "component" refers to the material used for the lane of the gel and components 2, 3, and 5 are the soluble cell lysate, the insoluble cell lysate, and the portion of the insoluble cell lysate which is soluble in urea. Lane identification: (1) total cell lysate; (2) soluble portion of the cell lysate; (3) insoluble portion of the cell lysate; (4) component 3 suspended in urea; (5) portion of component 4 that is soluble in urea; (6) wash of component 2; (7) elution of component 2; (8) wash of component 5; (9) elution of component 5; (10) molecular weight standards.

The FHA2 from inclusion bodies was purified in urea and therefore denatured and would only be useful in structural or functional studies if it could be refolded. The best refolding was accomplished by: (1) rapid dilution of the FHA2/urea solution into a solution containing 1.0% decyl maltoside detergent, 10 mM Tris-Cl, 2 mM EDTA, and 1 M arginine at pH 8.0; (2) storage overnight at 4°C to begin the refolding process; and (3) two days of dialysis in buffer containing Tris-Cl and decyl maltoside to remove the urea and arginine, with one buffer change after the first day (*12*). Although arginine is required for refolding, its mode-of-action is not yet well-understood (*19, 21*). Very similar CD spectra were obtained for refolded FHA2 and for FHA2 from the native purification, see Fig. 5-18 A and B, which suggests that the refolding is quantitative. The typical yield of refolded FHA2 from inclusion bodies was 10 mg/L culture.

The effects of the rate and temperature of the dilution step in refolding were investigated using CD spectroscopy, see Figure 5-18 C. For the "rapid dilution" protocol, the FHA2/urea solution was pushed through a syringe with a narrow gauge needle into a stirring refolding solution. For the "slow dilution" protocol, the FHA2/urea solution was poured into the refolding solution and no attempt was made to quickly mix the solutions. For both protocols, the resultant mixture solution was dialyzed for two days to remove the urea. Both protocols were done at 4°C and at room temperature and the most helical and presumably most folded FHA2 was obtained with rapid dilution at 4°C. Lower helicity was observed for slow dilution at 4°C, and even lower helicity was obtained for rapid or slow dilution at ambient temperature.

CD spectroscopy was used to assess the overall folding of the purified FHA2, see Fig. 5-18 A. The shape of the spectrum is consistent with predominant helical secondary structure and the $\theta_{222nm} \approx -17{,}000$ deg-cm^2-dmol^{-1} corresponds to ~50% of the residues in helical conformation (*33*). There is a crystal structure for a soluble fragment representing residues 35-178 of FHA2 and a structure for a fragment in detergent representing residues 1-20 of FHA2. For these two fragments, ~60% of the residues are in helical conformation and the CD spectrum of FHA2 is therefore generally consistent with these previous structural studies.

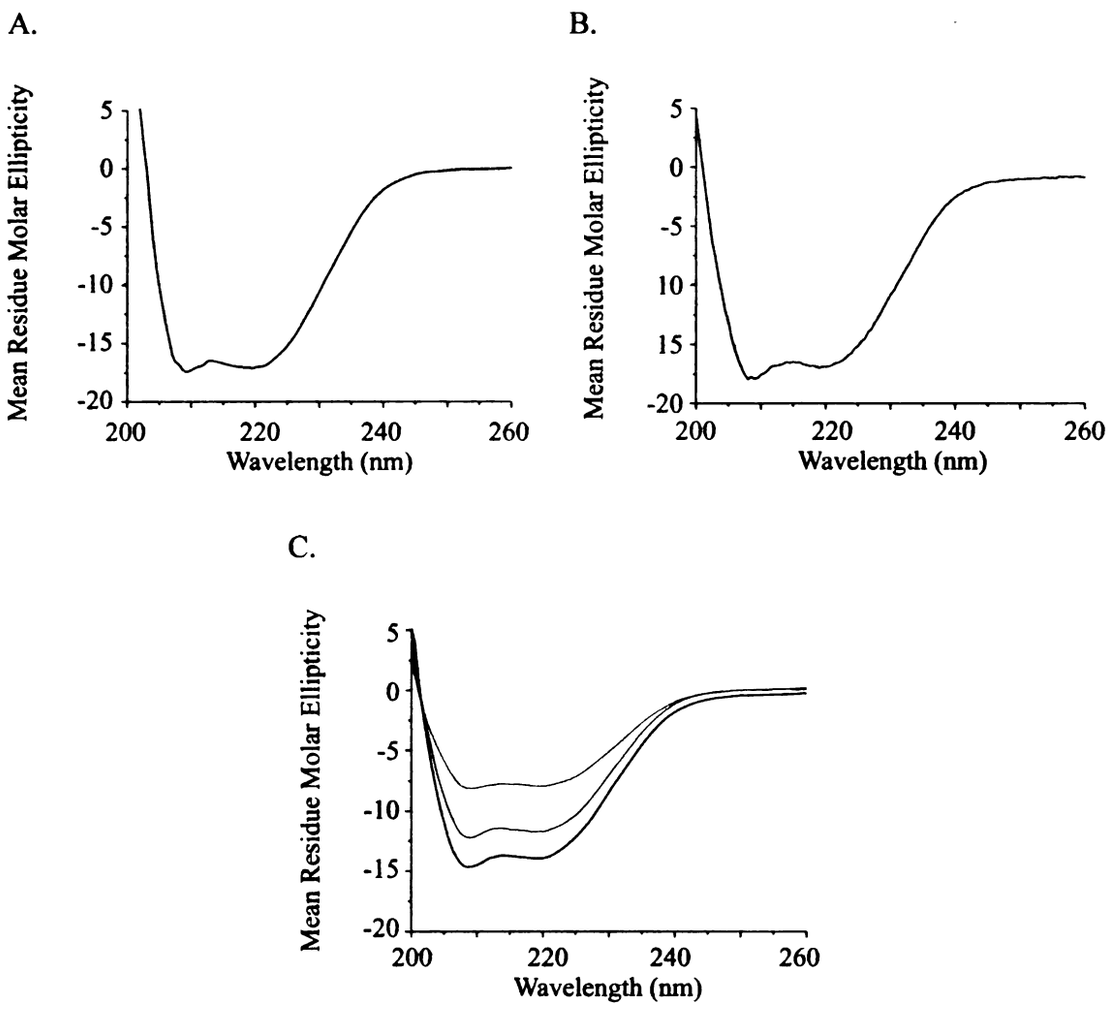


Figure 5-18. Circular dichroism structural analysis of FHA2. The temperature was 4°C and the FHA2 concentrations were 30, 9, and 16 μM in A, B, and C, respectively. (A) FHA2 obtained from purification of native protein from the soluble cell lysate. The θ_{222} value corresponds to ~50% helical conformation. (B) FHA2 obtained from the denaturing purification of the inclusion body protein and refolded using the optimized protocol. The θ_{222} value corresponds to ~50% helical conformation. (C) Investigation of refolding with different dilution protocols: fast dilution at 4°C (bottom line); slow dilution at 4°C (middle line); and fast dilution at room temperature (top line). These spectra were obtained after two days of dialysis. Mean residue molar ellipticity is in units of 10^3 deg-cm^2-dmol^{-1}.

VI. Membrane Reconstitution

One long-term goal of this research is high-resolution structural characterization of FHA2 bound to membranes, which is the most physiologically relevant state. Transfer of FHA2 from detergent to lipid was done by forming co-micelles of the lipid, detergent, and protein, and then removing the detergent via dialysis, the same reconstitution process used for the native protein as discussed in chapter 2. The suspension was centrifuged and FHA2 was only detected in the lipid pellet and not in the supernatant, see lanes 2 and 3 in the gel of Fig. 5-19. This was the result both for FHA2 obtained from purification of the soluble cell lysate and for FHA2 obtained from denaturing purification of the inclusion bodies. This was a successful result not just in the sense that the protein could be studied in the presence of membranes, but also indicates that the refolded protein reconstitutes into membranes as well as natively purified protein. There are still many details to be investigated in regards to the process of reconstitution, but the similar activity of the protein obtained from both purification methods is another piece of evidence that the protein is refolding into the native form.

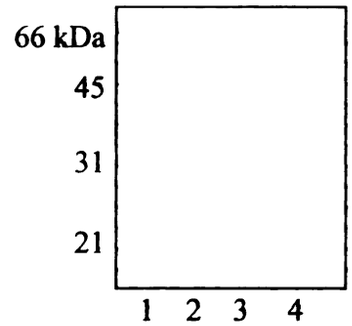


Figure 5-19. Membrane reconstitution of refolded FHA2. SDS-PAGE gel of the pellet (lane 2) and supernatant (lane 3) from centrifugation of membrane-reconstituted FHA2. Lanes 1 and 4 are molecular weight standards. The protein had been purified from inclusion bodies and refolded. The absence of FHA2 in the supernatant suggests that most of the FHA2 is membrane-bound. The gel for membrane reconstitution of FHA2 purified from the soluble cell lysate also showed no protein in the supernatant.

VII. Solid-State NMR Spectroscopy

Solid-state NMR spectroscopy in conjunction with isotopic labeling can provide information about conformation at specific residues in a large membrane-associated protein such as FHA2 (*29*). In this study, solid-state NMR was applied to probe the conformation at Leucine-98, see Fig. 5-20. The SS-NMR analysis is the same approach presented in chapter 3 for the study of native protein associated with membranes. The analysis relied on the well-known conformational dependences of ^{13}CO chemical shifts in proteins. For proteins with high-resolution structures, the distribution of Leucine ^{13}CO shifts is 178.3 ± 1.3 ppm for Leucine in helical conformation and 175.7 ± 1.5 ppm for Leucine in β strand conformation (*31*).

Fig. 5-20A displays the ^{13}C S_0 and S_1 REDOR spectra of a sample prepared with FHA2 obtained using purification of the soluble cell lysate. There appears to be substantial labeling as evidenced by the strong ^{13}CO carbonyl signal in the 170-180 ppm region. In addition, the S_1 ^{13}CO integrated signal intensity is ~6% smaller than the S_0

intensity and is consistent with ~80% incorporation of the 1-^{13}C, ^{15}N Leucine into FHA2 and with the loss of ^{13}CO signal from Leucine-98 which is one of thirteen leucines in the sequence. This loss is expected because: (1) signals from ^{13}COs directly bonded to ^{15}N are highly attenuated in the S_1 spectrum; (2) Leucine-98 residue is the only Leucine in the sequence that is followed by a Leucine; and (3) the natural abundance of ^{15}N is only 0.37%. The $S_0 - S_1$ difference spectrum is displayed in Fig. 5-20B and shows quantitative attenuation of natural abundance ^{13}C signals and a single sharp Leucine-98 ^{13}CO signal with peak shift of 178.3 ppm and a full-width-at-half-maximum linewidth of ~2 ppm. The peak shift is more consistent with Leucine-98 local helical conformation than with β strand conformation. Fig. 5-20C displays a similar difference spectrum for a sample for which the FHA2 was purified and refolded from inclusion bodies. The peak shift of 178.4 ppm is also consistent with helical conformation. In the crystal structure of the soluble fragment representing residues 35-178 of FHA2, Leucine-98 is part of a helix which extends from residue 38 to residue 105. The solid-state NMR data support retention of this conformation in membrane-associated FHA2. The solid-state NMR spectrum in Fig. 5-20 C also provides residue-specific support that the refolding of FHA2 was successful and are complementary to the CD analysis of refolding, Fig. 5-20 B.

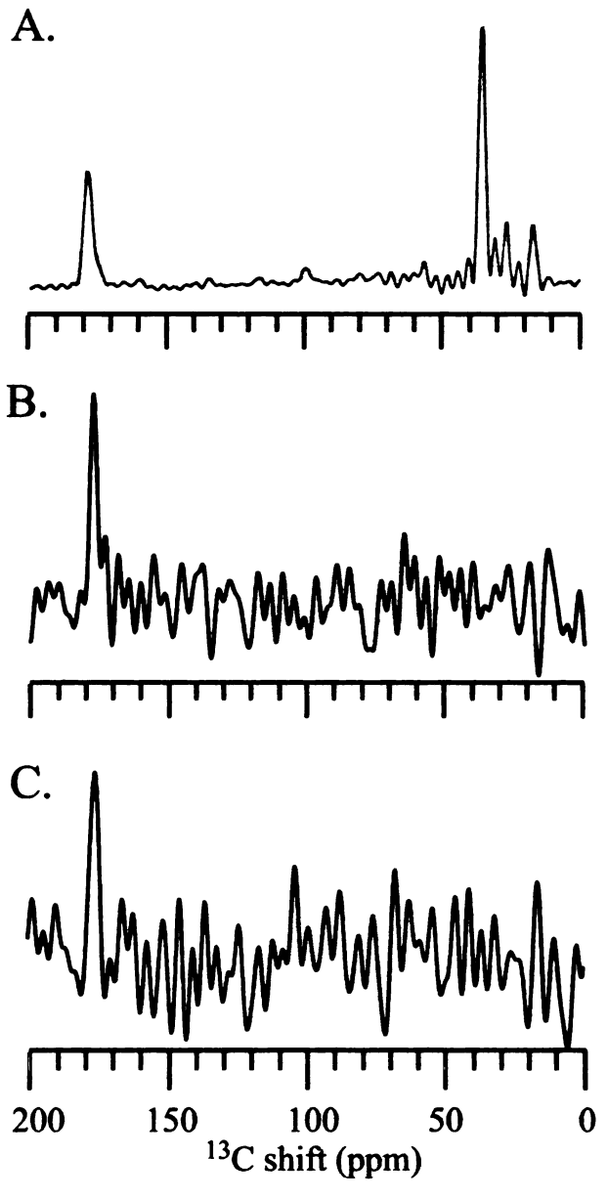


Figure 5-20. ^{13}C solid-state NMR spectra of membrane-associated FHA2. (A) REDOR S_0 (grey) and S_1 (black) spectra of a sample with FHA2 labeled with 1-^{13}C, ^{15}N leucine. The FHA2 was obtained from purification of the soluble lysate. (B) S_0 - S_1 difference of the two spectra from A. This difference is predominantly the Leucine-98 ^{13}CO signal. (C) S_0 - S_1 difference spectrum of a sample containing FHA2 purified and refolded from inclusion bodies. The peak chemical shifts in spectra B and C are 178.3 and 178.4 ppm, respectively, and indicate helical conformation at Leucine-98. The B and C spectra represent 16,000 and 90,000 total scans, respectively.

VIII. Conclusions and Future Work

Any inclusion body FHA2 solubilized by the 0.5 % Sarkosyl detergent was likely folded after exchange into more benign detergent and purification as evidenced by the following properties of the final FHA2 product: (1) no precipitation; (2) high helical content in the CD spectrum; (3) induction of vesicle fusion with much greater activity at low pH than at neutral pH which correlated with the low pH of intact influenza viral fusion; and (4) quantitative reconstitution into membranes at least at the level of comparison of band intensities in a gel similar to the one displayed in Figure 5-19 (*29*).

The FHA2 yield was doubled by solubilization, purification, and refolding of protein from inclusion bodies in the insoluble cell lysate. The refolded inclusion body protein could be isotopically labeled and was quantitatively incorporated into membranes. Refolding was evidenced by the strong similarity of the CD spectrum of the protein that had undergone the refolding protocol to the spectrum of the protein that had never been unfolded in urea, see Fig. 5-18 A and B. Most of the FHA2 structure is outside the membrane and folding is therefore driven in large part by the hydrophobic effect. The high helical content of folded FHA2 is primarily due to formation of a leucine zipper-like trimeric coiled coil where each FHA2 molecule in the trimer contributes one helix and the leucine zipper forms because of the hydrophobic effect (*34*). From the point of view of the hydrophobic effect, it is difficult to understand how the protein would be both soluble and retain high helical content without formation of folded trimers.

Quantitative reconstitution (from Fig. 5-19) of FHA2 into membranes provided some evidence that the FHA2 was not extensively aggregated as both membrane binding and aggregation would likely occur through the hydrophobic fusion peptide and

aggregated fusion peptide would be less likely to bind to membranes than non-aggregated peptide. Poorer membrane binding was observed for FHA2 in detergents other than BTOG detergent and the poorer binding was interpreted as due to more extensive aggregation in these other detergents. In Fig. 5-18 B and C, detection of helical conformation for Leucine-98 in membrane-reconstituted FHA2 by solid-state NMR does provide some evidence that the ectodomain is correctly folded. Leucine-98 has helical conformation in the native fold and is not in the membrane-binding N-terminal domain of FHA2.

Comparison of CD spectra under different conditions showed that the best refolding was obtained from rapid dilution of the FHA2/urea solution at 4°C, see Fig. 5-18 C. The refolding literature proposes that there is competition between aggregation and refolding and this model can be applied to interpret some of the FHA2 refolding results (*19, 20*). For example, in the slow dilution protocol, diffusion of urea will likely be faster than diffusion of FHA2 and the locally high concentrations of unfolded FHA2 may result in aggregation of FHA2. Although aggregation was not visually observed, there was lower helical content with slow dilution. The successful refolding of the inclusion body protein suggests that even higher yields of FHA2 could be obtained with longer induction times and the consequent production of greater quantities of FHA2-containing inclusion bodies.

The cell densities and FHA2 yields described in this paper are much larger than what is typical in the current literature for non-bacterial recombinant membrane proteins and the FHA2 methods should be applicable to some of these proteins. Strengths of the methods include use of shake flasks rather than fermenters and having half the protein

folded in the bacterial membrane and subsequently purified using native methods. The short histidine tag did not interfere with folding and may be compared to larger chaperone proteins such as the maltose binding protein and GST. It is usually desirable to cleave the chaperone protein after purification and then refold the target protein. Each of these steps may be difficult and neither was required for the purification of native FHA2 with the His tag.

Overall the purification and refolding of FHA2 inclusion body protein was successful. The biggest difficulty faced in this pursuit was the yield of the purification. From the gel electrophoresis and electron microscopy it is clear that a much larger amount of protein is present in the form of inclusion bodies than was retrieved through purification. If methods can be developed to obtain a higher yield from this purification, these would potentially be transferrable to other inclusion body proteins. A first step in this pursuit may be to better quantify the amount of protein that is lost during purification, to more accurately track the progress of the method development.

IX. References

(1) Link, A. J., and Georgiou, G. (2007) Advances and chattenges in membrane protein expression. *Aiche Journal 53*, 752-756.

(2) Speed, M. A., Wang, D. I. C., and King, J. (1996) Specific aggregation of partially folded polypeptide chains: The molecular basis of inclusion body composition. *Nature Biotechnology 14*, 1283-1287.

(3) Baneyx, F., and Mujacic, M. (2004) Recombinant protein folding and misfolding in Escherichia coli. *Nature Biotechnology 22*, 1399-1408.

(4) Villaverde, A., and Carrio, M. (2003) Protein aggregation in recombinant bacteria: biological role of inclusion bodies. *Biotechnology Letters 25*, 1385-1395.

(5) Kopito, R. R. (2000) Aggresomes, inclusion bodies, and protein aggregation. *Trends in Cell Biology 10*, 525-530.

(6) Ventura, S., and Villaverde, A. (2006) Protein quality in bacterial inclusion bodies. *Trends in Biotechnology 24*, 179-185.

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

(8) Umetsu, M., Tsumoto, K., Ashish, K., Nitta, S., Tanaka, Y., Adschiri, T., and Kumagai, I. (2004) Structural characteristics and refolding of in vivo aggregated hyperthermophilic archaeon proteins. *Febs Letters 557*, 49-56.

(9) Rudolph, R., and Lilie, H. (1996) In vitro folding of inclusion body proteins. *FASEB 10*, 49-56.

(10) Lilie, H., Schwarz, E., and Rudolph, R. (1998) Advances in refolding of proteins produced in E-coli. *Current Opinion in Biotechnology 9*, 497-501.

(11) Gorzelle, B. M., Nagy, J. K., Oxenoid, K., Lonzer, W. L., Cafiso, D. S., and Sanders, C. R. (1999) Reconstitutive refolding of diacylglycerol kinase, an integral membrane protein. *Biochemistry 38*, 16373-16382.

(12) 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(2) refolds into the trimeric low-pH-induced conformation. *Biochemistry 43*, 5902-5911.

(13) Buck, B., Zamoon, J., Kirby, T. L., DeSilva, T. M., Karim, C., Thomas, D., and Veglia, G. (2003) Overexpression, purification, and characterization of recombinant Ca-ATPase regulators for high-resolution solution and solid-state NMR studies. *Protein Expression and Purification 30*, 253-261.

(14) Tian, C. L., Breyer, R. M., Kim, H. J., Karra, M. D., Friedman, D. B., Karpay, A., and Sanders, C. R. (2005) Solution NMR spectroscopy of the human vasopressin V2 receptor, a G protein-coupled receptor. *Journal of the American Chemical Society 127*, 8010-8011.

(15) Oberg, K., Chrunyk, B. A., Wetzel, R., and Fink, A. L. (1994) Native-Like Secondary Structure in Interleucinekin-1-Beta Inclusion-Bodies by Attenuated Total Reflectance Ftir. *Biochemistry 33*, 2628-2634.

(16) Zettlmeissl, G., Rudolph, R., and Jaenicke, R. (1979) Reconstitution of Lactic Dehydrogenase. Noncovalent Aggregation vs. Reactivation. 1. Physical Properties and Kinetics of Aggregation. *Journal of the American Chemical Society 18*, 5567-5571.

(17) Booth, P. J., and Curnow, P. (2006) Membrane proteins shape up: understanding *in vitro* folding. *Current Opinion in Structural Biology 16*, 480-488.

(18) Brinkmann, U., Buchner, J., and Pastan, I. (1992) Independent domain folding of *Pseudomonas* exotoxin and single-chain immunotoxins: Influence of interdomain connections. *Proceedings of the National Academy of Sciences of the United States of America 89*, 3075-3079.

(19) Baynes, B. M., Wang, D. I. C., and Trout, B. L. (2005) Role of arginine in the stabilization of proteins against aggregation. *Biochemistry 44*, 4919-4925.

(20) Tsumoto, K., Umetsu, M., Kumagai, I., Ejima, D., Philo, J. S., and Arakawa, T. (2004) Role of arginine in protein refolding, solubilization, and purification. *Biotechnology Progress 20*, 1301-1308.

(21) Liu, Y. D., Li, J. J., Wang, F. W., Chen, J., Li, P., and Su, Z. G. (2007) A newly proposed mechanism for arginine-assisted protein refolding - not inhibiting soluble oligomers although promoting a correct structure. *Protein Expression and Purification 51*, 235-242.

(22) Orsini, G., and Goldberg, M. E. (1978) The renaturation of reduced chymotrypsinogen A in Guanidine HCl. *The Journal of Biological Chemistry 253*, 3453-3458.

(23) Yasuda, M., Murakami, Y., Sowa, A., Ogino, H., and Ishikawa, H. (1998) Effect of additives on refolding of a denatured protein. *Biotechnology Progress 14*, 601-606.

(24) Cleland, J. L., Hedgepeth, C., and Wang, D. I. C. (1992) Polyethylene glycol enhanced refolding of bovine carbonic anydrase B. *The Journal of Biological Chemistry 267*, 13327-13334.

(25) Tandon, S., and Horowitz, P. (1986) Detergent-assisted refolding of guanidinium chloride-denatured rhodanese. *The Journal of Biological Chemistry 261*, 15615-15681.

(26) Carrio, M. M., Corchero, J. L., and Villaverde, A. (1998) Dynamics of in vivo protein aggregation: building inclusion bodies in recombinant bacteria. *Fems Microbiology Letters 169*, 9-15.

(27) Vera, A., Gonzalez-Montalban, N., Aris, A., and Villaverde, A. (2007) The conformational quality of insoluble recombinant proteins is enhanced at low growth temperatures. *Biotechnology and Bioengineering 96*, 1101-1106.

(28) Garcia-Fruitos, E., Aris, A., and Villaverde, A. (2007) Localization of functional polypeptides in bacterial inclusion bodies. *Applied and Environmental Microbiology 73*, 289-294.

(29) 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. *Journal of the American Chemical Society 129*, 11320-+.

(30) Rooney, S. A., Nardone, L. L., Shapiro, D. L., Motoyama, E. K., Gobran, L., and Zaehringer, N. (1977) Phospholipids of Rabbit Type-2 Alveolar Epithelial-Cells - Comparison With Lung Lavage, Lung-Tissue, Alveolar Macrophages, and a Human Alveolar Tumor-Cell Line. *Lipids 12*, 438-442.

(31) Zhang, H. Y., Neal, S., and Wishart, D. S. (2003) RefDB: A database of uniformly referenced protein chemical shifts. *J. Biomol. NMR 25*, 173-195.

(32) Morcombe, C. R., and Zilm, K. W. (2003) Chemical shift referencing in MAS solid state NMR. *J. Magn. Reson. 162*, 479-486.

(33) Oshea, E. K., Rutkowski, R., and Kim, P. S. (1989) Evidence That the Leucinecine Zipper Is a Coiled Coil. *Science 243*, 538-542.

(34) Chen, J., Skehel, J. J., and Wiley, D. C. (1999) N- and C-terminal residues combine in the fusion-pH influenza hemagglutinin HA(2) subunit to form an N cap that terminates the triple-stranded coiled coil. *Proc Natl Acad Sci U S A 96*, 8967-72.

S

I. Introductio

Expr

production c

little practic

or about 10

chemistry a

possible wi

purification

regarded as

level of na

purification

proteins is

protein. Th

but many r

with solid-

A. Targets

The

influenza v

the hemagg

Chapter 6:

Solid-State NMR Structural Analysis of Bacterial Inclusion Bodies

I. Introduction

Expression of recombinant protein in bacterial cells is notoriously plagued by the production of insoluble protein aggregates, termed inclusion bodies, that typically are of little practical use. Inclusion bodies can compose up to 50% of the total cellular protein, or about 10% of the total cell mass, making improved methods of inclusion body protein chemistry an attractive target (*1*). Obtaining functional protein from inclusion bodies is possible with denaturants, but often includes time consuming steps of solubilization, purification, and refolding to the native, active form. Often inclusion body protein is regarded as a useless byproduct of recombinant expression, but for some proteins the level of natively folded protein expression is so low that these tedious inclusion body purification methods must be used. A logical initial step in the exploration of these proteins is to determine the structure and how it relates to the natively folded form of the protein. The structural study of inclusion bodies within bacterial cells is a difficult task, but many new approaches have been presented in recent years, including previous work with solid-state nuclear magnetic resonance (SS-NMR) (*2*).

A. Targets of Study

The subjects of our study are fragments of the viral fusion proteins of the HIV and influenza virus. One protein is the FHA2 protein, the 185 residue ectodomain region of the hemagglutinin (HA) protein of the Influenza virus. FHA2 contains a ~20-residue N-

terminal "fu

infection. F

constructs

crystallized

micelles (4

Ar

carbonyl

membran

compared

regarding

helical s

protein.

the for

bacteria

purifica

folded

solubili

act as a

The rer

conditio

study, F

bacteria

terminal "fusion peptide" that binds to the host cell membrane and plays a key role in infection. FHA2 is a membrane protein because of the fusion peptide while shorter constructs which lack the fusion peptide are soluble in aqueous solution and have been crystallized (*3*). There is also an NMR structure of the HA2 fusion peptide in detergent micelles (*4*).

An initial study of FHA2 focused on four residues of this protein, comparing the carbonyl chemical shift of these positions in the whole cell, the insoluble portion, and the membrane reconstituted native protein. The observed carbonyl chemical shift can then be compared to known literature ranges for secondary structure to draw conclusions regarding the form of the protein at these positions. These results indicated that the native helical secondary structure was maintained at these residues in inclusion bodies of this protein, and that it was feasible to study the structure of inclusion bodies in whole cells.

Evidence for a high percentage of the expressed recombinant proteins existing in the form of inclusion bodies includes electron microscopy studies of the expressed bacterial cells, as well as the harsh conditions required to solubilized the protein during purification. Small fractions of both proteins express as the native, folded form. The folded FHA2 protein, a membrane protein requiring the presence of detergent, is solubilized upon sonication in sarkosyl detergent. While sarkosyl, in some instances, can act as a denaturing detergent this process only solubilizes a small fraction of the protein. The remainder of the protein, the inclusion bodies, must be sonicated in the denaturing conditions of 8M urea or 10% SDS for solubilization (*7*). The second protein of this study, Hgp41, also requires harsh conditions for full solubilization. Sonication of the bacterial cells in sodium phosphate buffer solubilizes a small fraction of the protein, but

complete purification of protein is also not possible without a harsh denaturant such as glacial acetic acid.

B. Previous Structural Studies

There is no consensus on the structure of proteins in inclusion bodies, but it is clear that they are non-crystalline solids. The study of solid proteins introduces many obstacles not faced when working with soluble proteins, but methods such as IR and SS-NMR have given insight into the secondary structure. While each method possesses its own strengths and weaknesses and each may have particular situations in which they are superior, we believe that overall SS-NMR overcomes many of the difficulties experienced by other structural study methods. A big concern in the study of inclusion body structure is the sample preparation required. Solid state NMR on whole bacterial cells avoids any purification of the protein and is able to study the most biologically relevant structure. Other methods require differing levels of purification, which could affect the protein structure. Since the structures of proteins in inclusion bodies are not known, it is impossible to predict the potential change of structure that could result from purification.

1. Infrared Spectroscopy. The most common structural method to date has been the study of dehydrated samples of purified inclusion body protein using infrared spectroscopy, which can provide information about the overall fractions of different types of secondary structure. The amide band at 1650-1660 cm^{-1} is attributed to alpha helical or irregular structure, and at 1620-1630 cm^{-1} to β-strand structure (*8*). This can therefore be a useful

tool in observi

these studies is

in most cases,

in the amount

form. This h

associated to

amyloid prote

An IR

native struct

irregular and

aggregates p

also the inc

prepared by

strong deter

dried prior

dehydration

obtained fr

support the

native prote

The

indicated th

by cell lysi

analysis. T

tool in observing the fraction of a particular secondary structure. The main downside to these studies is in the sample preparation, as most require some level of purification and in most cases, dehydration. Most of the results have suggested that there was an increase in the amount of β sheet structure in the inclusion body protein compared to the native form. This has led to the proposal that individual proteins in the inclusion bodies associated together to form non-native intermolecular β-sheets in a structure similar to amyloid protein (*9-14*).

An IR study of Interleukin-1β inclusion bodies revealed a structure similar to the native structure, which includes a large proportion of β-sheet and also fractions of irregular and turn structure (*10*). The preparation of the protein for this study included aggregates produced from thermal denaturation or byproducts of protein refolding, and also the inclusion body protein. The inclusion body protein for these samples was prepared by lysing the cells followed by resuspension in a solution containing triton, a strong detergent that solubilizes the denatured protein. The protein solutions were then dried prior to IR analysis. This method of sample prep, using strong detergents and dehydration, could potentially change the secondary structure of the protein. The results obtained from this study, retention of native β-sheet structure, does not necessarily support the argument for either amyloid formation in inclusion bodies, or the retention of native protein structure as either could be the case.

The study of four hyperthermophilic archaeon proteins by several methods indicated the presence of native helical structure (*8*). These samples were also prepared by cell lysis, then washed in a strong detergent solution, and also dried prior to IR analysis. This study indicated that the interleuken protein studied maintained alpha

helical struct

analyzed for

reach witho

sections 2 a

structure.

The

inclusion

inclusion

dried prio

protein re

in the inc

I

but also

inclusio

27 C st

aggrega

dehydra

spectru

spectru

aggrega

temper

a larger

helical structure in inclusion bodies, but that the other hyperthermophilic proteins analyzed formed non-native β-sheet structure. This conclusion may have been difficult to reach without also using CD and NMR to analyze the secondary structure (discussed in sections 2 and 5), and again the sample preparation methods may have affected secondary structure.

The IR study of a β-galactosidase protein strongly supports the theory that inclusion body proteins adopt an amyloid like structure (*14*). These IR samples were inclusion bodies purified washing the cell lysis pellet with detergent solution and then dried prior to analysis. Comparison of the IR results from soluble and inclusion body protein reveals a large change in secondary structure, the formation of amyloid structure in the inclusion body protein.

IR spectroscopy has been used to study not just the structure of inclusion bodies but also the rate of their formation in cells, which can important in overall understanding inclusion bodies. The study of inclusion body formation of a lipase protein at by 37°C and 27°C study the rate of formation as well as provided information on the structure of the aggregated protein (*11*). Again, these samples were prepared by cell lysis and dehydrating the sample, but in this study avoided the use of strong detergents. The spectrum of the recombinant protein produced was obtained by subtracting the IR spectrum of un-induced cells from the induced cells. These results indicated that while aggregation occurs at both temperatures, when protein production is induced at the lower temperature a greater proportion of the produced protein in the helical form as opposed to a larger aggregation fraction at physiological temperature.

Overall, this set of IR studies fails to reach a concise conclusion regarding the structure of protein in inclusion bodies, but overall support the formation of amyloid structure. The biggest weakness in these studies is the sample preparation methods used. All of these studies were conducted on protein that was to some extent purified from the bacterial cell followed by dehydration prior to IR analysis. While there is no direct evidence to indicate that this could change secondary structure, basic biochemistry principles support the theory that strong detergents and dehydration could affect structure. In cases where the effect is unknown, the safest approach is to minimize the number of steps that could potentially affect structure.

2. Circular Dichroism. The use of circular dichroism to assess protein secondary structure provides information similar to IR, i. e. the overall secondary structure composition without the ability to assess individual residues of the protein. CD must be conducted on a soluble protein solution, which makes the study of the protein in the native, insoluble form impossible. In combination with the previously discussed IR analysis of Interleukin-1β inclusion bodies, CD analysis of protein solubilized by the addition of denaturants was conducted and revealed that even in these harsh conditions the helical structure is maintained (*8*). The argument made by these researchers is that is these conditions change the structure of the inclusion body protein, it should be to unfold the protein and structural studies would therefore yield a lack of secondary structure. The observed presence of any amount of helical structure, albeit small relative to the amount of observed β-sheet structure, indicates that this helical structure must be retained from the inclusion bodies.

3. Microscop

preparation p

evaluate inc

method is th

secondary st

Gree

the formati

also be pro

use of the

proteins w

inclusion

observed.

inclusion

indicates

but are ac

4. Dye E

qualitati

microsc

amyloid

morphog

galactos

3. Microscopy. Another method of inclusion body analysis that avoids the sample preparation pitfalls of IR and CD is microscopy. Using this method, it is possible to evaluate inclusion bodies while still within the bacterial cell. The drawback to this method is the specificity of the information provided. It is not possible to directly analyze secondary structure, but the overall activity can indicate the level of protein folding.

Green fluorescent protein (GFP) is commonly used as a fusion protein to detect the formation and location of a particular protein in cells. Conveniently, this protein can also be produced as inclusion bodies. Folding of this protein is easily monitored by the use of the fluorescence microscopy, folded proteins would be visible while unfolded proteins would not (*15, 16*) . These studies revealed that a significant proportion of the inclusion body protein was actually folded to the extent that the natural fluorescence was observed. Interestingly, the amount of this active protein was high within the center of the inclusion bodies as well and not just associated to the outer layers of the aggregates. This indicates that fully folded protein are not simply associated with the aggregate surface, but are actually a part of the aggregation process.

4. Dye Binding. The specific binding of dye to protein in the amyloid form is another qualitative method that can assess the overall protein folding, similar to electron microscopy. The dyes Thio T and congo red have been used to determine the presence of amyloid in protein samples of the antigen protein ESAT-6, the secretory human bone morphogenetic protein-2, the ectodomain of myelin oligodendrocyte glycoprotein, and β-galactosidase (*14, 17*). Again, this method does not allow for the analysis of specific

residues, but

the formation

5. Solid-Sta

that was a

observing t

bodies exp

previously

within the

inclusion

used lim

conclusi

6. Adva

are stra

protein

in E. c

of the

^{13}CO

confo

confo

in se

sequ

residues, but when used in combination with other spectroscopic techniques can support the formation of amyloid structure.

5. Solid-State NMR. There has also been a solid-state nuclear magnetic resonance study that was able to provide insight into the secondary structure of several proteins by observing the presence of partial helical structure of hyperthermophilic protein inclusion bodies expressed in E. coli (*8*). This study is a major advance when compared to the previously discussed methods in the fact that it can assess structure at individual positions within the protein. While this study introduced the use of solid state NMR in the study of inclusion bodies, the samples were not of the full bacterial cells, and the labeling scheme used limited the amount of information obtained. This makes it difficult to form conclusions about the secondary structure at multiple positions throughout the protein.

6. Advances in Structural Studies. The methods used in the study discussed in this chapter are straightforward, inexpensive, and should be broadly applicable to a wide variety of proteins in inclusion bodies. The viral fusion proteins in inclusion bodies were produced in E. coli cells in a minimal media containing the 1-^{13}C amino acid and ^{15}N-amino acid of the respective N- and C-terminal residues of a unique sequential pair. The filtered ^{13}CO NMR signal of the N-terminal residue of the pair was used to determine its conformation (*5*).The methods were initially developed to study residue-specific conformation in membrane-reconstituted proteins and were based on the observation that in sequences of proteins of moderate size, a large fraction of the residues are in "unique sequential pairs"; e.g. there would only be one instance of a Leu followed by a Leu. A

significant aspect of these studies was the development of methods to produce large amounts of isotopically labeled protein in inclusion bodies, described in detail in previous papers (*5*). In addition, NMR methods such as "rotational-echo double-resonance" (REDOR) can selectively detect the signal of the ^{13}C carbonyl (^{13}CO) nuclei which are directly bonded to ^{15}N nuclei and there are well-known correlations between the backbone ^{13}CO NMR chemical shift of a residue and its local conformation (*5, 6, 18*).

II. Materia

A. Expres

Th

was simil

subsequer

Yeon-Kyu

the Lac p

BL21(DE

with amp

body pro

resistanc

containi

minima

for one

protein

Expres

inclusi

of rec

expre

expre

addi

follo

II. Materials and Methods

A. Expression and labeling of Inclusion Body Proteins

The protocol for expression and labeling of both the proteins in inclusion bodies was similar to that described for FHA2 incorporated in bacterial membranes and subsequently purified in detergent and reconstituted in synthetic membranes (*5*). Dr. Yeon-Kyun Shin at Iowa State University donated the FHA2 plasmid which contained the Lac promoter and kanamycin resistance and the plasmid was transformed into E. coli BL21(DE3) cells. The plasmid of the Hgp41 protein also contained the Lac promoter but with ampicillin resistance, and was transformed into E. Coli BL21 Cells. The inclusion body protein expression was identical for each protein with the exception of the antibiotic resistance. The cells were first grown overnight to an OD_{600} of ~8 in 50 mL media containing luria broth and 10 g/L glycerol. The cells were then switched into 50 mL minimal media that contained 0.25 g glycerol. After continuation of growth in this media for one hour, 10 mg of one or two labeled amino acids were added to the media and protein expression was induced for three hours with 1 mM isopropyl thiogalactoside. Expression was done at 37°C rather than room temperature to augment production of inclusion bodies (*16*). Because the bacterial membrane space is limited, a greater fraction of recombinant protein will be incorporated into membranes at the beginning of the expression period and a greater fraction will be in inclusion bodies at the end of the expression period. Complete labeling of the inclusion bodies was therefore optimized by adding an additional 10 mg dose of labeled amino acid at the one and two hours following the induction of protein expression. After three hours of expression, the cells

were centr

this protoc

B. Solid-S

So

using a 9.

MHz, ^{1}H

was ^{1}H-^{13}

with ^{1}H d

(S_0) and

represent

nuclei. I

differen

in the s

the sam

G-L R

decrea

the re

meth

of V

exp

me

were centrifuged and the NMR sample was taken from the cell pellet. The cell yield with this protocol was typically ~ 0.5 g.

B. Solid-State Nuclear Magnetic Resonance Analysis

Solid-state NMR rotational-echo double-resonance experiments were carried out using a 9.4 T spectrometer and a 4.0 mm MAS probe tuned to ^{13}C detection at 100.8 MHz, ^{1}H decoupling at 400.8 MHz, and ^{15}N dephasing at 40.6 MHz. The pulse sequence was ^{1}H-^{13}C cross-polarization followed by a REDOR dephasing period and ^{13}C-detection with ^{1}H decoupling during the latter two periods. Data were alternately acquired without (S_0) and with (S_1) the ^{15}N π pulses during the dephasing period and respectively represented the full ^{13}C signal and the ^{13}C signal minus ^{13}Cs directly bonded to ^{15}N nuclei. Because the proteins contained a $^{13}CO/^{15}N$ unique sequential pair, the $S_0 - S_1$ difference was predominantly the filtered signal of this pair (*19*). The amount of decrease in the signal observed, dephasing, correlates with the number of residues labeled within the sample. For example, the FHA2 protein contains 16 glycines, but when studying the G-L REDOR pair, only one of these residues would experience dephasing. The expected decrease in signal is ~6%, about what was observed in the experimental results.

The experimental MAS frequency was 8.0 kHz, the dephasing time was 2.0 ms, the recycle delay was 1.0 s, and the ^{13}C chemical shifts were externally referenced to the methylene peak of adamantane at 40.5 ppm which allows direct comparison to databases of ^{13}C chemical shifts as a function of conformation (*6, 20*). Other details of the NMR experiments have been previously described (*5*). Peak ^{13}C carbonyl chemical shifts were measured with ±0.3 ppm precision. During data acquisition, samples were cooled by

flowing

not be t

because

evidence

temperat

reconstit

experime

that minir

Th

residues

protein st

conforma

β amyloi

The chem

referenced

shifts fro

172.0, 17

174.4, 17

175.5, 17

flowing nitrogen gas at a temperature of –30°C. For these samples, the NMR probe could not be tuned when the cooling gas temperature was higher than –30°C presumably because of effects from high salt concentration in the bacterial cells. This reasoning is evidenced by: (1) no problems with tuning the membrane-reconstituted samples at temperatures above –30°C; and (2) the ~10 mM salt concentration in the membrane-reconstituted samples. In the future, it will be possible to do the whole cell NMR experiments at temperatures higher than -30°C using recently developed NMR probes that minimize electric effects in conductive samples (*21*).

The reported distributions of ^{13}CO chemical shifts for Gly, Ala, Val, Ser, and Leu residues in helical conformations are from a database of chemical shifts from known protein structures (*6*). The reported ^{13}CO chemical shift ranges for β sheet/amyloid conformation were based on ^{13}CO shifts of the membrane-associated HIV fusion peptide, β amyloid peptide, human prion amyloid fibrils, and HET-s prion protein (*13, 22-26*). The chemical shifts from the references papers were adjusted so that they would be referenced in the same way as the shifts in the helical distributions. The β sheet ^{13}CO shifts from these samples in ppm units were: Gly, 171.5, 170.3, 170.7, 171.4, 170.7, 172.0, 172.0; Ala, 174.2, 174.9, 175.6, 174.7, 176.0, 175.2, 173.3; Leu, 174.2, 174.7, 174.4, 174.8, 175.0, 173.0; Val, 175.3, 172.6, 174.1, 174.6, 175.7, 173.4, 174.8, 174.8, 175.5, 173.0; Ser, 172.3, 173.4, 172.2, 172.1, 173.8, 174.4, 173.2.

III. FHA2 In

As c

express in

refolding o

analysis of

study of the

Tw

bacterial c

lysed and

more puri

but the r

signal. C

similar r

analysis.

A

position

bacteria

3).

III. FHA2 Inclusion Body Analysis

As discussed in chapter 5, a high proportion of the FHA2 and Hgp41 proteins express in the form of inclusion bodies. We have made progress on purification and refolding of inclusion bodies, but the focus of this project is now shifting towards analysis of the protein structure. Solid-state NMR analysis was effective in the structural study of the inclusion body protein within bacterial cells.

Two types of samples were prepared, one in which the samples were whole bacterial cells with the expressed inclusion bodies, and the other in which the cells were lysed and the soluble portion removed. The insoluble fraction (IF) samples may be a more purified version of the inclusion bodies, as all soluble protein has been removed, but the resulting NMR spectra of each sample type reveal comparable strong, sharp signal. Comparison of positions in which both sample types were analyzed all show similar results, indicating that either sample prep method is adequate for structural analysis.

A total of 13 positions were studied in the FHA2 inclusion body protein. Of these positions, 4 were samples of only the insoluble fraction (figure 6-1), 4 were of the whole bacterial cells (figure 6-2), and 5 positions were analyzed in both sample types (figure 6-3).

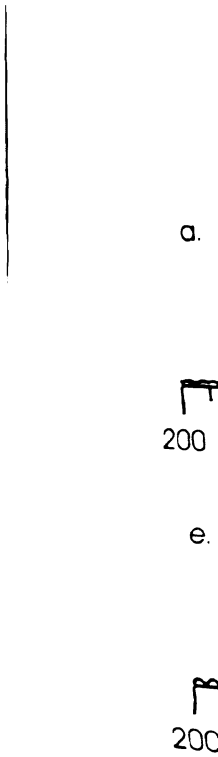


Figure 6-1
of the ba
abbreviate
^{15}N-amide
decreased
single "A
representi
(E) S_0/S_1
GV labeli

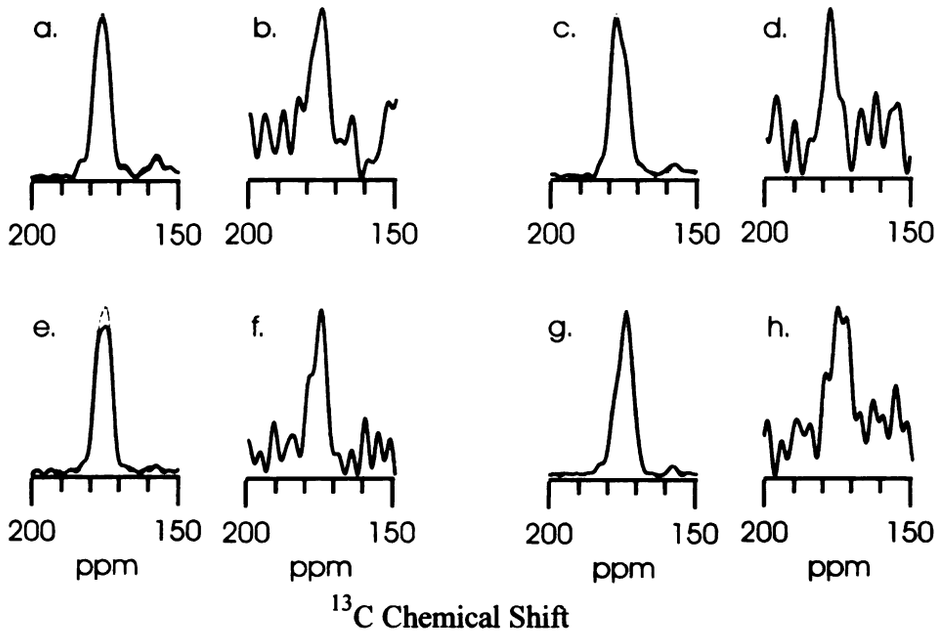


Figure 6-1. REDOR analysis of FHA2 inclusion bodies studied in the insoluble fraction of the bacterial cells. Labeling of selected amino acids in the protein sequence is abbreviated as "AB", where "A" is the ^{13}C-carbonyl labeled amino acid and "B" is the ^{15}N-amide labeled amino acid. This results in an S_0 signal, shown in grey, and a decreased S_1 signal, shown in black. The $S_0 - S_1$ difference represents the signal of the single "A" carbonyl. (A) S_0/S_1 signal from YG labeling and (B) difference signal representing Tyr-22. (C) S_0/S_1 signal from AD labeling and (D) resulting Ala-36 signal. (E) S_0/S_1 signal from VI labeling and (F) resulting Val-55 signal. (G) S_0/S_1 signal from GV labeling and (H) resulting Gly-175 signal.

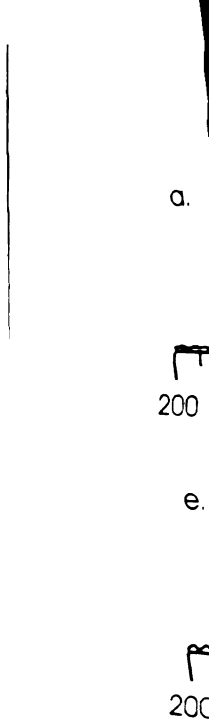


Figure 6-2
(A) S_0/S_1
119. (C)
from MG
(H) result

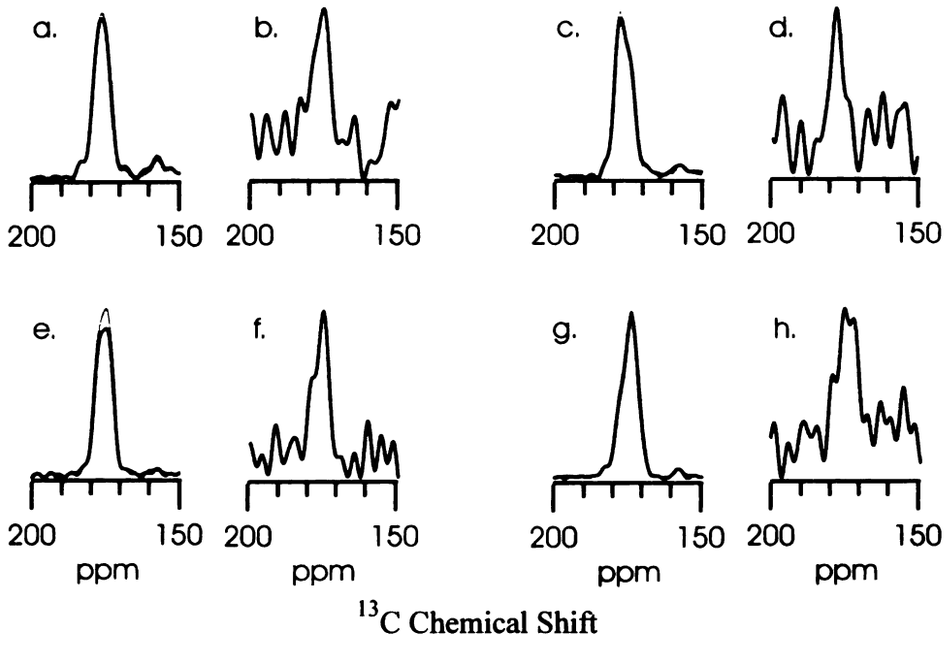


Figure 6-2. REDOR analysis of FHA2 inclusion bodies while still within bacterial cells. (A) S_0/S_1 signal from LF labeling and (B) difference signal representing Leu-2 and Leu-119. (C) S_0/S_1 signal from GM labeling and (D) resulting Gly-16 signal. (E) S_0/S_1 signal from MG labeling and (F) resulting Met-17 signal. (G) S_0/S_1 signal from VY labeling and (H) resulting Val-161 signal.

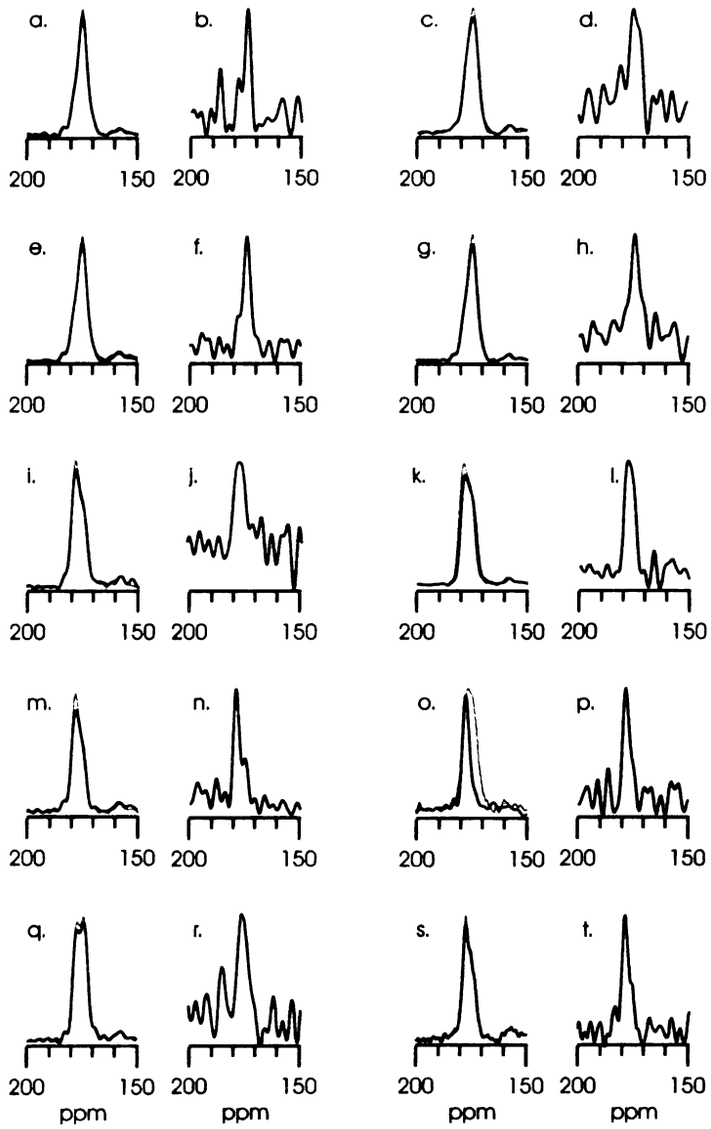


Figure 6-3. REDOR analysis of FHA2 inclusion bodies for positions studied in both the insoluble cell fraction and still within bacterial cells. Labeling of GL targeting gly-1 in the insoluble fraction (A) S_0/S_1 signal and (B) difference signal and the whole bacterial cells (C) S_0/S_1 signal and (D) difference signal. Labeling of GA targeting gly-4 in the insoluble fraction (E) S_0/S_1 signal and (F) difference signal and the whole bacterial cells (G) S_0/S_1 signal and (H) difference signal. Labeling of AG targeting ala-7 in the insoluble fraction (I) S_0/S_1 signal and (J) difference signal and the whole bacterial cells (K) S_0/S_1 signal and (L) difference signal. Labeling of LL targeting leu-98 in the insoluble fraction (M) S_0/S_1 signal and (N) difference signal and the whole bacterial cells (O) S_0/S_1 signal and (P) difference signal. Labeling of LV targeting leu-99 in the insoluble fraction (Q) S_0/S_1 signal and (R) difference signal and the whole bacterial cells (S) S_0/S_1 signal and (T) difference signal.

Co

membrane

sample ty

use of the

For these

carbonyl

will be fr

Leu-Phe

therefore

differenc

structure,

the positi

sequence

NMR str

residues

structure

ordered s

structure.

from resi

Ala-7, Gl

of backbo

1.3 ppm

Comparison of the ^{13}CO chemical shift in the inclusion body protein to that of the membrane-reconstituted protein, table 6-1, shows a general similarity between each sample type. Previous work studied the method of observing the chemical shift with the use of the REDOR filter method for positions that are not unique in the protein sequence. For these samples, the selective labeling will result in multiple positions where the carbonyl and amide labeled amino acids are adjacent, and the observed difference signal will be from all of these positions. For example, in the FHA2 sample the amino acid pair Leu-Phe exists twice in the sequence, at positions 2 and 119. Both of these positions therefore contribute to the observed $S_0 - S_1$ difference signal. If the observed carbonyl difference signal is sufficiently sharp, and lies within the known range of helical structure, then the determination can be made that both of the positions are helical. All of the positions studied in FHA2 were unique pairs (the pair only existed once in the protein sequence) with the exception of this Leu-2 and Leu-119 sample. In the existing liquid NMR structures of the detergent solubilized fusion peptide, at pH 5 the region from residues 2 to 9 are helical followed by a turn at residues 11 and 12, then a well defined structure including a short 3_{10}-helix. At pH 7.4 a change in structure is observed and the ordered structure of the 3_{10}-helix is no longer present and is replaced by disordered structure. Crystal structure information has revealed a second helical region that extends from residue 38 to residue 105. The observed peak ^{13}CO shifts for Gly-1, Leu-2, Gly-4, Ala-7, Gly-16, Leu-98, Leu-99, and Leu-119 exhibit better agreement with distributions of backbone ^{13}CO shifts in helical conformation (175.5 ± 1.2, 179.4 ± 1.3, and 178.5 ± 1.3 ppm for Gly, Ala, and Leu, respectively) than with characteristic shifts in β

sheet/amy

27). The

of Gly-16

results in

therefore

protein w

Si

at pH 5 a

observed

supports

the pH fr

in the hig

reporting

Th

helical str

the memb

types. Onl

the inclus

differences

structure.

the inclusi

179.1 ppm

chemical s

sheet/amyloid samples (170-172, 174-176, and 173-175 ppm, respectively) (*6, 13, 22, 23, 27*). These results are consistent with previous structural information with the exception of Gly-16. This residue was observed to be helical at pH 5 and extended at 7.4, and our results indicate helical structure. Our samples are of the whole bacterial cells and would therefore have a pH of near 7.4, but these results indicate the fusion peptide region of the protein while in inclusion bodies may be in the lower pH form.

Similar studies conducted by Yan Sun on the membrane associated fusion peptide at pH 5 and 7.4 also investigated the conformation of Gly-16. These results revealed an observed chemical shift of 175.2 ppm for pH 5.0 and 175.3 for pH 7.4. This result supports the model that the conformation at residue 16 does not change upon lowering the pH from 5.0 to 7.4. This result provides an alternative theory that the protein may be in the higher pH conformation, but that previous structural studies were erroneous in reporting a conformation change at this residue upon a rise in acidity.

The observed CO signal for these positions correlate well with known ranges of helical structure chemical shift, and also with the observed results of the same residues in the membrane reconstituted protein (*5*). Table 6-1 displays the results of both sample types. Only two of positions, Gly-1 and Leu-99, observed a decrease in chemical shift in the inclusion body sample when compared to the reconstituted protein, but the differences were 0.4 and 0.5 ppm and not large enough to indicate a change in secondary structure. Many of the other positions studied observed an increase in chemical shift of the inclusion body samples relative to the native protein, such as Ala-7 from 177.9 to 179.1 ppm or Gly-4 from 174.7 to 177.2 ppm. While this large of a change in the chemical shift could indicate a change in secondary structure, the observed shift is still

within the

random co

Th

compared

to a greate

inclusion

to larger o

within a r

known reg

samples m

known he

random co

within the range of helical structure and has moved further from the range of possible random coil.

The observed linewidth of the inclusion body samples are overall greater when compared to the membrane reconstituted samples discussed in chapter 3. This may be due to a greater distribution of conformation at the studied residues. The protein within the inclusion bodies may not all be adopting identical structures, and any variance could lead to larger observed linewidths. Typically the observed chemical shift for residues that lie within a random coil are broader and fall in the region between, or slightly overlap, the known regions of helical and strand chemical shifts. The results for the inclusion body samples may have broader lines, but the chemical shifts still consistently lie within the known helical range, indicating the change in linewidth is most likely not due to a random coil conformation.

Table 6-1: Ch
of the FHA2 p

Residue
Gly-1
Leu-2
Gly-4
Ala-7
Gly-16
Ala-36
Val-55
Leu-98
Leu-99
Val-10(
Leu-11(
Val-16
Gly-17

[a]The helic
distributio
are: Ala, 1
strand dist
were dete
structure H
reported C

IV. Hgp41

An
consistent
type. 5 onl
chemical sh
were studie

Table 6-1: Chemical shift information from the bacterial cells containing inclusion bodies of the FHA2 protein and previous analysis of the membrane reconstituted native protein.[a]

Residue	^{13}CO peak shift in cells (ppm)	^{13}CO peak shift in IF (ppm)	^{13}CO peak shift of reconstituted protein (ppm)	Helical shift (ppm)	β strand shift (ppm)
Gly-1	175.2	174.7	174.8	175.51	172.55
Leu-2	178.6	--	177.7	178.53	175.67
Gly-4	174.7	174.5	177.2	175.51	172.67
Ala-7	177.9	177.5	179.1	179.40	176.09
Gly-16	176.5	--	177.7	175.51	172.55
Ala-36	--	178.1	179.3	179.4	176.09
Val-55	--	174.9	--		
Leu-98	178.4	178.6	178.8	178.53	175.67
Leu-99	179.0	177.5	178.5	178.53	175.67
Val-100	--	173.5, 177.4	178.2		
Leu-119	178.6	--	177.7	178.53	175.67
Val-161	176.8	--	176.8	177.65	174.80
Gly-175	--	172.3, 175.4	176.8	175.51	172.55

[a]The helical and β strand shifts are the peaks of the respective literature chemical shift distributions for the residue. The standard deviations of the helical distributions in ppm are: Ala, 1.32; Gly, 1.23; Leu, 1.30; Ser, 1.39; Val, 1.38. The standard deviations of the β strand distributions in ppm are: Ala, 1.51; Gly, 1.58; Leu, 1.47; Val, 1.39. Distributions were determined based on a database of chemical shifts for proteins in which the structure has been solved.(*6*) The amyloid shift ranges were obtained from previously reported CO chemical shift values for known β sheet/amyloid conformation (*13, 22-26*).

IV. Hgp41 Inclusion Body Analysis

Analysis of the bacterial samples containing the Hgp41 inclusion body protein is consistent with the results from FHA2. A total of 16 positions were studied in this sample type, 5 only as whole cells (figure 6-4), and 11 as both sample types (figure 6-5). The chemical shifts observed for these samples are outlined in table 6-2. All of these residues were studied as part of a multi-target sample, with the exception of Ala-24 and Leu-89.

Despite the

samples if

secondary s

Sev

signal, suc

fraction pr

fraction p

attributed

different s

multiple p

not possil

chemical

structure

Despite the possibility of inconclusive results that can come from these multiple target samples if the observed signal is broad or if the chemical shift lies close to multiple secondary structure ranges, our results were convincing of highly helical structure.

Several of the observed chemical shifts appear to be composed of more than one signal, such as the sample targeting the Gly-13 and Gly-38 residues in the insoluble fraction protein and sample targeting the Ser-67 and Ser-70 residues also in the insoluble fraction protein, both shown in figure 6-5. The appearance of two peaks could be attributed to either a distribution of conformation, or that the two positions studied exhibit different secondary structure. These results illustrate the potential difficulty of studying multiple positions within one sample. When it appears that multiple signals present, it is not possible to determine the individual contribution from each residue. Overall, the chemical shifts still lie within the known helical range, this indicates that while the structure may vary at these positions, overall the protein is still helical at these residues.

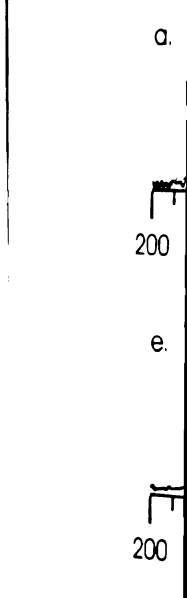


Figure 6-4. R
(A) S_0/S_1 si
S_0/S_1 signal
signal from I

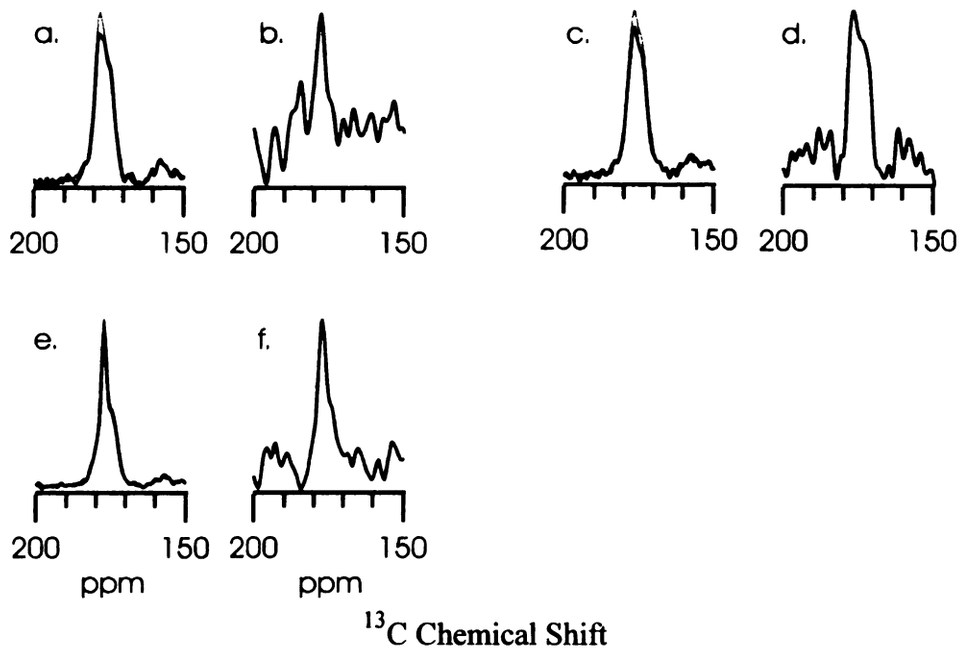


Figure 6-4. REDOR analysis of Hgp41 inclusion bodies while still within bacterial cells. (A) S_0/S_1 signal from AI labeling and (B) difference signal representing ala-24. (C) S_0/S_1 signal from GG labeling and (D) resulting gly-49 and gly-52 signal. (E) S_0/S_1 signal from LS labeling and (F) resulting leu-11 and leu-47 signal.

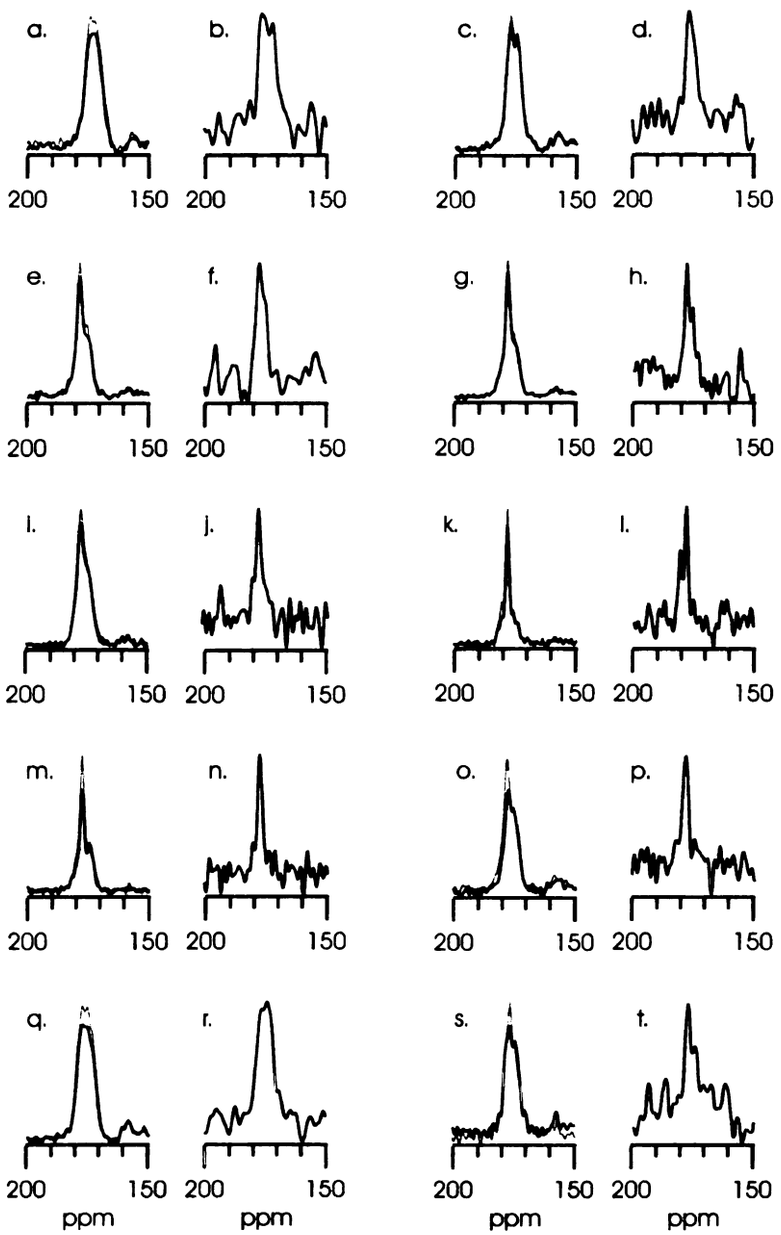


Figure 6-5. REDOR analysis of Hgp41 IB for positions studied in both the insoluble cell fraction (IF) and still within bacterial cells (BC). Labeling of GI targeting gly-13 and gly-38 in the IF (A) S_0/S_1 signal and (B) difference signal (DS) and the BC (C) S_0/S_1 signal and (D) DS. Labeling of LD targeting leu-89 in the IF (E) S_0/S_1 signal and (F) DS and the BC (G) S_0/S_1 signal and (H) DS. Labeling of LI targeting leu-67 and leu-71 in the IF (I) S_0/S_1 signal and (J) DS and the BC (K) S_0/S_1 signal and (L) DS. Labeling of LL targeting leu-10, leu-21, leu-31, and leu-86 in the IF (M) S_0/S_1 signal and (N) DS and the BC (O) S_0/S_1 signal and (P) DS. Labeling of SL targeting ser-67 and ser-70 in the IF (Q) S_0/S_1 signal and (R) DS and the BC (S) S_0/S_1 signal and (T) DS.

Table 6-2: Chemical shift information from the bacterial cells containing inclusion bodies of the Hgp41 protein.[a]

Residue	^{13}CO peak shift in Cells (ppm)	^{13}CO peak shift in IF (ppm)	Helical shift (ppm)	β strand shift (ppm)
Leu-10	178.3	178.5	178.53	175.65
Leu-11	178.0	--	178.53	175.65
Gly-13	176.9	177.7/173.2	175.51	172.55
Leu-21	178.3	178.5	178.53	175.65
Ala-24	178.2	--	179.40	176.09
Leu-31	178.3	178.5	178.53	175.65
Gly-38	176.9	177.7/173.2	175.51	172.55
Leu- 47	178.0	--	178.53	175.65
Gly-49	176.8	--	175.51	172.55
Gly-52	176.8	--	175.51	172.55
Ser-66	177.0	175.2/177.5	175.94	173.66
Leu-67	178.4, 180.9	178.5	178.53	175.65
Ser-70	177.0	175.2/177.5	175.94	173.66
Leu-71	178.4, 180.9	178.5	178.53	175.65
Leu-86	178.3	178.5	178.53	175.65
Leu-89	178.4	178.1	178.53	175.65

[a]The helical and β strand shifts are the peaks of the respective literature chemical shift distributions for the residue. The standard deviations of the helical distributions in ppm are: Ala, 1.32; Gly, 1.23; Leu, 1.30; Ser, 1.39. The standard deviations of the β strand distributions in ppm are: Ala, 1.51; Gly, 1.58; Leu, 1.47; 1.50. Distributions were determined based on a database of chemical shifts for proteins in which the structure has been solved (*6*). The amyloid shift ranges were obtained from previously reported CO chemical shift values for known β sheet/amyloid conformation (*13, 22-26*).

To date, there have been no site-specific structural studies of the membrane associated gp41 protein to which we can compare these results, but the observed ^{13}CO chemical shifts are consistent with the highly helical nature in crystal structures of similar constructs. Based on these similar structures, it would be expected that all residues except the 48 to 53 loop region would be a helical structure, connected by a turn. Residues Leu-

10, Leu-11, Gly-13, Leu-21, Ala-24, Leu-31, Gly-38, Leu-47, Leu-67, Ser-70 Leu-71, Leu-86, and Leu-89 and all exhibit carbonyl chemical shifts that are consistent with helical structure and the relatively sharp linewidths (typically ~2 ppm of less) indicate highly ordered structure at these positions. Compared to other sample types, such as microcrystalline samples, this linewidth may seem broad. When compared with other similar samples, such as amyloid fibrils where a linewidth of 2.5 ppm is considered to indicate well-ordered structure, these results are relatively sharp (*12*). The two residues within the loop region, Gly-49 and Gly-52, were studied in a single sample, as both are part of a Gly-Gly pair in the primary sequence. The chemical shift of this sample, 176.8 ppm, was above the known average of helical glycine residues, but the linewidth is relatively broad (~6 ppm). This broadness is most likely due to at least some disorder or distribution of conformation at one or both of these positions. Since this is a study of a multiple target sample, it is not possible to determine the chemical shift and line broadness contribution of each individual residue, but the broadness of the signal could be due to the two positions exhibiting different chemical shifts that when added together results in broad peak. While no direct conclusions can be drawn from this result, there is no evidence to indicate that these positions are not in the coil conformation as observed in crystal structures.

V. Conclusions and Future Work

A total of 13 positions throughout FHA2 and 16 positions of Hgp41 were analyzed while still within the bacterial cell, in the insoluble fraction, or both, and results in this study reveal that at least a portion of the secondary structure in viral fusion proteins within bacterial inclusion bodies retains native, helical structure. These studies were made possible due to the high fraction of the both the expressed FHA2 and Hgp41 protein that exists within inclusion bodies, as opposed to the native, soluble form. Electron microscopy results visually illustrate the large fraction of the cell that is composed of the protein aggregates. Gel electrophoresis results, described in chapter 5, show the difficulties encountered with solubilizing the expressed protein (*2, 7*). The high fraction of protein that remains in the insoluble portion of the cell after lysis by sonication is also shown in these gels, as the amount of FHA2 or Hgp41 protein in both the whole cells and insoluble fractions of the cells are equivalent, relative to the background from other cellular proteins. Typical yields of native FHA2 protein are up to 20 mg per liter of fermentation, typically ~10 g of cells, indicating that ~0.2% percent of the cellular mass is native protein. Gel electrophoresis of the bacterial contents indicate that up to 10% of the cell is composed of the expressed protein, or that greater than 95% of the protein is in the form of insoluble inclusion bodies.

The use of the REDOR filter method to observe the chemical shift of more than just unique sequential pairs became highly useful with the study of the Hgp41 protein. This protein had few unique positions containing amino acids that were commercially available for less than $200 per gram, making the study of solely unique REDOR positions financially unfeasible. By studying multiple positions in one sample, such as the

sample labeled with ^{13}CO-Leu and ^{15}N-Ser that targeted the two LS pairs in the sequence at residues 11 and 47, it was possible to draw conclusions about many positions in a small number of samples. This particular sample resulted in a sharp ^{13}CO signal at 178.0 ppm, a helical chemical shift. It can therefore be concluded that both Leu-11 and Leu-47 are part of a helix. This approach makes it possible to study positions that are not unique. As discussed earlier, if the observed signal from multiple positions lies definitively within the known range of helical chemical shifts, and is sufficiently sharp (~2 ppm or less), the conclusion can be drawn that the signal from each individual position must also be sharp and within the expected range, and each position is therefore helical. If the observed signal is relatively broad or lies near the range of multiple secondary structures, it becomes impossible to deconvolute the signal into the individual contributions from each position and structural determination cannot be made. This method makes it possible to study multiple residues in one sample, saving both time and financial resources.

Overall, the results presented in this study support a model of native, helical conformation being retained for a significant fraction of the protein within bacterial inclusion bodies for both the FHA2 and Hgp41 protein. Interestingly, there was no indication of β-sheet structure in either protein, which is not consistent with the amyloid model of inclusion body structure. Particularly in the case of the FHA2 fusion peptide, as this hydrophobic region would be a good candidate for the formation of amyloid aggregates when not associated with detergent or a membrane.

For both proteins, a theory exists that the formation of the secondary structure is driven by the hydrophobic effect and trimerization. If in actually the protein structure due

to the hydrophobic effect, this would indicate the protein found within inclusion bodies is actually exists within the aqueous environment of the cell where folding occurs, and is then incorporated into the inclusion bodies. If the produced protein does not fold due to a lack of space in the cell membrane and aggregates due to the exposed hydrophobic regions, as previous thought, this driving force for the formation of helical structure would not exist.

While ^{13}CO chemical shifts are not absolutely definitive in determining conformation and are correlated with regions of the Ramachandran plot rather than precise dihedral angles, this information does provide significant conformational information. This is particularly the case when helical shifts at several nearby residues are obtained (as in the fusion peptide region of FHA2 or all of the positions studied in Hgp41) or when sharp signals (indicating structural order) with helical shifts are obtained in the inclusion body protein (*6*).

Previous studies of the FHA2 protein also included SS-NMR structural analysis of the insoluble fraction of bacterial cells, in which the soluble portion of the cell was separated by centrifugation after cell lysis by sonication (*2*). The intent was to remove any soluble expressed protein that would contribute to the observed difference signal. The result of the CO chemical shift analysis for four positions showed good correlation between the whole cell, insoluble fraction, and membrane reconstituted samples. For example, the position Leu-98 observed carbonyl shifts for these samples were 178.4, 178.9, and 178.8 ppm, respectively. Initially, analysis of both the whole cell and insoluble fractions for several Hgp41 residues also yielded the same results. For the labeling scheme which included the leucines at positions 10, 21, 31, and 86 the observed

carbonyl shifts were 178.3 and 178.5 ppm for the whole cell and insoluble fractions, respectively. This is an expected observation, if the protein in inclusion bodies is fully folded, then the observed chemical shift for both insoluble fraction and whole cell samples should be very similar. For the majority of the structural analysis, only the whole bacterial cell samples were analyzed due to the strong agreement in chemical shifts observed in these initial samples, and the observation in both gel electrophoresis and electron microscopy that a large portion of the expressed proteins in the cell were in the insoluble form. While there will still be trace amounts of the soluble protein present in the whole cell samples, the majority of the S_0-S_1 observed difference signal will be due to the insoluble protein.

A large number of specifically labeled samples are required to use this method for a high-resolution NMR structural model of inclusion body protein, but given the low volume of fermentation required to produce each sample, it is possible to prepare many samples in parallel. Even with the feasibility of this method, it would still be simpler to study the protein in a manner which requires the preparation of fewer samples. Depending on linewidths, it may be possible to obtain structural data for samples with more significant labeling using methods previously applied to NMR structure determination of microcrystalline, amyloid, and membrane-associated proteins (*12, 13, 23, 24, 27*). Two dimensional SS-NMR experiments would allow for the study of proximity of individual amino acid residues to further explore secondary, and possibly even tertiary structure. This type of study would most likely be limited by the linewidth of the observed signal, as broad, overlapping signals can make it difficult to assign individual residues.

A structural aspect that is impossible to evaluate with the methods presented in this paper is the tertiary structure of the protein. Comparison between the natively folded membrane reconstituted protein and the inclusion body protein for FHA2 is consistent with the protein maintaining secondary structure while in the inclusion bodies. Since tertiary structure would be affected by changes in the secondary structure of the protein, a plausible theory that the proteins may also maintain their tertiary structure within the inclusion bodies as well, but to date we have not developed experimental methods to test this hypothesis.

Along with tertiary structure, the question of quaternary structure within inclusion bodies still exists, especially given that the results presented in this paper do not support amyloid formation. The results also do not support a completely unfolded model of the protein, in which it would be reasonable to hypothesize that the proteins randomly aggregate with no overall structure. The question still exists, if each individual protein is folded, what are the inter-protein interactions of the inclusion bodies and what type of order exists within the inclusion bodies. Natively, both the HA2 and gp41 fusion proteins exist as trimers and constructs of the ectodomain of these proteins have been observed in this formation as well. It is possible that this formation still exists in the inclusion bodies, and that they are overall composed of an aggregate of many trimer units. Most likely these aggregates would still have some sort of higher order structure, and one potential form could be protein micelles. If this model were true for FHA2, the hydrophobic fusion peptide region would likely aggregate to the center of the micelle, leaving the remainder of the hydrophilic ectodomain exposed to the intra-cellular environment. An observation not accounted for by this model is that the aggregates are large enough to be visualized

by electron microscopy, indicating that this model may help in beginning to understand inclusion body structure but does not tell the full story. Considering tertiary and quaternary structure clearly shows that studying higher order structure is the next logical step in the overall investigation of inclusion body formation and structure.

This paper presents an expanded study of secondary structure in bacterial inclusion body protein. This method is able to provide structural information about specific residues. This chapter presents the study of 29 total residues in two proteins, as opposed to IR, which can only provide information about the overall fraction of different types of structure. Another benefit is sample preparation. The procedures used in this study required little or no purification of the inclusion bodies, the study was simply conducted on the bacterial cell pellet following protein expression. Any purification method or dehydration of sample could potentially alter the secondary structure. Purification methods used to obtain inclusion bodies include washing the insoluble fraction of the cell with strongly denaturing detergents such as triton in order to solubilize the protein. A concern regarding this type of sample preparation is that these strong detergents are altering the secondary structure of the inclusion bodies (*8, 10, 28*). A common method of analysis is IR spectroscopy, which is typically conducted on samples that are dried, a process that could effect protein structure (*8, 10, 11*).

This method of studying whole cells involves no purification and allows for the study of fully hydrated samples. Inclusion body protein while still within intact bacterial cells is the most biologically relevant form. In summary, this study describes the expanded application of a general approach for residue-specific structural analysis of recombinant protein in inclusion bodies including those in whole E. coli cells as well as

evidence for native conformation at specific residues distributed throughout both the FHA2 and Hgp41 proteins, studying 5 and 17% of the total residues in each protein, respectively.

Further studies in the pursuit of inclusion body structural information will obviously include applying these methods to additional inclusion body forming proteins and to study more positions within each protein, but will likely take a new direction as well. The future focus will inevitably be on the interaction between protein units. This method of studying the secondary structure of individual positions has proven successful, but still leaves the question of higher order structure to be investigated. Previous studies have proposed amyloid formation within inclusion bodies, and accounts for the structure in the protein aggregates. Our theory is that the protein maintains native conformation, but if that is the case, the protein aggregation most likely is ordered to some extent. The hydrophobic regions of the membrane protein would not favor exposure to the cellular cytoplasm and would likely drive the formation of structured aggregates.

The most reasonable model given the current understanding of inclusion bodies is a protein micelle structure. The proteins could be arranging in a form similar that of lipids when exposed to aqueous solution. The hydrophobic fusion peptide region is mimicking the tail of the lipid and will occupy the center of the micelle, while the ectodomain region is forming the shell, similar to the hydrophilic lipid head groups. While this is a reasonable model and there is no evidence to dispute the formation of protein micelles, there is currently no experimental evidence for their existence.

Designing experiments to probe inter-protein interactions is much more difficult than secondary structure analysis. Simple evaluation of chemical shift does not give

useful information in this regard, and any experiments conducted on whole bacterial cells severely limits the possibilities. If the inclusion bodies are purified, as in many of the previous structural studies, hydrogen/deuterium exchange could be useful in determining which regions of the protein are exposed to the surrounding solvent. Residues which exchange rapidly likely lie on the outer edges of the inclusion body, while those that react more slowly are likely buried within the inclusion body. The main difficulty with this method is the sheer size of the inclusion bodies. Bacterial cells typically express only one or two inclusion bodies, each composed of many protein units. The extent that an inclusion body is porous is still unknown. The aggregates are clearly densely packed and water may not easily flow through the inclusion bodies In this case, only a small fraction of the proteins would lie near the shell of the inclusion body and could potentially participate in H/D exchange. If the inclusion bodies are somewhat porous, this would still introduce difficulty in hydrogen/deuterium exchange as the protein units on the interior of the aggregates would not be exposed to the new solvent as quickly as those on the surface and this would affect the observed exchange rate.

An approach to studying the interaction between synthetic peptides is to selectively label specific residues and observe the efficiency of magnetization transfer between peptide units. This has been successful at giving information on peptide structure, but becomes much more difficult in recombinant proteins. The most specific labeling scheme available is to isotopically label all residues of a particular amino acid. For example, the FHA2 protein has 13 glycines, and with no current methods to label an individual residue all glycines would possess the isotopic label and participate in the interaction. This would make it difficult to determine which regions of the protein are

near in the inclusion bodies, if an interaction is observed it could be from a number of different positions.

It is clear that there is no easy answer to this question, no simple method to determine the inter-protein interactions within inclusion bodies, especially while still within bacterial cells. The study of whole cells is more complicated due to the background signal from the remainder of the cell. Methods could potentially be developed that would filter this background signal or enhance the signal from the inclusion bodies to a level where structural analysis is feasible, but this situation is unquestionably more difficult than the study of purified proteins. For this reason, initial studies will likely be conducted on protein that has been removed from the cell and purified to some extent, hopefully working backwards to eventually develop methods applicable to study within the whole bacterial cell.

VI. References

(1) Baneyx, F., and Mujacic, M. (2004) Recombinant protein folding and misfolding in Escherichia coli. *Nature Biotechnology 22*, 1399-1408.

(2) Curtis-Fisk, J., Spencer, R. M., and Weliky, D. P. (2008) Native conformation at specific residues in recombinant inclusion body protein in whole cells determined with solid-state NMR spectroscopy. *Journal of the American Chemical Society 130*, 12568-12569.

(3) Chen, J., Skehel, J. J., and Wiley, D. C. (1999) N- and C-terminal residues combine in the fusion-pH influenza hemagglutinin HA(2) subunit to form an N cap that terminates the triple-stranded coiled coil. *Proc Natl Acad Sci U S A 96*, 8967-72.

(4) 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 Biology 8*, 715-720.

(5) 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. *Journal of the American Chemical Society 129*, 11320.

(6) Zhang, H. Y., Neal, S., and Wishart, D. S. (2003) RefDB: A database of uniformly referenced protein chemical shifts. *J. Biomol. NMR 25*, 173-195.

(7) 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.

(8) Umetsu, M., Tsumoto, K., Ashish, K., Nitta, S., Tanaka, Y., Adschiri, T., and Kumagai, I. (2004) Structural characteristics and refolding of in vivo aggregated hyperthermophilic archaeon proteins. *Febs Letters 557*, 49-56.

(9) Przybycien, T. M., Dunn, J. P., Valax, P., and Georgiou, G. (1994) Secondary Structure Characterization of Beta-Lactamase Inclusion-Bodies. *Protein Engineering 7*, 131-136.

(10) Oberg, K., Chrunyk, B. A., Wetzel, R., and Fink, A. L. (1994) Native-Like Secondary Structure in Interleukin-1-Beta Inclusion-Bodies by Attenuated Total Reflectance Ftir. *Biochemistry 33*, 2628-2634.

(11) Ami, D., Natalello, A., Gatti-Lafranconi, P., Lotti, M., and Doglia, S. M. (2005) Kinetics of inclusion body formation studied in intact cells by FT-IR spectroscopy. *Febs Letters 579*, 3433-3436.

(12) Petkova, A. T., Ishii, Y., Balbach, J. J., Antzutkin, O. N., Leapman, R. D., Delaglio, F., and Tycko, R. (2002) A structural model for Alzheimer's beta-amyloid fibrils based on experimental constraints from solid state NMR. *Proceedings of the National Academy of Sciences of the United States of America 99*, 16742-16747.

(13) Ritter, C., Maddelein, M. L., Siemer, A. B., Luhrs, T., Ernst, M., Meier, B. H., Saupe, S. J., and Riek, R. (2005) Correlation of structural elements and infectivity of the HET-s prion. *Nature 435*, 844-848.

(14) Carrio, M., Gonzalez-Montalban, N., Vera, A., Villaverde, A., and Ventura, S. (2005) Amyloid-like properties of bacterial inclusion bodies. *Journal of Molecular Biology 347*, 1025-1037.

(15) Garcia-Fruitos, E., Aris, A., and Villaverde, A. (2007) Localization of functional polypeptides in bacterial inclusion bodies. *Applied and Environmental Microbiology 73*, 289-294.

(16) Vera, A., Gonzalez-Montalban, N., Aris, A., and Villaverde, A. (2007) The conformational quality of insoluble recombinant proteins is enhanced at low growth temperatures. *Biotechnology and Bioengineering 96*, 1101-1106.

(17) Wang, L., Maji, S., Sawaya, M., Eisenberg, D., and Riek, R. (2008) Bacterial inclusion bodies contain amyloid-like structure. *PLoS Biology 6*, 1791-1801.

(18) Gullion, T., and Schaefer, J. (1989) Rotational-echo double-resonance NMR. *J. Magn. Reson. 81*, 196-200.

(19) Yang, J., Parkanzky, P. D., Bodner, M. L., Duskin, C. G., and Weliky, D. P. (2002) Application of REDOR subtraction for filtered MAS observation of labeled backbone carbons of membrane-bound fusion peptides. *J. Magn. Reson. 159*, 101-110.

(20) Morcombe, C. R., and Zilm, K. W. (2003) Chemical shift referencing in MAS solid state NMR. *J. Magn. Reson. 162*, 479-486.

(21) Stringer, J. A., Bronnimann, C. E., Mullen, C. G., Zhou, D. H. H., Stellfox, S. A., Li, Y., Williams, E. H., and Rienstra, C. M. (2005) Reduction of RF-induced sample heating with a scroll coil resonator structure for solid-state NMR probes. *Journal of Magnetic Resonance 173*, 40-48.

(22) Petkova, A. T., Ishii, Y., Balbach, J. J., Antzutkin, O. N., Leapman, R. D., Delaglio, F., and Tycko, R. (2002) A structural model for Alzheimer's beta-

amyloid fibrils based on experimental constraints from solid state NMR. *Proc. Natl. Acad. Sci. U.S.A. 99*, 16742-16747.

(23) Siemer, A. B., Ritter, C., Ernst, M., Riek, R., and Meier, B. H. (2005) High-resolution solid-state NMR spectroscopy of the prion protein HET-s in its amyloid conformation. *Angewandte Chemie-International Edition 44*, 2441-2444.

(24) Castellani, F., van Rossum, B., Diehl, A., Schubert, M., Rehbein, K., and Oschkinat, H. (2002) Structure of a protein determined by solid-state magic-angle- spinning NMR spectroscopy. *Nature 420*, 98-102.

(25) Helmus, J. J., Surewicz, K., Nadaud, P. S., Surewicz, W. K., and Jaroniec, C. P. (2008) Molecular conformation and dynamics of the Y145Stop variant of human prion protein. *Proceedings of the National Academy of Sciences of the United States of America 105*, 6284-6289.

(26) Heise, H., Hoyer, W., Becker, S., Andronesi, O. C., Riedel, D., and Baldus, M. (2005) Molecular-level secondary structure, polymorphism, and dynamics of full-length alpha-synuclein fibrils studied by solid-state NMR. *Proceedings of the National Academy of Sciences of the United States of America 102*, 15871-15876.

(27) Qiang, W., Bodner, M. L., and Weliky, D. P. (2008) Solid-state NMR Spectroscopy of human immunodeficiency virus fusion peptides associated with host-cell-like membranes: 2D correlation spectra and distance measurements support a fully extended conformation and models for specific antiparallel strand registries. *Journal of the American Chemical Society 130*, 5459-5471.

(28) Carrio, M. M., Corchero, J. L., and Villaverde, A. (1998) Dynamics of in vivo protein aggregation: building inclusion bodies in recombinant bacteria. *Fems Microbiology Letters 169*, 9-15.

Chapter 7:

Implementing an Online Homework Program in a Large Organic Chemistry Lecture Course

I. Introduction

Higher education is continuously changing in the areas of pedagogical methods and student types. While it is becoming clear that the traditional lecture approach to teaching classes can be much improved upon, increased class size and limited funding for increases in teaching staff make changes difficult. Large lecture courses cannot simply be split into many smaller sections, so this it up to the creativity of instructors to develop improved teaching methods in large classrooms.

Two areas that have been heavily researched and developed in recent years include homework and the internet components of courses. The enrollment of many courses are simply too large to assign hand graded assignments on a regular basis, leaving educators to determine the effectiveness of homework in general, and alternative methods of administration. The drastic increase in the prevalence of the internet in every other aspect of our lives has also crept into the classrooms. Students are tech-savy and feel comfortable, some ever prefer, using the internet as a component of college courses. Fortunately, much research has been devoted to joining these areas through the development of internet based homework systems. In order to adequately discuss the emergence and proper implementation of online homework (OHW), consideration of each topic individually and how they relate to chemistry education.

A. Chemistry Homework

Few educators would disagree with the idea that homework is more effective when it is graded by an instructor (the professor or a teaching assistant), feedback is given in regards to why a question is graded as incorrect, and returned to the students in a timely fashion. Unfortunately, most would also agree that this is not feasible in large lecture courses. Teaching assistants are commonly responsible for several recitation sections, and with commitments to research and their own coursework cannot devote the time required for this sort of homework system.

The question then becomes, where can changes be made to make homework more reasonable? Does optional, or un-graded, homework serve the same purpose? How necessary is feedback? Does a delayed return of homework affect how much a student can learn from completing the assignment? Many of these questions are dependent on the student and their motivation. A highly dedicated student would complete homework assignments even if they were not required, would take it upon themselves to determine their mistakes in answering questions incorrectly, and would re-evaluate homework returned to them even when returned several weeks after completion (*1*). The focus then needs to be how to target the remainder of the class.

B. Online Component to Chemistry Courses

There is no question; higher education is in the process of shift towards a more web-based environment. This comes along with the potential for many great advances in our ability to educate students, especially as resources such as classroom space and instructor time become more limited. The use of the internet allows students to attend

class without leaving home, to complete homework assignments that don't require a teaching assistant to grade, and to participate more in cooperative learning.

Some students prefer the flexibility of taking a course online, and many learn better through this method than the traditional lecture. Students generally appreciate the extra flexibility of the online portions of the course. A study of a traditional course taught in parallel with a hybrid course analyzed the effectiveness of the teaching methods. Surprisingly, the students reported similar levels of quality in regards to interaction with the instructor and fellow classmates, even though it may be expected to be lower for a hybrid class in which the students do not attend class as regularly. In regards to student learning in these environments, post-tests revealed that while the hybrid student performed better for all students, the upperclassman saw a greater discrepancy between course styles. This is an important observation; students need to be highly motivated to do well in an online course setting. The flexibility allowed in online courses also leads to increased opportunities to fall behind on the material. Upperclassman may have more developed study and motivation habits, while freshman may still lack the self-motivation required to make full use of online courses (*2*).

A new challenge facing college professors is how to make large lecture courses feel small. It is becoming clear that not all students are of the learning style that benefit from lectures, but alternatives can be hard to implement. Cooperative learning is gaining in popularity, but the logistics of group work while in class is still an issue. Many classrooms are not designed for this format, and class time is so precious that devoting time to group work may result in sacrificing the amount of material to be covered. Students' busy schedules can make it difficult to require outside of class meetings, but

this is where the internet can become useful. Students can work on group projects and even evaluate their own and their group members performance online (*3*). The third major impact of the internet on organic chemistry courses is in the area of homework. The remainder of this chapter will focus on OHW in organic chemistry.

C. Online Homework

A de facto opinion regarding college education suggests that students learn and retain material more effectively when they reinforce coursework situations in the context of homework or group projects. For practical class management, many general chemistry and organic chemistry courses at large universities are primarily structured as lecture-only classes, due to the limited resources and low teaching assistants/student ratios that make assigning traditional homework unfeasible. Since electronic media-based technology is becoming more pervasive, it is likely that many students presently attending college have experienced the flexibility and immediate feedback of digital processing. Therefore, an emerging solution to managing homework in large university classes is implementing online, computer-based homework systems that tap into the ever-growing tech-savvy student population. Many new programs have been introduced in the last decade, including, for example, LON-CAPA (Learning Online Network with a Computer-Assisted Personalized Approach), WE_LEARN (Web-based Enhanced Learning Evaluation And Resource Network) and EPOCH (Electronic Program for Organic Chemistry Homework) (*4-6*). Each program has its unique components, but they generally offer similar key features, such as randomizing questions, recording student performance, and instantly reporting the correctness of the answer. It is expected that

these program aspects will enhance the students' learning skills while providing the freedom for them to multi-task and log onto the system around their work schedules. It is clear that the popularity of this sort of learning program will likely continue to expand to complement several different types of classes. Thus, a new challenge emerges to address how best to apply the programs into the curriculum so that they effectively aid student learning. Several aspects, such as the number of attempts allowed, feedback, and randomization of questions differ between programs and implementation of OHW.

Any new technique is sure to be the target of criticism, and OHW will have to prove itself as a potentially useful tool in large chemistry classrooms. Most of the situations in which OHW would be used are those for which it is not possible to assign traditional written homework, that the online version is the only option. But the question will still exists, if written homework could also be administered in these classes, would it be more effective at teaching students the material? A very detailed study conducted at North Carolina State University provided evidence that this would not be the case. In a study comparing online vs. written homework in large physics classes with both calculus and algebra based sections, no significant difference was observed in student performance between the two types of homework. While there have been few studies of this type, the initial results indicate that when the course is under certain constraints (size, teaching staff), OHW can be just as effective as written homework (*7*).

Multiple Attempts. Most OHW programs are set up to allow students unlimited, or a high number, or attempts to answer a question and still receive credit. The general approach is that "students must attack the question repeatedly until they answer it correctly" (4). While in theory this may be good, it should force students to eventually

learn the concepts, in practice this promotes guessing. If there is no punishment for guessing at an answer, a student who is simply concerned with earning homework points will surely take this approach.

This is a key difference from traditional homework. Even in courses where instructors allow students to resubmit homework, this is typically still on a limited basis. The approach for each type of homework assignment therefore becomes very different. On the positive side, students may learn the material through the process of submitting an incorrect answer repeatedly until learning the concept. Students may be less likely to become discouraged and frustrated by their inability to answer the question correctly on the first attempt.

On the negative side, the motivation to study and learn the material prior to attempting homework questions is no longer present. There were consequences for not putting effort into the first response, but with unlimited responses this, and now the motivation, is no longer there. Another aspect of traditional homework lost in the multiple attempt system is desire to "double check" answers. When the students are concerned about the correctness of their answer, and potentially losing points, they are more likely to approach the solution from many directions in order to confirm their answer. In essence, the homework no longer becomes a "pre-test," where the students can determine how well they really understand the material prior to an in-class examination.

An interesting approach used at Louisiana State University is to simply offer the students a second chance at homework questions missed on the first attempt. All questions were multiple choice, and feedback was given regarding a common error that may have been made to cause them to select that answer. They are then allowed up to 48

hours to take a second chance at a question that is similar in nature, but not exactly the same. Overall, the program received positive feedback from the students and over 70% of the students made use of this option (*1*).

This unique approach to the OHW changes the attitude that the students have towards completing the assignments correctly. If the decision is whether to give the students 5, 10 or even unlimited attempts, then the message being sent to the students is that they are expected to use multiple attempts and careful thought prior to the initial submission is not necessary. The "second chance" approach tells the students that it is important that they take the homework seriously on the first attempt, but still allows them the opportunity to rework problems and learn from their mistakes.

Randomization. With any OHW program, there is the potential for classmates to work together and to share answers. This is especially probable if the students can return to the homework set at a later time and attempt the same exact questions. If the students approached this as an opportunity to study the concepts, then this collaborative effort would be ideal. Another possibility is that students will use this opportunity to compare answers to those of classmates simply to earn the homework points. There are a variety of approaches to lessen this possibility, including randomizing the order of questions or by varying specific parts of a question (*5*).

Many homework programs, including CAPA and OWL, allow for randomization of certain components of a question. A question with a numerical component may be available in several versions, so that the calculation completed in each case would be the same, but with different values. A conceptual question (such as true/false or multiple choice) may probe the same concept in many ways by simply change the wording of the

question. A question involving chemical reactions could test the knowledge of that specific reaction by providing several different starting materials.

This randomization of questions limits the sharing of answers amongst students, at least without the students also having to consider the process at how the answer is arrived. While it is still possible to share simple "plug and chug" methods of solving problems, it is a step in the right direction by limiting direct copying.

Hints/Feedback. One of the biggest benefits of OHW, as opposed to written homework, it is the possibility for immediate feedback in regards to the correctness of the answer and in some cases, hints as to how to solve an incorrectly answered problem. With traditional homework, many students will forget the method that they used to arrive at an answer or the thought process used to solve the problem by the time it is returned to them, if the student even makes any attempt at all to review their work. Regardless of the number of attempts allowed, immediate information regarding the correctness of an answer is the most beneficial for the students learning (8).

Some programs offer preset feedback after incorrect answers are entered. This type of feedback mostly takes the form of "hints," and typically is independent of the answer entered by the student. Authors of the questions may try to predict common mistakes made by the students and offer advice through hints, or pieces of the answer could be given in order to get the students started on the problem. This can be a benefit in giving the students a place to start on a problem for which they would otherwise have no ideas, and the feedback may happen to actually happen to address the specific misconception of the student. The downside is that students may become reliant on these

sorts of hints, and if they are not designed well, may actually give away the answer to the student before they are forced to really think through the problem.

Other programs, such as EPOCH, look for common mistakes in the answer and tailor the feedback to the specific concept with which the student is struggling. The program is able to analyze the student's answer not only for simply the presence of the correct answer, but also for potential mistakes. For example, if a student submits an answer for a substitution reaction that still has a bromine in the product structure, the program would recognize this and prove the feedback, "Br is a leaving group; it should not be present in your product." This type of feedback could be extremely useful for students struggling with the material, and may make it easier for the students to actually learn the core concepts and not simply pick up on patterns. In this example, if the student was just told that they were incorrect and shown the expected answer, in the future they may simply remember that products of substitution reactions should not have a bromine. They may not know the reason why, or to be able to extrapolate this is other related situations (for example, chlorine is also a leaving group, so should not be in a substitution product.) This type of feedback is an improvement over preset feedback, but still takes away the need for serious consideration of the problem prior to submitting answers (*4*).

Tutorials. Tutorials, either built into the program or available through another web resource, are also good options for students struggling with specific concepts to receive immediate assistance (9). A study on the use of a website with tutorials, "Visualization and Problem Solving for General Chemistry," assessed the students comfort level with learning through online tutorials and their level of understanding after use (10).

Surveys at the end of this class revealed that the students overall felt at minimum neutral in regards to the effectiveness, with most feeling that they especially learned from the tutorials involving visualization, which is particularly difficult to learn from class notes or a textbook. While the entire class did not participate in the tutorials, most that did not take advantage of the opportunity did so because they already had an understanding of the material, or did not know the option existed, and overall agreed that they *may* have been able to learn from the tutorials (*10*).

Collaborative Learning. Another facet of the online systems incorporates aspects of collaborative learning, where the students work together in small groups, and is becoming more prevalent in the classroom. This interactive model of learning has influenced the design of features used for web-based instruction (3, 11). Several systems also employ chat rooms for students to discuss difficulties with questions, or provide opportunities for students to contact instructors with questions about a specific problem (12). When the students can communicate in an open forum about the homework problems, this allows the instructor insight into how well the students are learning the material and they may be able to make changes as necessary to the course.

While designed with good intentions, the collaborative learning aspects of programs also have the potential to be abused by students. Even with randomized questions, students may still communicate general steps to complete a problem, without necessarily discussing the concepts or thought process of the problems. The use of student posts in a CAPA program was analyzed to determine the actual student use of this feature. The questions were divided based on type (essay, qualitative, etc.) and the difficulty level of each class was determined by determining the ratio of correct responses

to total attempts. The discussion related to each question was then classified as emotional, surface, procedural, or conceptual, and also by unrelated, solution-oriented, mathematical, or physics. This type of classification aims at diving communication as desirable or undesirable problem solving.

The study found that overall 65% of the students had posted at least once, with an average of 5 posts per student, indicating that a large proportion of the class was using this feature. While no correlation was observed between the number of posts and overall performance in the class, there was a correlation between the type of posts. Higher scoring students had a much lower proportion of solution based posts and a higher level of conceptual or physics related comments, while the opposite is true for lower scoring students (*12*).

A thorough search of the current literature yielded no studies on the effective of student learning with or without this type of chat room feature. It would be interesting, and useful for the development of OHW programs, to have more insight into how students are actually using and perceiving this feature. Are most students making good use of the tool by using feedback from classmates to learn the material, or is there a significant fraction of the class that simply posts procedural answers to questions, or makes use of these posts to quickly answer homework problems with little thought.

Chemical Drawing. A special feature added to programs used for teaching organic chemistry, such as OWL (Online Web-based Learning), is the chemical structure drawing module. A common difficulty that students new to organic chemistry encounter is the system of drawing chemical structures. Most general chemistry classes use structures with the atoms still present, so switching to the "stick form" is often a difficult process. A

student m
ability to
allows stu
possess th
practice
examinati

So
structures
combinati
that of th
weight, n

Cl
incorpora
meeting.
draw che
learning.
homewor
draw the
which ca
creative s
still worki

Ins
instructor

student may know the concepts meant to be tested in a particular question, but lack the ability to properly communicate the product structure. Hand graded traditional homework allows students the opportunity to practice this skill, and now many OHW programs possess this feature in a variety of ways. These kinesthetic exercises enable the student to practice drawing chemical structures, which is an inevitable skill needed on examinations.

Some programs take advantage of well-established software for drawing chemical structures, such as MarvinSketch. The EPOCH homework program uses this program in combination with Jchem to compare the properties of the student submitted structure to that of the expected answer to determine correctness, by analyzing molecular formula, weight, number of multiple bonds, etc (*4*).

Chemistry professors at Lee University presented an interesting approach to incorporating chemical drawing in their homework assignments at the spring 2008 ACS meeting. The homework program used by the university is not designed to have students draw chemical structures, but they realized how important this could be for student learning. Instead of asking students to draw the answer to questions directly into the homework program, the students used ChemSketch, a free structure drawing program, to draw the solutions to questions. This program will then convert the structure into a name, which can be copied and pasted into the homework program for grading. This is a creative solution, which allows students the practice in chemical structure drawing while still working in the limits of programs available to the university.

Instructor Information. In order for the homework programs to be successful, the instructor of the course needs to be able to continuously evaluate the status of the class.

Beyond

instruct

This co

time sp

This is

through

individu

was usi

answers

while o

dealing

correct

when th

stuck on

advice.

usefulne

the class

simply a

feedback

"love it"

to expect

Beyond simply the score each student is earning, many programs will provide the instructor with information about the class as a whole and about students specifically. This could include the number of attempts required to answer questions, the amount of time spent on each question, or even the incorrect answers submitted by the students. This is useful information if the instructor is willing to re-evaluate the use of the program throughout the semester and make changes as warranted. It is also useful in advising individual students, as the instructor will now possess proof of the way that each students was using the program, and this could help in giving advice to struggling students.

One particular aspect we found useful was the ability to view the students' answers to questions. Some programs will present the students last submitted answers, while others will make all answers available (*4*). This was an extremely useful tool in dealing with student complaints regarding the homework. "I'm sure I submitted the correct answer, but it didn't give me credit," can easily be taken care of. It also helps when the instructor or teaching assistant receive emails from students stating that they are stuck on a problem. Seeing what they did incorrectly means that we can provide specific advice.

Student feedback. While not all students can be expected to appreciate the usefulness of these new OHW programs, a successful program should convince most of the class of its usefulness (4). This can be evaluated through student surveys, or by simply an observed change in the amount of students participating (5). Overall, student feedback to online aspects of courses is more polarizing than traditional lecture, either a "love it" or "hate it" set of opinions. Students are accustomed to lectures, they know what to expect, and it may take a particularly good or bad class to elicit strong opinions from

the stud

opinion

simply

overall

this dire

D. Onlin

possible

struggle

students

actively

wide vari

answer a

answers,

about the

intentions

the systen

strategy,

the dilig

collaborat

the materi

the students. Online aspects, however, are unique enough that students often have more opinions. The approach is so different from the traditional, that it is rare that student is simply indifferent to the changes. Even when a program is implemented well and has overall approval from the course, those who disagree will likely have strong opinions in this direction (2).

D. Online Homework Administration

While the aforementioned options were designed to benefit the student, it is possible to have too many aids, which consequently detract from the need to conceptually struggle through answering a question. There must be a balance between giving the students the aid that they need to work through the problems, and still forcing them to actively participate in the homework. A survey of current literature shows that there is a wide variety of ways in which instructors are administering the programs.

A typical approach to OHW is to allow the students multiple opportunities to answer a finite set of questions, to provide suggestive hints that essentially feed the answers, and in limited cases, to offer a chat room feature for students to communicate about the homework problems. Unfortunately, these learning aids added with great intentions to the programs also supply an equally potential opportunity for exploitation of the system. In essence, the student could complete assigned work by a "guess-and-check" strategy, without ever needing to critically think about how to solve a problem. However, the diligent student could ideally find these helpful offerings of hints, student collaborations, and multiple submissions advantageous towards learning and practicing the material.

T

progress

students.

traditiona

method t

parameter

study beg

University

whose sc

significan

factors, i

correlatio

actually te

He

large Org

correlated

measure t

subject. P

were ider

incorporat

between t

Teaching traditional courses online has been studied from many aspects, and progress has been made in determining how to best use this type of medium to educate students. Many studies have previously compared the effectiveness of OHW against traditional written homework (or no assigned homework), but few have address the best method to implement the OHW program, especially in regards to setting the various parameters within the program to maximize the educational effects (*13, 14*). The present study began by analyzing the results of the OHW program used at Michigan State University (East Lansing, MI) in General Chemistry courses. Intriguingly, students whose scores ranked near the bottom in General Chemistry had nevertheless completed a significant amount of the homework problems (~80%). Among several unidentifiable factors, including the similarity of problems on the examinations and OHW, this correlation likely suggests that the OHW program was not being applied effectively to actually teach the students the necessary information.

Herein, we describe how a relatively new OHW system was implemented in a large Organic Chemistry lecture course. Student performance on quizzes and exams was correlated to their successful completion of problems, which in turn was used as a measure to assess the effectiveness of the program towards helping them master the subject. Pertinent oversights in the application of the General Chemistry OHW system were identified and addressed in the present method, and our data indicate that incorporating key limits on the parameters can effectively improve the correlation between the completion of OHW and better test scores.

II. Study

Th

which is

pre-profes

~1000 enr

Online W

accompan

option wa

the total 5

of questio

instructor.

included s

each ques

tutorials) t

three atten

units that

together.

number of

credit was

was submi

were encou

the concep

was not in

II. Study Design

The subjects of this study were the students enrolled in 12 sections of CEM 251, which is the first of a two-semester Organic Chemistry course taken predominantly by pre-professional students. The enrollment in these sections was ~320 students out of ~1000 enrolled in 36 sections during the fall semester (2007). The program used was the Online Web-based Learning (OWL) program by Thomson Learning, designed to accompany the textbook "Organic Chemistry" by John McMurry, edition 7e. The OHW option was administered as extra credit for students to earn up to 10% more points over the total 500 possible points for the class. This particular program featured randomization of questions to benefit the student, included parameter options, and beneficial to the instructor, logged the specific answers entered by each student. Additional options included switches to limit the number of attempts and tally the amount of time spent on each question. Each student was also given access to the supplemental material (e.g. tutorials) through the program. The parameters were generally setup as follows: typically three attempts were allowed for individual question, while more attempts were given for units that contained multiple questions in which several separate responses were graded together. Points were only awarded for correct answers achieved within the limited number of attempts, after which the student were still allowed to submit answers, but no credit was awarded. Furthermore, no feedback was provided until after the correct answer was submitted or the maximum number of attempts was exceeded. Instead, the students were encouraged by the instructors to use the tutorial function of the program to clarify the concepts for each problem deemed challenging. A collaborative learning component was not included with the OWL program, so the effects of this option are not included.

The result
correlated
converting
Se
included
University
administer
to answer
comments
141 when
the correct
look for a
teaching a
have used
cannot pro
student wi
is not the s
effectiveness
Chemistry
this study.
population
The
mandatory

The results of student performance on in-class material (quizzes and examinations) were correlated with the completion of OHW. All data analysis was completed prior to converting the homework tallies to extra credit points that were added to the final score.

Several (38%) of the students enrolled in the 12 sections of Organic Chemistry included in this study had taken General Chemistry, CEM 141, at Michigan State University in fall of the previous year. The LON-CAPA online program was used to administer the homework for this introductory course and typically allowed 10 attempts to answer a question, and included a chat room for students to post questions and comments about the homework. The general approach observed for the students in CEM 141 when answering problems is to guess answers until most of the attempts are used, if the correct answer is not found fortuitously beforehand. Alternatively, these students will look for an answer posted in chat room thread. Anecdotal evidence is supplied by teaching assistants who commonly report that students who visit help sessions ordinarily have used 9 out of 10 attempts to answer a question. Regrettably, these students often cannot provide a coherent basis for their problem solving strategy. In extreme cases, the student will submit a printout of a chat room thread to demonstrate how a posted answer is not the solution to their question. Thus, our analysis also includes a comparison of the effectiveness of the OHW system used in General Chemistry against that used in Organic Chemistry; the latter only included students enrolled in the 12 sections of CEM 251, for this study, who were previously enrolled in CEM 141 the previous year; the overall population difference of each class was large enough that bias could be introduced.

The low-end of the percentage (about 80-85%) of students who completed the mandatory homework in CEM 141 on the LON-CAPA system provided an arbitrary, yet

achievabl

homewor

the prese

minimum

The supp

and plotte

of the assi

achievable target for the students in this study. Thus, a completion of at least 80% of the homework was prescribed as the benchmark to prepare students for quizzes and exams in the present survey. The results of this investigation demonstrate that the utility of the minimum homework recommendation was effective for preparing students for the exams. The supporting data for each quiz and exam was split into groups based on exam scores and plotted against the fraction of students within each group who completed at least 80% of the assigned homework.

III. General Chemistry Students

Many of the students in this organic chemistry class had taken CEM 141 at Michigan State University in the previous year. The CAPA program was used to administer this homework and typically allowed the students 10 attempts to answer a question and included a chat room for students to post questions and comments about the homework.

Analysis of the system used in general chemistry for implementing OHW, shown in table 7-1 and figure 7-1, clearly indicates that the students are not really learning from completing the assignments. Students scoring as low as 50% on in-class exams successfully completed over 90% of their OHW, a strong sign that completing these problems did not have a significant affect on their learning. This could be for several reasons beyond simply stating that the program was ineffective. The concepts of the program may be solid, but just not used in a manner that is most effective. The students may not be required to complete enough problems to fully understand the concepts, or the problems given may lead students to believe that they can simply learn a process of steps for solving a problem, but then find on the exam that this simplified set of sets does not work in all situations. The goal of our study is determine not only if we can improve the correlation between homework completion and performance, but which factors affect this correlation and if any simple changes can be made to existing programs to increase the effectiveness.

Exam Score	OHW	Students
0	52	161
30	83	181
40	89	233
50	93	437
60	95	478
70	97	384
80	98	147
90	99	23

Table 7-1. General chemistry OHW performance. Student data was divided into grade levels based on total exam performance in general chemistry, and the average amount of OHW completed and the number of students at each grade level determined.

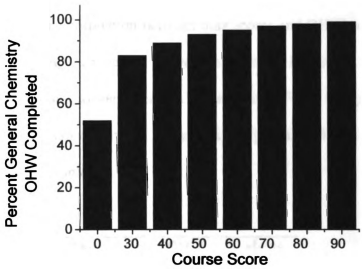


Figure 7-1. General chemistry OHW performance. Average student OHW completion for each grade level in general chemistry.

IV. Organic Chemistry

A. Quiz 1

The first quiz in CEM 251 was 3 weeks into the semester and covered a brief review of general chemistry, hybridization, resonance, naming of organic compounds, and newman projections. The first set of OHW was due at the beginning of class that day. The average quiz score was 72% and the average amount of OHW completed was 60%. The data was divided into groups based on quiz score, and the average amount of OHW completed for each group was calculated. Table 7-2 shows the results, with the quiz score shown as the bottom range of that group (i.e. 0 is the 0-19 group, 20 is the 20-29 group, etc.) Figure 7-2 shows the correlation between quiz scores and OHW completion, which clearly indicates a strong connection between successful homework completion and ability to correctly respond to the questions on the quiz. From this point forward the students told of the correlation between homework completion and quiz performance and were urged to increase completion.

Quiz Score	OHW	Students
0	0	2
20	29	6
30	30	2
40	39	29
50	55	42
60	61	38
70	67	70
80	71	53
90	85	41

Table 7-2. OHW completion for quiz 1. The number of students and the average amount of OHW completed for each grade level on quiz 1.

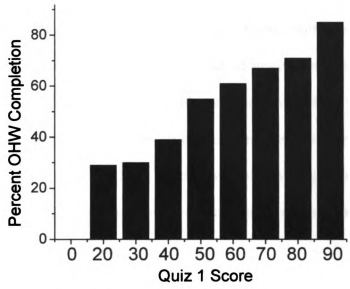


Figure 7-2. OHW completion for quiz 1. The average amount of OHW completed for each grade level.

Throughout the semester, the students were told that a recommended level of OHW completion was 80%. This allowed them room for answering questions incorrectly, but overall they would need to have a good understanding of the material to achieve this level. Teaching staff felt confident that this level would be adequate preparation for in-class evaluation. The data was again divided based on quiz score, and then the proportion of each group that completed the recommended 80% of the homework problems was determined, shown in table 7-3 and Figure 7-3. The results show a strong correlation between completing that minimum recommended level and quiz performance. The most assuring aspect of the analysis was that there were very few students completing that level and performing poorly on the quiz. These lower scoring students were not making full use of all of the study options available, and the instructors could make this simple suggestion to any student seeking advice in the course.

While the correlation was strong, overall the level of completion seemed low. The highest scoring group of students only had 60% with the recommended completion level. This could be simply due to an overall low level of motivation amongst the class for completing the homework. Another possible explanation is that the higher scoring students were choosing not to complete all of the homework not due to motivation, but because they were comfortable with the material without needing to complete the full set of homework. If the students can perform well on quizzes and exams, completing all of the OHW becomes less of an issue.

Quiz Score	Percent With 80% Completion
0	5
50	0
55	6
60	7
65	11
70	20
75	19
80	18
85	24
90	41
95	50
100	59

Table 7-3. Recommended OHW completion for quiz 1. It was recommended to the students that completing 80% of the OHW should prepare them adequately for quizzes and exams. The percentage of students at each grade level completing this amount is shown.

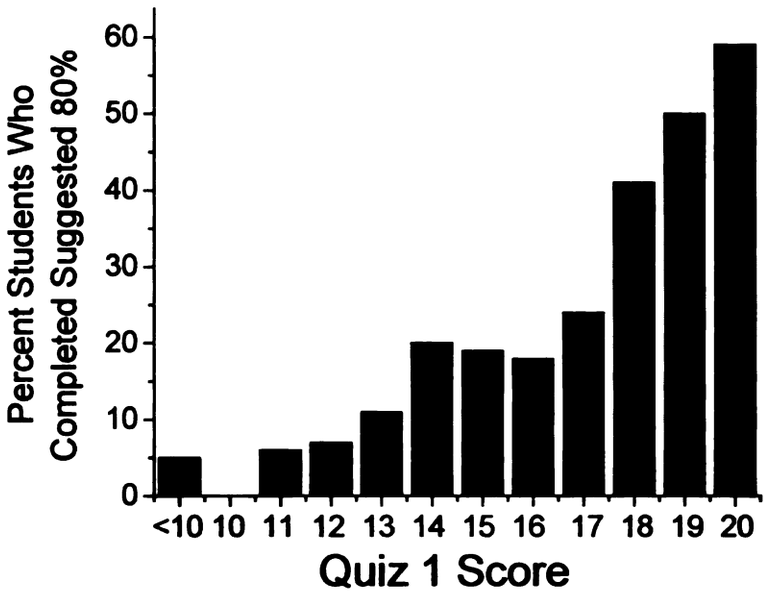


Figure 7-3. Recommended OHW completion for quiz 1. The percentage of students in each grade level completing the recommended level of OHW for quiz 1.

B. Quiz 2

Quiz 2 was administered at 5 weeks into the course and covered chair structures, chemical reactions, and naming/properties of alkenes. The average quiz score was 78% and the average OHW completion was 53%, slightly lower than quiz 1. The quiz scores were analyzed in the same manner as quiz 1 and the results shown in table 7-4 and figure 7-4, again showing a positive correlation between OHW and quiz scores.

By this point in the semester the students and trying to balance multiple classes and other obligations and typically spend less time on studying. At the same time the material becomes more difficult as this is a course where the material builds on itself and the class is inherently cumulative. This is the point in the course where students should be exploring and using all resources available to them to do well in the course, but quiz 2 results show this is not the case. The highest scoring group on each quiz decreased the average homework completion by about 10%.

Quiz Score	OHW	Students
0	17	7
10	10	5
20	27	8
30	30	16
40	36	10
50	29	30
60	50	33
70	54	59
80	64	82
90	74	77

Table 7-4. OHW completion for quiz 2. The number of students and the average amount of OHW completed for each grade level on quiz 2.

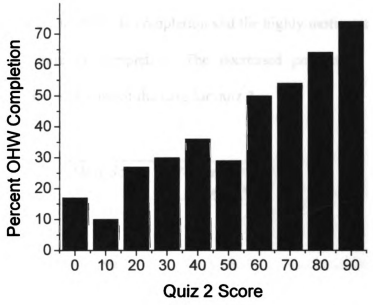


Figure 7-4. OHW completion for quiz 2. The average OHW completion for each grade level on quiz 2.

To determine if the minimum homework recommendation was still a good predictor of quiz success, the data was analyzed the same as for quiz 1, shown in figure 7-5 and table 7-5. In this case there appeared to be more students completing the suggested homework, but still scoring poorly on the quiz. This is likely a side effect of the data analysis, there were only 4 students scoring less than 60% on the quiz that met the minimum recommendation, but due to the overall low number of students in these grade divisions, the percentages make it appear greater. Similar to the overall completion level on quiz 2, the average amount of students completing the recommended level of OHW also decreased. This shows that all students were completing less homework, even the higher scoring students. One may expect that the middle to low scoring students would be the most likely to decrease completion and the highly motivated students would maintain this lower rate of completion. The decreased percentage completing the recommended level shows this is not the case for quiz 2.

Quiz Score	Percent with 80% Completion
0	5
30	6
40	0
50	6
60	18
70	20
80	35
90	53

Table 7-5. Recommended OHW completion for quiz 2. The percentage of students in each grade level completing the recommended amount of OHW for quiz 2.

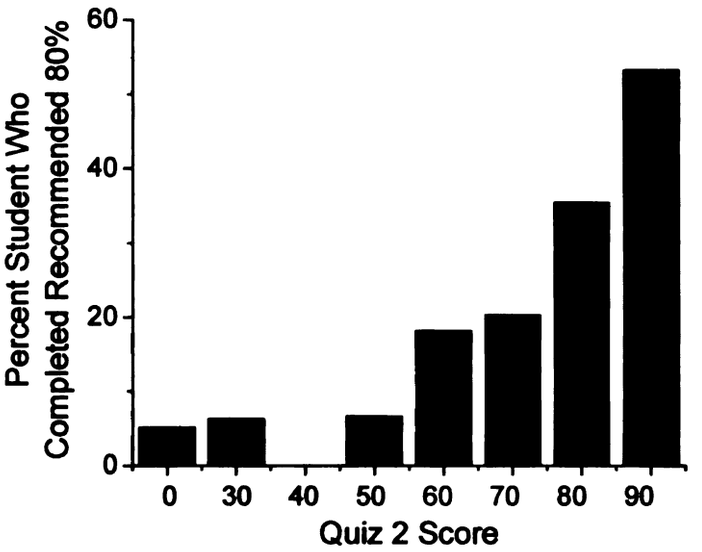


Figure 7-5. Recommended OHW completion for quiz 2. The percentage of students at each grade level completing the recommended amount of OHW for quiz 2.

C. Exam 1

Exam 1 was administered in week 7 of class and covered all material to date in the course: general chemistry review, properties and naming of alkanes and alkenes, and introduction to alkene reactions. The format of the test and level of difficulty was same as the student had seen on quizzes, and overall their individual performance seemed to be similar to previous quiz performance. The last set of OHW covering the exam material was due on the day of the exam, but the majority of the related homework had been due before for the previous quizzes. The average exam score was 72% and average amount of OHW completion for the unit was 55%.

Effect of OHW. The data was analyzed as previous and shown in table 7-6 and figure 7-6. The strongest correlation between OHW completion and in-class performance to date was observed. While it should be noted that again the lowering scoring groups had

very few students and therefore the data is not as significant, the results still fit well with the larger groups of higher scoring students.

The strong correlation observed for exam 1 indicates that the OHW is reinforcing the material in a way that the students are still performing well weeks after the quizzes on the tests. This could be due to the students returning to the repeat problems in preparation for the exam, or that the design of the question and the method of implementation allows the students to retain more of the information while completing the homework compared to simply cramming the material before the quiz.

Exam Score	OHW	Students
0	0	1
10	0	1
20	7	8
30	15	12
40	27	32
50	44	41
60	51	51
70	64	79
80	75	65
90	82	34

Table 7-6. OHW completion for exam 1. The number of students and the average amount of OHW completed for each grade level on Exam 1.

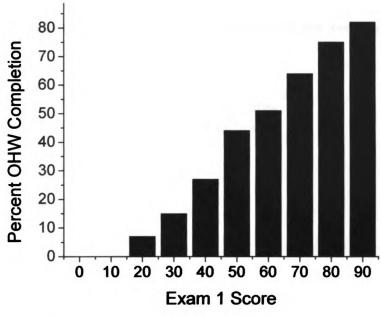


Figure 7-6. OHW completion for exam 1. The average OHW completion for each grade level for exam 1.

Recommended OHW Completion. The next step in data analysis was the recommended level of homework completion, which was analyzed in the same fashion as the quiz results, table 7-7 and figure 7-7. Overall, the portion of the class completing the minimum recommendation scored an average of 82.5% on exam 1, a very respectable score. The students completing less than the minimum scored an average of 61.7%, over 20% lower than the other group. Analysis of each grade level shows a strong correlation between each grade division and the fraction of each group completing the recommended homework. No student scoring lower than 50% on the exam had completed the suggested level of homework, while the highest scoring divisions had the majority of the group completing this level. This reinforced the effectiveness of the OHW program, and this finding was passed along to the students as motivation to increase their participation.

Students seeking advice on how to increase their score in the class were directed to the homework program, and those that aimed to maintain high scores were urged to continue.

Exam Score	Percent with 80% Completion
0	0
10	0
20	0
30	0
40	0
50	10
60	18
70	36
80	58
90	73

Table 7-7. Recommended OHW completion for exam 1. Percentage of students in each grade level completing the recommended level of homework completion on exam 1.

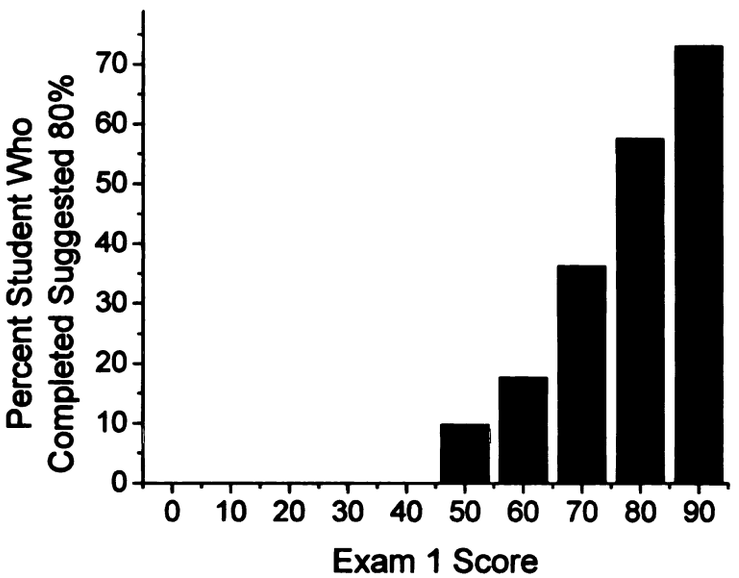


Figure 7-7. Recommended OHW completion for exam 1. The percentage of students in each grade level completing the recommended level of OHW on exam 1.

Oxidative Cleavage Concepts. Along with overall homework completion and quiz/exam performance, the teaching staff wanted to determine how effectively the homework program was at teaching specific topics. One subject in particular that seemed to be difficult for the students to grasp was oxidative cleavage. Most were able to understand the concept that certain reagents would break an alkene, producing two carbonyls in its place. The actual process of drawing those two products, or working a reaction backwards to provide the starting material from given products of a reaction, was more difficult. The students were still becoming familiar with the system of drawing chemical structures and developing useful habits such as counting carbons in the starting materials and products, and now complicated reactions were added.

The OHW set required the students to complete two questions related to oxidative cleavage, the questions illustrated in figure 7-8. The first question asked students to work backwards from a given product to determine the starting material, with the added twist that only a single molecule was formed (indicating the that initial alkene was a ring). The second question is a bit more complicated, as the students are given the formula for the starting material along with several clues about the reaction and asked to provide a structure. Overall, these were challenging questions relative to the level of difficulty which the students were expected to learn the material.

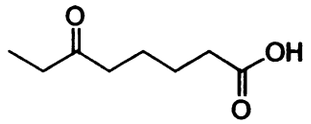
Exam 1 contained a three part question, a fill in the blank style question regarding the same concepts, shown in figure 7-9. Again, the students had to work backwards to provide the initial starting material, and also know which specific reagent would complete the reaction. Once the initial molecule was determined, it must be used to complete a second reaction. The instructor for the course felt that the OHW should have

prepared the students well for the exam question, performance on this specific exam question was correlated to their success in completing the OHW. The student data was divided into groups based on the number of exam points earned, and the average OHW points earned for each group determined (out of the maximum two points), table 7-8 and figure 7-10. There is a positive correlation between performing well on the exam and completing the related OHW questions. The students who scored no exam points earned an average of only .4 of the 2 possible points, meaning that the typical student did not answer either question correctly. Of the students scoring perfectly, the average homework points earned was 1.6, the typical student in this group answered between 1 and 2 questions correctly in the allotted number of attempts.

Attempts per Question. Throughout the semester, the students frequently asked for more attempts to answer the homework questions, as they had become used to general chemistry. Convinced that these extra attempts were not actually aiding in learning, the instructors maintained the level of three attempts for the average question. For this set of two questions, the minimum number of attempts for correct answers was two, and the maximum number to still receive credit was six. The data for students who arrived at the correct answer within the allotted number of attempts was divided into groups based on the number of attempts, and the average exam points earned calculated for each group, table 7-9 and figure 7-11. While the trend is not strong, a minor decrease in score is observed as more attempts are required. Clearly the increased number of attempts did not aid in the students understanding of the material. The students who either knew the answers initially from previous studying or paying careful attention in class and answered correctly on the first attempt faired the best on the exam, while the students who required

more attempts (or guesses) scored lower. This supported the argument that increasing the number of attempts to high levels such as 10 per question would not actually help student learning. It may actually hinder by allowing them to guess towards the correct answer without actually learning the material.

OHW question 6910
Compound **A** gives the product(s) below on oxidative cleavage with $KMnO_4$ in acidic solution. Propose a structure for **A.**



OHW Question 6940
Compound **A, $C_{10}H_{16}$** reacts with **1** molar equivalent(s) of hydrogen on catalytic hydrogenation. **A** undergoes reaction with ozone, followed by Zn treatment, to give a symmetrical diketone, $C_{10}H_{16}O_2$.
Propose a structure for **A.**

Figure 7-8. Oxidative cleavage OHW questions. OHW questions assigned to students in preparation for oxidative cleavage questions they may encounter on a quiz or exam.

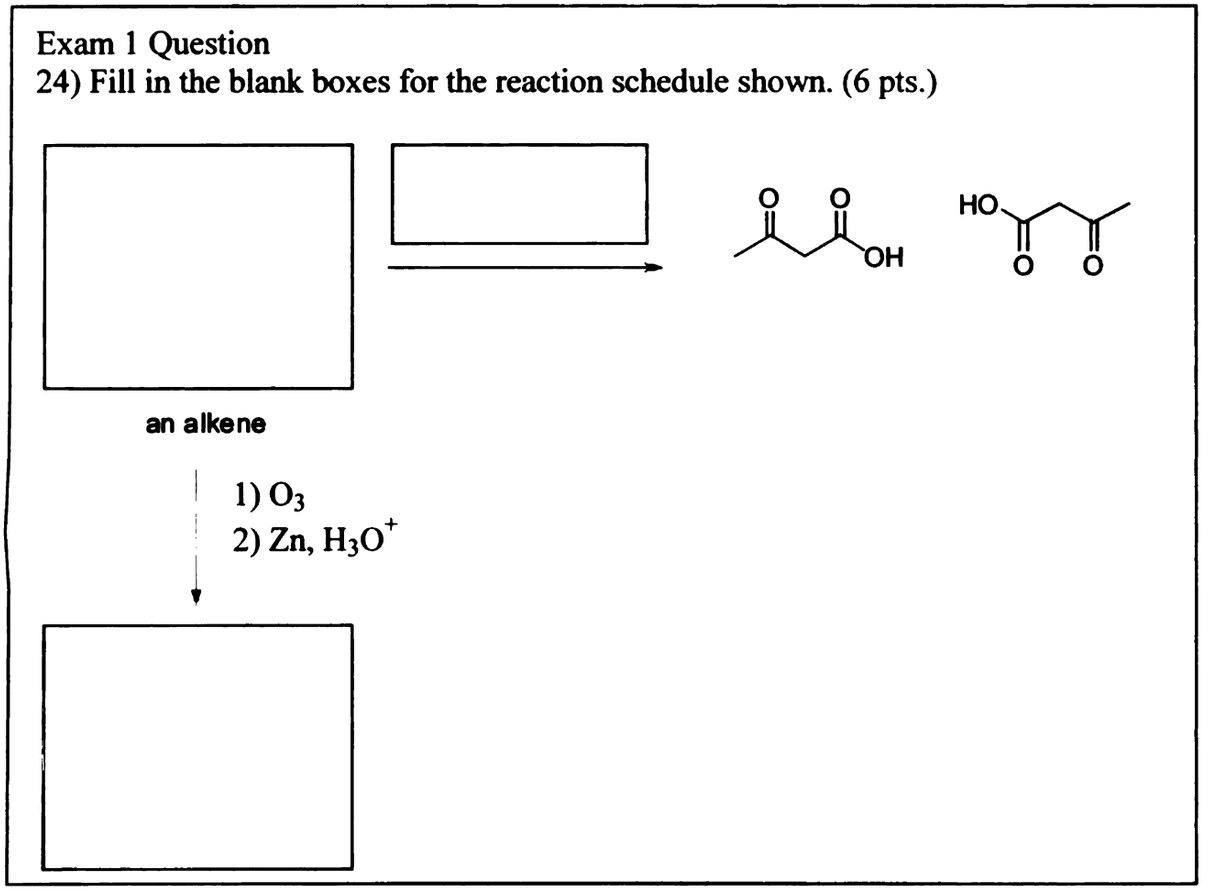


Figure 7-9. Oxidative cleavage exam 1 questions.

Oxidative Cleavage Exam Points (of 6)	Related OHW Points (of 2)	Related OHW Percent
0	.4	20
1	.5	25
2	.8	40
3	.9	45
4	1.1	55
5	1.1	55
6	1.6	80

Table 7-8. Oxidative cleavage exam performance vs. OHW completion. The average number of OHW points (of the two possible points) and percentage for each level of performance on the exam questions.

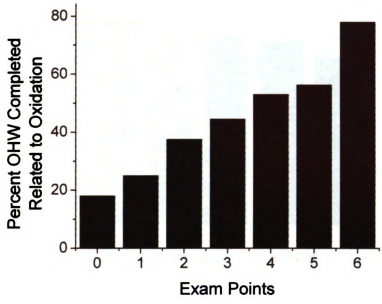


Figure 7-10. Oxidative cleavage exam performance vs. OHW completion. Average level of OHW completion for each level of performance on exam 1 oxidative cleavage questions.

OHW Attempts	Related Exam Points
2	5.1
3	4.7
4	4.7
5	4.5
6	4.1

Table 7-9. OHW attempts vs. exam performance. Average number of points scored on exam 1 oxidative cleavage questions for each number of attempts made on the related OHW questions.

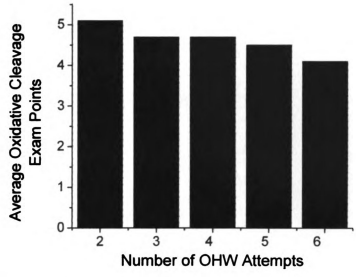


Figure 7-11. OHW attempts vs. exam performance. The average numbers of exam points earned on the oxidative cleavage questions on exam 1 for each number of attempts made at related OHW questions.

D. Quiz 3

Quiz three was given at week 10 in the semester and covered material including stereochemistry along with reactions of alkynes and halogens. The average quiz score was 58% and the average amount of OHW completed was only 49%. By this point the students were settled into the class and into their study habits. They have been evaluated in class on two quizzes and an exam, and had plenty of opportunities to meet with the professor or teaching assistant for guidance on how to increase their score in the class. If the OHW was not useful to the students, at this point we would expect to see participation dwindle and the correlation between quiz scores and OHW completion to fade away. This is not what was observed, as tables and 7-10 and 7-11 and figures 7-12 and 7-13 indicate. The students who are scoring higher on the quiz are still the group that is completing more of the homework. The recommended level of completion was still a

good indicator of success, as only a few students in the class completing that level and still scoring poorly on the quiz.

Quiz Score	OHW	Students
10	7	5
20	17	22
30	31	34
40	39	36
50	50	51
60	64	54
70	67	51
80	71	39
90	72	12

Table 7-10. OHW completion for quiz 3. The number of students and the average amount of OHW completed for each grade level on Quiz 3.

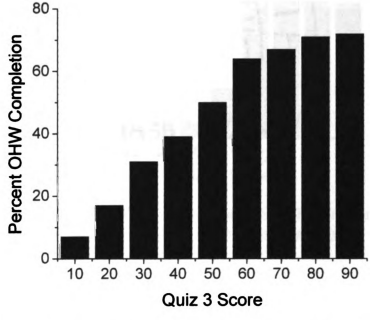


Figure 7-12. OHW completion for quiz 3. The average amount of OHW completed for each grade level on quiz 3.

Quiz Score	Percent with 80% Completion
10	0
20	5
30	6
40	6
50	18
60	50
70	49
80	49
90	67

Table 7-11. Recommended OHW completion for quiz 3. Percentage of students of each grade level completing the recommended level of OHW for quiz 3.

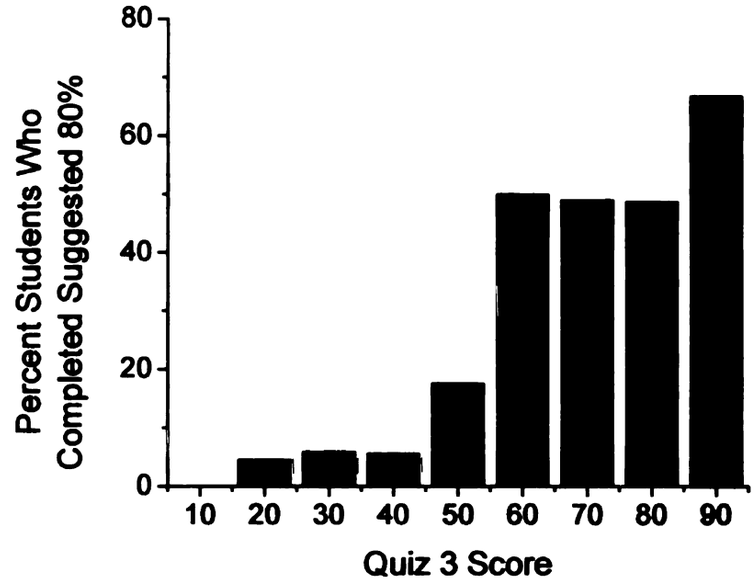


Figure 7-13. Recommended OHW completion for quiz 3. The percentage of students completing the recommended level of OHW completion for each grade level on quiz 3.

E. Quiz 4

Quiz 4 was given 12 weeks into the course, and covered material including substitution and elimination reactions. The average score on quiz 4 was 67%, and average homework completion was down to 54%. This may explain the data analysis of quiz 4, which indicated a little correlation between completing OHW and quiz performance, or

no correlation between completing the recommended level of homework, tables 7-12 and 7-13 and figure 7-14 and 7-15. There was no obvious reason for this lack of correlation, other than overall low homework participation that makes it difficult to accurately analyze the affects on course performance.

Quiz Score	OHW	Students
20	42	5
30	39	9
40	32	22
50	45	50
60	62	50
70	54	59
80	58	45
90	74	41

Table 7-12. OHW completion for quiz 4. The number of students and the average amount of OHW completed for each grade level on Quiz 4.

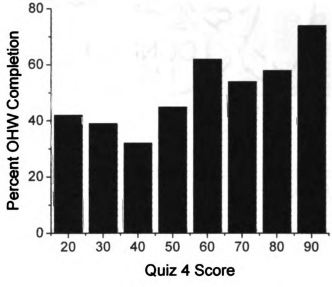


Figure 7-14. OHW completion for quiz 4. The average OHW completed for each grade level on quiz 4.

Quiz Score	Percent with 80% Completion
20	20
30	11
40	23
50	34
60	44
70	39
80	31
90	37

Table 7-13. Recommended OHW completion for quiz 4. Percentage of students completing the recommended level of OHW for each grade level on quiz 4.

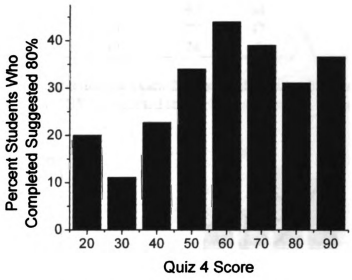


Figure 7-15. Recommended OHW completion for quiz 4. The percentage of students completing the recommended level of completion for each grade level on quiz 4.

F. Exam 2

Affect of OHW Completion. Exam 2 was given 13 weeks into the course and included the material mainly from quiz 4 and 5. The average score on exam 2 was 64%, with an average OHW completion level of 52%. Initial analysis of the data showed a restored correlation between OHW and performance, table 14 and figure 16. This was a

reassuring finding given the results from quiz 4. While not quite as strong as for exam 1, there is still an obvious relationship between completing the homework and the ability to do well on the exam. The 80% recommended completion level still appeared to be a good standard for exam 2, table 15 and figure 17.

Exam Score	OHW	Students
0	11	11
30	16	27
40	35	31
50	46	55
60	63	65
70	62	64
80	77	47
90	74	6

Table 7-14. OHW completion for exam 2. The number of students and the average amount of OHW completed for each grade level on Exam 2.

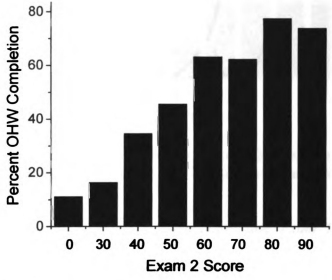


Figure 7-16. OHW completion for exam 2. The average amount of OHW completed for each grade level on exam 2.

Exam Score	Percent with 80% Completion
0	0
30	0
40	3
50	11
60	34
70	38
80	60
90	50

Table 7-15. Recommended OHW completion for exam 2. Percentage of students completing the recommended level of OHW completion for each grade level on exam 2.

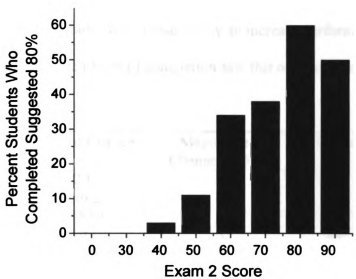


Figure 7-17. Recommended OHW completion for exam 2. The percentage of students completing the recommended amount of OHW for each grade level on exam 2.

Correlation of Change in OHW Completion and Exam Performance. After analyzing overall trends of the second exam, we were beginning to question if the trends observed were simply due to the fact that students who completed homework are also more likely to study for exams in other ways, and therefore score higher, but that the two may not be directly related. Even if completely unrelated material was covered in

homework than on exams, one would expect to still see a slight correlation between homework completion and exam scores. Student who scored poorly on the first exam were urged to complete more of the OHW in preparation for the second exam, while it was observed that there was a portion of the class that was decreasing in homework completion. The data was grouped by the change in exam scores, and the average change in OHW was determined for each group, shown in figure 7-18 and table 7-16. It was observed that each group with a decrease in exam scores had also decreased homework completion, and those with an increase had increased the amount of homework, with a slight correlation between the magnitude of each. This shows that students who turned to the OHW to increase their grades were rewarded by an increase performance, while those that neglected to maintain their level of completion saw this on exam scores as well.

Magnitude of Change in Exam Score	Magnitude of Change in OHW	Students
Less 20	-14	49
Less 10 to 20	-6	83
Less 0 to 10	-7	93
1 to 10	1	65
10 to 20	4	14
Over 20	9	5

Table 7-16. Change in exam performance. The student data was divided based on the change in magnitude of exam 1 to exam 2 scores, and the average change in OHW completion determined.

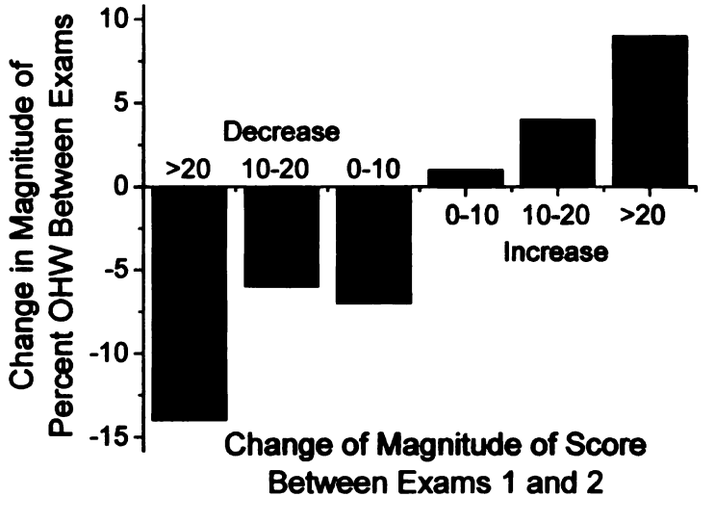


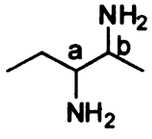
Figure 7-18. Change in exam performance. Average change in OHW completion for each division of change in exam scores between exam 1 and 2.

Concept Analysis. Again with exam 2 we attempted to find a relationship between specific concepts covered on the homework and then on the exam. The way that a question is asked to the student can probe different levels of understanding. Simple true/false questions are more likely to test the basic level of knowledge, possibly just the student's ability to memorize the material presented in lecture. Asking students to pull together multiple concepts, or to expand upon the material presented in class tests the highest level of knowledge. As part of our study of OHW, we wanted to determine if the type of questions we were providing the students was preparing them all levels of thinking on the exams.

Stereochemistry Concepts. The first question analyzed was stereochemistry, the exam and homework questions are shown in figure 7-20. On the homework, the students were asked to identify stereocenters as either R or S. The exam had questions of very similar nature, simply asking students to identify as R or S. This data analysis was

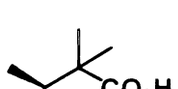
different than any previous in the course due to the multiple-choice nature of these questions, something not typically encountered in the class. It was expected that there would less of a distribution in grades. Simply guessing would give the students a chance at points, but there was no partial credit awarded. Even with this, there was a correlation observed between homework completion and test performance, as shown in figure 7-21 and table 7-17.

9.5-9.6a Homework: Specifying Absolute Configuration in JMOL Images
Specify the configurations (**R** and **S**) of chiral center **a** and **b** in the Jmol structure below.
(Color scheme: Light Red=oxygen; Blue=nitrogen; Bright Green= chlorine; Pale Green=fluorine; Yellow=sulfur; Brown=Bromine)



Carbon a: R S
Carbon b: R S

9.5-9.6b Homework: Determining R/S Configurations
Identify the stereochemistry in each of the following compounds as **R** or **S**. Note: if multiple stereocenters are present, indicate the stereochemical designations as: **RS, SS, RS,** or **SR.**





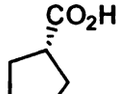


Figure 7-19. Stereochemistry OHW questions. OHW questions provided to students regarding the assignment of stereochemistry to chiral carbons.

Exam 2 Question
4) Assign R or S configurations to each indicated center of chirality in the molecules below (3 pts).

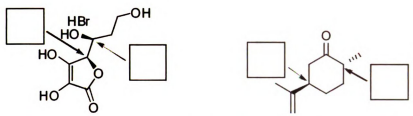


Figure 7-20. Stereochemistry exam 2 questions.

Exam Points Earned Related to Stereochemistry	Related OHW Points (of 4)	Percent Related OHW	Students
0	.3	6	4
1	1.5	30	22
2	2.4	48	66
3	2.9	58	120
4	3.1	62	81

Table 7-17. Stereochemistry OHW completion vs. exam performance. Number of student and average amount of related OHW questions completed for each level of performance on related exam questions.

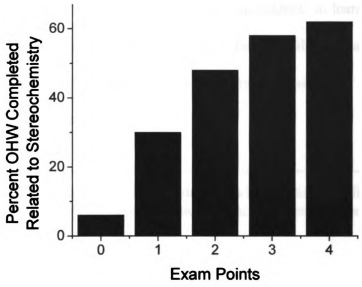


Figure 7-21. Stereochemistry OHW completion vs. exam performance. The average amount of related OHW questions completed for each level of performance on stereochemistry exam questions.

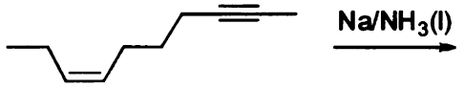
Alkyne Reaction Concepts. The next question type analyzed on exam two was a fill-in-the-box style set of problems regarding oxidation, reduction, and alkylation reaction of alkyles. This is a way of testing an intermediate knowledge of the material. The students must provide the answers, they are not provided with possible choices, but each question is only regarding a simple, one-step reaction. The OHW asked students to draw the products of reactions, figure 7-22, and were asked to answer test questions in the same manner, figure 7-23. While the OHW did not award partial credit, as typical, the students did have the possibility of earning some exam partial credit if their incorrect answer demonstrated enough understanding of the concepts.

Again, the students who performed well on the exam questions had also completed the related OHW questions, figure 7-24 and table 7-18. Unfortunately, even

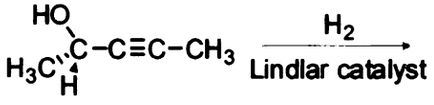
the lowest scoring group of students had completed some of the OHW. While we would like to see that any level of OHW completion can help students in learning material and therefore perform better on quizzes and exams. It is quite possible that a threshold exists. It may be the case that student must complete a minimum amount of the OHW before benefiting on exam performance.

8.4-8.5 Homework: Hydroboration-Oxidation and Reduction of Alkynes
Draw the product(s) of the following reactions including stereochemistry when it is appropriate.

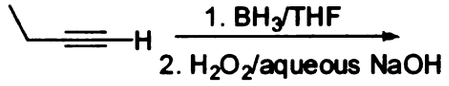
1.



2.

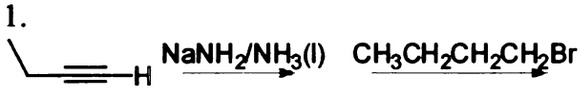


3.

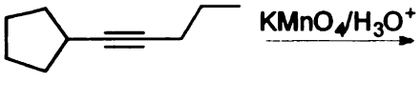


8.7-8.8 Homework: Oxidation and Alkylation of Alkynes
Draw the major organic product(s) of the following reaction.

1.



2.



3.

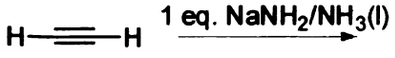


Figure 7-22. Alkyne OHW questions. Oxidation, reduction, and alkylation questions assigned to students on OHW in preparation for quiz and exam questions.

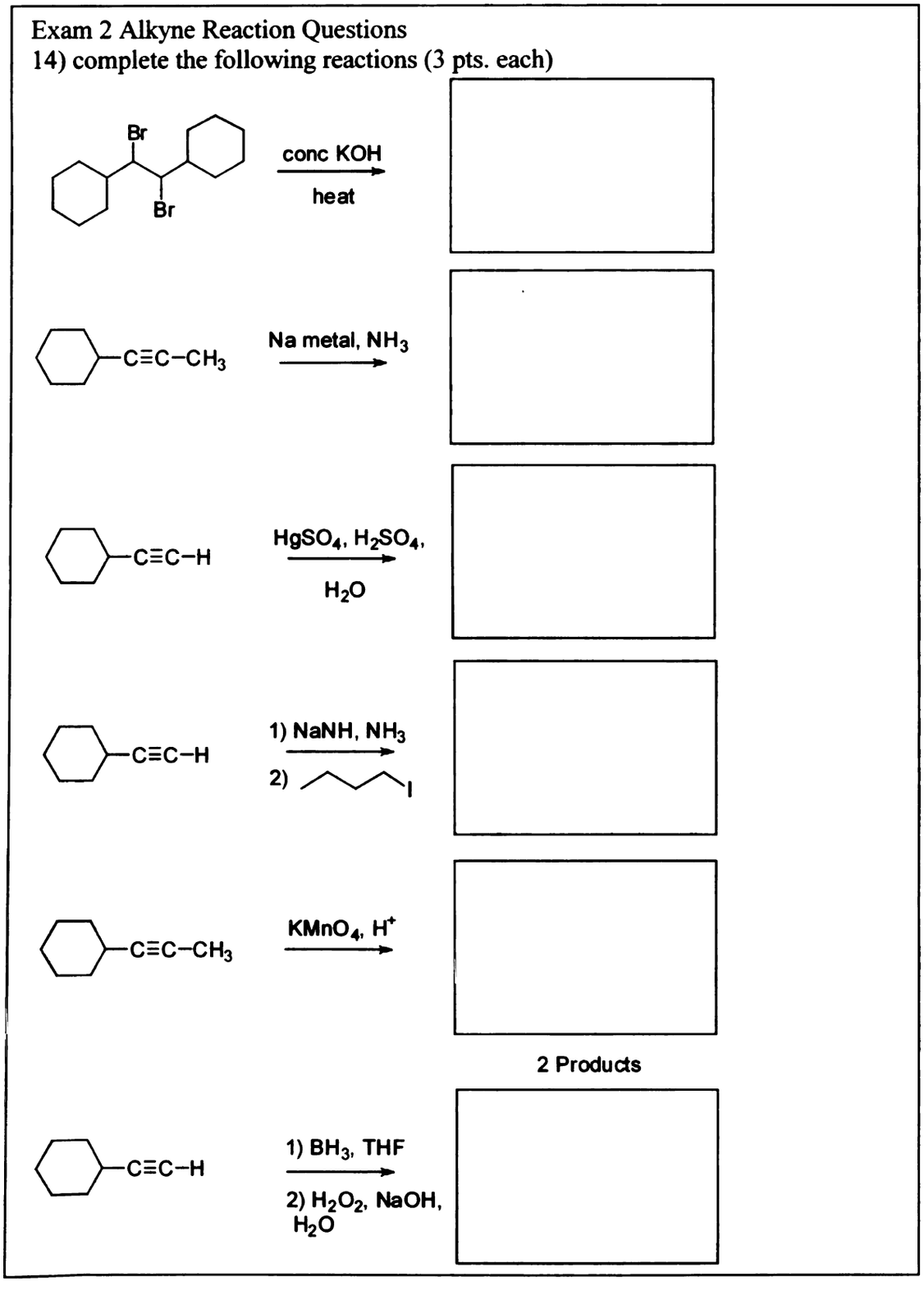


Figure 7-23. Alkyne exam 2 questions. Alkyne reaction questions encountered by students on exam 2.

Exam Points Related to Alkyne Reactions	Percent OHW Related to Alkyne Reactions	Students
0-20%	19	22
21-40%	41	43
41-60%	44	37
61-80%	60	71
81-100%	62	105

Table 7-18. Alkyne OHW completion vs. exam 2 performance. Average amount of related OHW completed for each level of exam performance on alkyne reaction exam questions.

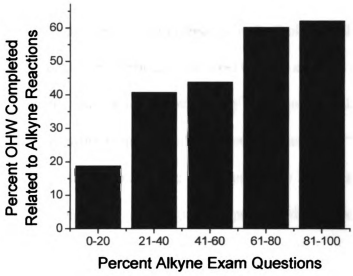


Figure 7-24. Alkyne OHW completion vs. exam 2 performance. The average amount of OHW completed for each performance level of alkyne reaction exam questions.

Synthesis Reaction Concepts. The last type of question analyzed is conceptually the most difficult for students, synthesis problems. For these questions the student must recognize the key transformations occurring between the starting material and product, know the reagents necessary to complete these changes, and have an understanding of how the order of these reagents could affect the outcome. Pulling together multiple concepts is very difficult for students. If they cannot answer the short answer problem

types, these become impossible. It is the most difficult to give students these types of questions to practice, as they take the longest to grade and unless designed well, may have several correct answers. Providing students with practice problems for which more than one possible solution exist is not intrinsically a bad idea, as it can allow students to think about problems in more than one way or learn more by comparing these different answers with classmates, it does pose a problem when administered by a computer program that does not acknowledge more than one potential answer. This seems to arise most often for the particularly talented students who look past the obvious, simple answer to question why other alternatives are not correct as well. Problems also arise when the material is presented in lecture at a basic enough level that the students could legitimately arrive at multiple answers, that based on what they were told in class all should be correct, but the online homework system is only prepared to accept the "real life" answer.

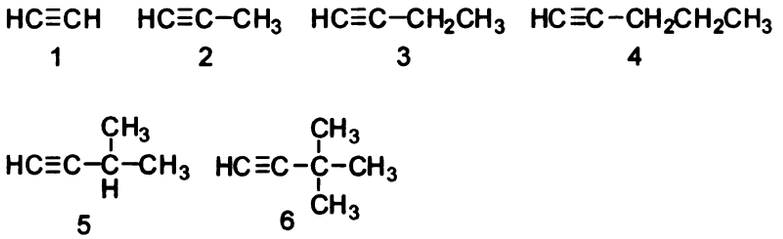
The OHW synthesis question, figure 7-25, asked the students to select from a bank of potential starting materials and reagents to produce the desired product. While this is structured as a multiple choice question, it is very unlikely that a student could simply guess the right combination, basically turning this into a free response question. The students were graded simply as correct or incorrect, with no partial credit for individual answers. Figure 7-26 shows a similar question given on exam 2, where the students are provided with both the starting material and product and asked what reagents are required to complete the reaction. For this question, the students were awarded partial credit for answers that were near the correct answer, and did not have any major conceptual flaws.

To analyze the effect of OHW completion on exam performance, the percentage of students correctly completing the OHW synthesis question was determined for each level of performance on the exam synthesis question. Shown in figure 7-27 and table 7-19, the trend was not smooth, but overall showed that more of the students who scored better on the exam had been able to complete the OHW question.

OHW Question

8.9 Homework: Organic Synthesis Using Alkynes
Devise a synthesis of **(Z)-3-octene** using one of the starting materials and any of the reactants below using the fewest steps possible. If you need fewer than the 3 steps allowed, enter "none" for reagents in the remaining unused steps.

Starting materials:



Reagents

a) $NaNH_2/NH_3(l)$	b) $NaOH/H_2O$	c) iodomethane
d) iodoethane	e) 1-bromopropane	f) 2-bromopropan
g) 1-bromo-3-methylbutate	h) *t*-butyl bromide	i) H_2/Pd on carbon
j) H_2/Lindlar catalyst	k) $H_2/NH_3(l)$	I) $Na/NH_3(l)$

Starting material?
Reagent for step 1?
Reagent for step 2?
Reagent for step 3?

Figure 7-25. Synthesis OHW questions. The question was graded as correct or incorrect, with no partial credit awarded.

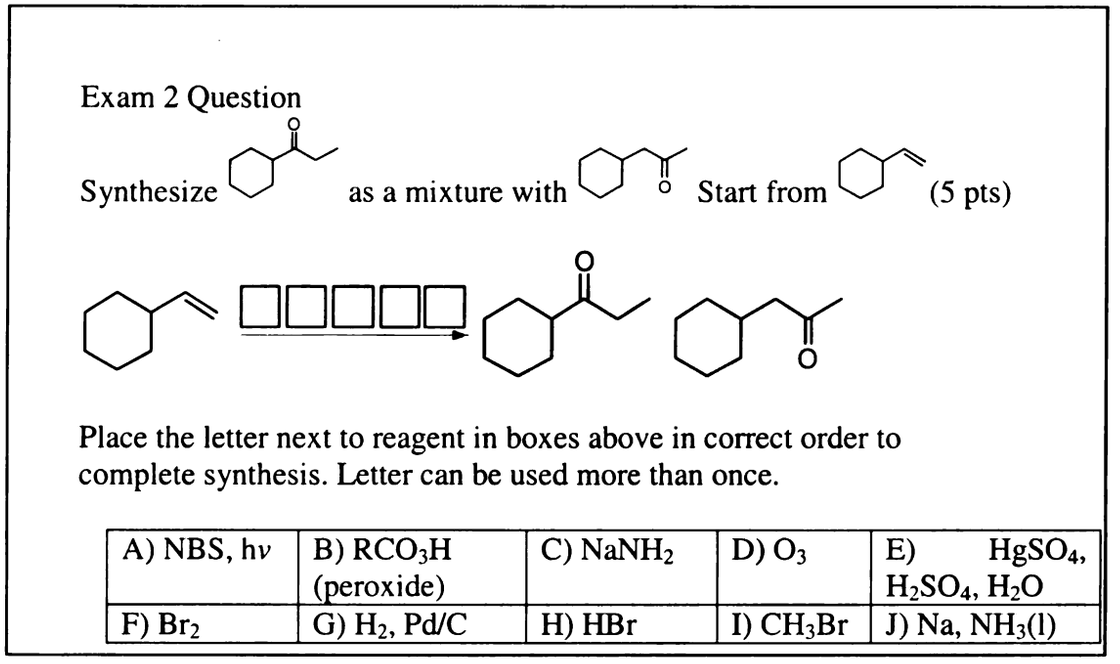


Figure 7-26. Synthesis question from exam 2. This question was scored out of 5 possible points with partial credit awarded.

Exam Points Related to Synthesis	Total Students	Students with Synthesis OHW Completed	Percent Students Completing Synthesis OHW
0	18	2	11
1	64	23	36
2	114	38	33
3	48	24	50
4	47	22	47
5	13	8	62

Table 7-19. Total numbers of students and number successfully completing the synthesis OHW question for each level of performance on the exam 2 synthesis question.

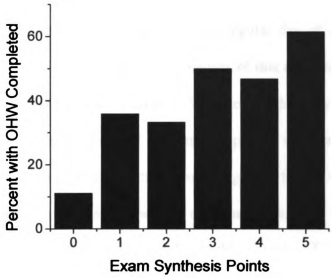


Figure 7-27. Synthesis OHW completion vs. exam 2 performance. The percent of students successfully completing the OHW synthesis question for each level of exam performance.

Analysis of specific question types from exam two showed that the OHW program has the potential to help students learn a variety of material. It worked best for simple problems, such as assigned the R or S stereochemistry of a chiral center, or for the most complex problem types, such as multi-step synthesis questions. The intermediate question type, the short answer questions, did not prepare the students as well for the exam as more were completing the OHW but still scoring low. This could have been for multiple reasons, such as including too few of questions. Or questions on either the homework or the exam that deviated from the most standard case, confusing those students who studied this material by simply memorizing the generic example reaction given in class or in the textbook. An obvious first step in attempt to remedy this problem would be to determine if assigning more questions, to give the students more practice with a wider variety of situations, would improve their performance.

G. Quiz 5

Quiz five was administered during the last regular day of class, covering the information learned in the spectroscopy unit. Analysis of this quiz would be particularly interesting, as the students made no attempt to hide the fact that they were pre-occupied by other commitments. Tests and end of the semester project in other classes, as well as studying for the final in organic chemistry were taking their time. Overall, the students were completing a relatively small amount of the homework, with the highest scoring group of students with only about 60% completing the recommended amount. Figure 7-28 and table 7-20 show the results of the data analysis comparing homework completion to quiz performance. Figure 7-29 and table 7-21 show the exam performance vs. the recommended completion level. Again it was observed that lower scoring students were completing OHW, but overall were not to the recommended level.

Quiz Score	OHW	Students
0	40	8
20	43	16
30	28	33
40	46	37
50	56	59
60	49	64
70	57	51
80	67	32
90	67	9

Table 7-20. OHW completion for quiz 5. The number of students and the average amount of OHW completed for each grade level on Quiz 5.

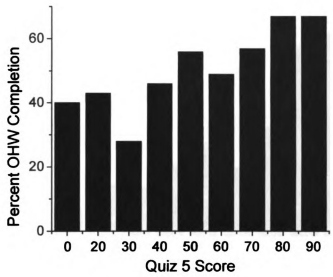


Figure 7-28. OHW completion for quiz 5. The average OHW completion for each level of quiz 5 performance.

Quiz Score	Percent with 80% Completion
0	13
20	19
30	18
40	24
50	29
60	23
70	33
80	53
90	56

Table 7-21. Recommended OHW completion for quiz 5. Number of students and percent completing the recommended level of OHW homework for each performance level on quiz 5.

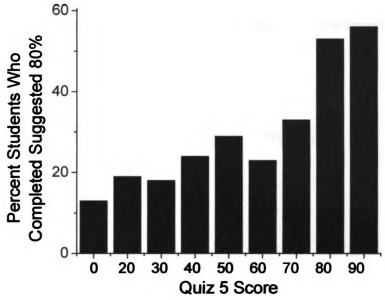


Figure 7-29. Recommended OHW completion for quiz 5. The average amount of students completing the recommended level of OHW completion for each level of quiz 5 performance.

H. Final Exam

The final exam was administered during the standard, university assigned time and covered all of the material learned in the semester. Overall, our goal in designing the OHW program was to teach the students the course material in a way that they were practicing and learning the concepts, that they would be able to demonstrate this on in-class evaluation, and ultimately retain the information. The data analysis throughout the semester indicated that the OHW was doing just that, and now the real test was to see if the trend continued on the final exam.

Figure 7-30 and table 7-22 show the results of the exam performance vs. OHW completion throughout the whole semester. Figure 7-31 and table 7-23 analyzed the recommended completion level, again for the whole semester. Both show strong trends, especially the recommended completion level. These results indicate that the students are

most likely not simply using the homework to memorize the material before the tests. They are learning the material in a way that they can still recall the knowledge at the end of the semester.

Final Exam Score	OHW	Students
0	14	5
20	14	16
30	34	25
40	42	35
50	56	51
60	64	65
70	76	72
80	76	34
90	92	3

Table 7-22. Total OHW completion vs. final exam performance. Number of students and average level of OHW completion for each level of final exam performance.

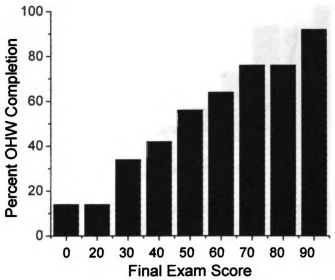


Figure 7-30. Total OHW completion vs. final exam performance. The average OHW completion for each level of final exam performance.

Final Exam Score	Percent with 80% Completion
0	0
20	0
30	0
40	3
50	12
60	29
70	51
80	50
90	100

Table 7-23. Recommended OHW completion vs. final exam performance. Number of students and percentage of students completing the recommended level of OHW for each level of final exam performance.

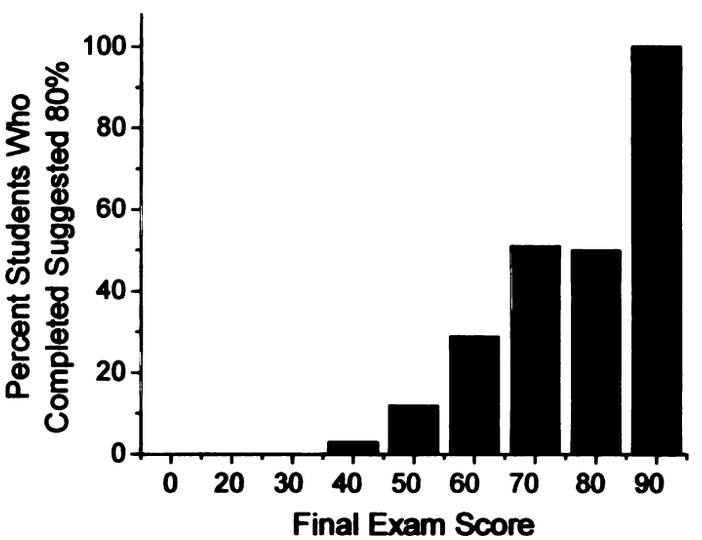


Figure 7-31. Recommended OHW completion vs. final exam performance. The percentage of students completing the recommended level of homework completion for each level of final exam performance.

I. Course Analysis

OHW in Organic Chemistry vs. General Chemistry. Overall, an excellent set of correlations was observed throughout the semester in regards to the data analysis throughout the semester that indicated this method of implementing homework was more successful than methods used in the general chemistry courses at Michigan State. To fully test this theory, the overall OHW completion was compared to the in-class performance from each class. For the organic chemistry course, all students were included in the data analysis, including those that did not register for the OHW program in organic chemistry. This was done to account for students who may have done no OHW (those who never registered) but may have still scored well on in class assessment. If this affects the data analysis in any way, it would lessen the apparent correlation between OHW completion and course score.

To most accurately compare the two sections, the general chemistry data was only considered for students who were enrolled in this section of organic chemistry. The general chemistry course is taken by many non-science majors as a general requirement, and that group of students would not be a good comparison for the mostly pre-professional students in our class. Table 7-24 and figure 7-32 show the results of this analysis. There is no correlation observed between homework completion and exam scores for general chemistry, but an excellent correlation exists for organic chemistry. Using the same group of students as a control, this can only lead to the conclusion that this method of implementation was superior than that used in previous classes. A higher level of participation was observed overall for general chemistry, and this was most likely due to the fact that it was part of the course grade instead of extra credit, as in organic

chemistry, forcing more students to participate. This strengthens the argument that students were learning more by completing the organic chemistry homework, as this is evidence of a large number of students completing the general chemistry homework and clearly not learning the material.

Course Score	Organic OHW	Students	General Chemistry OHW	Students
0	14.0	27	96.9	3
40	37.2	31	96	22
50	45.4	48	96.3	18
60	62.2	71	96.9	38
70	73.0	88	95.3	29
80	79.8	40	99.3	14
90	94	3	100	2

Table 7-24. OHW in general vs. organic chemistry. The average amount of OHW completed in Organic Chemistry and General Chemistry for each performance level on in-class evaluations.

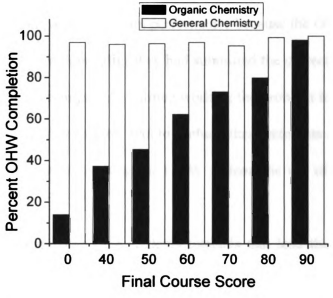


Figure 7-32. OHW in general vs. organic chemistry. The average amount of OHW completed for each grade level in both Organic Chemistry and General Chemistry.

Tutorial Participation. Throughout the semester, all of the data analysis considered homework completion as the submission of the correct answer within the number of allotted attempts. This was used as an overall measure of student participation in the OHW, considering that with three attempts the students who were dedicated to working the problems should have been able to complete most with little difficulty. There was still the chance that some of the students were working hard at learning the material through the OHW, but did not get to the correct answer before using their maximum number of attempts. These students would then essentially be falling through the cracks of our data analysis. They could be the lower scoring group of students who consistently showed a lower level of homework participation, but not because they were actually using the program less but because they weren't able to learn the material in the way our program was designed.

It is difficult to analyze the amount of attempts that the students required to get to the correct answer in comparison to exam performance because the OHW system allowed the students to return to problems after they had submitted the correct answer, or reached the maximum number of attempts, to continue working the problem for practice. The data that the instructors were able to view would include these attempts as well. As another measure of student dedication to using the OHW system, the use of the tutorial feature was analyzed.

The OWL program offers the students tutorials that lead them through difficult concepts. These are not graded assignments, and completion is not required. It was decided to use these in place of automatic feedback (which in certain cases essentially gave the students the answers). The students were directed to use these tutorials if they

were stuck on a certain concept, in addition to reviewing class notes, the textbook, or taking advantage of office hours. The only measure of completion that was made available to the instructors was if the students opened the tutorials, since they were not graded assignments. The student data was divided by the overall score in the course from in-class assessment, and the average amount of tutorials participated in was calculated. Students who did not register for the OWL program were not included in the analysis. Table 7-25 and figure 7-33 display the results, clearly the lower scoring students were not taking advantage of the tutorials as the higher scoring students. Even the highest scoring group did not participate in all of the tutorials. This is a reasonable finding, if they were able to learn the material through other methods and could adequately answer the assigned questions, they would have no need to use the tutorials. This was a re-assuring find in that the method of analyzing the effect of OHW completion on student performance was not simply missing a group of lower achieving students.

Course Score	Percent Student Participation in Tutorials	Students
0	20	4
30	25	16
40	33	27
50	37	46
60	46	69
70	56	89
80	63	39
90	67	3

Table 7-25. Tutorial participation. Number of students and the percentage of tutorials completed for each level of performance in the course.

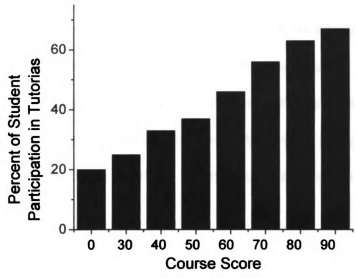


Figure 7-33. Tutorial participation. The average amount of tutorials completed for each performance level in the course.

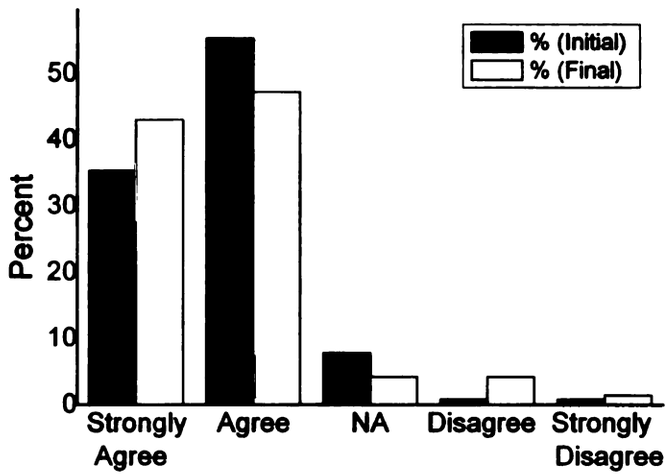
V. Surveys

Actual student performance was a large part of our study, but we were also concerned with the students' perception of the homework program. Realizing that students are much more likely to participate in a program which they feel is practical and beneficial, this became a secondary focus of the study. The students were asked survey questions at the beginning and end of the course. The questions generally referred to OHW, with the students initially answering the questions with previous courses in mind (like general chemistry), but by the end of the semester these opinions would include organic chemistry as well. The students were also given the opportunity to answer the questions in a free-response form, which gives even more insight into the students opinions.

Survey Question One. Question one was regarding the students overall opinion of OHW and how it helped students learn the material. The initial vs. final results showed

little change, shown in figure 7-34 along with example comments from the students. Specific comments regarding the questions varied from students who greatly appreciated the program, those that felt it was a nice change from a teaching style they didn't appreciate, to those who disagreed with specific aspects of the administration. Given that this course showed an increased correlation between homework completion and course performance, it was an interesting result that the students didn't feel this difference.

Question 1: I feel that OHW problems help me to learn the course material and achieved higher scores in my chemistry classes.



"I found the logical, methodical presentation of concepts and exercises a profound relief after the spasmodic, disjointed lectures."
"The OWL questions were really helpful. Even if you didn't get them right you learned something."
"If it weren't for OWL, I do not think I would have passed this class."
"Questions failed to give feedback unless the question was correct or all the attempts were used up."

Figure 7-34. Survey question results regarding the ability to score higher in chemistry courses by completing OHW.

Survey Question 2. Question 2 was similar in nature, asking the students if they would complete OHW if it did not affect their grade, figure 7-35. Surprisingly, the students again did not change their response in any significant amount between the beginning and end of the semesters. Even after seeing the correlations, and for most of the students personally experiencing the effects, they would not be any more likely to complete the OHW. Both questions 1 and 2 indicate that this program was not successful in the sense of motivating students to complete the problems. Since this is a critical aspect of OHW, future studies should consider what changes need to be made to improve the student impression of the homework.

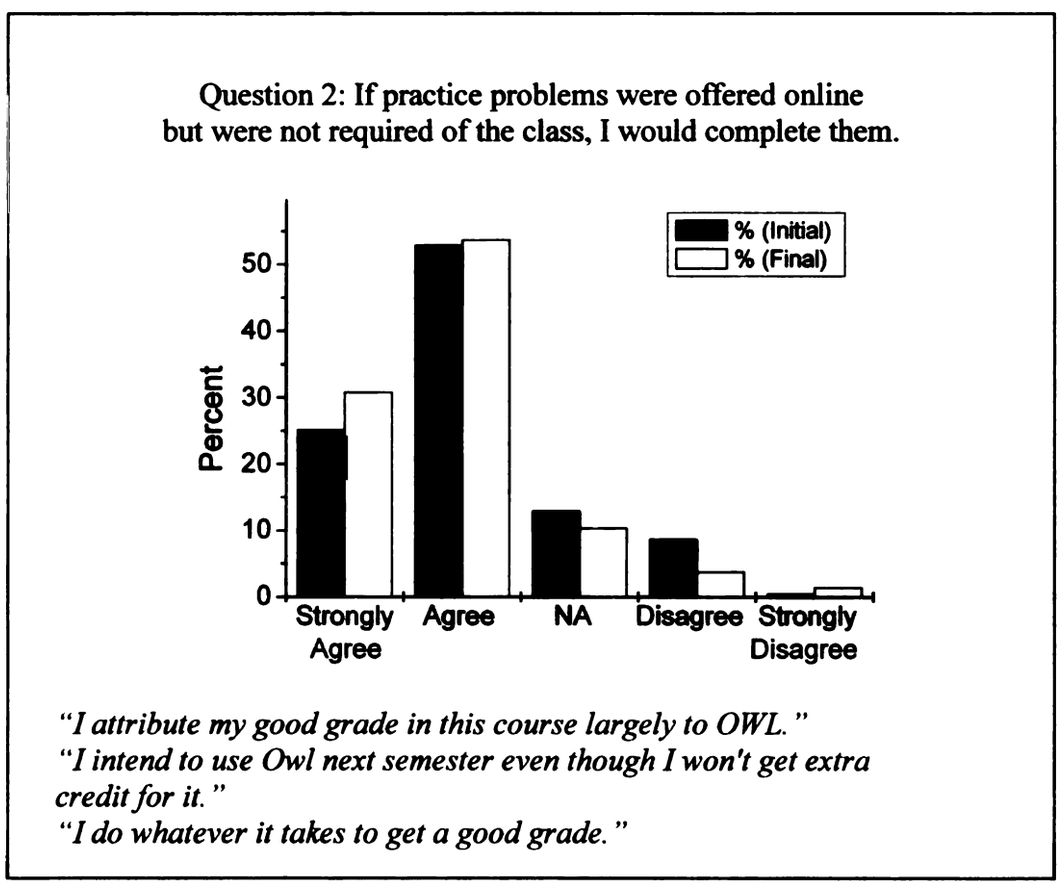


Figure 7-35. Results of the survey question regarding the students likelihood to complete OHW questions in the future.

Survey Question Three. The third question of the survey was regarding the students comfort level with taking computerized tests, figure 7-36. Most of the students in this course are pre-professional students that will be taking a standardized test in preparation for graduate school, most likely administered on a computer. Not the top priority of the course, but a beneficial side effect of OHW could be to prepare students for this type of testing. Initially, the students did not seem to feel ready for this type of testing, even after completing OHW in general chemistry. At the end of the semester the students overall felt an increased level of comfort. The free response questions also agreed with this result, with many students commenting on an increased level of confidence. This point could be used to motivate the students to complete the homework.

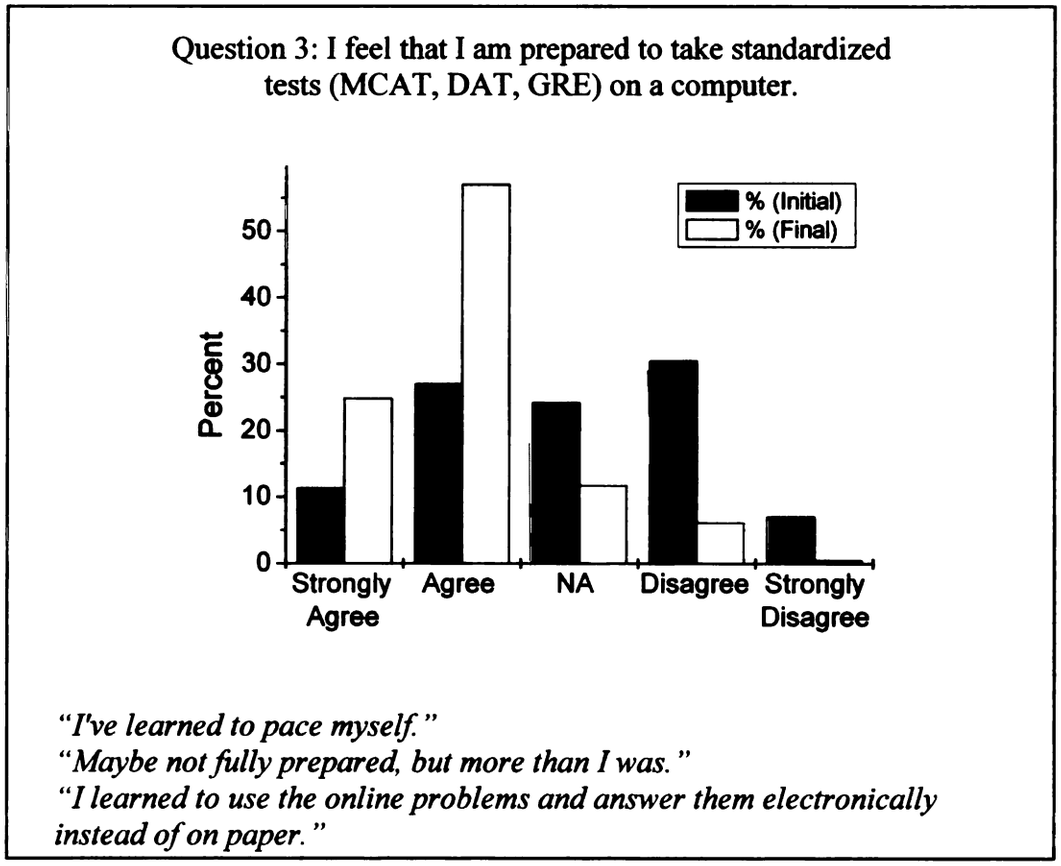


Figure 7-36. Survey question results regarding the students comfort level with completing chemistry problems on a computer.

Survey Question Four. OHW should be able to help students practice material learned in lecture, and optimally would help students who were not able to grasp the concepts in class by leading them through the proper thought process. Question 4, figure 7-37, targeted this. Any instructor knows that not all students will understand the material fully as it is presented in lecture, and the students should have resources outside of class to solidify the material, or explain the concepts in a different way that may make more sense to them. Through the use of tutorials related to the concepts and questions that build on themselves, this type of program has the ability to fulfill this goal as well.

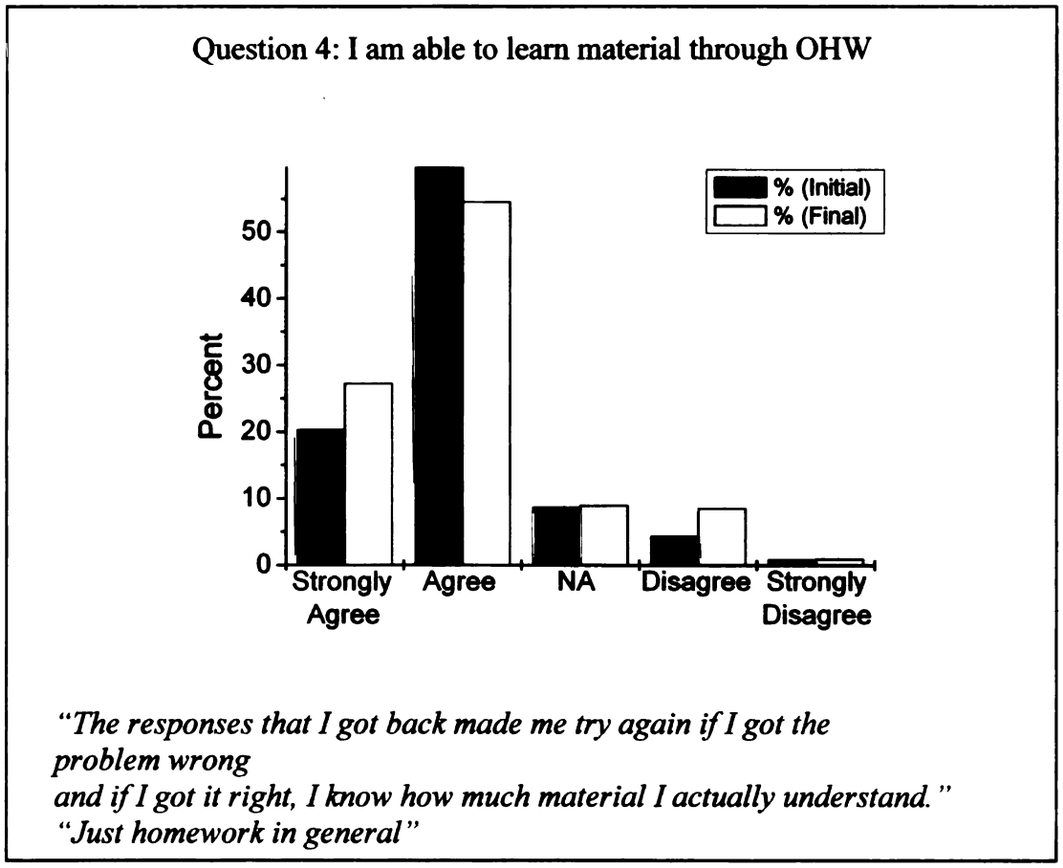
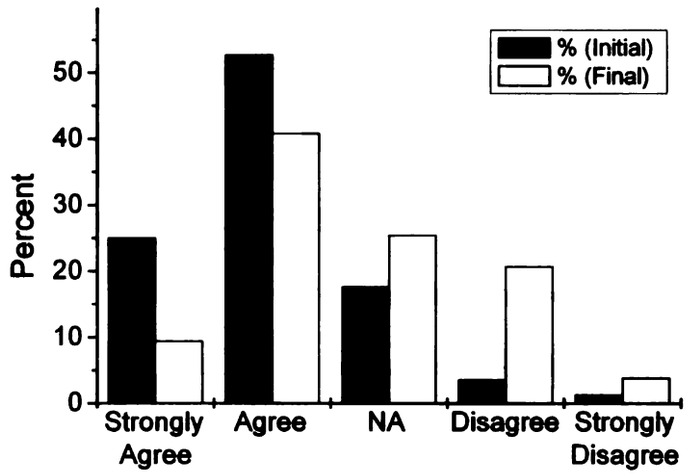


Figure 7-37. Results of the survey question regarding the students ability to learn material by working through OHW problems.

Survey Question Five. This fifth question of the survey was more regarding this specific course than OHW in general, by asking if the homework prepared them for the specific material that they would see on the quizzes and exams, figure 7-38. Students learn best and feel most comfortable when the goals and objectives of the lessons are clearly outlined. Most instructors will agree that homework questions should not simply be repeated on in-class evaluations, but concepts covered should be familiar. Students need to feel that completing the homework will be a benefit to them. Interestingly, the students felt that this applied less after the semester than before. The conclusion to be drawn from this results is that the students did not feel that the homework was consistent with the quiz and exams questions. This sort of opinion can affect the students overall appreciation of the homework programs, and in future courses will be addressed.

Question 5: OHW has prepared me for taking course quizzes and exams by allowing me to practice the specific material that will be covered and the form that the questions will take.



"The homework was a lot like the quizzes, and if I was confused I would go on OWL to see how it was done."
"I am usually always on my computer and by having online problems it was easier for me to do them rather than by reading the text and doing the text problems."

Figure 7-38. Survey results regarding the question about how prepared students felt for quizzes and exams after completing OHW.

VI. Conclusions

This study overall shows a very strong correlation between completing online homework and performance in the class. This is can been seen by comparing overall quiz/exam scores and homework completion, as well as specific topics covered on the in class evaluation. A potential explanation for these correlations could simply be that good students are highly motivated and will do homework, and that the same good students score well on exams, but that the two are unrelated. This means that we would actually be observing the effect of student motivation vs. performance. In order to address this concern, we were able to look further into our data in an attempt to really get at the effect of homework completion, or how much the "good student" affect.

The most convincing piece of evidence for us was shown in figure 7.18, which correlated the change in homework completion and performance on exams 1 and 2. We observed that the students who completed less homework scored lower on exams, and those completing more scored higher. This data should not simply be the "good student" effect, as the good students would be scoring well on exams irregardless of homework completion. This comparison was made between individual students habits and performance, and not simply the overall trend of the class.

The potential that the observed results are influenced, at least to some level, by the "good student" affect is something that we are conscious of and aim to investigate further in future studies. Our initial attempt at addressing this by comparing changes in homework completion and exam scores indicates that while this effect may be contributing to our results, it is no the sole reason for the observed correlations.

Despite the reason for the observed results, they do demonstrate the need to thoughtfully plan an OHW system in a large lecture class. Many instructors want to assume that the students exercise the best intentions, and that they will make proper use of all available resources. Our report indicates that is not always the case. Some students often resort to the path of least resistance, despite the consequence of gaining only a marginal understanding of the material. The best way to implement an OHW program is to anticipate places were students might take advantage of the system, such as a chat room feature or unlimited answer attempts. In certain situations these features may work well, but instructors need to seriously consider how the average student is going to use the program.

The instructors received much resistance from the students towards the new program. Students openly critiqued the new rules for OHW, and frequently requested that the system parameters be changed to those they had experienced in previous semesters using the LON-CAPA system at MSU. Most of the requests regarded features that could be easily exploited, such as increasing the number of attempts per question from three to unlimited, or including a chat room component where answers could be cherry-picked. The conscientious selection of the parameters chosen by the instructors precluded attempts for the students to passively answer all of the homework questions correctly.

Pre- and post-course surveys indicated that the students felt that OHW prepared them for exams equally well (or equally poorly) in both General and Organic Chemistry, although our results clearly contradict this perception. A clear inclination is that once students become accustomed to one system of homework administration, it is difficult for them to embrace another, similar system in which the parameters have been altered to

lead them towards comprehension and mastery of the concepts. The results described in this chapter disclose a simple solution to these shortcomings. Limiting the number of attempts will encourage the students to think through questions more thoroughly before submitting answers, and omission or curtailing of an online collaborative learning feature will instead direct students more to their textbook or class notes. By taking the time to thoughtfully plan the implementation of the features included in a web-based homework program, one can design a system that will lessen the contact hours required by the instructors and will likely enhance the independent cognitive skills required in the learning experience.

VII. Future Work

While this study gave great insight into the use of OHW in large lecture classes, it also introduced many new questions. The design of this study was to pick a set of conditions, maintain constant use of OHW throughout the semester that was the same for all students, and compare to the use of homework in previous courses. We observed that by changing a set of variables, an increase in correlation between completion and performance occurred. Future studies should look at each of these variables specifically, including the use of collaborative learning, feedback, and number of attempts allowed for each question. As a whole, these affected the outcome but we gained no insight into each specifically.

As the introduction of this chapter outlined, there are many opinions and various levels of use for each of these features. The use of each specific feature should be studied in detail. For example, we chose to lower the number of allowed attempts from 10 (used

in general chemistry) to 3, but it is not yet clear if this is the optimal value or if the student would learn more from having only one "second chance" or if an increase to 5 would help learning. We did not offer any type of feedback (at least until the maximum number of attempts had been used) or collaborative learning, but as also addressed in the introduction, there are methods that introduce these features in a limited way that could benefit students.

Any future studies on this subject will require careful thought into the use of a relevant control and the ethical issues that arise when working with human subjects. With this particular study, we were able to use the same group of students as our control, we simply wanted to compare the results of using OHW in general chemistry and organic chemistry. We felt comfortable conducting the study in this manner because we felt that the changes being made to the homework system were in the students' best interest and that a positive outcome would occur.

Future studies may address issues such as the number of attempts allowed per question. In order to accurately analyze the results, a second group of students must be included to compare the results. It is difficult to use the results of courses from past semesters, as the instructor and teaching assistants change, a different book may be used, or the concepts may be presented differently in lecture. A more accurate comparison would be within the same semester, but the decision to split the students into two groups raises ethical concerns. We are conducting these studies with the intent that one method of implementation will benefit the students learning more than another, and as instructors we have a duty to our students to provide them all with equal opportunities to learn. Dividing the students into groups randomly or based on decisions of the instructors is not

fair to the students, and they will surely voice opinions once they realize that the whole class is not being treated equal. A safer option is to allow the students to self-select the study group they prefer to be in, but this would introduce a tremendous amount of bias. Asking students to pick whether they prefer to have two attempts to answer a question or five would most definitely not result in equal, random groups of students.

These issues make education research much more difficult than traditional chemistry research. For this reason, many of the studies published in science education journals are not easily approached in a quantitative manner in which control and experimental groups can be built into the study. Many times education students rely on the instructors to make changes to their courses and then draw conclusions based on their experience and perceptions of the past and present systems. This is a great way to develop new teaching methods at the higher education level, but really instigating change will become easier if the research supporting the new ideas are definitive and certain. Instructors that are opposed to change, or simply do not put much thought into the outcomes of their teaching, are not going to be as affected by hearing these qualitative reviews of new teaching methods as they would be if given quantitative data the proves the effectiveness. We work in an environment of qualitative thinkers that rely on hard evidence to make decisions and change their methods. The need to incorporate this sort of evidence in chemical education research, combined with the inherent difficulties of working with human subjects, leaves the design of future research on this topic a difficult task.

The next logical step in the evaluation of how to best implement OHW programs is address the variables identified in this study (number of attempts, use of feedback,

collaborative learning, etc.) all within the same semester. For example, the number of attempts can be analyzed by carefully designing quiz and exam questions that will test the knowledge learned from specific homework problems. In the homework program, the students will be allowed a varying amount of attempts to answer the questions, possibly from two to ten attempts. After the quiz or exam, the correlation observed between performance and OHW completion can be analyzed for each number of attempts specifically. If multiple questions are offered with each level of attempts, it may become clear from each set of correlations which is the most effective, and this value could be used for the remainder of the class. This approach could be used for any of the other aspects discussed.

A second approach to furthering the study would be to not look at the level of homework completion compared to the in-class performance directly, but to instead analyze factors that may indicate how students are using the program. Again using the example of the number of attempts, it would be expected that if students were seriously considering a question, the concepts involved, and the problem solving process required to arrive at the correct answer, that the amount of time spent on that question would be greater than for students just guessing. This should be even more evident for the second attempt at a question. The students may open a question, feel they initially know the correct answer, and submit that answer just as quickly as a student who is guessing. A bigger difference should be seen in how long it takes the students to submit the second answer to the question. Those who are seriously considering the question would stop to re-evaluate after learning that their initial answer was wrong. They may turn to class notes, online tutorials, or even wait until office hours or lecture to ask for advice before

submitting a second answer. The students who are simply guessing would likely submit that next answer very shortly after the first. Some OHW programs, such as OWL, will provide this information to the instructor in for the form of an excel spreadsheet where the data can be easily analyzed for a very large class and conclusions rapidly drawn from the results.

Another option for comparing the indicators of homework effectiveness would be to assess the use of collaborative learning through a chat room type feature by evaluating the students' specific responses to the questions. With parameterized homework problems, the students receive a variety of versions of a specific question. The chat room feature would ideally allow them to post comments regarding problem-solving techniques or conceptual misunderstanding. Unfortunately this can also turn into a place to simply post the answer to questions. A student looking for the easy way out may take note the answers posted on the chat room and use those in attempt to complete their own homework. Fortunately, this doesn't usually work as well when the questions are designed in a way that the students cannot simply input numbers into a provided set of steps. While many students are still willing to put more time and effort into obtaining an answer from their classmates than it would take to simply solve the problem on their own, this makes it more difficult for students to avoid thinking through problems. Questions could be provided which on the surface appear to be very similar, but actually are very different questions. The instructors could evaluate the students individual responses to questions to identify how frequently students are submitting answers to other versions of the question, answers that would be unreasonable to arrive at without referring to the chat room. This would take more time as each students responses would

need to be evaluated separately, but it would give great insight into what proportion of the class is abusing the collaborative learning feature.

A third issue that will need to be addressed in future studies is the use of feedback for incorrectly answered questions. A good place to start this analysis should be for the instructors to begin the course without using any feedback or hints, but to evaluate the options available with the specific program they are using. Some programs tailor the feedback to the specific response submitted by the student, but for others this is pre-determined. For the programs in which the feedback is pre-determined, the instructors could evaluate the initial sets of OHW in regards to the students submission that were incorrect in comparison to what the feedback would have been. For example, if the pre-set feedback for a substitution question were that $^{-}$SH is the nucleophile and therefore be present in the product, but many of the students were instead providing an elimination product for the answer then this feedback would not be affective. The students would take the hint that their product needs to contain a sulfur but not understand why their elimination product was incorrect. A better piece of feedback in this case would be regarding how the decision is made whether a reaction is substitution or elimination. The first step in assuring that the feedback is benefiting the students and not promoting guessing is confirming that the available feedback actually addresses the student misconceptions.

In conclusion, this study provides an excellent starting point for the evaluation of how to best use OHW programs. It initiated a quantitative approach to studying the situation that will hopefully be continued in future work.

VIII. References

(1) Hall, R. W., Butler, L. G., McGuire, S. Y., McGlynn, S. P., Lyon, G. L., Reese, R. L., and Limbach, P. A. (2001) Automated, Web-based, second-chance homework. *Journal of Chemical Education 78*, 1704-1708.

(2) Riffell, S., and Sibley, D. (2005) Using web-based instruction to improve large undergraduate biology courses: An evaluation of a hybrid course format. *Computers & Education 44*, 217-235.

(3) Glaser, R. E., and Poole, M. J. (1999) Organic chemistry online: Building collaborative learning communities through electronic communication tools. *Journal of Chemical Education 76*, 699-703.

(4) Chamala, R. R., Ciochina, R., Grossman, R. B., Finkel, R. A., Kannan, S., and Ramachandran, P. (2006) EPOCH: An organic chemistry homework program that offers response-specific feedback to students. *Journal of Chemical Education 83*, 164-169.

(5) Penn, J. H., Nedeff, V. M., and Gozdzik, G. (2000) Organic chemistry and the Internet: A web-based approach to homework and testing using the WE_LEARN system. *Journal of Chemical Education 77*, 227-231.

(6) Kashy, E., Sherrill, B. M., Tsai, Y., Thaler, D., Weinshank, D., Engelmann, M., and Morrissey, D. J. (1993) Capa - an Integrated Computer-Assisted Personalized Assignment System. *American Journal of Physics 61*, 1124-1130.

(7) Bonham, S. W., Deardorff, D. L., and Beichner, R. J. (2003) Comparison of student performance using web and paper-based homework in college-level physics. *Journal of Research in Science Teaching 40*, 1050-1071.

(8) Cole, R. S., and Todd, J. B. (2003) Effects of web-based multimedia homework with immediate rich feedback on student learning in general chemistry. *Journal of Chemical Education 80*, 1338-1343.

(9) Freasier, B., Collins, G., and Newitt, P. (2003) A web-based interactive homework quiz and tutorial package to motivate undergraduate chemistry students and improve learning. *Journal of Chemical Education 80*, 1344-1347.

(10) Donovan, W. J., and Nakhleh, M. B. (2001) Students' use of Web-based tutorial materials and their understanding of chemistry concepts. *Journal of Chemical Education 78*, 975-980.

(11) Carpenter, S. R., and McMillan, T. (2003) Incorporation of a cooperative learning technique in organic chemistry. *Journal of Chemical Education 80*, 330-332.

(12) Kortemeyer, G. (2006) An analysis of asynchronous OHW discussions in introductory physics courses. *American Journal of Physics 74*, 527-536.

(13) Thoennessen, M., and Harrison, M. J. (1996) Computer-assisted assignments in a large physics class. *Computers & Education 27*, 141-147.

(14) Cheng, K. K., Thacker, B. A., Cardenas, R. L., and Crouch, C. (2004) Using an OHW system enhances students' learning of physics concepts in an introductory physics course. *American Journal of Physics 72*, 1447-1453.

Appendix 1: FHA2 NMR Data Files

^{13}C	^{15}N	Residue	Sample Comp	Temp (°C)	PH[1]	^{13}C Chemical Shift (ppm)	File Name
G	L	1	PC/PG	-50	5	174.7	IFProt/101006GLHA2-2[2]
			PC/PG	-10	5	174.8	GL/102907redor
			Cells	-20	-	175.2	GL/Cells/011508redor
			Insol	-20	-	174.4	GL/IB/011007redor
			Insol	-20	-	174.7	GL/IB/011707redor
			D Insol	-20	-	173.8	GL/IB/030608redor
L	F	2, 119	PC/PG	-10	5	177.7	LF/030708redor
			Cells	-20	-	178.6	LF/Cells/062008redor
G	A	4	PC/PG	-50	5	177.8	AGHA2/020707redor2
			PC/PG	-10	5	177.2	GA/103007redor
			Insol	-20	-	174.5	GA/IB/010407redor
			Cells	-20	-	174.7	GA/cells/020808redorsum
			D Insol	-20	-	175.2	GA/IB/030508redor
A	I	5, 44	PC/PG	-10	5	178.1	AI/022908redor2
A	G	7	PC/PG	-10	5	179.1	AG/100707redor
			Insol	-20	-	177.5	AG/IB/011607redor
			Insol	-20	-	178.9	AG/IB/032508redorsum
			Cells	-20	-	177.9	AG/Cells/030308redor
			D Cells	-20	-	177.9	AG/Cells/030408redor
G	F	8, 23	PC/PG	-10	5	176.3	GF/022708redor
G	M	16	PC/PG	-10	5	177.7	GM/060207/GMREDOR3
			PC/PG	-50	5	176.9	GM/052807/ GMREDORminus50
			Cells	-50	-	174.4	incell/042507Gmredor3
			Cells	-30	-	176.9	GM/Cells/071808redor2
M	I	17	PC/PG	-10	5	178.1	MI/101107redor2
Y	G	22	PC/PG	-10	5	176.5	YG/101607redorsum
			Insol	-20	-	175.2	YG/IB/011107redor2
A	D	36	PC/PG	-10	5	179.3	AD/073007redor2
			Insol	-20	-	178.1	AD/IB/111908redor
V	I	55	PC/PG	-10	5	179.1, 176.8	VI/120107redor
			Insol	-20	-	174.9	VI/IB/021208sum
L	L	98	PC/PG	-10	5	178.8	LL/062707redor2
			PC/PG/Chol	-10	5	178.4	LL/070207redor2

Appendix 1: FHA2 NMR Data Files

			D Cells	-50	-	179.4/176.0	LL/cell/041807redor
			D Soluble	-50	-	178.3	LL/cell/041807redorlysate
			PC/PG Refold	-10	5	178.4	LL/091607redorsum-2
			Insol	-20	-	178.6	LL/IB/010707redor
			D Insol	-20	-	178.0	LL/IB/030608redor
			Insol	-20	-	179	LL/Cells/010707redor
L	V	99	PC/PG	-50	5	178.0	IFProt/091406redor2[2]
			PC/PG	-10	5	178.5	LV/120307redor
			Cells	-30	-	178.8	LV/Cells/053008redorsum
			Cells	-30	-	178.2	LV/Cells/032008
			Insol	-20	-	177.5	LV/IB/031108redorsum
V	A	100	PC/PG	-10	5	178.2	VA/070807redorfinal
			PC/PG/Chol	-10	5	177.6	VA/072606redorsum
			PC/PG	-10	7.4	178.1	VA/072707redor
			PC/PG/Chol	-10	7.4	178.1	VA/072907redor
			Insol	-20	-	173.5, 177.4	VA/IB/020508redor
A	L	102, 167	PC/PG	-10	5	178.4	AL/050508redorsum
M	G	133	PC/PG	-10	5	177.6	MG/102607redor
			Cells	-10	-	177.7	MG/cell/100207redor
G	S	136	PC/PG	-10	5	176.4	GS/110207redor
S	F	137	PC/PG	-10	5	178.4	SF/120507redor
V	Y	161	PC/PG	-10	5	176.6	VY/062807redor2
G	V	175	PC/PG	-10	5	176.8	GV/120707redor
			Insol	-20	-	172.3, 175.4	GV/IB/021408redor3

D=dehydrated, Insol=Insoluble fraction of the cell lysate, Cells=whole bacterial cell sample

Unless noted all files stored in mb4b/data/Jaime/ folder

[1]pH of reconstituted samples

[2]file stored in curtis43/data folder

Appendix 2: Hgp41 NMR Data Files

^{13}C	^{15}N	Residue	Sample Comp	Temp (°C)	PH[1]	^{13}C Chemical Shift (ppm)	File Name
G	L	1	PC/PG	-50	5	174.7	IFProt/101006GLHA2-2[2]
			PC/PG	-10	5	174.8	GL/102907redor
			Cells	-20	-	175.2	GL/Cells/011508redor
			Insol	-20	-	174.4	GL/IB/011007redor
			Insol	-20	-	174.7	GL/IB/011707redor
			D Insol	-20	-	173.8	GL/IB/030608redor
L	F	2, 119	PC/PG	-10	5	177.7	LF/030708redor
			Cells	-20	-	178.6	LF/Cells/062008redor
G	A	4	PC/PG	-50	5	177.8	AGHA2/020707redor2
			PC/PG	-10	5	177.2	GA/103007redor
			Insol	-20	-	174.5	GA/IB/010407redor
			Cells	-20	-	174.7	GA/cells/020808redorsum
			D Insol	-20	-	175.2	GA/IB/030508redor
A	I	5, 44	PC/PG	-10	5	178.1	AI/022908redor2
A	G	7	PC/PG	-10	5	179.1	AG/100707redor
			Insol	-20	-	177.5	AG/IB/011607redor
			Insol	-20	-	178.9	AG/IB/032508redorsum
			Cells	-20	-	177.9	AG/Cells/030308redor
			D Cells	-20	-	177.9	AG/Cells/030408redor
G	F	8, 23	PC/PG	-10	5	176.3	GF/022708redor
G	M	16	PC/PG	-10	5	177.7	GM/060207/GMREDOR3
			PC/PG	-50	5	176.9	GM/052807/ GMREDORminus50
			Cells	-50	-	174.4	incell/042507Gmredor3
			Cells	-30	-	176.9	GM/Cells/071808redor2
M	I	17	PC/PG	-10	5	178.1	MI/101107redor2
Y	G	22	PC/PG	-10	5	176.5	YG/101607redorsum
			Insol	-20	-	175.2	YG/IB/011107redor2
A	D	36	PC/PG	-10	5	179.3	AD/073007redor2
			Insol	-20	-	178.1	AD/IB/111908redor
V	I	55	PC/PG	-10	5	179.1, 176.8	VI/120107redor
			Insol	-20	-	174.9	VI/IB/021208sum
L	L	98	PC/PG	-10	5	178.8	LL/062707redor2
			PC/PG/Chol	-10	5	178.4	LL/070207redor2
			D Cells	-50	-	179.4/176.0	LL/cell/041807redor

Appendix 2: Hgp41 NMR Data Files

			D Soluble	-50	-	178.3	LL/cell/041807redorlysate
			PC/PG Refold	-10	5	178.4	LL/091607redorsum-2
			Insol	-20	-	178.6	LL/IB/010707redor
			D Insol	-20	-	178.0	LL/IB/030608redor
			Insol	-20	-	179	LL/Cells/010707redor
L	V	99	PC/PG	-50	5	178.0	IFProt/091406redor2[2]
			PC/PG	-10	5	178.5	LV/120307redor
			Cells	-30	-	178.8	LV/Cells/053008redorsum
			Cells	-30	-	178.2	LV/Cells/032008
			Insol	-20	-	177.5	LV/IB/031108redorsum
V	A	100	PC/PG	-10	5	178.2	VA/070807redorfinal
			PC/PG/Chol	-10	5	177.6	VA/072606redorsum
			PC/PG	-10	7.4	178.1	VA/072707redor
			PC/PG/Chol	-10	7.4	178.1	VA/072907redor
			Insol	-20	-	173.5, 177.4	VA/IB/020508redor
A	L	102, 167	PC/PG	-10	5	178.4	AL/050508redorsum
M	G	133	PC/PG	-10	5	177.6	MG/102607redor
			Cells	-10	-	177.7	MG/cell/100207redor
G	S	136	PC/PG	-10	5	176.4	GS/110207redor
S	F	137	PC/PG	-10	5	178.4	SF/120507redor
V	Y	161	PC/PG	-10	5	176.6	VY/062807redor2
G	V	175	PC/PG	-10	5	176.8	GV/120707redor
			Insol	-20	-	172.3, 175.4	GV/IB/021408redor3

D=dehydrated, Insol=Insoluble fraction of the cell lysate, Cells=whole bacterial cell sample

Unless noted all files stored in mb4b/data/Jaime/ folder

[1]pH of reconstituted samples

[2]file stored in curtis43/data folder

MICHIGAN STATE UNIVERSITY LIBRARIES
3 1293 03062 7396