Proteins Inside Cells and On the Surface of Membranes:

Developing Three Dimensional In-Cell NMR and Structural Characterization of a Trimeric Membrane Proximal External Region Construct from HIV-1 gp41

by

Patrick Neal Reardon

Department of Biochemistry
Duke University

Date:_______________________

Approved:

_________________________________
Leonard D. Spicer, Supervisor

_________________________________
David C. Richardson

_________________________________
Pei Zhou

_________________________________
Michael C. Fitzgerald

Dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Biochemistry in the Graduate School of Duke University

2011
ABSTRACT

Proteins Inside Cells and On the Surface of Membranes:

Developing Three Dimensional In-Cell NMR and Structural Characterization of a

Trimeric Membrane Proximal External Region Construct from HIV-1 gp41

by

Patrick Neal Reardon

Department of Biochemistry
Duke University

Date: ______________________

Approved:

____________________________________
Leonard D. Spicer, Supervisor

____________________________________
David C. Richardson

____________________________________
Pei Zhou

____________________________________
Michael C. Fitzgerald

An abstract of a dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Biochemistry in the Graduate School of Duke University

2011
Abstract

Our ability to understand complex biological processes is enhanced by studying the components of those processes at atomic resolution. Major advancements in technology and methodology have been made that continue to improve our ability to determine atomic resolution structures of proteins. These advancements have also enabled new ways to study difficult protein systems. In this dissertation, I will discuss our efforts to apply NMR spectroscopy and other biophysical techniques to better understand two problems. The first problem, and the focus of the first part of this work, is to understand the behavior of proteins in the complex milieu of a living cell. We have developed 3D In-Cell protein NMR and demonstrated its use to assign the backbone resonances of a protein in living E. coli cells. I will also discuss our application of In-Cell NMR to show that the methionine repressor, MetJ, is generally associated non-specifically with DNA inside living cells. In the second part of this dissertation I will discuss our efforts to better understand the biophysical behavior of the membrane proximal external region (MPER) of the HIV-1 envelope protein, gp41. This region is membrane associated and an important target for HIV-1 vaccine development because it contains the epitopes for several of the broadly neutralizing antibodies against HIV-1. In addition to studies of the monomeric peptide, we have developed a new trimeric
MPER construct, designated gp41-M-MAT, that associates with detergent micelles and lipid bilayers. This construct is a stable trimer, which binds the broadly neutralizing antibodies 2F5 and 4E10, as shown by equilibrium analytical ultracentrifugation and surface plasmon resonance. Finally, we have solved the structure of this construct bound to detergent micelles using NMR spectroscopy. The structure shows that the MPER adopts an α-helix conformation and suggests that each helix in the symmetric trimer associates directly with the surface of detergents and lipid membranes. This association suggests that the MPER might contribute to the mechanism of viral fusion by inducing strain in the viral membrane.
Contents

Abstract ................................................................................................................................. iv

Contents............................................................................................................................... vi

List of Tables ........................................................................................................................... ix

List of Figures ........................................................................................................................ x

Acknowledgements ............................................................................................................. xiii

1. Multi-Dimensional NMR in Living Cells ........................................................................ 1
   1.1 Experimental Procedures .......................................................................................... 6
   1.2 Multi-Dimensional In-Cell NMR on GB-1 ................................................................. 9
       1.2.1 GB-1 Expression and Initial 2D HSQC Spectrum ............................................. 11
       1.2.2 Determining Time Limits for Viable E. coli Cells in a Tube ......................... 12
       1.2.3 Projection Reconstruction NMR for Backbone Assignments .................... 17
   1.3 Application of In-Cell NMR to a Biological Problem ...................................... 20
   1.4 In-Cell NMR on the Large Protein HCA-II, and Unexpected Results ............ 25
   1.5 Conclusions and Future Directions ..................................................................... 29

2. A brief Introduction to HIV-1 and Broadly Neutralizing Antibodies .................. 33

3. Studies on SP62, a GP-41 Derived Peptide Containing the 2F5 Epitope ........... 46
   3.1 Experimental Procedures ....................................................................................... 47
   3.2 Liposome Bound SP62 ........................................................................................... 51
   3.3 NMR of SP62NK in DHPC micelles ..................................................................... 60
4. Biophysical Characterization of New MPER Constructs .................................................. 69
   4.1 Experimental Procedures .......................................................................................... 72
   4.2 Constructs .............................................................................................................. 79
   4.3 Analytical Ultracentrifugation ............................................................................... 81
      4.3.1 Trimerization of MPER Peptides in Various Detergents ............................... 82
      4.3.2 Determination of Partial Specific Volume of Detergents .............................. 90
   4.4 Screening for Feasibility of Structural Analysis by NMR ................................. 94
   4.5 Measurement of 2F5 and 4E10 Binding by SPR ............................................. 102
   4.6 Summary .............................................................................................................. 104
5. Structure of GP41-M-MAT ......................................................................................... 105
   5.1 Experimental Procedures ...................................................................................... 109
   5.2 NMR Structure of gp41-M-MAT .......................................................................... 115
   5.3 Binding of Non-Neutralizing Antibodies to gp41-M-MAT ................................. 129
   5.4 Comments on the Acquisition of RDC’s ............................................................ 131
      5.4.1 Strained Polyacrylamide Gels ....................................................................... 132
      5.4.2 Guanosine Dinucleotide Alignment ............................................................... 135
      5.4.3 Collagen Gel Alignment ............................................................................... 136
      5.4.4 Charged Strained Polyacrylamide Gels ....................................................... 137
      5.4.5 Conclusions .................................................................................................. 138
   5.5 Implications of gp41-M-MAT Structure .............................................................. 139
Appendix A....................................................................................................................... 148
References ........................................................................................................................................... 151

Biography ........................................................................................................................................... 169
List of Tables

Table 1. Complex Molecular Weights of 30 μM 2R2 Peptide in Various Detergents......84
Table 2. gp41-M-MAT Complex Molecular Weight in DPC Detergent. .........................86
Table 3. Partial Specific Volumes of Several Detergents. ........................................92
Table 4. Summary of Structural Statistics and Restraintsα ........................................117
Table 5. Correlation and Rhombicity for RDC’s fit to the Foldon Domain....................132
Table 6. Modified M9 Minimal Media Composition ..................................................148
Table 7. Load/Wash Buffer ....................................................................................149
Table 8. Elution Buffer ..........................................................................................149
Table 9. Size Exclusion Buffer ............................................................................149
Table 10. 16.5% Tris-Glycine Polyacrylamide Gel ..............................................150
Table 11. Tris-Tricine Polyacrylamide Gel ............................................................150
Table 12. Tris-Glycine Running Buffer .................................................................150
Table 13. Tris-Tricine Running Buffer ................................................................150
List of Figures

Figure 1. Schematic of Conventional vs Reduced Dimensionality data acquisition. .......4

Figure 2. Cartoon of GB-1. ...........................................................................................................9

Figure 3. SDS-PAGE Gel of GB-1 Overexpression. .................................................................10

Figure 4. \(^{15}\)N-HSQC Spectra of GB-1. ..................................................................................11

Figure 5. Schematic of Cell Viability Assay..................................................................................13

Figure 6. *E. coli* Cell Viability in an NMR Tube. ...........................................................................14

Figure 7. Settling of *E. coli* Cells in an NMR Tube.....................................................................15

Figure 8. Comparison of \(^{15}\)N-HSQC Spectrum of GB-1 with and without IgG....................16

Figure 9. Strip Plot of an In-Cell 3D-HNCA Spectrum.................................................................17

Figure 10. 3D Representation of a 3D-HNCO In-Cell NMR Spectrum...........................................19

Figure 11. Comparison of MetJ In-Cell NMR Spectrum and Spectrum of Un-Induced Cells.................................................................................................................................21

Figure 12. Recovery of the MetJ Spectrum from non-specific DNA..............................................23

Figure 13. Recovery of the MetJ Spectrum from Cell Lysate...........................................................24

Figure 14. Comparison of HSQC spectra of HCAII In-Cell, in lysate, and in solution...27

Figure 15. Schematic of HIV-1 Virus Particle.................................................................................34

Figure 16. Schematic of HIV-1 gp41. ............................................................................................36

Figure 17. Model of HIV-1 Membrane Fusion.............................................................................37

Figure 18. Structure of gp41 Core Six-Helix Bundle. ....................................................................38

Figure 19. Model of Peptide Inhibited Prefusion Intermediate....................................................39
Figure 20. Sequence Logo of MPER Region of HIV-1 gp41. ......................................................... 41
Figure 21. NMR Spectra of Liposomes and SP62-GTH1 Peptides. .................................................... 52
Figure 22. Effect of SP62-GTH1 on Choline Head Group of POPC Liposomes. ................................. 53
Figure 23. $^{15}$N-HSQC Spectra of SP62-GTH1 With and With Out Liposomes. ................................. 54
Figure 24. 3D-NOESYHSQC Comparison of free and Liposome bound SP62-GTH1. ....................... 56
Figure 25. Comparison of Methods to Prepare Liposome-Peptide Conjugates .................................. 57
Figure 26. SPR of SP62-GTH1-liposome conjugate prepared using Method 1. ................................. 58
Figure 27. Assigned $^{15}$N-HSQC Spectrum of SP62-NK in DHPC Detergent. ................................. 61
Figure 28. Strip Plot from an HNCA on SP62-NK in DHPC Detergent. ............................................. 62
Figure 29. Chemical Shift Indexing and NOE analysis of SP62-NK. .................................................... 63
Figure 30. pH Dependence of $^{15}$N-HSQC Spectra of SP62-NK in DHPC Detergent ..................... 65
Figure 31. Schematic of MPER Constructs. ......................................................................................... 80
Figure 32. Structures of Trimerization Domains Used. ....................................................................... 81
Figure 33. Trimerization Depends on Detergent Concentration. ....................................................... 86
Figure 34. Total Mass of Detergent Bound to gp41-M-MAT. ............................................................... 89
Figure 35. Partial Specific Volumes from AUC. .................................................................................... 91
Figure 36. Partial Specific Volumes Determined by Densimetry. ....................................................... 93
Figure 37. HSQC Spectra of 2R2 in Various Detergents. .................................................................... 95
Figure 38. HSQC spectrum of 2R2 in Low LMPC Concentration. ...................................................... 96
Figure 39. Sample Spectra for Assignment of 2R2 in LMPG. ............................................................. 97
Figure 40. Chemical Shift Indexing of 2R2 in LMPG. ......................................................................... 98
Figure 41. Trosy HSQC Spectrum of FR2 ................................................................. 100
Figure 42. TROSY Spectrum of gp41-M-MAT in DPC Micelles.............................. 101
Figure 43. SPR sensograms of 2F5 and 4E10 binding M-MAT-liposomes............... 103
Figure 44. Three Cryo-EM Structures of ENV. ....................................................... 106
Figure 45. Structures of MPER Not in a Six-Helix Bundle .................................... 107
Figure 46. Selected Spectra of gp41-M-MAT ......................................................... 118
Figure 47. Structural Overview of gp41-M-MAT .................................................... 119
Figure 48. Overlay of Monomeric MPER Peptide and gp41-M-MAT ..................... 120
Figure 49. Structure of the Epitopes for 2F5 and 4E10 ......................................... 122
Figure 50. 16-Doxyl Stearic Acid Mediated Paramagnetic Relaxation .................. 125
Figure 51. Model of gp41-M-MAT with DPC Micelle ............................................. 126
Figure 52. gp41-M-MAT Modeled into Cryo-EM Volumes ..................................... 128
Figure 53. Interaction of Non-Neutralizing Antibodies with gp41-M-MAT............. 130
Figure 54. Schematic of Membrane Fusion with gp41-M-MAT Inspired Intermediate. 141
Figure 55. Model of Membrane Fusion with Multiple ENV's. ............................. 142
Figure 56. Comparison of Antibody Bound MPER and gp41-M-MAT ................. 144
Acknowledgements

Scientific research of any kind rarely occurs as a purely individual effort. I would like to thank those who have contributed their time and expertise to the successful completion of the projects discussed in this work. Most of the NMR spectroscopy was done at the Duke NMR Center with the help of Ron Venters, Tony Ribeiro, and Don Mika. Additional NMR experiments were done at the David H. Murdoch Research Institute NMR facility with the help of Kevin Knagge. I am grateful to them for the training I received operating the NMR instruments and for the valuable discussions that helped guide the evolution of these projects.

The Research conducted on HIV-1 was done in collaboration with the Duke Human Vaccine Institute. I would like to thank our collaborators at the Human Vaccine Institute for the contributions to this project, especially S. Moses Dennison, S. Munir Alam, and Barton F. Haynes, who were instrumental in acquiring the surface plasmon resonance data on our constructs. This work was funded as part of the Collaboration for Aids Vaccine Discovery from the Bill and Melinda Gates Foundation.

Much of this work would not have been possible without the help of my friends and colleagues in the Duke Biochemistry Department. The fast NMR methods were developed by Brian Coggins, Ron Venters, and Pei Zhou, who were kind enough to help
me use these methods as they were being developed. I’d also like to thank Vincent Chen and Gary Kapral for help using King to visualize and analyze our structures, as well as helpful discussions about these projects. I also thank my current and former colleagues in the Spicer group, Anne Marie Augustus, Kirsten Robinson, and Karel Base, with whom I collaborated on various aspects of these projects. I’d like to thank my advisor Leonard Spicer for his support and advice throughout my time at Duke. His helpful suggestions and ideas have greatly contributed to the successes achieved in the work that is discussed in this dissertation. Finally, I thank my wife, Carrie Marean-Reardon, for her support and help throughout our time at Duke.
1. Multi-Dimensional NMR in Living Cells

The chemical and physical environment within cells is crowded and complex [1]. It has been estimated that even the simplest free living organisms utilize hundreds of small molecules and genetically encode up to 1000 different proteins in a single cell. Higher animals and humans encode 10 to 100 times more proteins than these simple organisms [2,3]. This environment likely influences the chemistry of the cell and the behavior of biological macromolecules. Understanding the influence of the numerous large and small molecular components on the stability and behavior of individual proteins and the biological processes in which they participate is one of the major challenges of contemporary biomedical science. The term systems biology is often used to describe the network of potential interactions affecting the behavior and function of individual molecules or complexes, as opposed to more traditional approaches that focus on isolated macromolecules or complexes.

Traditional structural biology techniques, such as NMR and X-Ray crystallography, are done under strictly controlled in vitro conditions that are often very different from the crowded and complex environment inside living cells. However, NMR is a non-invasive solution technique that has the potential to study biological macromolecules inside living cells. Indeed, there has been growing interest in detection of specific heteronuclear fingerprint spectra of individual proteins in cells and cell lysates [4,5,6,7,8]. To observe these spectra, the target protein is differentiated by being
over-expressed and labeled with the stable isotopes $^{15}$N or $^{13}$C. The initial reports that high resolution 2D HSQC (Heteronuclear Single Quantum Correlation) fingerprint spectra could be observed well over background for an over-expressed $^{15}$N labeled protein in a processed cell lysate without further purification [9] demonstrated that protein over-expression and stable isotope labeling could be used to selectively probe a target protein in a complex mixture, including other macromolecular components, by virtue of its higher concentration. This concept was later extended to include proteins in living cells when the first in-cell 2D HSQC spectra were reported in 2001 [7].

Further 2D HSQC spectra acquired in-cell have been published for a few proteins [4,5,6,7,8,10] and an initial evaluation of some of the experimental parameters which contribute to success in these experiments has been reported [11]. Based on comparisons with in vitro HSQC spectra of purified proteins, these investigators have interpreted their results in terms of conformational variations and structural differences attributed to the cellular environment [4,5,6,10]. In addition, evidence for in vivo binding interactions with externally delivered metal ions and small binding ligands has been observed [5].

Multi-dimensional heteronuclear experiments are the cornerstone of modern applications of NMR spectroscopy with purified biological macromolecules in vitro. They are essential for de novo assignment of resonances of uncharacterized proteins and proteins whose spectra change significantly due to structural rearrangements or binding
interactions. When run conventionally, 2D experiments can be completed in minutes to hours. In contrast, conventional 3D experiments can take days to acquire even with ≥ 1mM concentrations due to the data collection method and limited probe sensitivity. While these acquisition times are acceptable for typical solution NMR samples of purified proteins, living cells are unlikely to survive under these conditions in a 5mm NMR tube for the duration of such an experiment.

Two new technologies provide the means to overcome this hurdle. The development of cryogenically cooled probes has increased the signal to noise 2 to 4 times that of conventional room temperature designs. This is achieved by cooling the RF coils to 15-30 K, which results in reduced resistance in the coils, lower thermal noise, and increased probe quality factor. These probes can extend the range of protein concentrations that can be detected with high field NMR down to levels of 100 μM or less, or reduce the acquisition time by the sensitivity gain factor squared. Thus, cryo-probes have already impacted NMR studies by extending the lower limits of the concentration range needed for reliable spectral acquisition in case of limited sample availability or stability.

The second important technology is the development of fast acquisition methods for multidimensional NMR data collection. The data for a conventional 3D experiment is collected by setting one indirect dimension, $t_1$, constant and incrementing across the other, $t_2$ (see Figure 1), then $t_1$ is incremented once and the process is repeated, thus each
t₁ and t₂ combination is sampled independently. This yields a standard sampling pattern where all data points collected completely fill out a Cartesian grid.

The resulting 3D data set is Fourier transformed in three dimensions to yield the 3D frequency spectrum. The resolution in an indirect dimension is determined by the number of increments collected in that dimension and the sampling theorem \[12\].

Conventional 3D NMR data must be sampled at every t₁ and t₂ combination, and increasing the resolution requires increasing the number of t₁ and t₂ points that must be sampled. Thus, 3D and 4D experiments are often collected at relatively low resolutions in order to reduce the time required to acquire the data. In contrast, fast NMR methods use various data collection methods that result in greatly reduced time while

![Figure 1. Schematic of Conventional vs Reduced Dimensionality data acquisition](image-url)

In the schematics, F3 is the directly detected dimension. In the conventional experiment (A), data are collected by setting the F₂ dimension to a constant and incrementing across the F₃ dimension. Projection Reconstruction (B) has two orthogonal planes, F₁,F₃ and F₁,F₂ with radial projections from the origin formed by the intersection of the two orthogonals. The angle α is defined by the rate of incrementation of the indirect evolution times.
maintaining high resolution [13]. For in-cell NMR experiments, we used Projection Reconstruction (PR) techniques to generate 3D spectra from radially collected 2D planes, where the indirect dimensions are acquired simultaneously by incrementing the $t_1$ and $t_2$ evolution times together [14]. In addition, orthogonal planes are collected where one indirect evolution time is set to the initial evolution time and the other is acquired conventionally. Fourier transformation of the 2D time domain radial planes results in a projection of the 3D spectrum onto the 2D frequency domain plane. These projections can either be mathematically analyzed as in GFT [15] to directly obtain chemical shift information, or used to reconstruct the 3D spectrum as is the case in PR NMR [14,16]. The time required to acquire a PR experiment is much less than is required for a conventional experiment, since much less data is required to accurately represent the 3D spectrum with radial planes. High-resolution 3D and 4D PR experiments useful for characterization of proteins have been reported and applied to a range of proteins with masses up to 30 kDa [17]. The time required for data collection in these studies was reduced to about 1/8 the time for conventional 3D Fourier transform methods.

In the first section of this chapter I will present our studies on the protein GB-1. GB-1 is a commonly used model protein for NMR spectroscopy because of its small size, ease of production in *E. coli*, and good behavior in solution. In the last two sections of this chapter I will present our studies on two additional proteins, MetJ and human carbonic anhydrase II. Our studies on MetJ used NMR to better understand the
behavior of the protein in cells and its mechanism of function. Finally, we used human carbonic anhydrase II to test the feasibility of studying a large protein with in-cell NMR. Instead we rediscovered that human carbonic anhydrase II is more than just a model protein.

1.1 Experimental Procedures

The GB-1 expression vector was transformed into BL21(DE3) E. coli and transformants were selected by plating on Luria Broth (LB) agar plates, supplemented with 100 μg/ml ampicillin. An overnight culture was started in LB containing 100 μg/mL ampicillin (LB-amp) from a single colony off of the LB plate. After overnight growth, a glycerol stock was prepared from the overnight culture by adding 200 μL of 80% glycerol to 800 μL of the overnight culture, followed by flash freezing in dry ice/ethanol, and storage at -80°C. All subsequent overnight cultures were started from this glycerol stock.

GB-1 was expressed by growing an overnight culture in LB-amp. The following morning, 2x50 ml of modified M9 minimal media (see appendix A) were inoculated to an OD_{600} of approximately 0.05 from the overnight culture. The cultures were grown at 37°C until they reached an OD_{600} of ~ 0.5. The cultures were harvested by gentle centrifugation at 1000 x g for 10 minutes, then resuspended in fresh modified M9 minimal media containing 1 g/L ¹⁵N ammonium chloride and either 2 g/L ¹³C-glucose or 2 g/L ¹²C-glucose. Cells were allowed to grow for an additional 10 minutes, then
induced with Isopropyl β-D-1-thiogalactopyranoside (IPTG). One culture was harvested four hours after induction by centrifugation at 1000xg for 20 minutes. The cell pellet was resuspended in modified M9 minimal media supplemented with 10% D₂O at approximately 20% v/v, and loaded into a 5mm NMR tube. This culture was used to determine NMR experimental parameters, such as proton pulse width and initial shims. After the parameters were determined, the second culture was harvested in the same manner, and the NMR experiment started immediately using the experimental parameters determined on the previous sample.

MetJ samples for in-cell NMR were expressed using the same protocol as for GB-1. Since overexpression of MetJ results in repression of methionine biosynthesis, all cultures where MetJ was expressed were supplemented with 1 mM methionine. For whole cell lysates containing MetJ, 50 mL of cells induced using the above protocol were harvested and lysed in 4 mL/g Bugbuster (Pierce), supplemented with 50 μg/mL lysozyme, 25 μg/mL DNase, 5mM MgSO₄, 1mM dithiothreitol DTT, 1mM phenylmethylsulfonyl fluoride (PMSF), and 10% D₂O. For experiments that required unlabeled cell lysates, cells were grown overnight in LB and lysed in the same manner.

Human Carbonic Anhydrase II samples were prepared similarly to GB-1 in-cell NMR samples. Lysates of the protein were prepared using the same protocol as was used for MetJ lysate samples.
SDS-PAGE analysis was carried out using either a Bio-Rad Protean II or Protean III mini gel apparatus. Gels were cast in the provided gel caster with the compositions listed in Appendix A. Either Tris-Glycine or Tris-Tricine gels were used, depending on the molecular weight range desired. Tris-Tricine gels yield better separation of lower molecular weight proteins and peptides. After electrophoresis, gels were washed three times for 5 minutes in H₂O, then stained with Gelcode Blue (Pierce).

Cell viability was determined using a modified micro drop technique [18]. The assay was performed by making three 1:100 serial dilutions followed by three 1:10 serial dilutions into fresh LB media. Three 10 μL drops from each of the four most dilute samples and two 10 μL drops from the other samples were placed on a sectored LB-amp agar plate as shown in Figure 5. The plate was incubated at 37°C overnight. The number of colonies in each sector was used to back calculate the total number of colony forming units per mL.

NMR spectra were collected on either a Varian 600 MHz Innova spectrometer equipped with an H,C,N cryogenically cooled triple resonance probe, or a Varian 800 MHz spectrometer equipped with a room temperature H,C,N triple resonance probe. All spectra were processed with nmrPipe and visualized with nmrview [19,20]. Projection reconstruction data were generally collected using 5 tilt angles and two orthogonals, for a total of 12 radial planes. Projection reconstruction data were further
processed using a suite of programs developed by Dr. Brian Coggins and Dr. Pei Zhou at Duke University [14].

1.2 Multi-Dimensional In-Cell NMR on GB-1

Protein G is a multi-domain protein that is found on the surface of streptococcus bacteria. The protein helps the bacteria hide from the host immune system by tightly binding the constant Fc domains of immunoglobulin G (IgG) [21]. This activity is mediated by two to three similar IgG binding domains. One of these domains is the B-1 domain which we denote as GB-1. GB-1 is a 56 amino acid protein that is easily expressed to very high levels in E. coli cells. GB-1 folds into a highly stable and compact structure that consists of four β-strands and one α-helix [21]. The Tₘ of GB-1 has been measured to be 87.5°C [22].

Figure 2. Cartoon of GB-1
The image was generated using VMD (PDB: 1GB1).
The favorable properties exhibited by GB-1 make it a good model system to further develop in-cell NMR techniques. Initial in-cell NMR studies have shown that high levels of overexpression of the protein of interest is required in order to detect the protein signals over the noise and background signals [11]. GB-1 can be expressed at very high levels, often near 100 mg/L of E. coli culture when grown in shaker flasks[22]. Nearly all of the GB-1 overexpressed in E. coli cells is well folded and soluble. The protein’s small size is an additional advantage of GB-1 since it results in a short rotational correlation time. The rotational correlation time influences a solution NMR experiment because a long rotational correlation time increases line broadening and reduces overall sensitivity. This effect may be especially important given the increased viscosity and molecular crowding inherent in the in-cell environment. In addition to the favorable biophysical properties of GB-1, the resonance assignments have already been made as part of the determination of the NMR structure of the protein [21].

Figure 3. SDS-PAGE Gel of GB-1 Overexpression
CL is the total cell lysate and MW are molecular weight markers. The arrow indicates the overexpressed GB-1
1.2.1 GB-1 Expression and Initial 2D HSQC Spectrum

The protein GB-1 overexpressed well in BL-21(DE3) E. coli cells. The typical overexpression levels that are achieved are shown in Figure 3. The first step to study any protein with NMR is to collect a 2D $^{15}$N-HSQC spectrum. This spectrum is used to evaluate the feasibility of NMR studies. An individual protein gives rise to a unique $^{15}$N-HSQC, since chemical shift is dependent on the unique local chemical environment around each nucleus. Thus, this spectrum is a fingerprint for a specific protein that has adopted a particular structure. We found that GB-1 in living cells gave rise to the same $^{15}$N-HSQC spectrum as that of GB-1 in vitro (Figure 4). This indicates that GB-1 inside living cells has the same 3D macromolecular structure as GB-1 in vitro.

![Figure 4. $^{15}$N-HSQC Spectra of GB-1](image)

A. In-cell NMR spectrum of GB-1. B. Spectrum of purified GB-1 in solution. Spectra were collected on a Varian 600 MHz spectrometer equipped with a cryogenic probe.
1.2.2 Determining Time Limits for Viable *E. coli* Cells in a Tube

It has already been shown that short $^{15}$N-HSQC spectra can be obtained on overexpressed proteins in living cells. Most of these experiments (including the experiment shown above) were collected in less than 15 minutes [6,7,11]. However, even using fast NMR methods, three dimensional NMR spectra can take several hours to acquire. In order to determine the feasibility of acquiring these types of spectra, it is important to determine what factors limit the time available to perform the experiment.

The most obvious factor that can affect the time available to perform an NMR experiment is cell viability. The environment inside an un-perfused 5mm NMR tube is harsh. The cells are present at very high densities, are not generally aerated or mixed, and no additional nutrients are added to support the cells during the experiment. Therefore, it is important to determine the viability of the cells as a function of time inside the 5mm NMR tube. To do this we used a modified micro drop technique [18]. A schematic of the technique is shown in Figure 5 and the method is described in the experimental methods section of this chapter. As shown in Figure 6, the constant value of the number of colony forming units indicates that *E. coli* cells overexpressing GB-1 remain viable in an NMR tube for at least 12 to 18 hours. This is sufficient time to complete many of the commonly used 3D NMR experiments with fast NMR methods.

A second factor that can limit the total amount of time available to conduct an in-cell NMR experiment is the time it takes for the cells to settle to the bottom of the NMR
tube. The NMR signal is detected by the receiver coil located inside the NMR probe. The bottom of the NMR tube is located well below the receiver coil. A standard NMR tube is typically loaded with sample such that the sample extends significantly above and below the probe coils. Thus, as the cells settle in the NMR tube they will slowly move out of the receiver coil, eventually leading to the loss of the NMR signal. This settling time was measured by preparing an in-cell NMR sample and measuring the height of the cells from the bottom of the tube over time. Figure 7 shows that approximately 3 hours following sample preparation, the cells had begun to settle below the top of the receiver coil. While more limiting than the time

![Diagram](image-url)

Figure 5. Schematic of Cell Viability Assay
A. Schematic of the serial dilution process used in the assay. Numbers at the top of the tube indicate the exponent for the total dilution after plating. Numbers at the bottom indicate the final dilution in each tube. B. Schematic of the organization of 10 μL drops on the agar plates. The numbers indicate the exponent for the total dilution after plating.
limit found for viability, 3 hours is still sufficient to acquire some 3D NMR spectra using fast methods.

The third factor that can limit the time available to acquire an NMR spectrum on proteins in living cells is protein leakage. If protein escapes the cell and is not adequately scavenged, it will be detected in the media, which is a different environment than that in the cell. To assess the importance of this potential problem, an in-cell NMR sample was prepared as previously described, and then incubated for 2 hours at 25° C.

Figure 6. *E. coli* Cell Viability in an NMR Tube
E. coli cells were allowed to incubate in an NMR tube for various times, after which viability was assayed by the modified microdrop technique. The data points for 0-3 hours are the average of three experiments and the 6-18 hour time points are for one experiment. For the 0-3 hour time points, the error bars indicate the standard deviation between the average colony forming units determined from each experiment. For the 6-18 hour time points, the error bars indicate the standard deviation between the colony forming units determined from each drop within one experiment.
Figure 8 A shows the spectrum of the supernatant of the sample after the cells have been spun out. Importantly, the spectrum of GB-1 can be observed in this supernatant. The leakage of GB-1 would lead to ambiguous results for higher dimensional experiments, which take an hour or more even with fast methods, since the signals observed could come from protein outside the cell. In this case, the biological function of GB-1 provided a simple solution to this problem. GB-1 binds IgG, which is a large molecule that would not be observable using standard HSQC pulse sequences due to its slow rotational correlation time. By adding IgG to the NMR sample, it can act as a scavenger for any

Figure 7. Settling of *E. coli* Cells in an NMR Tube
*E. coli* cells prepared as previously described for an in-cell NMR experiment were allowed to settle in an NMR tube. The height was measured at various time points. The dashed lines indicate the location of the receiver coil in a Varian triple resonance probe.
GB-1 that leaks from the cell during the first few hours. To demonstrate the utility of this method, IgG was added to the previous sample to a final IgG concentration of 2 mg/ml. Figure 8 B shows that addition of IgG to the supernatant eliminates most of the resonances from GB-1. Since IgG is not transported into the E. coli cells it can have no effect on GB-1 located inside the living cells. Given this data, we included 5 mg/ml IgG in all subsequent in-cell NMR experiments on GB-1. While we did not test this, it also may be possible to include paramagnetic relaxation agents in the media to suppress extra-cellular protein signals.

Figure 8. Comparison of $^{15}$N-HSQC Spectrum of GB-1 with and without IgG
A. $^{15}$N-HSQC spectrum of the supernatant of an in-cell NMR sample after incubating in an NMR tube for 2 hours. B. Sample from panel A, after addition of 2 mg/mL IgG.
1.2.3 Projection Reconstruction NMR for Backbone Assignments

The assignment of backbone resonances in NMR is usually done with a suite of triple resonance experiments. In this case we used a total of 3 projection reconstruction 3D NMR experiments, HNCO, HNCA, and HA(CA)NH. The HNCA experiment gives rise to resonances at the chemical shifts for $\text{HN}_i$, $\text{N}_i$, and $\text{C}_\alpha_i$, as well as the neighboring residue $\text{C}_\alpha_{i-1}$ (see Figure 9). Normally this experiment is used in conjunction with an HN(CO)CA, which detects the resonances for $\text{HN}_i$, $\text{N}_i$, and $\text{C}_\alpha_{i-1}$. This uniquely identifies the $\text{C}_\alpha_{i-1}$ resonance from the $\text{C}_\alpha_i$ resonance in the HNCA. An alternative approach is to compare the intensity of the two $\text{C}_\alpha$ resonances at a given $\text{HN}, \text{N}$ chemical shift in the HNCA. Typically, the weaker peak will correspond to the $\text{C}_\alpha_{i-1}$ resonance. This method
was sufficient to differentiate the majority of the Cα resonances from the Cαᵢ₋₁ resonances in the HNCA.

The Cα resonance pairs are used to sequentially assign the backbone resonances. This is done by finding matching Cαᵢ and Cαᵢ₋₁ resonances. These can be aligned using a strip plot as shown in Figure 9. One of the difficulties in this method is that several residues may have very similar Cα chemical shifts, giving rise to multiple possible pairings. This problem is solved by using an additional experiment that gives similar pairs of resonances, but uses different nuclei. In our case, we used an HA(CA)NH experiment that yields resonances for HNᵢ, Nᵢ, Hαᵢ, and Hαᵢ₋₁. While the Hα can have chemical shift overlaps as well, the residues with overlapping Hα chemical shifts are often not the same as the residues with overlapping chemical shift in the HNCA experiment. For a small protein like GB-1, a combination of two nuclei will typically yield unambiguous results, as was the case in our experiments. The HNCACB is a more commonly used experiment that supplies Cβ chemical shifts, however it did not yield good spectra for the in-cell sample, necessitating the use of the HA(CA)NH. While this experiment worked well for the small protein GB-1, it is unlikely to be as useful for larger systems where Hα relaxation will result in substantial signal attenuation. The HNCA and HA(CA)NH experiments provided several large fragments of consecutive resonances, however the exact amino acid specific assignments are not known unless these fragments can be put into the context of the amino acid sequence for the protein.
Certain amino acids have characteristic chemical shifts that can be used as possible reference points to align the sequential fragments. This method was sufficient to align the sequential fragments for GB-1. Finally, backbone carbonyl assignments were made using an HNCO experiment (Figure 10) guided by the assigned 2D HSQC spectrum.

In order to keep the acquisition time within the limits dictated by the factors previously discussed, it was necessary to use a minimal number of planes for the projection reconstruction experiments. We found that a total of seven planes, 2 orthogonal and 5 equally spaced angular slices, resulted in acceptable results for our experiments. While this worked well for us, more planes may be necessary for larger

Figure 10. 3D Representation of a 3D-HNCO In-Cell NMR Spectrum
proteins. Studies conducted on Human Carbonic Anhydrase II in solution utilized up to 100 planes to acquire 4D and 5D projection reconstruction NMR data [23].

1.3 Application of In-Cell NMR to a Biological Problem

MetJ is a homodimeric transcription repressor protein that regulates methionine biosynthesis in *E. coli*. This is accomplished by the regulation of a minimum of 12 genes involved in methionine biosynthesis and transport which are located at several different sites in the bacterial genome. MetJ performs its function by binding to regulatory sequences called metboxes that have the consensus sequence AGACGTCT [24]. The promoter regions of genes regulated by MetJ contain at least two, and up to five, metboxes.

When methionine is plentiful in the cell, the downstream activator S-adenosylmethionine (SAM) builds up in the cell, and binds to MetJ which results in its activation [25,26]. The MetJ then finds the metbox sequences where it binds tightly and oligomerizes, leading to inactivation of the gene. As the concentration of methionine in the cell is depleted, the concentration of SAM is reduced, eventually leading to the deactivation of MetJ, and activation of the genes. The number of MetJ homodimers that bind a particular promoter is dependent upon the number of metboxes in that promoter [27]. In addition, the metboxes themselves exhibit sequence variability, which when combined with the variation in the number of metboxes, leads to a highly regulated repression system that can be tuned for individual members of the regulon.
Since the MetJ system was already under study in our lab, we decided to apply in-cell NMR techniques to gain a better understanding of the behavior of MetJ in living cells. We began by collecting an in-cell HSQC spectrum on *E. coli* cells overexpressing MetJ. This spectrum did not show evidence for the typical MetJ spectrum observed *in vivo*. 

![Figure 11. Comparison of MetJ In-Cell NMR Spectrum and Spectrum of Un-Induced Cells](image)

A. NMR spectrum of un-induced *E. coli* cells. B. NMR spectrum of cells overexpressing MetJ.
vitro (see Figure 11 A). The only resonances observed were those that were similar to the background signals typically observed from uniformly labeled *E. coli* cells [11]. The lack of an in-cell spectrum could be due to insufficient MetJ overexpression for detection by NMR. However, SDS-PAGE analysis of the cells following the in-cell NMR experiment revealed that the protein was overexpressed with an estimated concentration of ~300 μM. This concentration is more than sufficient to observe MetJ in dilute solution using cryo-probes. Since the MetJ was present at concentrations sufficient for NMR detection, the lack of observable protein resonances suggested a second possibility. MetJ could be interacting with cellular components that would result in a greatly increased tumbling time and loss of the solution NMR resonances. MetJ is a DNA binding protein, which suggested the hypothesis that MetJ could associate non-specifically with DNA in the cell, resulting in the loss of the NMR spectrum.

To test this hypothesis, we first determined if non-specific DNA could cause the loss of the MetJ NMR spectrum in dilute solution. To do this we added sonicated salmon sperm DNA, which contains 300 bp to 2000 bp DNA fragments, to purified $^{15}$N labeled MetJ. We found that the addition of DNA to a solution of MetJ resulted in the loss of the MetJ NMR spectrum, as shown in Figure 12. It took 120 μg of non-specific DNA to essentially eliminate the spectrum of 7.2 mg of MetJ. More importantly, the spectrum of MetJ returned when we added DNase to digest the DNA. In addition, the spectrum of MetJ could also be rescued from non-specific DNA by a short
oligonucleotide that contained two tandem metboxes. This data demonstrated that MetJ can interact with non-specific DNA and that this interaction can result in the loss of the NMR spectrum that was observed by in-cell NMR.

Given that MetJ could interact with DNA in dilute solution, we decided to test our hypothesis by adding unlabeled cell lysate to purified MetJ. We found that it only required the lysate obtained from 4 mL of cell culture grown overnight to an OD₆₀₀ of
approximately 1 to completely eliminate the NMR spectrum of 7 mg of MetJ (see Figure 13). The spectrum of MetJ could be rescued from the lysate by the addition of DNase, however 10 fold more DNase was required to elicit the same result as observed in the *in vitro* experiment with sonicated salmon sperm DNA.

These observations shed some light on the potential molecular mechanism of MetJ. It has been shown that several DNA binding proteins can locate their binding sites faster than would be expected if the process were dependent upon 3D diffusion [28,29,30]. Facilitated diffusion brought about by close interaction with non-specific DNA has been proposed as a possible mechanism to explain these results [31,32,33]. Several single molecule studies have been conducted in both *E. coli* and eukaryotic cells.
that lend support to this model [34,35]. Our observation that MetJ does not give rise to NMR spectra in lysates and in living cells suggests that it is primarily associated with DNA in vivo and only makes limited excursions away from DNA. Thus it is possible that MetJ makes use of a linear diffusion mechanism along DNA to locate its binding sites, similar to that proposed for other DNA binding proteins [32]. It should be noted that MetJ does not exhibit a strong non-specific binding affinity for DNA [27]. However, the high concentration of potential non-specific DNA binding sites presented by the E. coli chromosome will drive the equilibrium to the bound state. Importantly, the relatively weak binding of any one MetJ dimer to non-specific DNA will allow the protein to rapidly move out of the way of other critical proteins which must interact with DNA. This explains why overexpression of MetJ does not cause any other growth defects besides the constitutive repression of methionine biosynthesis. In summary, weak interaction of MetJ with DNA allows it to find its binding site rapidly while not inhibiting the important biological functions of other proteins that interact with DNA.

1.4 In-Cell NMR on the Large Protein HCA-II, and Unexpected Results

In addition to our studies of GB-1 and MetJ, we also wanted to determine if in-cell NMR could be applied to larger proteins. At the time we embarked on these studies, the proteins observed by in-cell NMR were generally limited to about 10 kD or less in total molecular weight. To determine the feasibility of in-cell NMR as applied to larger proteins, we chose to study the 29 kD protein Human Carbonic Anhydrase II
(HCAII). This protein had already been extensively studied in our lab using solution state NMR, where it was used as one of the first demonstrations of the value of protein perdeuteration as a method to improve the quality of spectra obtained from large proteins [36]. It has also been used to demonstrate the power of reduced dimensionality fast NMR techniques like PR NMR [23]. The protein expresses well in *E. coli* and is reasonably stable, making it a good candidate for these studies.

We performed a 2D in-cell NMR experiment and found that we could not observe any NMR resonances that were consistent with the spectrum expected for HCAII (see Figure 14). While HCAII expresses well for a large protein, it does not express as well as GB-1, so we also acquired an NMR spectrum on a cell lysate prepared from the in-cell sample. Interestingly, we could observe some NMR resonances that were consistent with the solution spectrum of HCAII.

The observation of HCAII in lysate, but not in-cell, was somewhat puzzling. Further review of the literature suggested a possible explanation for our observation. During respiration, carbon dioxide is produced and must be eliminated from the tissue. As carbon dioxide diffuses out of tissues and into capillaries it eventually enters red blood cells. In the red blood cells it is rapidly converted into bicarbonate by carbonic anhydrases, including HCAII. The resulting bicarbonate is then transported out of the red blood cell by the membrane bound chloride/bicarbonate anion exchanger (AE1). Since bicarbonate is more soluble in plasma than carbon dioxide, this process increases
the amount of carbon dioxide that can be transported by the blood plasma. In the lungs, the process is reversed and the bicarbonate is transported back into the red blood cell, converted to carbon dioxide and eventually exhaled. It has been shown that HCAII interacts directly with AE1 via a basic patch that is formed by several histidines located in the first 40 residues of HCAII [37]. The complex formed by these proteins has been

Figure 14. Comparison of HSQC spectra of HCAII In-Cell, in lysate, and in solution
A. In-cell NMR spectrum of HCAII overexpressed in BL21(DE3) E. coli cell. B. 2D-HSQC spectrum of cell lysate obtained from sample (A). C. 2D-HSQC spectrum of purified HCAII (courtesy of Ron Venters)
dubbed a metabolon, which is a complex formed to facilitate the rapid transfer of metabolites between enzymes.

The involvement of HCAII in a metabolon may have important consequences for its study by in-cell NMR spectroscopy. Since our experiments were conducted in *E. coli* cells, AE1 is not present to interact with HCAII. HCAII has been shown to interact with erythrocyte membranes, which is likely facilitated by the presence of AE1, however a weak direct interaction with membrane has not been ruled out. Thus it is possible that HCAII may interact weakly with the *E. coli* membrane in cell, but is released upon cell lysis. Alternatively, the basic patch at the amino terminus of HCAII may also be weakly interacting with other cellular components in the *E. coli* cell. Once the cells are lysed, the subsequent dilution may result in shifting the equilibrium toward the free in solution state, and allow the observation of the NMR spectrum from the fraction of HCAII in solution.

We attempted to observe NMR spectra of the large protein HCAII in living cells. We did not observe the HCAII spectrum in cells, but were able to observe the spectrum in cell lysates. The spectrum that we observed in the lysate was relatively weak. While it is possible that the increased tumbling time in the cell may have been enough to eliminate the spectrum of HCAII, there are other possible causes which are consistent with the biology and function of HCAII *in vivo*. Further studies of this enzyme and the
related enzyme human carbonic anhydrase I, which lacks the AE1 binding activity, may shed some light onto this problem.

1.5 Conclusions and Future Directions

In the first part of this chapter we presented our contributions to extending the in-cell NMR technique to include 3D data collection. At the time, these data were the first of their kind [38]. They demonstrated that 3D NMR was possible on an in-cell NMR sample. We have gone on to publish much of the methodology outlined in this chapter in the Methods in Molecular Biology book series [39]. One of the primary applications for in-cell NMR that has been proposed is the study of protein-ligand and protein-protein interactions. While these types of studies have been done using 2D NMR experiments, our new methods provide a way to study complexes that undergo significant structural rearrangements that alter the 2D-HSQC spectrum, since 3D experiments could be employed to reassign the spectrum of the complex. In addition, 3D NMR is the backbone of modern atomic resolution structure determination. Our methods have shown that it could be possible to determine the structure of a protein in living cells. Indeed, work published several years later by another lab used a different fast NMR method, maximum entropy, to determine a protein structure in living cells [40].

In addition to our contributions to the development of the in-cell NMR technique, we have also applied in-cell NMR to study important biological problems.
Our studies of MetJ have demonstrated that the protein likely interacts continuously with DNA while searching for its binding site, thus increasing the efficiency of the search process. Importantly, these insights were gained by an in-cell experiment that did not result in an observable NMR spectrum. Thus, in-cell NMR can provide important information about biological problems even in the absence of a spectrum.

Similarly, our initial studies on HCAII did not have the outcome we expected. However, with further experimentation it may be possible to better understand the results we have obtained. Since HCAI does not include the AE1 binding site, and does not interact with erythrocyte membranes, it would be beneficial to perform the same in-cell experiment on this protein. The observation of HCAI in cell would indicate that HCAII is likely interacting with either membranes or other cellular components. In addition, it is possible to remove the N-terminal region of HCAII without affecting its enzymatic activity. Therefore it would also be informative to study the in-cell behavior of an N-terminal deletion mutant of HCAII.

In-cell NMR as a technique is still in its infancy. The methods for executing an in-cell NMR experiment are continuously evolving. Some of the current physical limitations on in-cell NMR, such as cell settling and protein leakage, still present significant barriers. However, there may be ways to overcome these barriers. Cell settling problems may be overcome by increasing the viscosity of the medium in which the cells are suspended, such as by adding glycerol. Alternatively, the influence of cell
settling may be reduced by raising the bottom of the NMR tube so that it is just below the receiver coil. This could be accomplished using NMR tubes that incorporate susceptibility matched plugs, such as Shigemi tubes. The concept of increasing viscosity was recently taken to an extreme when it was demonstrated that cells could be suspended in a gel for in-cell NMR experiments [41]. In addition, this method was also shown to help prevent protein leakage. In our own work, we used extra-cellular scavenging by IgG to eliminate the NMR spectrum from GB-1 that leaked out of the cell. If an antibody against the protein of interest is available, this same method could be employed to suppress the extra-cellular spectrum from the protein of interest. If an antibody is not available, it may be possible to include other scavengers in the media, such as membrane impermeable paramagnetic relaxation agents, to provide a more general extra-cellular scavenger.

This work was performed in *E. coli* because it is a well understood and easily handled model system. However, *E. coli* is not the only model system that could be amenable to in-cell NMR techniques. Since these initial studies were performed, in-cell NMR experiments have been conducted using cells other than *E. coli* [42,43]. Importantly, these studies have not relied on overexpression of the protein of interest in the cell to be studied. Instead, they have relied on the introduction of pre-purified protein into cells using either microinjection or cell penetrating peptides [42,43]. These methods have extended in-cell NMR into eukaryotic cells, such as cultured mammalian
cells and frog oocytes, by overcoming the need to endogenously overexpress the protein of interest. These methods also have the added advantage of eliminating the background signals associated with uniformly labeled cells.

Since the extension to eukaryotic cells is of great value, further work in our lab is now focused on the extension of in-cell NMR to yeast. Yeast shares many of the desirable characteristics of \textit{E. coli}, such as relatively rapid growth and straightforward genetic manipulation. These characteristics make yeast easier to work with than either the oocytes or HeLa cells used in other studies. Both cell penetrating peptides and overexpression techniques are currently being evaluated in yeast. In addition, cell penetrating peptides are being evaluated as a method to deliver isotope labeled protein into \textit{E. coli} and frog oocytes.

Finally, since we conducted these experiments, the field of fast NMR has continued to advance. The methods we used to reconstruct our PR NMR data have been supplanted by filtered back projection, which has been shown to yield equivalent results to a polar Fourier transform [44]. In addition, it has also been shown that many of the artifacts that arise from projection reconstruction NMR are a result of the radial sampling pattern [45]. Further advances in the sampling patterns employed and the reconstruction methods have resulted in reduced artifacts and increased signal to noise [45]. These new fast NMR methods may provide new ways to conduct in-cell NMR experiments.
2. A brief Introduction to HIV-1 and Broadly Neutralizing Antibodies.

There are approximately 34 million people worldwide currently infected with the Human Immunodeficiency Virus (HIV) [46]. On average, about 2 million people die annually of the disease caused by the virus, Acquired ImmunoDeficiency Syndrome (AIDS) [46]. There are about 2.6 million new infections reported annually. The virus has spread to nearly every country in the world. In some parts of Africa over 15% of all adults are infected with HIV. In the US, between 0.5 and 1.0% of all adults are infected with the virus.

HIV shares a similar architecture to that of other lentiviruses and retroviruses [47]. Like all retroviruses, HIV stores its genetic information as an RNA genome, rather than a DNA genome. Two single stranded copies of the RNA genome are found in a virus particle. These are chaperoned by the nucleocapsid protein p7. The viral genomes, nucleocapsid protein, integrase, protease, and reverse transcriptase are associated with the viral capsid, which is a cone shaped structure formed by the capsid protein p24. The capsid is surrounded by the matrix, formed by p17, which is in turn encased in the viral membrane. The viral membrane is derived from the host cell membrane during the process of viral budding. Finally, the trimeric envelope complex (ENV) is located on the surface of the viral particle, and is responsible for host cell recognition and membrane fusion. A cartoon of the overall viral architecture is shown
Since the viral membrane is derived from the host cell, host cell proteins are also incorporated onto the virus particle, but are not shown in the figure for clarity.

The HIV lifecycle has been reviewed elsewhere [48], therefore only a brief summary of the process will be presented here. Beginning when the virus encounters a host cell, such as a CD4+ T-cell, the ENV complex on the surface of the virus membrane interacts with the CD4 and either of the co-receptors CCR5 or CXCR4 [49,50,51]. These interactions lead to fusion of the viral membrane with the host cell membrane resulting in the entry of the virus into the host cell. Once the virus has entered the host cell, its RNA genome is converted into DNA by viral reverse transcriptase [52]. After the DNA
genome is generated, it is imported into the host cell nucleus where it is integrated into the host cell DNA by the integrase protein [53]. The integrated viral genome is known as a provirus, and serves as the template for further viral gene transcription.

The RNA encoded by the provirus is transcribed by a combination of host cell and viral proteins. The RNA can be unspliced, partially spliced, or fully spliced, each of which are translated to give rise to different viral proteins. This allows the virus to utilize the same sequences of RNA to encode multiple proteins. Besides being translated, the unspliced RNA is packaged into new virions as viral genomes. After the synthesis of the required viral proteins is complete, the components are assembled into new virus particles which eventually bud from the host cell.

The ENV complex is the macromolecular machine that is responsible for carrying out fusion of the viral membrane with the host cell membrane [54]. ENV is a trimer of the glycoprotein gp160, which is expressed in the host cell endoplasmic reticulum and transported via the golgi to the cell surface where it is incorporated into newly forming virions [55]. During transport to the cell surface, each gp160 is cleaved by a host cell protease to form gp120 and gp41. Gp120 contains five constant regions (C1-C5) and five variable loop regions (V1-V5). A schematic diagram of the gp41 sequence is shown in Figure 16. The N-terminus of gp41 contains the fusion peptide that is inserted into the host cell membrane during viral fusion. HR-1 and HR-2 are the N and C-terminal heptad repeat regions that form the helices that make up the core six-helix bundle.
The MPER region is juxtaposed to the transmembrane domain. MPER is also called the tryptophan rich region because of its high concentration of tryptophan’s that have been shown to play an important role in membrane fusion [58]. The transmembrane domain anchors the entire ENV complex to the viral membrane. In addition, biochemical evidence has shown that the transmembrane domain can undergo trimerization [59] and that certain amino acids in the transmembrane domain are important for viral infectivity [60]. Even though gp41 functions as a trimer, free MPER peptides have not been shown to trimerize strongly without the addition of other trimerization domains.

The model of membrane fusion by ENV is hypothesized to involve at least three major ENV conformational states [54]. A schematic representation of the current fusion model is shown in Figure 17. Prior to host cell encounter, ENV is in the native pre-triggered state. Upon host cell encounter gp120 binds CD4, resulting in a conformational change that increases the affinity of the co-receptor binding site for the

---

**Figure 16. Schematic of HIV-1 gp41**
The schematic is oriented N-terminal to C-terminal from left to right. The following abbreviations were used: FP = Fusion Peptide, HR-1 = Heptad Repeat 1, CC = CC-loop region, HR-2 = Heptad Repeat 2, MPER = Membrane Proximal External Domain, TM = Transmembrane domain, CD = Cytoplasmic Domain.
co-receptor. The co-receptor binding site is specific for either CCR5 or CXCR4, which are members of the G-coupled protein receptor family. The increased affinity of the co-receptor binding site results in binding of the co-receptor. Binding of CD4 and the co-receptor allows the virus to identify the target host cell. Changes in the co-receptor specificity lead to changes in the type of host cell that is infected, such as CD4+ macrophages vs. CD4+ T-cells [50]. Following co-receptor binding, gp41 inserts its N-terminal fusion peptide into the host cell membrane. Between co-receptor binding and viral fusion, gp120 dissociates from gp41, resulting in the second major conformational state, a pre-fusion transitional intermediate. Further conformational rearrangements take place that result in the fusion of the viral membrane with the host cell. During this

![Figure 17. Model of HIV-1 Membrane Fusion](image)

In the current model of HIV-1 fusion, gp120 interacts with CD4 and the co-receptor which allows gp41 to insert its fusion peptide into the host cell membrane. This leads to the pre-fusion intermediate state (middle) which folds back to result in the post-fusion six helix bundle. Fusion is thought to occur during six-helix bundle formation. For clarity, the cytoplasmic domain of gp41 is omitted.
process, the heptad repeat regions (HR-1 and HR-2) of gp41 fold back to form an anti-parallel six-helix bundle, resulting in the third major conformational state, the post-fusion six-helix bundle [56,57]. The process of forming the six-helix bundle is believed to catalyze the membrane fusion event, although the exact mechanism by which this occurs is not well understood.

A variety of evidence has been reported to support this fusion model. Several crystal structures of the core six-helix bundle of the post-fusion complex have been reported [56,57,61]. These structures all reveal an anti-parallel six-helix bundle with HR-1 forming an inner core three-helix bundle and HR-2 packing into the grooves formed by the HR-1 three-helix bundle. Figure 18 shows one of the structures of the six-helix bundle conformation (PDB: 1AIK) [56]. One six-helix bundle structure has been solved with a naturally occurring non-neutralizing HIV-1 antibody bound, further supporting

![Figure 18. Structure of gp41 Core Six-Helix Bundle](image)

HR-1 is shown in Red and HR-2 is shown in yellow. (PDB: 1AIK)
that this conformational state is present after fusion has taken place [61]. Support for the pre-fusion transitional intermediate state has been provided by peptide inhibition of viral fusion. Peptides derived from both HR-1 and HR-2 potently inhibit the viral fusion process [62,63,64,65]. One HR-2 derived peptide, T-20 or Fuzeon, was the first clinically approved inhibitor of viral fusion [66]. It is hypothesized that these peptides bind to the pre-fusion transitional intermediate by mimicking the interactions in the six-helix bundle, which in turn prevents gp41 from folding into the six-helix bundle (see Figure 19) [64]. Thus, the strong inhibition of membrane fusion by these peptides is evidence for the existence of the pre-fusion transitional intermediate state of gp41.

Figure 19. Model of Peptide Inhibited Prefusion Intermediate
HR-2 derived peptides (blue) bind HR-1 on the prefusion intermediate, preventing association of HR-1 and HR-2 (green).
The ENV complex is an important target for vaccine development because it is located on the surface of the virus and is responsible for viral entry into the host cell. Unfortunately, HIV-1 has developed many strategies to elude the human immune response against the ENV complex. These strategies include extensive and variable glycosylation patterns on gp120 that block immunogenic sites [67,68] while not inhibiting receptor binding [67]. Conformational changes in both gp120 and gp41 during the fusion process allow potentially immunogenic sites to remain hidden for much of the viral fusion process, only being exposed when required for fusion to take place [68]. Finally, these proteins have regions which are exposed but exhibit high degrees of mutational variability, which can act as a decoy to the adaptive immune system, leading to antibody responses that are narrow and lag behind viral evolution [69].

Even with all of these protective measures, broadly neutralizing antibodies against HIV have been discovered [70,71,72,73,74]. Most of these antibodies target various epitopes on the ENV complex, including the CD4 binding site, the V2 and V3 loop regions, glycosylation sites, and the membrane proximal external region (MPER) of gp41 [75,76,77,78]. Some of the most potent and broadly neutralizing antibodies against HIV-1 are 2F5 and 4E10, which target the MPER of gp41 [71,79]. One likely contributing factor to the breadth of neutralization by these antibodies is the high peptide sequence conservation observed in the MPER [79]. The high level of conservation in the MPER
sequence can be seen in Figure 20, which is a sequence logo obtained from the alignment of all HIV-1 sequences in the HIV sequence database. While the discovery of broadly neutralizing antibodies has inspired hope for a vaccine, an effective vaccine that elicits broadly neutralizing antibodies has yet to be developed.

Extensive structural and immunological studies have been conducted on the broadly neutralizing antibodies in order to better understand their mechanism and how to induce them. High resolution crystal structures of 2F5 and 4E10 have been solved bound to short, monomeric peptides that contain the core epitopes [80,81]. The peptide conformation in 4E10 is an α-helix. In contrast, the 2F5 peptide adopts a relatively extended conformation with the core epitope buried deeply into the antibody binding pocket. Both of these antibodies have unusually long and hydrophobic CDR-H3 loops.

Figure 20. Sequence Logo of MPER Region of HIV-1 gp41
The frequency of a particular amino acid at the given position is proportional to the height of the one letter code of that amino acid. The epitopes for 2F5 and 4E10 are residues 664-666 and 672-676.
In the crystal structures of 2F5 this loop is either stabilized by crystal packing contacts or no electron density is observed.

Recent studies have shown that both 2F5 and 4E10 are polyreactive antibodies. They interact with a diverse range of antigens besides their respective peptide epitopes, including histones, centromeres, and phospholipids such as cardiolipin and phosphatidylserine [82]. This characteristic is not exclusive to 2F5 and 4E10, since other broadly neutralizing antibodies such as the CD4 binding site antibody 1B12 also display polyreactivity [83]. Further surface plasmon resonance studies of the MPER directed antibodies have shown that they bind their peptide epitopes when displayed on the surface of liposomes as well as or better than free peptides in solution [84]. Finally, mutations in the long CDR-H3 of 2F5 and 4E10 have been shown to eliminate the lipid reactivity of these antibodies [85]. While these mutations did not significantly alter the binding to the peptide epitopes, the mutant antibodies no longer neutralized the virus in infection assays [85]. Taken together, these data revealed that polyreactivity, particularly lipid reactivity, likely plays an important role in viral neutralization by the MPER specific antibodies.

The model for viral fusion postulates at least three conformations of ENV (see Figure 17) [64]. These conformational changes are thought to play an important role in the mechanism of neutralization by the MPER antibodies 2F5 and 4E10. Neither 2F5 nor 4E10 bind with high affinity to ENV complexes in the native pre-triggered state [86].
The antibodies also do not bind with high affinity to the post-fusion six-helix bundle of gp41 [86]. This has led to speculation that the neutralizing MPER antibodies bind the pre-fusion transitional intermediate state proposed in the model for viral fusion. By interacting with the MPER in the pre-fusion intermediate, the antibodies prevent the formation of the six-helix bundle and viral fusion, similar to HR-1 and HR-2 peptides. MPER constructs that attempt to mimic the pre-fusion intermediate state have been produced, and in some cases bind with high affinity to both antibodies [59,86,87].

The hypothesis that 2F5 and 4E10 target the pre-fusion intermediate is not without controversy. There is some data in a recent report that suggests the MPER antibodies can induce shedding of ENV, although with relatively slow kinetics, suggesting that the MPER antibodies may interact directly with ENV prior to CD4 binding [88]. Additional studies have shown that there are differences in the exposure of the MPER depending upon the strain of HIV-1 virus examined [89]. Data has also been presented that suggests the MPER is likely in a conformation that is compatible with binding by 2F5 and 4E10, and the reduced affinity of binding to certain ENV complexes is likely caused by masking of the epitope by gp120 [89]. Thus, the transition to the pre-fusion intermediate state may not form the MPER epitopes per se, but may increase exposure of the MPER epitopes, leading to high affinity antibody binding.

Structure based vaccine design has been proposed as an alternative approach to induce MPER specific antibodies. These studies have used the crystal structure of MPER
antibodies bound to short epitope peptides to engineer protein scaffolds that present the MPER peptide epitopes in the conformations observed in the crystal structures [90,91]. These designed vaccines have been used to immunize animals and have given rise to antibodies [91,92]. Some of the antibodies produced bound the epitopes in a conformation very similar to that seen in the high resolution crystal structures of 2F5 [91]. However, these antibodies did not exhibit significant neutralization of HIV-1 [93].

The above observations suggest that viral neutralization by the MPER antibodies 2F5 and 4E10 is likely more complex than simply achieving high affinity antibody binding. Antibodies have been induced that bind the MPER region, including the core 2F5 epitope, however none of these antibodies yielded significant HIV-1 neutralization [92,94]. Notably, these vaccine induced non-neutralizing antibodies have not been shown to possess the significant polyreactivity observed for other potent neutralizing antibodies like 2F5 and 4E10. The lack of polyreactivity, and specifically lipid binding, may provide an explanation for the lack of neutralization [93]. In addition, the structure based vaccine design efforts have been focused on crystal structures of short monomeric peptides bound to the neutralizing antibodies [90,91]. Since gp41 is a native trimer, it is also possible that these crystal structures may not capture the conformation that is normally displayed in the native ENV trimer or pre-fusion transitional intermediate.

The last two decades have witnessed an enormous increase in our understanding of HIV-1 biology. However, it is clear that there is still a great deal we do not yet
understand about both HIV-1 and its interactions with the human immune system. The mechanism of viral fusion, as well as the mechanism by which MPER antibodies function to prevent fusion, are still not completely understood. Our understanding of the behavior of peptide antigens, such as MPER derived peptides, in lipid environments is limited. Insufficient understanding of this behavior impacts our ability to rationally design vaccines that can induce both potent and broadly reactive neutralizing MPER antibodies. In the following chapters, I will discuss several studies that provide new insights into the biophysical behavior of the MPER domain in lipid environments.
3. Studies on SP62, a GP-41 Derived Peptide Containing the 2F5 Epitope

Broadly neutralizing antibodies 2F5 and 4E10 interact with lipids, and evidence suggests that this lipid interaction is integral to the neutralization of HIV by these antibodies [83]. Vaccine candidates have been proposed and developed based upon this observation [84]. These vaccine candidates are conjugates where peptides containing the epitopes of 2F5 and/or 4E10 are conjugated to liposomes [84,94]. Peptide-liposome conjugates of this type have been shown to bind 2F5 and 4E10 with nanomolar affinity [84,94]. Additionally, kinetic studies of 2F5 and 4E10 binding to the peptide-liposome conjugates suggested that the antibodies bind with a two-step process, consistent with an encounter-docking model [95]. In contrast, the two-step model was not observed for MPER peptides that were not presented on a lipid membrane [95].

In these peptide-liposome conjugates, the peptide is anchored to the liposome via a hydrophobic peptide anchor, referred to as GTH1 [95]. The MPER peptides were anchored using GTH1 because they did not remain stably associated with liposomes without it (Dennison, personal communication). The GTH1 sequence used in these vaccine candidates is derived from the HIV-1 capsid protein. A crystal structure of the capsid protein where this peptide originated shows that the peptide adopts a helix in its native context (PDB: 1GWP) [96]. Even though the GTH1 peptide anchor is not a transmembrane helix, it still stabilized the interaction between MPER peptides and liposomes [95].
Given the apparent importance of lipid to the function of the broadly neutralizing antibodies 2F5 and 4E10 [85], it is important to understand the biophysical behavior of the peptide epitopes in the presence of lipids. When we began our studies, only one structural study of the MPER peptide had been published. The peptide structure was solved by NMR spectroscopy at pH 3.0 in dodecylphosphocholine (DPC) detergent micelles [97]. The peptide formed a well ordered helix under these conditions. Unlike the peptide-liposome conjugates, the MPER peptide used in these structural studies did not contain an additional membrane anchor.

We initiated studies to better understand the structure of the GTH1 anchored peptides in the presence of detergent micelles and liposomes. Since the GTH1 anchored peptide displayed increased affinity for lipids and altered binding kinetics compared to MPER peptides without lipids, we hypothesized that the GTH1 anchored peptide may adopt a different conformation than the unanchored peptides. Additional biophysical characterization of the conformation of the GTH1 anchored peptide in the presence of detergent micelles and liposomes may provide insight into the lack of neutralizing antibodies produced to date by these conjugates. These insights will help guide further iterations of vaccine design based upon MPER peptide-liposome conjugates.

3.1 Experimental Procedures

The SP62-GTH1 peptide

(QQEKNQELLELDKWASLWNYKRIILGNKIVRMYS) was purchased from Primm
scientific. Specifically labeled SP62-GTH1, with $^{15}$N labeled amino acids incorporated at residues L9, A16, and L18, (LAL-SP62-GTH1) was also purchased from Primm. The SP62-NK peptide (QQEKNEQELLELDKWASLWNKYKRWIILGNK), was produced by expression in BL21(DE3) E. coli. The peptide DNA sequence was incorporated into a modified pet41b vector (Novagen) that had all methionines removed, except the start codon and one immediately prior to the SP62-NK sequence. The final expressed protein contained a GST-tag, S-tag, and 6-histidine tag, followed by the SP62-NK sequence.

Protein expression of SP62-NK was carried out by growing the cells in modified M9 minimal media (appendix A). For labeling, $^{15}$N labeled ammonium chloride was used as the sole nitrogen source and uniformly labeled $^{13}$C glucose was used as the sole carbon source. Cells containing the expression plasmid were grown at 37° C to an OD$_{600}$ of ~0.6, when protein expression was induced by the addition of 1 mM IPTG. The cells were allowed to grow for a further 4 to 6 hours, after which they were harvested by centrifugation.

Cells were lysed in Bugbuster reagent (Pierce) with 6M Guanidine-HCl added. Insoluble debris was removed by centrifugation at 15,000xg for 30 minutes. The clarified sample was applied to a Ni-Sepharose column (GE Healthcare). The column was washed with 4 column volumes of Ni-wash buffer (Appendix A), then eluted with 4 column volumes of Ni-elute buffer (Appendix A). The fractions containing the fusion protein were pooled then dialyzed against ddH$_2$O overnight. This resulted in the
precipitation of the fusion protein, which was harvested by centrifugation at 15,000xg.

The fusion protein was dissolved in 70% Trifluoroacetic acid (TFA) and mixed with 0.1 M cyanogen bromide (CNBr) to initiate the cleavage reaction. The reaction was allowed to proceed overnight (approximately 16-18 hours). The cleavage reaction was stopped and reaction by-products were removed by lyophilization with an apparatus designed to safely trap the toxic reaction by-products.

The cleaved peptide was separated from larger peptides and proteins by size exclusion chromatography. The sample was dissolved in size exclusion buffer (Appendix A), and run over a 1.6cm x 60cm Sephacryl S-100 column. Fractions containing the SP62-NK peptide were pooled and concentrated using a 2,000 molecular weight cutoff spin concentrator (Sartorius). The concentrated sample was dialyzed against ddH₂O using a 2,000 molecular weight cutoff dialysis cassette (Pierce). Sufficient TFA and acetonitrile were added to the dialyzed sample to make the final solvent 30% acetonitrile and 0.1% TFA. The sample was further purified using reverse phase chromatography on a Resource RPC column (GE Healthcare). The sample was injected in 30% acetonitrile, 0.1% TFA, and eluted with a gradient from 30% to 70% acetonitrile, 0.1% TFA. Fractions containing the SP62-NK peptide were pooled and lyophilized to a dry powder.

1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine (POPC) liposomes were prepared by drying the appropriate amount of lipid from chloroform under vacuum.
The dried lipid was hydrated in buffer, and vortexed, which resulted in large multi-
lamellar vesicles (LMV’s), indicated by an opaque solution. LMV’s were converted to
small uni-lamellar vesicles (SUV’s) by sonication in a bath type sonicator. Sonication
was carried out until the opaque solution turned to a clear slightly blue solution,
indicating the formation of SUV’s, which usually took 45 minutes. Two methods were
employed to prepare peptide-liposome conjugates. The first method (method 1)
prepared the liposomes as above in ddH₂O. Peptide in ddH₂O was added to the
liposomes after SUV’s were formed. The second method (method 2) incorporated the
peptide into the lipid prior to chloroform removal by dissolving the peptide and lipid in
chloroform and methanol. The dried lipid-peptide mixture was hydrated ddH₂O, then
vortexed and sonicated as indicated above.

Samples of SP62-NK for NMR were prepared by dissolving the peptide in 50 mM
sodium phosphate, pH 7.0, containing 100 mM 1,2-dihexanoyl-sn-glycero-3-
phosphocholine (DHPC) detergent. Sufficient D₂O was added to make a final
concentration of 10% D₂O. The final pH of the sample was between 3.0 and 4.0.
Samples were placed in Shigemi NMR tubes. 3D NMR experiments, including HNCA,
HNCACB, and HN(CO)CA, were used to assign the backbone resonances of the SP62-
NK peptide. NOE data were collected using an ¹⁵N NOESY-HSQC. Chemical shift
indexing was carried out using previously determined random coil chemical shifts for
Cα and Cβ carbons [98]. We employed the ΔCα-ΔCβ method, which has been shown to
provide reliable prediction of secondary structure [36]. NMR experiments were carried out using either a Varian Inova 600 or 800 MHz spectrometer equipped with cryogenically cooled triple resonance probes at the Duke NMR Center.

### 3.2 Liposome Bound SP62

It has been postulated that the membrane surface plays an important role in viral neutralization by the broadly neutralizing antibody 2F5 [85]. It is clear from previous SPR studies that 2F5 binds differently to MPER peptides anchored on the surface of the liposomes than to free MPER peptides [95]. However, the structure of the SP62-GTH1 peptide when bound to liposomes had not been studied. To date, these liposome conjugates have not given rise to neutralizing antibodies when administered to animals (Dennison et al, unpublished results). Therefore, it is important to understand what the structure of SP62-GTH1 might be on the surface of the liposome, and what may actually be presented to the immune system in a liposome-peptide conjugate based vaccine.

In order to initially characterize the peptide on the surface of liposomes, we prepared liposome-peptide conjugates using method 1. 1D proton spectra on these liposome-peptide conjugates are shown in Figure 21. Even though liposomes are generally large, we were able to observe peaks in the 1D proton spectrum that corresponded to peptide, as well as the expected signals from the liposome. This observation suggested that the peptide may not be completely constrained on the liposomes, and was not exhibiting the increase in rotational correlation time that is
expected for a peptide completely buried in the lipid bilayer of a liposome. Importantly, a comparison of the peptide spectrum alone and the peptide spectrum in POPC liposomes do show specific chemical shift perturbations that likely indicate interaction with the liposome (shown in Figure 21). Further perturbations could be observed in the
spectrum of the liposome. The choline head group of POPC gives rise to a strong sharp proton resonance at approximately 3.24 ppm. In liposomes, this resonance is split into two populations that arise from the choline on the inner and outer leaflets of the phospholipid bilayer. This is caused by the difference in the packing of the head groups due to the difference in curvature between the inner and outer leaflets [99]. The choline head group broadens in the presence of SP62-GTH1, indicating some interaction between the phospholipid head groups and the peptide. Figure 22 is an expanded view of the spectrum showing the choline head group peaks.

We conducted multi-dimensional NMR analysis on these liposome-peptide

![Figure 22. Effect of SP62-GTH1 on Choline Head Group of POPC Liposomes](image)
The top spectrum shows the choline head group resonance (~3.25 ppm) of POPC liposomes without SP62-GTH1. The bottom spectrum shows the choline head group resonance of SP62-GTH1 peptide-liposome conjugate prepared using method 1.
conjugates. In these experiments we used LAL-SP62-GTH1 peptide labeled with $^{15}$N at positions L9, A16, and L18. Comparing the 2D-HSQC spectrum of LAL-SP62-GTH1 free in solution vs. with POPC liposomes reveals that there is at least one perturbation in the chemical shifts of the labeled residues by the POPC liposomes (see Figure 23). This chemical shift perturbation indicates that there is some interaction with the liposome, although the labeled amino acids may not be part of the interacting surface. In the SP62-GTH1 peptide, the GTH1 segment is used to anchor the peptide to the liposome. Therefore, it is possible that part of the GTH1 segment is more closely interacting with the liposomes, while the SP62 is interacting more with the solvent. If this were the case,

![Figure 23. $^{15}$N-HSQC Spectra of SP62-GTH1 With and With Out Liposomes](image)

Spectrum of LAL-SP62-GTH1 (red) overlaid with LAL-SP62-GTH1 conjugated with POPC liposomes (black).
then it would also agree with the above observation that the peptide has sharper lines than would be expected for a peptide buried in the liposome bilayer.

We used LAL-SP62-GTH1 to further characterize the liposome-peptide conjugate by collecting 3D-NOESY-HSQC spectra. We collected data on free peptide and peptide with POPC liposomes (see Figure 24). While the spectra are similar, noticeable peak shifts occurred. Additionally, there appeared to be significant broadening of some of the peaks in the liposome-peptide spectrum as compared to the peptide alone. Finally, new peaks emerged in the liposome-peptide spectrum that were not present in the spectrum of the peptide alone. This suggests that the liposome is inducing some conformational alterations in the peptide as compared to peptide free in solution. Taken together, the above data strongly suggests that the peptide is interacting with the POPC liposome and that the liposome is influencing the conformation of the peptide.

The method used to prepare these liposomes was somewhat different than the method used to prepare liposomes for the previous SPR studies [95]. Those liposomes were prepared using the second method outlined in the experimental section. Liposome-peptide conjugates prepared using method 2 gave rise to a different NMR spectrum than those prepared using method 1. With method 2, the 1D proton spectrum showed very broad peaks for the peptide that were essentially indistinguishable from the background (see Figure 25). This would suggest that the peptide is more buried in the lipid bilayer
and tumbling with the same rotational correlation time as the liposome. Interestingly, a sample made initially using method 1, then lyophilized and dissolved in water yielded an intermediate spectrum (Figure 25 C).

Figure 24. 3D-NOESYHSQC Comparison of free and Liposome bound SP62-GTH1
A. NOESY-HSQC spectrum of LAL-SP62-GTH1 free in solution. B. NOESY-HSQC spectrum of LAL-SP62-GTH1 conjugated to POPC liposomes. Spectra are shown with all nitrogen dimension planes overlaid.
Since the two methods of liposome preparation yielded dramatically different NMR results, it was important to determine if the liposome-peptide conjugates interacted in a similar fashion with 2F5. Liposome-peptide conjugates prepared with

Figure 25. Comparison of Methods to Prepare Liposome-Peptide Conjugates
The top set of spectra emphasize the peptide region, while the bottom set of spectra emphasize the POPC liposome region. A. SP62-GTH1 in solution. B. SP62-GTH1-liposome conjugate prepared using method 1. C. Sample from B lyophilized to a dry powder, then resuspended in buffer and sonicated. D. SP62-GTH1-liposome conjugate prepared using method 2.
method 1 were analyzed for 2F5 binding using surface plasmon resonance. We found that 2F5 bound specifically to the liposome-peptide conjugate vs. liposomes alone (see Figure 26). In addition, these experiments also show that the peptide does not rapidly exchange between the liposome and solvent. The experiment is carried out by first binding the liposome-peptide conjugate to the lipophilic L1 chip. The sensogram is allowed to stabilize before the antibody is injected. Thus, there is a significant time period available for any rapidly exchanging peptide to wash away, prior to the antibody binding assay.

Figure 26. SPR of SP62-GTH1-liposome conjugate prepared using Method 1
Samples of POPC liposomes (Blue) or SP62-GTH1-liposome conjugates (Red) were prepared using method 1. The curves show the association and dissociation of the monoclonal antibody 2F5.
These observations suggest that the method of liposome-peptide conjugate production can lead to conjugates with different properties. While many properties of the conjugates could be affected by the method of preparation, two properties which our data suggest could be altered are the peptide conformation and the extent to which the peptide is buried in the liposome membrane. These are important properties for vaccine design. In order for a B-cell to produce an antibody, it must be stimulated by specific binding of an antigen to its B-cell receptor. The stimulation, along with other factors, results in clonal expansion of the B-cell and affinity maturation of the antibody. The conformation of the peptide, or the way in which it is displayed on the surface of a membrane, could influence the interaction with the B-cell receptor, resulting in different B-cells being stimulated. In this case, the epitope conformation may be different depending on preparation method, and therefore one preparation method may not result in an antigen that stimulates the same B-cell as the other method, which could give rise to different antibodies. We believe it is important to consider how vaccine candidates are prepared, and how the chosen preparation method may influence the antibody repertoire eventually produced.

The sensitivity of the conjugates to preparation method may also indicate sensitivity to other environmental characteristics. A vaccine is exposed to many different conditions as it is formulated and injected into a subject. It is not clear if the conformation of the peptide remains the same when the conjugates are mixed with
adjuvants and other components that are typically used in vaccine formulations. Such sensitivity may result in changes to the peptide conformation during formulation and after administering the vaccine.

3.3 NMR of SP62NK in DHPC micelles

While our studies of the SP62-GTH1 peptide revealed insights about the preparation of liposome-peptide conjugates and how different preparation methods may alter the conformation of the peptide, it does not give insight into the possible conformation of the peptide. Therefore we initiated studies to determine the conformation of the SP62 peptide in the presence of detergent micelles using NMR spectroscopy. This information, along with other biochemical, biophysical, and immunological data, could then be used to help guide subsequent rounds of vaccine design.

In order to determine the conformation of the peptide in the presence of lipids using multi-dimensional NMR spectroscopy techniques, we labeled the peptide with $^{15}$N and $^{13}$C. Synthesizing a uniformly labeled peptide the size of SP62-GTH1 is cost prohibitive; therefore, we expressed the peptide in E. coli cells. We found that a truncation of SP62-GTH1 that we refer to as SP62-NK expressed better in E. coli and was somewhat easier to purify (see experimental methods section for the sequences of these peptides). This peptide was used for all subsequent NMR studies of the linear peptide monomer conjugated with detergent micelles.
We studied this peptide in DHPC detergent micelles. Detergent micelles have successfully been used to solubilize membrane proteins for NMR studies by several groups [100]. It has been found that integral membrane proteins that have enzymatic activity can retain that activity in some detergent micelles, indicating that detergent micelles can support correctly folded membrane proteins [101]. In this case, DHPC was chosen because it is similar to the phosphatidylcholine based lipids that make up a major fraction of mammalian membranes, differing only by its shortened acyl chain length. The SP62-NK peptide was soluble in water, but only gave rise to good NMR spectra in the presence of DHPC detergent (see Figure 27). We used a standard suite of

![Figure 27](image_url)

*Figure 27. Assigned $^{15}$N-HSQC Spectrum of SP62-NK in DHPC Detergent
The region in the blue box is expanded in the inset.*
triple resonance NMR experiments, including HNCA, HNCO, HNCACB, HN(CO)CA, and HN(CO)CACB to assign the backbone amide resonances of SP62-NK in DHPC micelles for all but the first residue in the peptide. A strip plot from an HNCA experiment is shown in Figure 28. This peptide has one glycine and one serine, which have chemical shifts noticeably different from most other amino acids. These served as useful starting points during the assignment process.

Once the backbone resonances were assigned we could use the data to gain insight into the conformation of the SP62-NK peptide. Chemical shift indexing is a

![Figure 28. Strip Plot from an HNCA on SP62-NK in DHPC Detergent](image)
The residue identity and nitrogen chemical shift are listed above each strip. Resonances arising from the same $\text{Ca}$ are connected by a dashed line.
common method used to identify secondary structural elements in proteins [102]. It has been shown that the major secondary structural elements lead to systematic perturbations in the chemical shifts of the Cα and Cβ carbons from the shifts observed in unstructured peptides [98,103]. We used a chemical shift indexing method that combines the perturbations of both Cα and Cβ to determine the secondary structure adopted by the SP62-NK peptide [36]. As shown in Figure 29, nearly the entire peptide adopts a helical conformation. There is a slight dip in the region where the MPER derived SP62 sequence ends, and the membrane anchoring GTH1 sequence begins. The

![Chemical Shift Indexing and NOE analysis of SP62-NK](image)

**Figure 29.** Chemical Shift Indexing and NOE analysis of SP62-NK
A. Chemical shift indexing using the ΔCα-ΔCβ method, where values greater than zero indicate α-helix and values less than zero indicate β-sheet. B. Diagram of NOE’s observed that indicate α-helix formation where each line indicates the specified NOE between two amide protons.
secondary structure trends towards an unfolded protein near the C-terminus of the peptide, as indicated by the chemical shift differences near zero. Generally the SP62 portion of the peptide shows a strong helical tendency.

In addition to the experiments used to assign the backbone resonances of the peptide, we also collected an $^{15}$N-NOESY-HSQC. NOESY experiments provide inter-proton distance information and form the primary data traditionally used to determine atomic resolution structures by NMR. In this case, we used the NOE data to identify regular secondary structure [104]. This is possible because helical segments and sheet segments give rise to characteristic backbone inter-proton distances. Consecutive stretches of these distances are strong indicators of secondary structure type [104]. Again, the backbone inter-proton distances we observed were consistent with a helical conformation for the majority of the peptide.

Our chemical shift indexing and NOE pattern analysis shows that the SP62-NK peptide adopted a helical conformation under the conditions we used for NMR. This conformation is consistent with the conformations observed for other MPER peptides that contain the SP62 sequence [97,105,106,107]. The MPER sequence has been shown to adopt an $\alpha$-helix conformation in DPC detergent micelles at pH 3.5, consistent with our observation in DHPC micelles at a similar pH [97]. A more recent study of the MPER peptide in DPC micelles at pH 6.0 also showed that the peptide adopts an $\alpha$-helix conformation, but with a kink following the SP62 sequence [106]. The region of this
structure that included SP62 was also a helix, consistent with our data. Finally, a third publication showed that the N-terminal residues of the MPER sequence (residues 650-662) tended to be sensitive to pH [105]. In this study, the MPER peptide was found to adopt a helical conformation at low pH, as we and others observed. However, as the pH was raised to 6.0, they found that the N-terminal region (residues 650-662) of the MPER became disordered. The peptide only retained regular helical structure from the core DKW of the 2F5 epitope through the C-terminus. The peptide sequence used in these experiments extended into the HR-2 region further than the other sequences studied, including ours. We found that increasing the pH in our samples resulted in broadened and shifted resonances (see Figure 30), further indicating that the conformation of these peptides is dependent on their environment.

Figure 30. pH Dependence of 15N-HSQC Spectra of SP62-NK in DHPD Detergent 15N-HSQC spectra of SP62-NK were collected at pH 3.5 (A.) and pH 5.0 (B.).
Our data and that of others have shown that the HIV gp41 MPER likely adopts a helical conformation for at least part of the sequence. However, the data also suggests that the environment can have a dramatic effect on the conformation that these peptides adopt. This is shown by the pH dependency of the NMR spectrum that we observed for our construct, and the similar dependency found by Coutant et al. [105]. These observations, along with our previous observation that preparation method can result in different liposome-peptide conjugates, suggests that liposome-peptide conjugates may behave differently depending on their local environment. This poses an important problem for evaluating a vaccine design since vaccines can be administered into very different environments, such as blood and mucosa. Our observations suggest that the liposome-peptide conjugate may not adopt the same conformation in these different environments, potentially influencing the overall antibody response. This data suggests that short, monomeric MPER peptides may be able to adopt several different conformations, making it difficult to predict which conformation will give rise to antibodies.

The crystal structures for both 2F5 and 4E10 have been determined bound to short peptide epitopes [80,81]. In the structure of 4E10, the epitope is in a helical conformation. However, the structure of the peptide epitope in 2F5 does not adopt a helical conformation in the antibody peptide complex. Instead, it adopts an extended conformation with the core epitope residues forming a β-hairpin-like conformation that
is deeply inserted into the antibody binding site [80]. Our construct only contained the 2F5 epitope and was in a helical conformation similar to the other MPER peptide constructs [97,105,106]. The significance of the difference in the structures of the 2F5 epitope between the 2F5 bound crystal structure and 2F5 containing peptides is not well understood. Clearly, 2F5 can bind specifically to peptide constructs that adopt helical conformations in solution, as evidenced by our SPR data, which shows binding to SP62-NK. Others have also shown that a helical MPER peptide in complex with detergent micelles will bind 2F5 with high affinity [106]. Thus, 2F5 may induce substantial changes in the conformation of its epitope upon binding. Alternatively, the crystal structure may represent a binding mode that is favorable for short, monomeric peptides, but may not be the same as might occur in the context of the entire trimeric gp41.

In summary, the data presented in this chapter demonstrate that liposome-peptide conjugates are sensitive to environmental conditions during preparation and use. We have also shown that the SP62 sequence adopts a helical conformation on DHPC micelles. These constructs are sensitive to their environment as exhibited by the pH sensitivity of the $^{15}$N-HSQC spectrum of SP62-NK. Our observations are also consistent with another study that demonstrated pH dependence on a longer MPER derived peptide [105]. While liposome-peptide conjugates still represent important vaccine candidates, it is important to consider what changes may be occurring to these conjugates due to environmental conditions, given the role of lipid reactivity in
neutralizing antibodies that bind MPER epitopes. Short monomeric peptides may exhibit significant conformational heterogeneity which may influence the antibodies induced by these conjugates.
4. Biophysical Characterization of New MPER Constructs

The HIV-1 envelope protein complex (ENV) is located on the surface of the native HIV-1 virion and is responsible for fusion of the viral membrane with the host cell membrane. The complex is trimeric, made up of 3 copies each of gp120 and gp41. The gp120 subunits contain the receptor binding sites that are responsible for host cell recognition. The gp41 subunits are responsible for the membrane fusion process. The architecture of gp41 was previously described in chapter 2 and in Figure 16. High resolution crystal structures of trimeric HR-1 and HR-2 show that gp41 forms a six-helix bundle [56,57]. It is thought that the six-helix bundle conformation is representative of the post-fusion complex [64]. The broadly neutralizing antibodies 2F5 and 4E10 do not interact strongly with the post-fusion six-helix bundle, but are hypothesized to target the pre-fusion transitional intermediate state of gp41 [86].

Several constructs have been designed to study the putative pre-fusion transitional intermediate state of gp41. Chen et al developed two constructs where the TM and CD domains are replaced by the foldon domain from T4 fibritin to facilitate trimerization. To expose the HR-2 epitopes, the construct incorporated a second copy of HR-2 so that HR-1 would fold back into the six-helix bundle with one copy of HR-2, and leave the other copy free to mimic the intermediate state [86]. This construct, referred to as gp41-inter, binds with nanomolar affinity to both 2F5 and 4E10 [86]. A modified version of this construct has also been presented that replaces the HR-1 and extra HR-2
with GCN4, a three-helix bundle trimerization domain. A separate trimeric MPER construct, called HA-gp41, was designed independent of gp41-inter but is similar to the modified version of gp41-inter. HA-gp41 contains HR-2 and the MPER which are trimerized by incorporating GCN4 on the C-terminus and influenza HA on the N-terminus [87]. HA-gp41 bound 2F5 and 4E10 with high affinity, and gave rise to non-neutralizing MPER specific antibodies in rabbits [87]. A third independent construct used a C-terminal GCN4 sequence to trimerize the MPER peptide. A crystal structure of this construct was solved and showed that it formed a three-helix bundle [107]. However, the construct did not bind either 2F5 or 4E10 with high affinity, suggesting that this construct does not present the MPER in a manner likely to induce 2F5 or 4E10-like antibodies [107].

These trimer constructs do not incorporate lipid binding into their design, nor is there evidence reported that would indicate they interact with lipids to a significant extent. Lipid binding has been shown to be an important part of viral neutralization by MPER specific broadly neutralizing antibodies [85]. The absence of a lipid component may contribute to the lack of neutralizing antibodies induced by the previous trimeric constructs. At least one membrane associated trimeric gp41 construct (referred to as gp41ctm) has been reported [59]. This construct consisted of HR-2, MPER, the gp41 transmembrane domain, and a C-terminal addition derived from a read-through mutation of the expression vector [59]. While the construct was a trimer and bound 2F5
and 4E10 with high affinity, it did not give rise to antibodies that significantly neutralized HIV-1. Circular dichroism analysis on gp41ctm suggested that ~30% of the construct retained regular helical secondary structure, which the authors attributed to the TM domain [59]. This conclusion was at least partially based on previous circular dichroism studies that showed that short HR-1 and HR-2 peptides do not adopt regular secondary structure in aqueous solution [108]. Thus, it is unclear if the HR-2 region and/or the MPER in gp41ctm retained regular secondary structure or if those regions were trimerized, which could adversely influence the immunogenicity of the construct.

In order to better understand the biophysical characteristics of the MPER, we initiated the development of a new MPER construct. This new construct is designed to bridge the gap between the trimers presented above and the peptide-liposome conjugates discussed earlier in this dissertation. To do this, a minimum of three primary design goals needed to be met. First, the construct must interact with lipid in a similar manner as might occur during membrane fusion. Second, to enable detailed structural analysis, the construct must form a stable trimer in the presence of lipid membranes and detergent micelles. Finally, the construct must bind with high affinity to the broadly neutralizing MPER antibodies. Since we were interested in studying the structure of these constructs by NMR spectroscopy, we set an additional goal, that the constructs give rise to good NMR spectra. This chapter will discuss the production and characterization of several constructs designed to satisfy these criteria. It culminates
with the characterization of a new construct called gp41-M-MAT, which stands for gp41 MPER containing Membrane Associated Trimer.

4.1 Experimental Procedures

Proteins were expressed in C41(DE3) E. coli cells using pCLEH, which is a T7 based expression vector that expresses the protein or peptide of interest fused to TrpLE [109]. TrpLE is useful for small peptides and potentially toxic proteins because it directs the over-expressed protein to insoluble inclusion bodies [109,110]. Purification under denaturing conditions is simplified by the inclusion of a 6-hisitidine tag. A methionine is included prior to the start of the protein of interest which allows cleavage by cyanogen bromide (CNBr) to remove the TrpLE and histidine tag. Samples prepared for ultracentrifugation or SPR analysis were grown in LB, while samples for NMR were grown in modified M9 minimal media (see appendix A) supplemented with the appropriate stable isotope labels. Generally, 1 liter cultures were inoculated to an OD$_{600}$ of ~0.05 from a 50 ml LB overnight growth. The cultures were grown at 37° C and induced with 1mM IPTG at an OD$_{600}$ of ~ 0.4-0.6. The cultures were induced for 4 to 6 hours and then harvested by centrifugation.

Cell pellets were lysed by incubating in Bugbuster (Pierce Scientific) reagent with 100 µg/mL lysozyme and 200 µg/mL DNase for 30 minutes. Cell clumps were broken up by sonication using a Misonix 3000 sonicator equipped with a microtip. After lysis, the inclusion bodies were separated from the soluble protein by centrifugation at
15,000xg for 30 minutes. Residual soluble proteins were removed by a second wash with Bugbuster and subsequent centrifugation. The washed inclusion bodies were dissolved in Ni wash buffer (see Appendix A). Any remaining insoluble components were removed by centrifugation at 15,000xg for 30 minutes. Fusion protein was eluted from the column using Ni elution buffer (see Appendix A). The eluted sample was stored overnight at 4°C in the Ni elution buffer with 1% B-Mercaptoethanol (B-ME), which prevented the oxidation of methionines. To concentrate and desalt the fusion protein, it was dialyzed against ddH₂O until the protein precipitated. The precipitated protein was pelleted by centrifugation at 15,000xg for 20 minutes, washed with ddH₂O, and dried under vacuum.

TrpLE was cleaved from the fusion protein by incubation with 1 M CNBr in 70% trifluoroacetic acid (TFA) for 2 hours at a final fusion protein concentration of 10 to 20 mg/mL. The cleavage reaction was stopped by drying the sample under vacuum until all liquid was removed. The cleavage reaction products were dissolved in Ni wash buffer and the pH readjusted to 8.0. Cleaved MPER constructs were separated from free TrpLE and un-cleaved fusion protein using a Ni-sepharose column under the same conditions as above. Cleaved MPER constructs will not bind the Ni-sepharose column and will be present in the flow through and wash, while uncleaved fusion protein and cleaved TrpLE retain the 6-histidine tag, and therefore elute with 250 mM imidazole. This procedure generally resulted in > 95% pure MPER constructs.
MPER constructs in Ni buffer were desalted by dialysis against ddH₂O for 12-18 hours with at least 2 changes of the water. The MPER constructs often precipitated during dialysis. Precipitated protein was dissolved in the appropriate detergent and buffer. For occasions where the protein did not precipitate, it was either precipitated by the addition of buffer or detergent was added prior to the buffer and the sample concentrated using an appropriate MW cutoff spin concentrator (Vivaspin). There was no difference between the precipitated and non-precipitated protein observed by NMR, therefore precipitation was used as an alternative method to the spin concentrator to concentrate the samples. Samples for ultracentrifugation and NMR were prepared using these procedures.

Liposomes were prepared from a chloroform solution containing 5 mg of 1,2-dimyristoyl-sn-glycero-3-phosphocholine (DMPC). The chloroform was lyophilized leaving dried lipid which was resuspended in 1 mL of ddH₂O, vortexed, then extruded through a 0.1 micron membrane to form small unilamellar liposomes. These liposomes were mixed 10:1 with 1.5 mg/mL purified MPER construct in Ni wash buffer. Liposome samples were dialyzed against ddH₂O for 12 – 18 hours with at least 2 changes of the water. Dialyzed liposomes were sonicated using a Misonix 3000 probe tip sonicator, and then extruded a final time through a 0.1 micron membrane following dialysis. The liposomes were analyzed immediately.
The oligomerization state of MPER constructs in detergent micelles was determined by equilibrium analytical ultracentrifugation. The method of density matching introduced by Tanford and Reynolds [111,112] was used to remove the contribution of the detergent to the sedimentation equilibrium data. The solvent density was adjusted to be the reciprocal of the partial specific volume ($\bar{\nu}_d$) of the detergent by the addition of D$_2$O and D$_{2}^{18}$O, taking into account the additional density change from the buffer. All ultracentrifugation experiments were performed on a Beckman XL-A ultracentrifuge equipped with absorbance optics. Protein concentration was monitored by absorbance at 280 nm. Data were collected at 25° C and three rotor speeds, 16,000, 20,000, and 24,000 rpm. The best fit buoyant molecular weight ($W_b$) was calculated using XL-A IDEAL 1 software. Buoyant molecular weights were related to the molecular weight of the protein ($P$) according to the following equation.

$$W_b = P(1 - \bar{\nu}_p \cdot \rho)$$  \hspace{1cm} (1)

Solvent density ($\rho$) and partial specific volume of the protein ($\bar{\nu}_p$) were calculated using SEDNTERP (J. Philo, D.B. Hayes, and T.M. Laue, University of New Hampshire). No correction to $\bar{\nu}_p$ for the small change in mass expected for hydrogen/deuterium exchange was made. To obtain equilibrium binding constants for trimerization, several datasets at two or three rotor speeds were globally fit to various equilibrium models using HeteroAnalysis (J. L. Cole and J.W. Lary, University of Connecticut).
To estimate the amount of detergent bound to the construct, we used an extrapolation method based on a rearrangement of equation 1. The buoyant weight of a protein detergent complex \( W_{bc} \) can be expressed as the sum of the buoyant weight terms for each component.

\[
W_{bc} = D(1 - \bar{\nu}_d \cdot \rho) + P(1 - \bar{\nu}_p \cdot \rho) \tag{2}
\]

In equation 2, \( D \) is the total mass of the bound detergent, \( P \) is the total mass of the protein, \( \bar{\nu}_d \) and \( \bar{\nu}_p \) are the partial specific volumes of the detergent and protein respectively, and \( \rho \) is the solvent density. This equation can be rewritten in linear form as equation 3, where the y-intercept corresponds to the total molecular weight \( (MW_c) \) of the complex.

\[
W_{bc} = MW_c - \rho(\bar{\nu}_d \cdot D + \bar{\nu}_p \cdot P) \tag{3}
\]

This analysis was only performed for the most promising construct, gp41-M-MAT in DPC detergent micelles. Buoyant weights were determined at several solvent densities and plotted as a function of solvent density. The data were fit to a linear equation using Kaleidagraph, and the molecular weight of the complex was estimated from the y-intercept of the line (shown in Figure 34). The total mass of the bound detergent could be estimated by taking the difference between the total mass of the complex determined from the y-intercept of equation 3 and the total protein mass determined from the density matched experiments described above.
For some detergents, the partial specific volume had not been determined. Therefore, we determined the partial specific volume for these detergents using two methods. The first method utilized density matching and analytical ultracentrifugation. The density of the solvent was varied using D$_2$O and D$_2$H$^{18}$O. The buoyant molecular weight was plotted as a function of density. Inspection of equation 1 reveals that if the buoyant molecular weight is equal to zero, then solvent density is equal to the reciprocal of the partial specific volume. The data were fit to a linear equation and the reciprocal of the x-intercept was taken to be the partial specific volume of the detergent. We also measured the partial specific volume of these detergents using a density meter. The density meter was calibrated using the density of air and the density of H$_2$O. The temperature of the density meter was maintained at 25° C using a recirculating water bath with insulated tubing. Proper operation and calibration of the instrument was confirmed by determining the density of D$_2$O. The densities of several concentrations of the detergent in H$_2$O were determined, and the data was plotted as a function of detergent concentration. The partial specific volume is related to the slope of the density vs. concentration plot according to the following equation[113].

$$ \lim_{c \to \text{cmc}} \left( \frac{\partial \rho}{\partial c} \right) = 1 - \bar{v} \rho_{H2O} $$

(4)

Where $\rho_{H2O}$ is the density of H$_2$O, $\bar{v}$ is the partial specific volume of the detergent, and the limit is the slope of the line as it approaches the critical micelle concentration of the detergent. The slope is obtained by performing a least squares fit of the data to a linear
equation. As a control, we determined the partial specific volume of DPC to be 0.944 cm$^3$/g, which is in good agreement with the reported partial specific volume of 0.937 cm$^3$/g [114].

Binding of antibodies to gp41-M-MAT conjugated to DMPC liposomes (M-MAT-liposomes) was assayed using surface plasmon resonance following a protocol previously used to measure the MPER peptide liposome binding of 2F5 and 4E10 antibodies [84,95]. Briefly, M-MAT-liposome conjugates were immobilized (~500 RU) on a Biacore L1 chip that had been conditioned by immobilizing ~2500-3000 RU of BSA. Monoclonal antibodies (MAb) were flowed over the chip for 2 minutes at a flow rate of 20 µL/min. MAb’s were allowed to dissociate for 5 minutes while buffer continued to flow at 20 µL/min. For gp41-M-MAT in DPC detergent micelles, the antibodies were immobilized on a Biacore CM-5 chip to a final capture level of 1000 RU. Detergent solubilized gp41-M-MAT was flowed over the chip for 2 minutes at a flow rate of 30 µL/min. The construct was allowed to dissociate for 250 seconds while buffer continued to flow at 30 µL/min. Assays were carried out on a Biacore 3000 instrument and data were fit using the BIAeval 4.1 software (Biacore).

NMR experiments were conducted at 600, 800, or 950 MHz. 2D TROSY HSQC experiments were used to initially determine feasibility of further NMR experiments. In cases where the 2D spectrum was of sufficient quality, we proceeded to assign the backbone resonances using standard triple resonance experiments, such as the HNCA
and HNCACB. For some constructs, we found a recently developed hNCAnH to be particularly helpful for assignment [115].

4.2 Constructs

Several constructs were screened as part of this work, however only three were found to express well in *E. coli*. Those constructs are shown in schematic form in Figure 31. All of the constructs contain an N-terminal trimerization domain. The two trimerization domains used on the N-terminus of our constructs were GCN4 and foldon from T4 fibritin. GCN4 is an isoleucine zipper that adopts a three-helix bundle structure [116]. We used a modified GCN4 sequence that has been previously used to trimerize a 17 amino acid fragment of HR-1 from HIV-1 gp41. The resulting GCN4/HR-1 trimer was extremely stable, having a Tm of >100° C [65]. The foldon domain of bacteriophage T4 fibritin is a 27 amino acid sequence that rapidly folds into a β-propeller structure and is believed to nucleate the formation of fibritin fibers [117]. The foldon domain has previously been used to trimerize and form several types of fibers, as well as trimerize HIV-1 gp41 and gp140 [86,118]. The atomic resolution structures of each of these trimerization domains have been determined and are shown in Figure 32.

In addition to the N-terminal trimerization domains, two of the three constructs incorporated a C-terminal transmembrane trimerization domain. This domain was engineered from GCN4 by inverting the hydrophobic and hydrophilic residues, resulting in a hydrophobic exterior and a hydrophilic interior that contained amino acid
residues that could form a hydrogen bonding network [119,120]. This peptide has been shown to trimerize in the presence of SDS micelles, as well as other detergents [121,122]. We incorporated this peptide into our designs in order to provide a stable trimeric membrane anchor. In addition, it had been shown that the point at which the peptide entered the membrane could be controlled by the location of arginine residues in the sequence [121]. This could provide a valuable tool to reposition the MPER relative to the lipid membrane.

Figure 31. Schematic of MPER Constructs
A. Sections in gray are removed by cyanogen bromide cleavage. GSSG refers to the sequence gly-ser-ser-gly. B. The peptide sequences for each block are shown. The epitopes for 2F5 and 4E10 are shown in the MPER sequence in red and blue respectively. In the MPER sequence, the residues in lower case were included in FR2 and gp41-M-MAT.
4.3 Analytical Ultracentrifugation

We screened the constructs for trimerization in several detergents. The detergents screened were dodecylphosphocholine (DPC), 1-myristoyl-2-hydroxy-\textit{sn}-glycerol-3-phosphocholine (LMPC), 1-myristoyl-2-hydroxy-\textit{sn}-glycerol-3-phosphoglycerol (LMPG), and 1,2-dihexanoyl-\textit{sn}-glycerol-3-phosphocholine (DHPC). DPC is commonly used for the study of membrane proteins by NMR spectroscopy [100]. However, it has been noted that care must be taken when using DPC because it is somewhat more denaturing than other detergents [100]. Lyso-lipid based detergents like LMPG and LMPC are generally less denaturing than DPC and may be less likely to lead to non-native protein structures. A recent study of the solubility of several membrane proteins in various detergents has shown that lyso-lipid detergents can lead
to increased solubility and stability of membrane proteins [123]. DHPC has been used in
the study of membrane proteins by NMR spectroscopy [124]. The primary advantage of
DHPC is that it is essentially identical to the phospholipids that are the most abundant
component of mammalian cell membranes, differing only in its shortened acyl chain
length. The detergents were screened for solubility of the protein and NMR spectral
quality.

The trimerization state of the peptides was determined by analytical
ultracentrifugation. We employed the method of Tanford and Reynolds that utilizes
density matching to eliminate the contribution of the detergent to sedimentation. As
discussed in the experimental procedures section of this chapter, this method requires
prior knowledge of the detergent partial specific volume. The partial specific volumes
of some of the detergents used in these studies had not been reported. We
experimentally determined the partial specific volumes for these detergents, which is
discussed in section 4.3.2.

4.3.1 Trimerization of MPER Peptides in Various Detergents

Initial estimates of the trimerization state of the various peptide constructs were
obtained from equilibrium analytical ultracentrifugation (AUC) by fitting the data to
equation 1, which represents a single ideal species model. If there is more than one
species present, the value obtained from this fit will represent a mass average of all
species in solution. Thus, the oligomerization state can be estimated by noting how
close the mass average molecular weight is to either the expected molecular weight of
the monomer or of the trimer. An intermediate result indicates that multiple species are
significantly populated. We used AUC to determine the oligomerization state of 2R2
and gp41-M-MAT. We did not use AUC to study FR2 because the NMR spectrum of
FR2 was of poor quality, which will be discussed later.

In most cases, the mass averaged molecular weight increased with decreasing
detergent concentration (see Table 1). The expected trimeric molecular weights of 2R2
and gp41-M-MAT were 25 kD and 21 kD, respectively. In general, the extent of
trimerization increased as the ratio of detergent micelles to putative trimers approached
one. The concentration of the micelles was estimated based on the aggregation number
obtained from either our experimental results or published data. It should be noted that
the aggregation number may change upon interaction with the peptide, which would
alter the actual number of micelles in solution. If the number of micelles were lower
than expected, then we may see particles with more than three peptides per micelle. We
noted a significant increase in apparent molecular weight of the 2R2 peptide between
12.88 mM LMPC and 6.44 mM LMPC, which may indicate more than three peptides per
micelle. No further analysis was conducted on constructs in DHPC because we found
that the constructs were either not soluble or formed large aggregates in this detergent.

2R2 (see Figure 31) in LMPG had a molecular weight near that expected for a
trimer at a detergent to protein trimer ratio near 1:1 (1.64 mM LMPG). At higher
detergent to protein ratios, the molecular weight measured was consistent with a significant monomer population (3.08 mM LMPG), indicating that the construct did not form a stable trimer. In LMPC, 2R2 remained primarily a trimer over a wider range of detergent concentrations, but also populated the monomeric state at high concentrations of detergent (50 mM). 2R2 in DPC did not seem to change significantly with detergent concentration, and the molecular weight observed was suggestive of a significant monomer population. These data are summarized in Table 1.

<table>
<thead>
<tr>
<th>Detergent</th>
<th>Concentration (mM)</th>
<th>Molecular Weighta</th>
</tr>
</thead>
<tbody>
<tr>
<td>LMPC</td>
<td>6.44</td>
<td>40719</td>
</tr>
<tr>
<td>LMPC</td>
<td>12.88</td>
<td>24065</td>
</tr>
<tr>
<td>LMPC</td>
<td>19.22</td>
<td>22248</td>
</tr>
<tr>
<td>LMPC</td>
<td>50</td>
<td>12148</td>
</tr>
<tr>
<td>LMPG</td>
<td>1.64</td>
<td>27214</td>
</tr>
<tr>
<td>LMPG</td>
<td>3.08</td>
<td>15667</td>
</tr>
<tr>
<td>DPC</td>
<td>2.5</td>
<td>12464</td>
</tr>
<tr>
<td>DPC</td>
<td>5</td>
<td>13625</td>
</tr>
<tr>
<td>DPC</td>
<td>10</td>
<td>13711</td>
</tr>
</tbody>
</table>

aMonomer, Dimer, and Trimer weights for 2R2 are 8.3 kD, 16.7 kD, and 25 kD respectively

We further analyzed the AUC data by fitting the data to various equilibrium models. Construct 2R2 in LMPC fit well to a monomer to trimer equilibrium model. At a ratio of ~2:1 detergent micelles to putative trimer, the best fit Ka was 1.95e10 M^-2. The Ka dropped by two orders of magnitude to 4.95e8 M^-2 when the detergent micelle to trimer ratio was increased to ~3:1. Since the results in LMPG showed that trimers were
only formed at very low concentrations of detergent, we did not attempt to calculate a
Ka.

Gp41-M-MAT trimerized well even in high concentrations of DPC (100 mM). However, it is clear that while most of the construct is trimerized, the population of monomer is increasing as the detergent concentration increases, as evidenced by the decrease in the mass averaged molecular weight. The overall molecular weights observed for gp41-M-MAT in DPC micelles is summarized in Table 2. Further analysis of this data by fitting to various models revealed that the construct fit well to a monomer to trimer equilibrium model. The Ka’s obtained varied with the concentration of detergent, similar to the data obtained for construct 2R2 in LMPC micelles, although the Ka’s for gp41-M-MAT were higher. The calculated Ka’s are plotted as a function of detergent concentration in Figure 33. The equilibrium dissociation constant of the foldon domain has been reported to be 7.4e-16 M^-2 [125]. The equilibrium constant for our construct is somewhat lower than that of foldon by itself, and is dependent upon the detergent concentration. At low detergent concentration, however, the gp41-M-MAT construct exhibits highly favorable trimerization characteristics, with a dissociation constant of 1.6e-13 M^-2. We did not study gp41-M-MAT in the other detergents because it was clearly trimeric in DPC. DPC is a commonly used detergent for NMR spectroscopy and perdeuterated DPC is readily available, making it a good choice for further NMR studies on gp41-M-MAT.
<table>
<thead>
<tr>
<th>[DPC] mM</th>
<th>Molecular Weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.5</td>
<td>21581</td>
</tr>
<tr>
<td>10</td>
<td>20099</td>
</tr>
<tr>
<td>100</td>
<td>18556</td>
</tr>
</tbody>
</table>

Variation of the apparent association constants with detergent concentration has been observed for some transmembrane peptides. In detergent solutions, transmembrane domains are confined to the detergent micelle phase, and are not found at high concentrations free in solution. Thus, the available volume in which the transmembrane peptide can remain soluble is only that of the detergent micelles, and not the bulk solvent. The effective concentration of the transmembrane peptide in the

Figure 33. Trimerization Depends on Detergent Concentration
Data were fit to a monomer to trimer equilibrium model and the resulting $K_a$ plotted as a function of DPC concentration. Concentration of gp41-M-MAT ranged from 20 μM to 50 μM.
detergent phase is different from the concentration as calculated from the total solution volume. Since association constants are generally calculated from concentrations based upon total solvent volume, these association constants will be dependent on the detergent’s effect on the effective protein concentration. Studies conducted on the transmembrane peptide SNGpA99 have been used to confirm this observation, and have shown that the results can be explained by the simple dilution of the peptide in the detergent phase [126]. We did not collect sufficient data for a similar quantitative analysis. However, our observation that the association constants increase with decreasing detergent concentration suggests that the association of gp41-M-MAT with the surface of the detergent micelles may contribute to our results via a similar mechanism to that suggested for dimeric transmembrane peptides.

The foldon trimerized constructs both showed evidence of stable trimer formation, while the GCN4 constructs were notably less stable in detergent micelles. This provides further evidence that the T4 fibritin foldon domain is a useful tool to form protein trimers. It has been employed to form several trimeric constructs, including fibers and gp41 trimers [86,118]. We show that the use of foldon as a trimerization domain can be extended to membrane associated proteins in detergent micelles. While the observed dissociation constant was lower for gp41-M-MAT than foldon alone, the foldon still conferred strong trimer association in the presence of detergent micelles.
Thus, the foldon domain is suitable for facilitating the formation of stable, detergent soluble protein trimers.

Finally, we estimated the amount of detergent bound to the gp41-M-MAT trimer by varying the density of the solvent. Tanford and Reynolds have noted that the buoyant molecular weight of a detergent protein complex ($W_{bc}$) can be written as the sum of the two components, as shown in equation 3. This equation can be rearranged into a new form, shown in equation 4. It is obvious that $W_{bc}$ is related in a linear fashion to the solvent density. This approach has been previously employed to estimate the number of detergent molecules bound to a small peptide [74]. A similar approach has been proposed to determine the partial specific volume of a detergent-protein complex [127]. When fit to a linear equation, the y-intercept of the equation will give an estimate of the total mass of the complex. Since we had already determined that the protein was a trimer with a molecular weight of 21 kD, we can subtract the trimeric mass of the protein from the total mass of the complex, yielding the total mass of detergent in the complex. We performed these experiments at two different detergent concentrations and found that the amount of detergent bound was similar at both detergent concentrations. Plots of these data, along with linear curve fits, are shown in Figure 34. The total complex mass in 100 mM and 10 mM detergent was 43,015 daltons and 40,765 daltons, which are within 5.5% of each other and near the error of equilibrium analytical ultracentrifugation methods. If we take the average of the above complex molecular
weights (41,890 daltons) and subtract the molecular weight of the gp41-M-MAT protein trimer, we arrive at 20,872 daltons of detergent bound to the trimer. This corresponds to ~60 detergent molecules bound per trimer, and is within the range of 50-60 molecules that corresponds to the reported aggregation number of a DPC micelle. Thus, we conclude that gp41-M-MAT binds approximately one micelle of DPC per protein trimer.

Figure 34. Total Mass of Detergent Bound to gp41-M-MAT
Buoyant molecular weights are plotted as a function of solvent density. Data were fit to equation 3, where the terms in parentheses are treated as a constant. The resulting y-intercept yields the total molecular mass of the detergent-protein complex. A. Plot and subsequent fit of data obtained at 16,000 RPM and 100 mM DPC detergent. B. Plot and subsequent fit of data obtained at 16,000 RPM and 10 mM DPC detergent. In both plots, the linear fit and equation are shown in red.
4.3.2 Determination of Partial Specific Volume of Detergents

The partial specific volume of a solute is defined as the change in volume of the solution associated with the addition of one gram of solute [113]. This value has been reported for some detergents, however it has not been reported for LMPG or DHPC [114]. Therefore, we determined the partial specific volumes for LMPG and DHPC detergents using either analytical ultracentrifugation or densimetry. We also determined the partial specific volume for LMPC using both methods when it became obvious that the published value was not correct for our experimental conditions.

To determine the partial specific volume using ultracentrifugation, we measured the buoyant molecular weights of micellar solutions at various solvent densities. It is clear from equation 1 that the solvent density is equal to the reciprocal of the partial specific volume when the buoyant molecular weight is zero. To accurately determine the density where the buoyant molecular weight was zero, we fit the data to a linear equation and determined the x-intercept. Plots of these data are shown in Figure 35. The partial specific volumes obtained from these data for LMPC and LMPG were 0.896 cm$^3$/g and 0.829 cm$^3$/g, respectively. The published partial specific volume of LMPC was 0.97 cm$^3$/g, which is significantly different from the value obtained in our experiments [114]. Finally, the Y-intercept of our curve fit provided an estimate of the total weight of the micelle. We estimated the aggregation number of two of the detergents by dividing
We also determined the partial specific volume of these detergents using densimetry. As noted in equation 4 of the experimental methods section, the partial specific volume of the detergent micelles is related to the limiting slope of the density as a function of detergent concentration as the detergent concentration approaches the critical micelle concentration (CMC). It should be noted that the density need not change linearly with concentration, however, in our case the density did change linearly with concentration over the range of concentrations where we conducted our studies.
As a control, we also determined the partial specific volumes of DPC by densimetry. Gratifyingly, the value of 0.945 cm$^3$/g is in good agreement with the published value of 0.937 cm$^3$/g [74].

The partial specific volumes obtained from these two sets of experiments are listed in Table 3. The data, along with the resultant curve fit for each experiment, is plotted in Figure 36. As noted above, we found that the previously reported partial specific volume of LMPC did not agree with our experimentally determined value. Importantly, we found good agreement between the two independent experimental methods we used to determine the partial specific volume of LMPC. Therefore, we have concluded that our experimentally determined value is a more accurate measurement of the actual partial specific volume of LMPC, in the conditions used for our experiments.

<table>
<thead>
<tr>
<th>Detergent</th>
<th>$\bar{\bar{\nu}}$ by AUC (cm$^3$/g)</th>
<th>$\bar{\nu}$ by Densimetry (cm$^3$/g)</th>
<th>Aggregation number</th>
</tr>
</thead>
<tbody>
<tr>
<td>DPC</td>
<td>n.d. [0.937]$^a$</td>
<td>0.945</td>
<td>n.d. [50-60]$^a$</td>
</tr>
<tr>
<td>DHPC</td>
<td>n.d.</td>
<td>0.876</td>
<td>n.d. [19-35]$^b$</td>
</tr>
<tr>
<td>LMPC</td>
<td>0.896</td>
<td>0.909</td>
<td>85</td>
</tr>
<tr>
<td>LMPG</td>
<td>0.824</td>
<td>0.818</td>
<td>45</td>
</tr>
</tbody>
</table>

$^a$ value previously reported in [114], $^b$ value previously reported in [100], n.d.- not determined
Figure 36. Partial Specific Volumes Determined by Densimetry
Plots show the sample density determined with an Anton-Paar DMA-02C densimeter vs. detergent concentration for LMPG (A), DHPC (B), LMPC (C), and DPC (D). Data were fit to a linear equation and the slope used with equation 4 to determine the partial specific volume of each detergent.
4.4 Screening for Feasibility of Structural Analysis by NMR

2D $^{15}$N-TROSY HSQC spectra of 2R2 in various detergents showed that LMPG gave rise to the best spectrum. Only a small number of broad peaks were observed in LMPC. The DPC spectrum yielded a smaller number of peaks than the LMPG spectrum, although the peaks observed were more intense. The spectra from these experiments are shown in Figure 37. Notably, the detergent concentrations used in these initial experiments were high and resulted in a significant amount of the monomeric form of the peptide as shown by analytical ultracentrifugation. LMPC was the only detergent that favored the trimerization of 2R2 over a significant detergent concentration range. Therefore, we also collected NMR spectra in LMPC at detergent to protein ratios that favored trimerization. As shown in Figure 38, the spectrum from this sample was of poor quality, and would not support further structural characterization.

Nevertheless, the spectrum of 2R2 solubilized in LMPG was interesting, and we decided to further investigate the monomeric form of this peptide with NMR spectroscopy. The spectrum was assigned using a combination of 3D-TROSY HNCA,
Figure 37. HSQC Spectra of 2R2 in Various Detergents
HSQC spectrum of 2R2 in LMPG (A), DPC (B), and LMPC (C).
HNCACB, and hNCAnH experiments. A strip plot from an HNCA on 2R2 in LMPG micelles is shown in Figure 39-A. The hNCAnH experiment is a recently developed experiment useful for assigning spectra with poor chemical shift dispersion, such as α-helical peptides and proteins [115]. This experiment gives rise to a unique pattern of positive and negative resonances that make assignment relatively straightforward. We found this experiment to be particularly useful for this assignment problem. Figure 39-B shows a plane from this experiment with the resonances for L29 and D30 labeled.

Only 75% of the backbone amide resonances for 2R2 in LMPG were observed. Of these, ~90% were assigned using the above 3D experiments. The missing resonances were all in the MPER region, between the 2F5 epitope and the transmembrane trimerization domain. Chemical shift indexing revealed that the GCN4 and transmembrane sequence both adopted helical conformations as expected.
Figure 39. Sample Spectra for Assignment of 2R2 in LMPG
A. Strip plot from an HNCACB experiment showing residues E28-A33, including the core DKW of the 2F5 epitope. B. Plane from an hNCAnH experiment showing the assignment of D30 and L29.

(see Figure 40). The MPER sequence N-terminal of the 2F5 epitope tended to be less helical in nature, while the 2F5 epitope itself appeared more helical. This data is
consistent with our observations that the MPER peptide tends to adopt a helical conformation, but that there can be considerable variation. Our experiments were conducted at pH 7.0. Previous studies have demonstrated that the region N-terminal of the 2F5 epitope can adopt a helix at low pH but also can adopt a more extended conformation at higher pH [105]. The 2F5 epitope tended to adopt a helical conformation at all pH’s previously studied [97,105,106]. Our observation shows that

![Chemical Shift Indexing of 2R2 in LMPG](image)

**Figure 40. Chemical Shift Indexing of 2R2 in LMPG**

Chemical shift indexing using Cα (A), and ΔCα-ΔCβ (B). Values greater than 1 indicate helix, while values less than -1 indicate sheet. * indicates no Cβ value available. The black line indicates that resonances were not observed for those residues.
the 2F5 epitope in LMPG solubilized 2R2 is likely to adopt a helical conformation consistent with previous structures of other MPER constructs. Notably, other studies have been conducted using the neutral detergent DPC, while LMPG is an anionic detergent. The HIV-1 viral membrane has been shown to contain a significant percentage of the anionic lipid phosphatidylserine [128], making the structure of the MPER epitopes in an anionic environment of some importance. Our data from this study and the previous chapter suggest that the 2F5 epitope can adopt a helical conformation in both an anionic and neutral lipid environment.

FR2 and gp41-M-MAT (see Figure 31) both use foldon from T4 fibritin as a soluble trimerization domain. This domain has been previously studied by NMR and a structure of the trimerized form has been determined [129]. Importantly, the trimerized and folded monomer forms of this domain have also been studied by NMR, and 2D HSQC spectra have been reported for both folded monomer and trimer [130]. The 2D HSQC spectra of the monomer and trimer differ from each other, and can be used as an independent confirmation of trimerization. The characteristic spectrum of trimerized foldon was clearly observed in FR2 (see Figure 41). Unfortunately, we did not observe the 82 expected resonances to account for the entire FR2 construct. Since we did not observe enough resonances in this construct, we did not pursue further studies on it, even though it was trimeric.
The final peptide we studied was gp41-M-MAT. This peptide is unique from the other designs in that it relies upon the natural lipophilicity of the MPER sequence to bind lipids, rather than incorporating an additional transmembrane anchor. Our analytical ultracentrifugation data indicated that this construct formed a stable trimer in DPC micelles. The C-terminus of the MPER in gp41-M-MAT is not restrained by a strong trimerization domain, unlike gp41-inter and HA-gp41. Thus, while the analytical ultracentrifugation shows that the construct is overall a trimer, it does not guarantee that the C-terminal region of gp41-M-MAT is closely associated. The strong trimerization

![Figure 41. Trosy HSQC Spectrum of FR2](image)

$^{15}$N-TROSY spectrum of 420 uM FR2 in 100 mM LMPC, 50 mM sodium phosphate pH 6.0. The red box indicates some of the resonances that are most characteristic for the trimeric form of foldon.
domains incorporated at the C-terminus of other MPER trimers, like gp41-inter and HA-gp41, are likely to force the C-terminus of MPER into a tight trimer.

To determine the feasibility of further structural studies of gp41-M-MAT we collected a 2D $^{15}$N TROSY spectrum on this sample in DPC micelles. This spectrum is of good quality as shown in Figure 42. Like the FR2 design, this construct also exhibited the characteristic resonances of the trimerized foldon domain. Unlike FR2, most of the resonances for the MPER are observed in this construct. In addition, we did not observe

![Figure 42. TROSY Spectrum of gp41-M-MAT in DPC Micelles](image)

$^{15}$N-TROSY spectrum of 1 mM gp41-M-MAT in 100 mM DPC, 50 mM sodium phosphate pH 6.0. The red box indicates some of the resonances that are most characteristic for the trimeric form of foldon.
significantly more than the predicted number of amide resonances for gp41-M-MAT, indicating that the construct behaves as expected for a symmetric homotrimer. If the construct were not symmetric, we would expect to observe multiple peaks for each backbone amide since the lack of symmetry would likely result in somewhat different chemical environments. The high quality of this spectrum suggested that it would be feasible to determine the structure of this construct with high resolution multidimensional NMR spectroscopy, which will be the topic of chapter 5.

4.5 Measurement of 2F5 and 4E10 Binding by SPR

Of the constructs designed, only gp41-M-MAT satisfied the design goals of lipid binding, strong trimerization and good NMR spectra. The final design criteria that our construct needed to satisfy was tight binding to the broadly neutralizing antibodies 2F5 and 4E10. This is especially important for trimerized MPER constructs, since at least one other MPER trimer did not bind the neutralizing antibodies [107]. Thus, it is important that we confirm that our construct does bind with high affinity to 2F5 and 4E10.

2F5 and 4E10 bound the M-MAT-liposomes with high affinity. The data were best fit to a two-step encounter docking model, with overall Kd’s of 0.18 nM and 27 nM for 2F5 and 4E10 respectively (see Figure 43). M-MAT-liposomes bound 2F5 and 4E10 with similar affinity when compared to the MPER peptide-liposome conjugates, which are derived from monomeric peptides [84]. The Kd’s observed for M-MAT-liposomes binding 2F5 and 4E10 were also similar to those observed for other soluble gp41 trimer
constructs [86,87]. Specific binding to 2F5 was observed for gp41-M-MAT in DPC detergent micelles. The calculated binding affinity was between 200 and 300 nM, however the quality of the data and resulting curve fit was poor. This may be due to excess DPC micelles adversely affecting the assay, or alterations to the antibody caused by its direct conjugation to the CM5 Biacore chip.

Previous studies of MPER peptide-liposome conjugates have also found that the binding of 2F5 and 4E10 were best fit with a two-step encounter-docking model [84,95]. In contrast, MPER peptides that are not displayed on lipid membranes do not exhibit two-step binding [95]. This observation has led to the hypothesis that the first step of MPER antibody binding is an initial interaction with the lipid membrane, followed by

Figure 43. SPR sensograms of 2F5 and 4E10 binding M-MAT-liposomes
Titrations of antibodies with M-MAT-liposomes to determine the Kd of 2F5 (A) and 4E10 (B). Data were fit to a two-step encounter-docking model, indicated by the dashed red line. The Kd’s calculated for 2F5 and 4E10 were 0.18 nM and 27 nM respectively.
tight binding to the peptide epitope [82]. Thus, the binding kinetics of M-MAT-liposomes with MPER antibodies is consistent with this hypothesis, although other models, such as peptide extraction from the lipid bilayer, have also been proposed to explain the two-step kinetics [106].

4.6 Summary

In summary, we have developed several HIV-1 gp41 derived peptides that included the MPER region. These peptides were designed to trimerize and interact with lipid membranes and detergent micelles, as well as be useful models for further biophysical characterization of the MPER. Of the constructs studied, gp41-M-MAT was found to satisfy all of our design criteria. It bound to 2F5 and 4E10 with sub-nanomolar and nanomolar affinity when displayed on DMPC liposomes. Importantly, it formed well defined trimers in DPC micelles as evidenced by analytical ultracentrifugation. The construct also showed promise for further structural characterization by NMR spectroscopy.
5. Structure of GP41-M-MAT

The MPER of HIV-1 gp41 contains the peptide epitopes for several broadly neutralizing antibodies against HIV-1, including 2F5 and 4E10. HIV-1 gp41, along with gp120, is part of the trimeric ENV complex located on the surface of the HIV-1 virion. It has been suggested that the broadly neutralizing antibodies that target the MPER bind to a pre-fusion intermediate state exposed during the process of host cell recognition and viral membrane fusion [86]. The role of MPER, both in viral fusion and as a target for broadly neutralizing antibodies, has made it the subject of extensive biochemical and immunological studies.

Several cryo-electron microscopy (cryo-EM) studies on Simian Immunodeficiency Virus (SIV) have resulted in proposed overall architectures for the ENV complex and possible conformations for the MPER. However, there is limited agreement between the various cryo-EM reconstructions in the region of the MPER. In one cryo-EM tomographic reconstruction, the MPER is splayed out with little or no apparent interaction between the individual gp41 subunits of the trimer [131]. Subsequent cryo-EM studies have proposed structures where the MPER is closely associated, with no evidence for significant separation between the gp41 subunits near the membrane interface [132,133]. Three of the current cryo-EM structures are shown in Figure 44.
Atomic resolution structures of the MPER have generally been limited to short MPER peptides, which have not been shown to trimerize. These structures indicate that the MPER adopts a helical conformation in the presence of detergent micelles [97,105,106]. Two of the reported MPER structures are shown in Figure 45 A and B [97,106]. While all of these structures display varying degrees of helicity, it is clear that there are significant differences. These structures were determined under different conditions.
sample conditions, which may give rise to the observed conformational heterogeneity.

In addition, these structures utilize short monomeric peptides that are free to adopt more conformations than may be expected if they were restrained in a trimer.

Crystal structures of trimerized forms of the MPER have been reported. At least two six-helix bundle structures that include some or all of the MPER sequence have been published [57, 61]. In these six-helix bundle structures, most of the MPER is a continuous
helix, with some deviation in the last few residues of the C-terminus [57]. Interestingly, in a six-helix bundle that included both the fusion peptide proximal region and the MPER, the MPER is splayed outward, leading to the hypothesis that it may induce strain in the membrane to facilitate the fusion process [57]. However, six-helix bundle constructs have been shown to interact weakly with either 2F5 or 4E10, suggesting that the conformation displayed in these structures may not be conducive for the induction of 2F5 or 4E10-like antibodies [86]. Six-helix bundle targeted antibodies generally do not neutralize HIV-1, although they may be able to mediate antibody dependent cell cytotoxicity [61,134,135]. Another reported MPER crystal structure used a C-terminal GCN4 trimerization sequence to facilitate the formation of an MPER trimer [107]. In this structure, the MPER adopts a coiled-coil motif that is somewhat separated in the MPER region (see Figure 45 C). However, almost no binding of this construct to either 2F5 or 4E10 was observed [107]. It was postulated that this construct may represent the conformation adopted by the MPER in the pre-triggered ENV complex. Subsequent studies have suggested that the MPER does adopt a conformation capable of 2F5 and 4E10 binding in the pre-triggered ENV, but that other ENV components block antibody access to the MPER [89]. Thus, it is unclear if this trimeric MPER structure represents the pre-triggered gp41 state.

In the previous chapter, we characterized a novel membrane associated, trimeric HIV-1 gp41 MPER trimer called gp41-M-MAT [136]. This construct exhibited stable
trimerization, high expression levels in *E. coli*, tight binding to the broadly neutralizing MPER antibodies, and was well folded in DPC micelles. These characteristics make this construct suitable for characterization with NMR spectroscopy. In this chapter, we will present the structure of this construct, solubilized in DPC micelles, as determined by multi-dimensional heteronuclear NMR spectroscopy in solution.

### 5.1 Experimental Procedures

Protein was expressed and purified as previously described in the experimental procedures in chapter 4 of this dissertation. The protocol for deuterated samples was modified to improve *E. coli* cell growth in D$_2$O media. An LB culture was inoculated from a frozen stock of *E. coli* cells transformed with the gp41-M-MAT expression vector and grown at 37° C for ~12 hrs. This culture was diluted 1:100 into LB prepared with D$_2$O substituted for the H$_2$O (LB-D$_2$O). The LB-D2O culture was grown at 37° C for ~12 hrs. The cells were harvested by gentle centrifugation at 1000xg and used to inoculate an appropriately isotope labeled modified M9 minimal media culture made with D$_2$O substituted for H$_2$O and supplemented with 10% labeled Bioexpress media (CIL). Further growth, induction, and purification of gp41-M-MAT was carried out as described in chapter 4. Purified gp41-M-MAT in 6 M guanidine was desalted by dialysis against ddH$_2$O for 12-18 hours with at least two changes of the water. Sodium phosphate, pH 6.0, was added to the desalted protein to a final concentration of 50 mM. This typically resulted in precipitation of the protein, which was centrifuged out at
15,000xg for 15 minutes. The precipitated protein was dissolved in buffer containing 50 mM sodium phosphate, sufficient 100 mM DPC, and 10% D$_2$O. It was often easier to handle the samples when the protein was dissolved in a larger volume of buffer than the minimum required for a Shigemi NMR tube, therefore the concentration of DPC was adjusted such that the final concentration in the Shigemi NMR tube was 100 mM after concentration. Samples were concentrated to the final volume of 350 µL using a 10,000 MW cutoff spin concentrator (Sartorius). This procedure generally resulted in between 1 mM and 2 mM samples of gp41-M-MAT. For experiments where smaller concentrations of protein were required, proportionally less detergent was used.

NMR experiments were conducted on a 600 MHz or 800 MHz Varian Inova spectrometer, or a 950 MHz Bruker Avance spectrometer. The experiments used for backbone assignment were HNCA, HNCO, HN(CO)CA, HNCACB, and hNCAhH. Side chain assignments were made using an HCCH-TOCSY. A small number of $^3$J-coupling constants were acquired from a 3D HNHA experiment. $^1$H-$^1$H NOE based distance restraints were derived from a 4D sparse sampled TROSY-NOESY-TROSY experiment [137] and a 3D NOESY-HSQC, both with 200 ms mixing times. The sparse sampled spectrum was reconstructed using dft4d software (Brian E. Coggins, Duke University). Inter-molecular $^1$H-$^1$H NOE’s were identified by making a mixed sample consisting of uniformly $^{13}$C-labeled protein and uniformly $^3$H, $^{15}$N-labeled protein. A NOESY-HSQC spectrum was collected on this sample with a mixing time of 300 ms and compared to an
identical reference spectrum collected on uniformly $^3$H, $^{15}$N-labeled protein. NOE’s found in the mixed sample, but not in the reference spectrum, were identified as intermolecular. All spectra were processed using nmrPipe [19] and visualized using either CARA (Rochus Keller, ETH Zürich) or nmrviewJ [20].

Residual dipolar coupling data were collected in a 4.5% strained polyacrylamide gel. A 4.5% gel was prepared from a 30% acrylamide stock solution with 29:1 (w/w) N,N’-Methylene-bis-acrylamide crosslinker, in a 6 mm casting tube. Polymerization was initiated by addition of 0.082% (w/v) ammonium persulfate (APS) and 0.64% (v/v) tetramethylethylenediamine (TEMED). Polymerization was allowed to proceed overnight at 30° C. The polymerized gel was rinsed overnight in ddH₂O and equilibrated overnight with sodium phosphate, pH 6.0. gp41-M-MAT in DPC detergent and 50 mM sodium phosphate, pH 6.0, was added to the gel and allowed to soak in over a period of two days. The transport of the protein into the gel could be monitored by the absorbance at 280 nm of the bath protein solution. Protein soaked gels were stretched into a 5 mm open ended NMR tube using a New Era gel stretching apparatus. Residual dipolar couplings (RDC) were measured using a 3D TROSY-HNCO based experiment [138]. The experiments were collected three times to improve the accuracy of the measurement of the dipolar coupling. Only HN RDC’s were measured because other dipolar couplings were too small to measure with high precision.
Collagen gels for RDC measurement were generated using methods similar to those previously published [139]. Briefly, rat tail type 1 collagen was purchased as an acidic stock from BD Bioscience. The acidic collagen was added to buffer (either sodium phosphate or Tris) containing 10% D$_2$O at 4° C, then placed in a 5 mm Shigemi NMR tube. The pH was adjusted by the addition of NaOH, or by sufficiently increasing the buffer concentration to achieve the desired pH. The sample was placed in the NMR instrument which had been pre-cooled to 4° C. The temperature was ramped up to 37° C in two phases. The first phase warmed the sample to 15° C at a rate of 1.5° C per 10 minutes. The second phase warmed the sample to 37° C at a rate of 1° C per 15 minutes [139]. Alignment of the resulting gel was assayed using splitting of the deuterium line. For samples containing gp41-M-MAT, the protein and detergent was added following pH adjustment.

Guanosine dinucleotide alignment media for RDC measurements were prepared by adding dinucleotide to gp41-M-MAT in potassium phosphate pH 6.0 at a final dinucleotide concentration of 13 mg/mL. The samples were vortexed to ensure complete mixing of the dinucleotide. Samples were loaded into a 5mm Shigemi tube and placed in the NMR instrument. The sample was allowed to align in the magnetic field for several hours prior to initiating data collection.

In addition to neutral strained gels, charged gels can be used as alignment media for RDC measurements [140]. Charged gels were prepared from either of the co-
polymers sodium acrylate (SA, Sigma-Aldrich) or (3-acrylamidopropyl)-
trimethylammonium chloride (APTMAC, TCI). Each co-polymer was prepared as a
30% stock solution with N,N’-methylenebisacrylamide ratio of 29.5:1 w/w. SA gels were
prepared by mixing SA stock and acrylamide stock 1:1 v/v to a final total concentration
of acrylamide and co-polymers of 4.5%. APTMAC gels were prepared by mixing
APTMAC stock and acrylamide stock 1:3 v/v to a final total concentration of acrylamide
and co-polymer of 4.5%. Polymerization was carried out using the same procedure as
for neutral gels above. 0.5 mL Gels were rinsed in ddH₂O overnight to remove reaction
by-products. In order to prevent excessive electroosmotic swelling, the gels were
contained in 6 mm, 10k cutoff dialysis tubing. Gels were partially dried and then
sufficient gp41-M-MAT, DPC, sodium phosphate pH 6.0, D₂O, and H₂O were added to
make a 0.5 mM gp41-M-MAT, 100 mM DPC, 50 mM sodium phosphate pH 6.0, and 10%
D₂O gel sample with a final approximate volume of 0.5 mL (estimated by gel mass).
Gels were allowed to absorb the protein and buffer overnight, after which they were
squeezed into an open ended NMR tube using the New Era gel apparatus.

Localization of the detergent micelle was assayed using paramagnetic relaxation
by 16-DOXYL stearic acid (DSA). The extent of paramagnetic relaxation enhancement
was determined by acquiring a TROSY spectrum in the absence of DSA (reference
spectrum) and comparing it to a TROSY spectrum with DSA (experimental spectrum).
Data are analyzed by plotting the ratio of the peak intensities between the reference and
experimental spectra. To optimize the paramagnetic relaxation enhancement, DSA was titrated into 0.5 mM uniformly $^{15}$N-labeled gp41-M-MAT in 40 mM DPC until the peak intensity ratios between reference and experimental spectra ranged between <0.1 and >0.9. This was achieved at a final DSA concentration of 0.5 mM.

Backbone chemical shifts were used to obtain backbone dihedral angle restraints using the program TALOS [141]. $^1$H-$^1$H NOE peak volumes were calculated using the elliptical volume calculation module built into nmrviewJ. Volumes were converted to distances using CYANA 2.1 [142], and separated into short, medium, and long range distances, corresponding to maximum distances of 3.5, 4.5, and 5.5 angstroms (Å) respectively.

The structure was calculated in two steps using XPLOR-NIH v. 2.26 [143]. An initial structure was calculated using standard simulated annealing with all of the data except the RDC restraints. The standard potentials for bond angles, improper angles, bond length, Van der Waals, and ramachandran plot were used. In the previous chapter, we noted that the gp41-M-MAT spectrum indicated that the trimer subunits were symmetric since only one set of resonances was observed. Therefore, symmetry was enforced using the non-crystallographic symmetry potential to minimize the RMSD between the subunits. The foldon structure has been determined by both NMR and X-ray crystallography and both techniques yielded the same structure. The HSQC spectrum of foldon in our construct was identical to the reported HSQC, therefore we
included an additional reference potential that minimized the RMSD between the
published foldon structure and the foldon in our structure. This helped to ensure
proper packing of the trimeric foldon, despite the limited side chain distance restraints.
Final structures were calculated using the structure from the initial calculation as a seed
and including the RDC restraints. The procedure resulted in 11 structures out of 400
that had no NOE violations greater than 0.5 Å and no dihedral violations greater than 5°.

Binding of antibodies to gp41-M-MAT conjugated to liposomes (M-MAT-
liposomes) was assayed using surface plasmon resonance following the protocol used
earlier for measuring the MPER peptide liposome binding of 2F5 and 4E10 antibodies
[84,95]. Briefly, M-MAT-liposome conjugates were immobilized (~500 RU) on a Biacore
L1 chip that had been conditioned by immobilizing ~2500-3000 RU of BSA. MAb’s
were flowed over the chip for 2 minutes at a flow rate of 20 µL/min and a concentration
of 100 µg/mL. MAb’s were allowed to dissociate for 5 minutes while buffer continued to
flow at 20 µL/min rate.

5.2 NMR Structure of gp41-M-MAT

We have determined the structure of gp41-M-MAT, a designed membrane
associated gp41 MPER trimer. The structure was calculated from NOE based distance
restraints, dihedral angles, \(^3\)J-coupling constants, and HN RDC’s. Selected spectra of
gp41-M-MAT in DPC micelles are shown in Figure 46 (the TROSY HSQC spectrum from
this construct is shown in chapter 4, Figure 42 ). The overall structure of gp41-M-MAT is
shown in Figure 47, and a table of restraints and structural statistics is presented in Table 4. The overall structure of gp41-M-MAT is shown in Figure 47. The MPER subunits each form an approximately 40 Å long α-helix. In turn, the helices form a three-fold symmetric left handed bundle, which is closely associated from residues 32-42, after which the helices splay out leading to 20-30 Å separation between the Cα atoms of the C-terminal residue. The structures selected from the XPLOR-NIH calculation, shown in Figure 47, did not contain any distance constraint violations greater than 0.5 Å or dihedral angle violations greater than 5°. The limited data available for the foldon domain was consistent with the reported structure of the foldon [129].
Table 4. Summary of Structural Statistics and Restraints

<table>
<thead>
<tr>
<th>Distance Restraints</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Intra-residue</td>
<td>16</td>
</tr>
<tr>
<td>Inter-residue</td>
<td>287</td>
</tr>
<tr>
<td>Sequential (</td>
<td>i-j</td>
</tr>
<tr>
<td>Medium-range (</td>
<td>i-j</td>
</tr>
<tr>
<td>Long-range (</td>
<td>i-j</td>
</tr>
<tr>
<td>Inter-Molecular</td>
<td>12</td>
</tr>
<tr>
<td>Total dihedral angle restraints</td>
<td>85</td>
</tr>
<tr>
<td>Phi</td>
<td>43</td>
</tr>
<tr>
<td>Psi</td>
<td>42</td>
</tr>
<tr>
<td>RDC restraints</td>
<td>52</td>
</tr>
<tr>
<td>$^3$J-HNHA coupling constant restraints</td>
<td>23</td>
</tr>
<tr>
<td>Total Restraints</td>
<td>463</td>
</tr>
</tbody>
</table>

Structure Statistics

<table>
<thead>
<tr>
<th>Violations (mean and standard deviation)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Distance restraints</td>
<td>0.072±0.005</td>
</tr>
<tr>
<td>Dihedral angle restraints</td>
<td>0.837±0.069</td>
</tr>
<tr>
<td>RDC</td>
<td>0.654±0.041</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Deviations from ideal geometry</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Bond length (Å)</td>
<td>0.003±0.000</td>
</tr>
<tr>
<td>Bond angles (°)</td>
<td>0.548±0.021</td>
</tr>
<tr>
<td>Impropers (°)</td>
<td>0.483±0.044</td>
</tr>
<tr>
<td>Main chain atom RMSD to the mean structure (Å)</td>
<td>1.653</td>
</tr>
<tr>
<td>Average main chain pairwise RMSD (Å)</td>
<td>2.359</td>
</tr>
</tbody>
</table>

MPER Ramachandran plot statistics

| Favored region (MPER residues 33-59)       | 98.3%  |
| Allowed regions (MPER residues 33-59)      | 100%   |

*Each restraint was used once for each subunit in the trimer*
Figure 46. Selected Spectra of gp41-M-MAT
A. Strip plot from an HNCA on a 2mM uniformly $^2$H, $^{15}$C, $^{15}$N labeled sampled of gp41-M-MAT. The dashed lines indicate resonances that arise from the same Cα.
B. 2D plane from a TROSY-NOESY-TROSY experiment on the same sample as (A), centered on S44. The * is an anti-TROSY peak from the S44 diagonal peak.
Figure 47. Structural Overview of gp41-M-MAT
A. Overlay of 11 low energy gp41-M-MAT structures that did not contain NOE violations or dihedral violations of greater than 0.5 Å or 5° respectively. B. Ribbon diagram of the minimized average structure from the ensemble in A, with the foldon domain colored in gray. C. Space filling diagram of the minimized average structure with the 2F5 and 4E10 epitopes colored in blue and red respectively.
The helical structure of the MPER observed in gp41-M-MAT is consistent with the overall observations of monomeric MPER peptides. However, there are some differences between our structure and some of the other monomeric MPER structures. Figure 48 shows an overlay of one of the gp41-M-MAT MPER subunits with another monomeric MPER structure (PDB: 2PV6) determined in DPC detergent micelles. The overlay was done using W51-K59, which reveals the difference in the bend between the structures. It has been proposed that the bend or kink seen in the structure 2PV6 is important for viral fusion [106]. It is clear from Figure 48 that 2PV6 has a severe kink that is not present in gp41-M-MAT. While the gp41-M-MAT does not adopt a kink like [Figure 48. Overlay of Monomeric MPER Peptide and gp41-M-MAT](image)
The monomeric MPER peptide from PDB 2PV6 is shown in red, gp41-M-MAT is shown in blue. The structures were aligned on residues 51-59 to emphasize the kink in 2PV6. There is not a severe kink in gp41-M-MAT.
2PV6, the helix does undergo a gentle bend in the same region. Other structures of the MPER region also do not adopt the severe kink seen in 2PV6 [97,105]. The data for 2PV6 was collected at pH 6.6 while the data for gp41-M-MAT was collected at pH 6.0. It is possible that the kink may be dependent on the pH, manifesting as a bend in our structure at pH 6.0. Alternatively, the kinked structure seen in 2PV6 may be a result of the monomeric nature of linear MPER peptides interacting with the DPC micelle.

One potential concern when making chimeric fusions between two different proteins is the potential of one protein to influence the structure of the other protein. Foldon terminates with a short sequence of helix at its C-terminus. We suspected that the MPER might form a helix based upon our own previous research (see chapter 3), as well as the structures determined by others [97,105,106]. However, we did not want to induce the MPER to fold by nucleating helix formation from the foldon. This is especially important since a previous study has suggested that the N-terminal region of the MPER may not adopt an α-helix at pH 6.0 [105]. In addition, if the MPER did form a helix, it was important that the MPER helices be allowed to orient themselves such that the correct helical face made up the trimer interface. In order to address this issue, we included a GSSG (gly-ser-ser-gly) linker in between the foldon and the MPER. In the structure of gp41-M-MAT, this linker does not adopt a regular helix which indicates that the MPER helix is not likely to be the result of propagation of the foldon helix. Furthermore, heteronuclear NOE data show that the linker region is dynamic.
Therefore, it is unlikely that the foldon domain is inducing any artificial structure in the MPER. Notably, the MPER helix begins immediately following the flexible linker, suggesting that the trimeric MPER prefers to adopt an $\alpha$-helix even at pH 6.0.

The core 2F5 epitope residues D40, K41, and W42 are shown in Figure 49 A. Residues D40 and K41 are well exposed on the surface of gp41-M-MAT. Part of the W42 side chain is oriented towards the axis of symmetry, but remains at least partially solvent exposed. Thus the 2F5 epitope is exposed on the surface of gp41-M-MAT and accessible for antibody binding. This is consistent with our observation in the previous chapter that 2F5 binds with 0.18 nM affinity to gp41-M-MAT. Interestingly, the least

![Figure 49. Structure of the Epitopes for 2F5 and 4E10](image)

A. Structure of gp41-M-MAT with 2F5 epitope in blue. B. Structure of gp41-M-MAT with 4E10 epitope in red.
conserved residue in the 2F5 core epitope, K41 (see Figure 20, residue K665), is oriented directly away from the trimer interface towards the bulk solvent.

The 4E10 epitope region, W48 – T52, is shown in Figure 49 B. Like the rest of the MPER, the 4E10 epitope region adopts a helical conformation. In the crystal structure of 4E10 bound to short epitope peptides, W48, F49, I51, and T52 account for the most contacts between the peptide and the antibody [81]. In our structure F49 is pointed inward toward the axis of symmetry, and presumably toward the micelle (discussed below). This is similar to the orientation proposed for F49 in 2PV6, which suggested that F49 was buried in the lipid or micelle [106]. The other residues are situated such that they might interact more with the phosphocholine head group region. N50 is pointed outward, where it could interact to a greater extent with solvent. Notably, this residue makes fewer contacts in the crystal structure, and is somewhat less conserved (see Figure 20) than the other residues in the epitope [81,144].

The regions of gp41-M-MAT that likely interact with the DPC detergent micelle were identified using paramagnetic relaxation enhancement by the detergent soluble probe, 16-DOXYL-stearic acid (DSA). DSA is a paramagnetic probe that localizes to the hydrophobic interior of detergent micelles and membranes [145]. Amino acid residues that are near the micelle will undergo paramagnetic relaxation enhancement that will result in a reduction of peak intensity. Residues that are near the micelle can be identified by comparing the peak intensities between a spectrum with DSA and one
without. This method has been successfully used by several groups to estimate the localization of both surface associated and transmembrane peptides with respect to a detergent micelle [145,146,147]. The intensities of residues 35 through 59 were significantly impacted by DSA, while residues 1 through 27 were not significantly impacted (see Figure 50). Importantly, there were no significant chemical shift perturbations observed in the spectrum with DSA, indicating that the DSA was not likely binding directly to the hydrophobic MPER region of gp41-M-MAT. This data indicates that most of the MPER peptide in gp41-M-MAT is associated with the detergent micelle. Since gp41-M-MAT interacts strongly with both 2F5 and 4E10, the MPER peptide is unlikely to be deeply buried in the hydrophobic interior of the micelle, but instead is likely associated with the surface of the micelle.

In the previous chapter, we determined that gp41-M-MAT bound the approximate mass of one DPC micelle per protein trimer. Previous studies of DPC micelles have suggested that they are likely prolate spheroids [74,148], although data indicating that they may be spherical has also been reported [149]. Our DSA data indicates that the majority of the MPER is associated with the micelle (see Figure 50). In our NMR structure of gp41-M-MAT, the distance between the C-termini of each MPER helix is ~30 Å [148]. The cavity formed by the splayed MPER helices is of similar size when compared to the reported size of the hydrophobic core of a DPC micelle. Therefore, it is possible that the micelle fills the cavity formed by the splayed MPER
helices, with the helices interacting with the head group region of the detergent.

Interestingly, the hydrophobic core of a DPC micelle based upon either the spherical model [149] or the prolate ellipsoid model [148] can be modeled into the cavity formed by the helices of gp41-M-MAT. Both models result in the MPER helices primarily occupying space in the phosphocholine head group region, with only small incursions into the hydrophobic acyl chain region. The MPER region contains several highly conserved tryptophan residues. Tryptophan is well known to be present in the interfacial regions between the hydrophobic core of lipid membranes and bulk solvent [150]. Thus, the association of the MPER with the head group region of the detergent micelle is consistent with the typical behavior of tryptophan rich regions of other

![Figure 50. 16-DoxyI Stearic Acid Mediated Paramagnetic Relaxation](image)

Data is shown as the ratio of amide proton peak intensities from 2D-TROSY spectra obtained with and without 0.5 mM DSA. * indicates residue that could not be measured due to peak overlap, or peaks that could not observed in either spectrum.
membrane proteins. Figure 51 shows a model of the DPC micelle inserted into the cavity formed by the splayed gp41-M-MAT helices using micelle dimensions intermediate between the two reported DPC micelle shape models.
The orientation of some of the critical residues for 4E10 binding could indicate that the epitope would not be available for antibody binding. However, our kinetic data show that 4E10 binds tightly to gp41-M-MAT with an overall Kd of 27 nM. This binding affinity is similar to that reported for 4E10 epitope peptides with or without lipid [95]. The binding is slightly weaker than was reported for other soluble MPER trimers such as gp41-inter and HA-gp41 [86,87]. This could indicate some additional occlusion of the epitope by lipid, but the differences were not dramatic and the experiments would have to be conducted side-by-side to make more detailed comparisons. Overall, the kinetic observations suggest that 4E10 can access its epitope, even though some important residues may be buried in the micelle. It has been proposed that the lipid reactivity of 4E10 allows it to extract its epitope from the lipid membrane [106]. The structure of gp41-M-MAT shows that reorientation of the 4E10 epitope out of the micelle is likely required for antibody binding. However this reorientation does not dramatically reduce the overall binding affinity of 4E10 for gp41-M-MAT when compared to soluble linear MPER peptides, which have been reported to bind to 4E10 with an affinity of 19 nM [95].

The various cryo-EM structures of ENV differ significantly in the MPER region. Figure 52 shows these cryo-EM structures with the gp41-M-MAT structure modeled into the general location of the MPER region using the program Chimera [151]. The splay observed in gp41-M-MAT is consistent with the size of the stalk region of the ENV for two of the three cryo-EM structures (EMDB: 1216 and 5272). One of the structures is
splayed much more than is observed in our gp41-M-MAT structure (EMDB: 1246 and 1247). In all cases the N-terminal region of the MPER appears to extend into the gp120 region of the ENV, suggesting that gp120 could mask much of the MPER prior to gp120 dissociation.
5.3 Binding of Non-Neutralizing Antibodies to gp41-M-MAT

We assayed the binding of two non-neutralizing antibodies that were raised against the 2F5 epitope. The first antibody was 13H11, which is a mouse antibody that binds the core DKW of the 2F5 epitope as well as an additional leucine and asparagine (L45 and N47 in the gp41-M-MAT structure) [134]. The crystal structure of 13H11 with and without a short linear epitope peptide bound has been reported [134]. The MPER peptide in the structure of this antibody-peptide complex adopts a well-defined α-helix conformation, which is similar to the post-fusion six-helix bundle [61,134]. 13H11 interacts with high affinity with free 2F5 linear epitope peptides [95]. However, this antibody does not bind M-MAT-liposomes (see Figure 53). In gp41-M-MAT, L45 is oriented such that the side chain is approximately parallel to the presumed transition interface between the hydrophobic interior and the hydrophilic head group region of the micelle. This could prevent 13H11 from gaining access to L45 and result in the loss of antibody binding.

The second non-neutralizing antibody tested was 11F10. This antibody was generated by immunizing rabbits with designed protein scaffolds that present the MPER in the same conformation that was seen in the 2F5 crystal structure [91,92]. The structure of the antibody-scaffold co-crystal has been solved, and confirms that the antibody binds the peptide in a conformation nearly identical to that observed in 2F5 [91]. The antibody lacks polyspecificity and lipid reactivity and does not neutralize
HIV-1. Interestingly, the antibody does not bind with high affinity to gp41-M-MAT, as shown in Figure 53. The binding of 11F10 was too low to determine an accurate Kd with the amount of 11F10 available.

Both 13H11 and 11F10 lack polyspecificity which may contribute their inability to bind gp41-M-MAT. In the case of 13H11, occlusion of additional epitope residues

Figure 53. Interaction of Non-Neutralizing Antibodies with gp41-M-MAT
SPR sensograms of various MPER specific antibodies are shown binding to M-MAT-liposomes. The non-neutralizing antibodies 13H11 (black) and 11F10 (blue) do not bind, or bind weakly to M-MAT-liposomes, respectively. For comparison, the neutralizing antibodies 2F5 (green) and 4E10 (red) are also shown. All antibodies were used at 100 μg/mL concentration.
such as L45 by the lipid or detergent may explain the lack of antibody binding. The mechanism by which 11F10 is excluded from gp41-M-MAT binding is less clear. Nevertheless, gp41-M-MAT appears to exclude the binding of some non-neutralizing antibodies while allowing high affinity binding of broadly neutralizing antibodies. This exclusion is not observed for gp41-inter where tight binding to 11F10 was observed (Dennison et al, unpublished result). Increased specificity for neutralizing antibodies vs. non-neutralizing antibodies, as found in gp41-M-MAT, may be a desirable property in new vaccine designs.

5.4 Comments on the Acquisition of RDC’s

RDC’s have become an important part of NMR structure determination [152]. However, many of the most common alignment media used to acquire RDC datasets, such as pf1 phage and bicelles, are not compatible with detergents. As part of the effort to determine the structure of gp41-M-MAT, we tested several different alignment media to determine which media would give reliable RDC’s for our construct. Of the media screened, only neutral stretched polyacrylamide gels produced reliable RDC data for gp41-M-MAT. The presence of foldon in our construct provided a convenient method to test the reliability of RDC’s measured in various alignment media, since the foldon structure had previously been reported. RDC datasets were fit to the known structure for foldon (PDB: 1RFO) using the program PALES [153], which uses singular value decomposition to determine the best fit alignment tensor for the dataset to the structure
and calculates expected RDC values based on the structure. Since our construct is a symmetric homotrimer, the rhombicity of the alignment tensor provides a useful measure of the reliability of the RDC dataset. It has been shown that symmetric multimers should have rhombicities of zero [154]. Thus, since foldon is a symmetric homotrimer, the RDC’s observed should also have rhombicity close to zero. While all of the alignment media we tested have been successfully used to collect RDC datasets on other protein samples, we found that the successful use of an alignment media is highly sample dependent. The Pearson correlation and rhombicity [153,155] for the foldon RDC’s are shown in Table 5. The results from each of the different alignment media screened will be discussed in the following sections.

### Table 5. Correlation and Rhombicity for RDC’s fit to the Foldon Domain

<table>
<thead>
<tr>
<th>Alignment Media</th>
<th>aCorrelation (R)</th>
<th>Rhombicity</th>
<th>bMaximum</th>
<th>D</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>4.5% stretched gel</td>
<td>0.968</td>
<td>0.052</td>
<td>28.5 Hz</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5.5% stretched gel, protein co-polymerized with gel</td>
<td>0.422</td>
<td>0.05</td>
<td>16.4 Hz</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Collagen gel</td>
<td>0.53</td>
<td>0.05</td>
<td>2 Hz</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4.5% stretched gel, containing 25% APTMAC</td>
<td>0.971</td>
<td>0.303</td>
<td>11.4 Hz</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

aPearson correlation coefficient bMaximum absolute value of the dipolar coupling
Correlation and Rhombicity were calculated using the program PALES [153]

#### 5.4.1 Strained Polyacrylamide Gels

The alignment media which gave the best RDC’s was strained polyacrylamide gels, which have been used for both soluble proteins and detergent solubilized membrane proteins [156,157,158]. Strained polyacrylamide gels can be prepared as
either radially compressed (also called stretched) or vertically compressed gels. Radial compression is achieved by casting a gel with a larger diameter than an NMR tube and squeezing the gel into the tube [156]. Vertical compression is achieved by casting a gel with a smaller diameter than an NMR tube and using a plunger to compress the gel until it fills the NMR tube [157,158]. We attempted both radial and vertical compression and found that radial compression yielded a reliable RDC dataset. All of the expected peaks were observed and we measured good RDC values, up to ~30 Hz. The method that produced good results was to polymerize the gel with high concentrations of TEMED and APS, rinse out reaction byproducts in H₂O, then soak the gp41-M-MAT sample into the gel. RDC’s measured from samples prepared in this manner exhibited high correlation (>0.9) with the known foldon structure and low rhombicity (<0.1). Gel samples of this type exhibited some shrinkage over time, but no change in RDC’s was observed over the one week data collection time. We arrived at this particular method after testing several other previously reported methods for strained gel sample preparation as discussed below.

While we were able to get reliable RDC’s from a radially compressed gel sample, we found that there were some important considerations for sample preparation. First, we found that the concentration of TEMED and APS used to polymerize the gel were extremely important for the success of the experiment. It had been reported that very low concentrations of TEMED and APS can be used in order to reduce the pH change
associated with the addition of the TEMED and allow for inclusion of the protein when the gel is cast [159]. However, we found that using low TEMED and APS concentrations resulted in significant broadening or loss of the resonances. This was significantly reduced if the TEMED and APS concentrations were increased. It is possible that the low concentration of TEMED and APS is insufficient to overcome the inhibitory effect of residual oxygen, resulting in incomplete polymerization of the gel. It was reported that extensive degassing was required for success of this method [159]. While we did degas the buffers, our degassing may not have been sufficient to overcome this problem. It may be possible to use low TEMED and APS if the reaction were carried out in an anaerobic chamber to eliminate oxygen from the reaction. Second, we found that co-polymerizing the gp41-M-MAT sample in the gel resulted in RDC’s that did not correlate well with the known foldon structure, even though the HSQC spectrum was essentially the same. Interestingly, the removal of excess reaction by-products by rinsing the gel for a short time resulted in good correlation (>0.9) but poor rhombicity (>0.50). Thus, we abandoned the use of co-polymerization.

We note that samples prepared using strained gels are often dried prior to addition of the protein solution [158]. We found that this was not required if the New Era gel apparatus is used, but may be useful to increase the final concentration of protein in the gel since the solution is not diluted by the buffer in the gel. In our case, sufficient protein concentration in the gel could be achieved without drying the gel.
We found that vertical compression of polyacrylamide gels was complicated by the tendency of the gels to coil in the Shigemi tube as pressure was applied with the plunger. Previous reports showed that a medium wall 524 NMR tube from Wilmad could be used as a cast to prepare gels for vertical compression [157]. Gels cast in this tube would have a 0.8 mm smaller diameter than a standard 5 mm Shigemi NMR tube. This gel diameter was too small, and resulted in coiling of the gel in the Shigemi tube and limited RDC’s. Another report suggested that a 0.5 mm difference in diameter could alleviate the coiling problem [158]. Unfortunately, a gel cast of this size is not commercially available, precluding further attempts to use vertically compressed gels in this study.

5.4.2 Guanosine Dinucleotide Alignment

Guanosine dinucleotides have been reported to form tetrads which, in the presence of potassium ions form a cholesteric liquid crystal phase (LC) [160]. The LC phase of these dinucleotides has been successfully used as an alignment media for the acquisition of RDC data in the presence of detergents, including DPC, making it a promising media for our application. We found that these dinucleotides aligned in potassium phosphate buffer at pH 6.0 as had been previously reported [160]. Splitting of the deuterium signal due to quadrapolar coupling indicated that the sample was adequately aligned. However, when we collected the initial spectra, we found that the peaks corresponding to the foldon domain were missing. This indicated that the
construct was likely interacting with the alignment media, which could result in unreliable RDC’s for the observed resonances. We set this data aside, since we did not have the foldon control to confirm the reliability of the dataset.

### 5.4.3 Collagen Gel Alignment

Collagen gels have been reported to align in a magnetic field during polymerization and that these aligned gels could be used as RDC alignment media [139]. The alignment of collagen gels was tested under various conditions by monitoring the splitting of the deuterium resonance. We found that while the gels would polymerize in phosphate buffers, they did not align. Collagen only aligned in Tris buffer pH 7.4. However, when gp41-M-MAT was incorporated into the samples in Tris buffer pH 7.4, we observed small RDC’s (absolute value less than 2Hz). Importantly, the HSQC fingerprint of gp41-M-MAT in Tris buffer pH 7.4 and collagen did not change significantly from that observed in the sodium phosphate buffer pH 6.0 used to obtain resonance assignments. The small RDC’s observed in collagen gel indicated that there was little alignment of the gp41-M-MAT even though the deuterium splitting observed (~14 Hz) was similar to that reported where there was significant alignment of other proteins [139]. We hypothesize that the DPC micelles likely interfere with the concurrent alignment and polymerization that collagen gels undergo, leading to inhomogeneous gels that did not yield alignment of gp41-M-MAT.
5.4.4 Charged Strained Polyacrylamide Gels

In addition to neutral strained polyacrylamide gels, strained gels that incorporate charged co-polymers have been shown to be useful as alignment media [140]. We screened a positive (APTMAC) charged gel and a negative (SA) charged gel. The spectrum of gp41-M-MAT in the SA only showed resonances for the MPER and not for the foldon, similar to the behavior of the guanosine dinucleotides. The spectrum of gp41-M-MAT in APTMAC showed resonances for the foldon and the MPER, however there were also additional resonances observed. The similarity between the behavior of gp41-M-MAT in SA gel and the dinucleotide suggests that there may be a common cause. The pI of gp41-M-MAT was calculated to be ~4.56 using the program protparam [161] resulting in an overall net positive charge at pH 6.0. It is possible that the construct is interacting with the negative charge that is present in both the SA gel and the dinucleotide, leading to reduced mobility of the foldon and increased line broadening. In this case, the flexible GSSG linker may allow sufficient mobility of the MPER to still observe those resonances. In APTMAC gels, even though resonances for both domains were present, the fit of the RDC’s to the foldon domain resulted in high rhombicity. The observation of additional peaks could indicate that the structure and/or symmetry of the construct are adversely affected by the positively charged gel. Alternatively, there could be additional unknown factors in the method of preparation that are contributing to the results. Notably, the correlation of the APTMAC RDC’s was actually very high (see
Table 5). However, including these data and the stretched neutral gel data in the structure calculation resulted in structures with significant violations of the RDC’s. This combined with the high rhombicity and additional peaks suggested that the APTMAC dataset was not reliable.

5.4.5 Conclusions

Our observations above raise an important consideration for the use of RDC’s as structural restraints. In all of the above cases, RDC’s could be measured for at least part of the protein, however the RDC’s could only be confirmed as reliable in the stretched neutral polyacrylamide gel. We made this determination based upon the agreement of the measured RDC’s with the known foldon structure. Importantly, it was not always readily apparent from the NMR spectrum that there was a problem with the RDC dataset, as was the case in the co-polymerized polyacrylamide gels. This begs the question of how to determine if an RDC dataset is reliable when the structure is not known, as is typically the case when a protein structure is determined by NMR spectroscopy. Normally, RDC’s are only incorporated as a refinement of a structure obtained using more traditional NMR restraints, such as NOE’s and dihedral angles [152]. This approach may mitigate this problem, since RDC datasets that don’t fit well to the NOE structure could be identified. RDC datasets in multiple alignment media are often collected which can reveal inconsistencies between datasets. It may be more difficult to detect unreliable datasets when methods that use RDC’s as the primary data
source are employed and data is collected in only one alignment media. We made these observations while studying only one protein construct, making it possible that this problem is unique to gp41-M-MAT. However, it is important that the extent of this problem be explored to determine the potential impact it could have on RDC based structure determination. For symmetric homo-oligomers, rhombicity may provide an additional check to determine the reliability of an RDC dataset.

**5.5 Implications of gp41-M-MAT Structure**

The HIV-1 transmembrane domain has been suggested to contribute to trimerization based upon computational data and the trimeric behavior of a gp41 construct that included HR-2 and TM [59,162]. The six-helix bundle structures all show that HR-1 forms the core three-helix bundle within the six-helix bundle [56,57,61]. In addition, a recent crystal structure of the six-helix bundle shows that the fusion peptide proximal region of HR-1 is also part of the core three-helix bundle, although it is not as tightly packed as the rest of the HR-1 [57]. Given these structures, it is probable that the fusion peptide of gp41 would extend from the N-terminus of HR-1 and occupy the central space of the six-helix bundle as it enters the membrane. If this is true, then it follows that the TM domain would extend from the MPER, continuing the outer portion of the six-helix bundle as it enters the membrane, and would no longer be self-associated in the post-fusion state. Thus, our construct may represent an intermediate state where the TM domain is no longer tightly bundled as a trimer in order to make room for the
fusion peptide and HR-1 to form the inner core of the six-helix bundle. This would also allow the MPER to interact with the membrane, similar to the interaction we observe with the micelle in the gp41-M-MAT structure.

It has already been shown that the MPER region contributes to viral membrane fusion [58,163]. Based upon a six-helix bundle structure of gp41 that included the MPER, it has been suggested that the MPER may splay out and induce strain in the outer leaflet of the membrane in order to promote the membrane fusion process [57]. In the gp41-M-MAT structure, the MPER splays out as previously proposed. In addition, our data suggests that detergent fills the cavity formed by the splayed MPER helices. The MPER helices are not deeply buried in the detergent micelle, but are likely located in the phospholipid head group region. Amphipathic helices located in this region of the outer leaflet of a lipid bilayer induce bulges in membrane bilayers [164]. Thus, the association of the splayed MPER helices with the outer leaflet of the viral membrane could induce significant bulging towards the host cell membrane. Bulging of the membrane in this manner could assist the formation of the membrane fusion stalk (where the outer leaflets of the two membranes have joined) believed to be involved in the fusion of phospholipid membranes [165]. Figure 54 incorporates this hypothesis and potential intermediate into the overall fusion model for HIV-1. The incorporation of a splayed MPER intermediate similar to the one depicted in Figure 54 has been previously proposed [166]. Whether this is the target for broadly neutralizing antibodies is not
clear, nor is the lifetime of such an intermediate obvious. Finally, Figure 54 incorporates only one ENV complex for clarity. However, many HIV-1 membrane fusion models, which are based upon the fusion model for influenza hemagglutinin, assume that more than one ENV is involved in membrane fusion and pore formation \([54,167,168]\). The MPER helices from several ENV trimers may interact cooperatively to induce sufficient strain in the membrane to promote membrane fusion, as depicted in Figure 55.

**Figure 54. Schematic of Membrane Fusion with gp41-M-MAT Inspired Intermediate**
The association of the MPER induces bulging in the membrane that assists the membrane fusion event.
Interestingly, the helical conformation of the 2F5 epitope in gp41-M-MAT is strikingly different from the conformation observed in crystal structures of 2F5 bound to short, monomeric epitope containing peptides. In the crystal structures, the 2F5 core epitope adopted a beta-turn conformation, while the rest of the peptide was extended [80]. However, several NMR structures of monomeric MPER peptides under various conditions also show that the 2F5 region adopts a helical conformation [97,105,106]. The structure of the 2F5 epitope from the 2F5 bound co-crystal structure (PDB: 1TJI) and

Figure 55. Model of Membrane Fusion with Multiple ENV's
Model of membrane fusion based upon influenza hemagglutinin that includes multiple gp41 trimers forming a fusion pore. The MPER (blue) is shown interacting with the membrane and aiding in the formation of a bulge that could result in a membrane pore.
gp41-M-MAT are shown overlaid in Figure 56 A-C. It is clear from Figure 56 A that the structure of the MPER would have to change significantly to adopt the same structure as observed in 1TJI. 2F5 binds to gp41-M-MAT with a Kd of 0.18 nM, which is a somewhat stronger interaction than the 6.6 nM Kd that has been reported for a free extended HR-2 peptide that contained the 2F5 epitope [94]. Thus, high affinity binding of 2F5 is not inhibited by the α-helical conformation adopted by the MPER in gp41-M-MAT, even though the crystal structure indicates that substantial conformational rearrangements of the MPER are required to recapitulate the linear peptide conformation observed in the crystal structure. Clearly, more structural data is required to better understand the differences between conformations of the 2F5 epitope and what they imply for membrane fusion and the induction of neutralizing antibodies.

The conformation of the 4E10 epitope in gp41-M-MAT and the conformation observed in the peptide bound conformation in the crystal structure are α-helices. Alignment of a gp41-M-MAT subunit on the 4E10 structure reveals this similarity (see Figure 56 D-F). The alignment also shows that the MPER helix must make an abrupt turn in order to avoid clashing with the light chain of 4E10. Like 2F5, 4E10 must make significant alterations in the structure of the MPER observed in gp41-M-MAT in order to bind the epitope in the way that is observed in the crystal structure. Again, more structural information is required to better understand how 4E10 interacts with gp41-M-MAT.
Figure 56. Comparison of Antibody Bound MPER and gp41-M-MAT
The epitopes for 2F5 (A–C) and 4E10 (D–F) in the structure of gp41-M-MAT are overlaid on to the same residues in the peptide-antibody co-crystal structures. The antibody light chains and heavy chains are shown in yellow and green respectively. Panels A and D only show gp41-M-MAT (blue) and the peptide epitope (red). Panel B and E add the antibodies with the heavy chains made transparent. Panels C and F are the same as B and E except the light chains are transparent. The PDB codes for 2F5 and 4E10 are 1TJI and 2FX7, respectively.
While gp41-M-MAT was developed to better understand the conformations of the MPER as a lipid associated trimer, it also represents a new immunogen. The splayed trimer observed in gp41-M-MAT may present the MPER epitopes in a manner different from other vaccine candidates. Most of the reported MPER trimers are either six-helix bundles or trimers where the C-terminus of the MPER is held together by a strong trimerization domain [57,59,61,86,87,107]. It is unlikely that these constructs could adopt the same splayed conformation observed in gp41-M-MAT because the C-terminal trimerization domain would prevent the separation of the individual MPER subunits. We have initiated studies that test gp41-M-MAT as a potential immunogen and vaccine candidate. It is important to determine if the construct can induce MPER specific antibodies and determine if those antibodies exhibit neutralizing activity. We are currently immunizing guinea pigs with a combination of adjuvants and M-MAT-liposomes to determine the immunogenicity of gp41-M-MAT.

This construct opens up new avenues of research to better understand the structure of the MPER of HIV-1 and its interaction with broadly neutralizing antibodies. As previously mentioned, there are significant differences between the MPER epitope in the 2F5 crystal structure and the structure of gp41-M-MAT. 2F5 binds with high affinity to gp41-M-MAT, suggesting that it should be possible to produce a stable antibody bound complex. Either NMR spectroscopy or X-ray crystallography might be employed to determine the structure of the gp41-M-MAT bound to antibodies. This would reveal
the structure of the 2F5 bound MPER in the context of a trimer and perhaps conjugated to a lipid membrane or detergent micelle. A structure of this complex could provide important insights into the interaction between 2F5 and trimeric MPER that could guide future structure based vaccine design, and may shed some light on the observed differences between the crystal structures of 2F5 bound MPER and the solution structures of lipid associated MPER.

The gp41-M-MAT construct discussed in this work included only the MPER region. HR-2 and the transmembrane domain of HIV-1 also likely influence the structure and organization of the MPER. Thus extension of the construct to include additional portions of HR-2 and/or the transmembrane domain could provide further insights into the molecular mechanism by which gp41 induces membrane fusion. It is likely that portions of HR-2 could easily be added to gp41-M-MAT without compromising its favorable expression and NMR spectral qualities. Alternatively, portions of the HIV-1 transmembrane domain could be added to gp41-M-MAT. We have already made a construct that includes the entire HIV-1 transmembrane domain to gp41-M-MAT and found that it does not express in E. coli cells. However, we have not tried truncating the HIV-1 transmembrane domain to improve expression. This represents another potential avenue to better understand how the MPER interacts with the membrane when the authentic transmembrane domain is present.
In summary, we have developed gp41-M-MAT, a membrane associated trimer of the MPER region from HIV-1 gp41, and determined its structure in DPC detergent micelles. The structure of gp41-M-MAT reveals that the C-terminal part of the MPER region splays out along the surface of the detergent micelle, suggesting that the MPER could at least partially lay on the surface of phospholipid membranes. Association of peptides with the surface of phospholipid membranes has been hypothesized to induce bulges that may assist with the process of membrane fusion. This may explain part of the function of the MPER during viral membrane fusion with the host cell. The structure of gp41-M-MAT shares similarities with other linear MPER peptide structures, but does exhibit significant differences such as the absence of the sharp kink seen in a previous structure of a linear MPER peptide on DPC micelles. The gp41-M-MAT construct is a new model trimeric membrane associated MPER that can be easily modified to study the overall structural behavior of this important part of HIV-1 gp41. It also represents a potentially novel immunogen that may elicit a novel antibody response, as well as provide a modifiable framework from which additional new immunogens could be developed. This in turn could result in improved immunogens that may eventually give rise to an HIV-1 vaccine.
Appendix A

Composition of modified M9 minimal media used for stable isotope labeling [36] is listed in Table 6. Phosphate buffer was prepared as a 10x stock solution. Sodium chloride was prepared as a 50x stock solution. Glucose and ammonium chloride were prepared as 100x stock solutions. All other solutions were prepared as 1000x stock solutions. Water and phosphate were autoclaved to sterilize. Other components were filtered through a 0.2 micron syringe filter to sterilize. Glucose and/or ammonium chloride were substituted with the appropriate stable isotope labeled forms, depending on the desired labeling pattern. For deuterated growths, 99% D$_2$O was substituted for H$_2$O, and phosphate was prepared in D$_2$O. In addition, triple labeled Bioexpress rich media (CIL) was used as a supplement to a final concentration of 10% which improved growth and protein yield.

<table>
<thead>
<tr>
<th>Component</th>
<th>Concentration</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sodium Phosphate (Dibasic)</td>
<td>80 mM</td>
</tr>
<tr>
<td>Potassium Phosphate (monobasic)</td>
<td>40 mM</td>
</tr>
<tr>
<td>NaCl</td>
<td>8.6 mM</td>
</tr>
<tr>
<td>MgSO$_4$</td>
<td>2 mM</td>
</tr>
<tr>
<td>FeCl$_3$</td>
<td>1 µM</td>
</tr>
<tr>
<td>ZnSO$_4$</td>
<td>25 µM</td>
</tr>
<tr>
<td>CaCl$_2$</td>
<td>100 µM</td>
</tr>
<tr>
<td>H$_3$BO$_3$</td>
<td>2 µM</td>
</tr>
<tr>
<td>CuSO$_4$</td>
<td>2 µM</td>
</tr>
<tr>
<td>CoCl$_2$</td>
<td>2 µM</td>
</tr>
<tr>
<td>MnCl</td>
<td>10 µM</td>
</tr>
<tr>
<td>Nickel (II) Sulfate</td>
<td>2 µM</td>
</tr>
<tr>
<td>Ammonium heptamolybdate</td>
<td>2 µM</td>
</tr>
<tr>
<td>Ammonium Chloride</td>
<td>0.1% (w/v)</td>
</tr>
</tbody>
</table>

Table 6. Modified M9 Minimal Media Composition
Glucose 0.2% (w/v)

Composition of chromatography buffers used for Ni-sepharose and Ni-NTA columns are shown in Table 7 and Table 8. Composition of buffer used for size exclusion chromatography is shown in Table 9. All buffers were filtered with a Nalgene 0.2 micron filter prior to use. The pH of the buffers was adjusted to 8.0 using 6 M NaOH or 37% HCl for Ni chromatography. The pH of the size exclusion buffer was not adjusted. Generally, only minor pH adjustments were required.

<table>
<thead>
<tr>
<th>Table 7. Load/Wash Buffer</th>
</tr>
</thead>
<tbody>
<tr>
<td>Component</td>
</tr>
<tr>
<td>Sodium phosphate (Dibasic)</td>
</tr>
<tr>
<td>Guanidine HCl</td>
</tr>
<tr>
<td>Imidazole</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table 8. Elution Buffer</th>
</tr>
</thead>
<tbody>
<tr>
<td>Component</td>
</tr>
<tr>
<td>Sodium phosphate (Dibasic)</td>
</tr>
<tr>
<td>Guanidine HCl</td>
</tr>
<tr>
<td>Imidazole</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table 9. Size Exclusion Buffer</th>
</tr>
</thead>
<tbody>
<tr>
<td>Component</td>
</tr>
<tr>
<td>Sodium phosphate (Dibasic)</td>
</tr>
<tr>
<td>Guanidine HCl</td>
</tr>
</tbody>
</table>

SDS-PAGE gels were prepared according to Table 10 for Tris-glycine gels, and Table 11 for Tris-Tricine gels. Gels were polymerized with 0.1% (w/v) APS and 0.125%
(v/v) TEMED. Running buffer components are shown in Table 12 and Table 13 for Tris-glycine and Tris-Tricine gels, respectively. Acrylamide was purchased as a 30% stock solution from Biorad, and was 37.5:1 acrylamide to N,N’-methylene-bis-acrylamide.

### Table 10. 16.5% Tris-Glycine Polyacrylamide Gel

<table>
<thead>
<tr>
<th>Component</th>
<th>Concentration</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tris, pH 8.8</td>
<td>375 mM</td>
</tr>
<tr>
<td>Sodium Dodecylsulfate</td>
<td>0.1% (w/v)</td>
</tr>
<tr>
<td>30% Acrylamide stock</td>
<td>51.2% (v/v)</td>
</tr>
</tbody>
</table>

### Table 11. Tris-Tricine Polyacrylamide Gel

<table>
<thead>
<tr>
<th>Component</th>
<th>Concentration</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tris, pH 8.45</td>
<td>1 M</td>
</tr>
<tr>
<td>Sodium Dodecylsulfate</td>
<td>0.1% (w/v)</td>
</tr>
<tr>
<td>30% Acrylamide stock</td>
<td>55 % (v/v)</td>
</tr>
<tr>
<td>Glycerol</td>
<td>10% (v/v)</td>
</tr>
</tbody>
</table>

### Table 12. Tris-Glycine Running Buffer

<table>
<thead>
<tr>
<th>Component</th>
<th>Concentration</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tris</td>
<td>25 mM</td>
</tr>
<tr>
<td>Glycine</td>
<td>192 mM</td>
</tr>
<tr>
<td>Sodium Dodecylsulfate</td>
<td>0.1% (w/v)</td>
</tr>
</tbody>
</table>

### Table 13. Tris-Tricine Running Buffer

<table>
<thead>
<tr>
<th>Component</th>
<th>Concentration</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tris</td>
<td>100 mM</td>
</tr>
<tr>
<td>Tricine</td>
<td>100 mM</td>
</tr>
<tr>
<td>Sodium Dodecylsulfate</td>
<td>0.1% (w/v)</td>
</tr>
</tbody>
</table>
References


52. Telesnitsky A, Goff SP (1997) Reverse Transcriptase and the Generation of Retroviral DNA.


Biography

Patrick Reardon was born in Portland, Oregon on February 26th, 1978. He grew up in Milwaukie, Oregon, and graduated from Rex Putnam High School in the class of 1996. His undergraduate career started at Clackamas Community College, after which he transferred to Portland State University, and then to Oregon State University where he graduated cum laude with a Bachelor’s of Science in Biochemistry and Biophysics. After receiving his B.S. degree, he took a two year hiatus from student life and worked as a technician for Dr. Joseph Beckman. During this time Patrick solidified his desire to be involved in scientific research at the Ph.D. level. He accepted an offer to join the Ph.D. program in the Biochemistry Department at Duke University in 2003. Together with his fiancé, Carrie Marean, whom he had met while attending Oregon State University, they packed up their worldly belongings in a trailer and drove ~3000 miles from Corvallis, OR to Durham, NC. Carrie returned to Corvallis to finish her final term at Oregon State, while Patrick began his time with the Duke Biochemistry Department. On December 27th, 2003, Patrick and Carrie were married in Milwaukie Oregon, after which she joined him permanently in Durham. Patrick affiliated with Dr. Leonard Spicer in the early summer of 2004. During his time in the Spicer lab he has been involved in several projects and publications, including one patent application. He has received several awards, including student poster awards at several Biochemistry Departmental retreats, as well as a best student poster award at the 25th Symposium of
the Protein Society. Patrick has also been the recipient of student travel awards to the Experimental NMR Conference and the 25th Symposium of the Protein Society. A current list of his publications follows.


