Spectroscopic Studies of Membrane Protein Structure, Dynamics, and Function

Ryan Haowen Lo
Paramus, NJ

B.A. Franklin and Marshall College, 2009

A Dissertation presented to the Graduate Faculty
of the University of Virginia in Candidacy for the Degree of
Doctor of Philosophy

Department of Chemistry

University of Virginia
May, 2014

To my advisor, who has helped me grow as both a better scientist and as a capable person.

To my mom, who has sacrificed more than I would think a person is capable of, to get me here today and who has encouraged me to nurture my own interests.

To my stepfather, Brian, who has always supported me in all my endeavors.

To my wife, who completes me.

Abstract

Biomolecular NMR spectroscopy and EPR spectroscopy are established methods for investigating the structure and dynamics of membrane proteins. However, there are various experimental challenges in working with membrane proteins that has hindered the progress of membrane protein investigations including structure determination. One of the primary bottlenecks is the selection of a membrane mimic that will stabilize a functional membrane protein fold. Mimic selection is largely empirical, and only a select few detergents have led to NMR membrane protein structure determination. NMR spectra of membrane proteins in different detergents are extremely variable in quality and often do not lead to structure determination. Using two eight stranded β-barrel model systems, Opa_{50} and Opa_{60}, solution conditions that influence the NMR spectra were systematically investigated in order to gain an understanding of the interactions between the bilayer mimic and membrane protein that stabilize a fold. Detergent and ionic strength had a significant impact on the quality of NMR spectra due to interactions between the extracellular loops within a monomeric protein-detergent complex. DMPC lipid nanodiscs were also used as a bilayer mimic, and the resulting fold was comparable to the micelle embedded structure.

In addition to structure, magnetic resonance investigations of membrane proteins can also provide information about membrane protein dynamics. The dynamics of model helical membrane protein, TM0026 were investigated using EPR spectroscopy and NMR relaxation. We have noted a correlation between EPR scaled mobility and inverse second moment data with ^{15}N NMR relaxation data. Both the NMR and EPR data indicate that both methods can reflect membrane protein backbone dynamics in the ns timescale. In the case of TM0026, a proline kink decouples the dynamics of the transmembrane helical backbone such that the N-terminal region of the helix is more dynamic than the C-terminal region. The results provide evidence that although the EPR lineshapes and interactions of the nitroxide are different in membrane proteins the lineshapes reflect backbone dynamics in the ns time regime.

Curriculum Vitae

Ryan H. Lo

Phone: (201)566-0816
University of Virginia
ryanlo@virginia.edu
Department of Chemistry

EDUCATION

University of Virginia – Charlottesville, VA 2009-2014
Doctor of Philosophy in Chemistry
Franklin & Marshall College – Lancaster, PA 2005-2009
Bachelor of Arts in Chemistry

RESEARCH

Interest

The primary focus of my research deals with investigating structure and dynamics of Opa_{50} and Opa_{60}, two integral β-barrel outer membrane proteins. These Opa proteins are crucially involved in the pathogenesis of *Neisseria gonorrhea* and *Neisseria meningitidis*, binding to human host cell receptors and causing engulfment of the bacteria. In working towards understanding the molecular determinants of these binding interactions I have subsequently investigated the roles of ionic strength, membrane mimic, and temperature on the protein-detergent system, as well as worked on NMR methods development for structural assignment of integral β-barrel membrane proteins. The information gained from these techniques may lead to a better understanding for conditions leading to structural determination; to date, only 6 of the 71 unique β-barrel membrane protein structures have been solved with solution NMR, underlying the difficulties associated with solution structure determination that must be overcome.

Experience

Research Assistant, University of Virginia 2010-current
I conduct research towards the determination of the structure and function of the Opa_{50} and Opa_{60} proteins under the supervision of Dr. Linda Columbus, using various biophysical characterization techniques such as NMR spectroscopy, CD spectroscopy, fluorescence polarization, and isothermal titration calorimetry.
Research Assistant, University of Virginia 2010-2012
I conducted research on the purification and biophysical characterization of the thrombospondin repeat domains of the brain angiogenesis-1 protein under the supervision of Dr. Linda Columbus and Dr. James Casanova, with collaborator Emily Billings, a PhD candidate in the Microbiology department.
Research Assistant, Franklin and Marshall College 2006-2009
I conducted research on the physical properties of various micelles using spectroscopic methods including UV/vis absorbance and fluorescence characterization under the supervision of Dr. Richard Moog.
Laboratory Researcher, Franklin and Marshall College Fall 2008
I developed a hypothesis for the active site mechanism of HPII catalase of *Escherichia coli* and tested my hypothesis by performing site-directed mutagenesis on the predicted active site residue, Tyr415, with unnatural amino acids, under the supervision of Dr. Ryan Mehl.

PUBLICATIONS AND PRESENTATIONS

Peer Reviewed Journals

Lo RH, Fox DA, and Columbus L. "Ionic strength modulates β-barrel membrane protein loop dynamics and interactions and dramatically affect NMR spectral quality." (*In Preparation*)

Lo RH, Kroncke BM, Solomon T, and Columbus L. "Mapping dynamics in membrane proteins: a comparison of site-directed spin labeling to NMR ^{15}N-relaxation measurements." (*In Preparation*)

Oliver RC, Lipfert J, Fox DA, **Lo RH**, Kim JJ, Doniach S, and Columbus L**.** "Modulating the physical properties of micelles using binary detergent mixtures for membrane protein investigations." *(In Preparation)*

Fox DA, Larsson P, **Lo RH**, Kroncke BM., Kasson PM, and Columbus L. "NMR and MD hybrid approach to the structure ensemble of Opa_{60}: Insights into host interactions that induce phagocytosis." *JACS* (*submitted)*

Kenwood BM, Weaver JL, Bajwa A, Poon IK, Byrne FL, Murrow BA, Calderone JA, Huang L, Divakaruni AS, Tomsig JL, Okabe K, **Lo RH**, Coleman GC, Columbus L, Yan Z, Saucerman JJ, Smith JS, Holmes JW, Lynch KR, Ravichandran KS, Uchiyama S, Santos WL, Rogers, GW, Okusa MD, Bayliss DA, Hoehn KL. "Identification of a novel mitochondrial uncoupler that does not depolarize the plasma membrane." *Molecular Metabolism 2(5) (December 2013)*

Oliver RC, Lipfert J, Fox DA, **Lo RH**, Doniach S, and Columbus L**.** "Dependence of Micelle Size and Shape on Detergent Alkyl Chain Length and Head Group." *PLoS ONE 8(5): e62488. (May 2013).*

Selected Talks

Lo RH, Fox DA., and Columbus L. "Ionic strength modulates β-barrel membrane protein loop dynamics and interactions and dramatically affect NMR spectral quality." *Southeastern Regional Meeting of the American Chemical Society, Inc* (Raleigh, NC, November 2012).

National Poster Presentations

Lo RH, Fox DA, and Columbus L."Structure, dynamics, and receptor binding of Opa proteins." *58th Annual Meeting of the Biophysical Society* (San Francisco, California. February 2014).
Lo RH, Fox DA, and Columbus L. "Ionic strength modulates β-barrel membrane protein loop dynamics and interactions and dramatically affect NMR spectral quality." *Keystone Symposia – Frontiers of NMR in Biology* (Snowbird, Utah, January 2013). This presentation was selected for a lightning talk.
Lo RH, Fox DA, and Columbus L. "Effects of detergent, ionic strength, and temperature on the quality of β-barrel membrane protein NMR spectra." *Federation of American Societies for Experimental Biology Science Research Conference* (Snowmass Village, Colorado, June 2012. Received an NSF Travel award). This presentation was selected for a lightning talk.
Lo RH, Fox DA, and Columbus L. "Investigation of the *Neisseria Gonorrhoeae* Opacity Associated Membrane Protein A Structure, Dynamics, and Host-receptor Interactions." *55th Annual Meeting of the Biophysical Society* (Baltimore, Maryland, March 2011). Published Abstract: Biophysical Journal, vol. 100 (3): 385a.
Lo RH and Moog RS. 2009. "Characterization of Environments in Triton X-100 Micelles." *237th ACS national meeting* (Salt Lake City, Utah, March 2009).

Regional Poster Presentations

Lo RH, Fox DA, and Columbus L."Effects of detergent, ionic strength, and temperature on the quality of β-barrel membrane protein NMR spectra." *University of Virginia's 12th Annual Robert J. Huskey Graduate Research Exhibition* (Charlottesville, Virginia. March 2012).
Lo RH, Fox DA, and Columbus L. "Investigation of the *Neisseria Gonorrhoeae* Opacity Associated Membrane Protein A Structure, Dynamics, and Host-receptor Interactions." *University of Virginia's 5th Biannual Biotechnology Training Program Symposium*. (Charlottesville, Virginia, May 2011. Received an Honorable Mention award).
Lo RH, Fox DA, and Columbus L. "Investigation of the *Neisseria Gonorrhoeae* Opacity Associated Membrane Protein A Structure, Dynamics, and Host-receptor Interactions." *University of Virginia's 11th Annual Robert J. Huskey Graduate Research Exhibition* (Charlottesville, Virginia. March 2011).
Lo RH and Moog RS. "Characterization of Environments in AOT Micelles." *Franklin & Marshall College Departmental Chemistry Research Fair* (Lancaster, Pennsylvania, July 2008).
Lo RH and Moog RS. "Characterization of Environments in AOT Micelles." *Franklin & Marshall College Hackman Student Research Fair* (Lancaster, Pennsylvania, October 2007).

Lo RH and Moog RS. "Characterization of Environments with SDS Micelles and Exploring Quantum Efficiency." *Franklin & Marshall College Departmental Chemistry Research Fair* (Lancaster, Pennsylvania, July 2007).
Lo RH and Moog RS. "Characterization of Environments in Triton X-100 Micelles." *Franklin & Marshall College Hackman Student Research Fair* (Lancaster, Pennsylvania, October 2007).
Lo RH and Moog RS. "Characterization of Environments in Triton X-100 Micelles." *Franklin & Marshall College Hackman Student Research Fair (*Lancaster, Pennsylvania, October 2006).

TEACHING EXPERIENCE

Interest

General Chemistry, Physical Chemistry, Biochemistry, Biophysics, NMR Spectroscopy, Pedagogy

Experience

University of Virginia **Charlottesville, VA**

***Biochemistry Intern Instructor* (CHEM4411, 4421)** *Fall 2012-Spring2014*
Served as an intern instructor in the Biochemistry4411 and 4421 courses, and engaged in the lab lectures with the 88 participating students. Review sessions were also hosted for students outside of normal class hours to ensure and enable student understanding and success in the laboratory course.

***From Your Lab Bench to Your Medicine Cabinet Intern Instructor* (CHEM4430)** *Spring2013, Spring2014*
Served as an intern instructor in the From Your Lab Bench to Your Medicine Cabinet course. Lead and facilitated discussion and analysis of contemporary primary literature articles. This course is rooted strongly in the CREATE (Consider, Read, Elucidate hypothesis, Analyze and interpret data, think of the next Experiment) method.

Biochemistry Teaching Assistant Mentor *Fall 2012-Spring2013*
Served as a mentor to the Biochemistry 4411 and 4421 TAs and acted as a resource for laboratory protocols, grading, and troubleshooting. Selection for assistant mentorship was based on previous performance as a TA and consultations were made on a weekly basis.

Undergraduate Student Mentor *Spring2012-Fall2012*
Acted as a mentor to undergraduate Sidney Bush('15) and instructed her towards the purification and biophysical characterization of the thrombospondin repeat domains of the brain angiogenesis-1 protein.

Chemistry 4411, 4421 Teaching Assistant *Fall 2010-Spring2011*
Instructed students in the Biochemistry 4411 and 4421 laboratory course investigating the putative functions of TM0542 and TM1385, two proteins from *Thermotoga maritima.* The TM proteins studied in this course had structures obtained by the Protein Structure Initiative (PSI) Centers but were of unknown function; students in the Biochemistry 4411 course use bioinformatics and design appropriate spectroscopic assays to functionally characterize the protein of interest, and systematically perturb the assay (e.g., temperature/pH dependencies, or cofactor requirements).

Chemistry 1411, 1421, 1611, 1612 Teaching Assistant *Fall 2009-Spring2010*
Instructed students in the College of Arts and Sciences in the General Chemistry laboratory course, as well as engineering students in the modified General Chemistry laboratory course tailored for students that have had less exposure to high school chemistry and/or laboratory experience.

Franklin and Marshall College **Lancaster, PA**

Chemistry 321 Teaching Assistant *Fall 2008*
Provided guidance and assistance to undergraduate students in the Thermodynamics and Kinetics course and laboratory.

Professional Development and Outreach

University of Virginia **Charlottesville, VA**

Tomorrow's Professors Today Fellowship *Summer2012-2014*

Participated in the *Tomorrow's Professor Today* program, which is hosted by the Teaching Resource Center at the University of Virginia. This program focuses on graduate student improvement in three key areas- student preparedness in teaching, professional development, and adjustment to a university career, to enhance the teaching skills and effectiveness of the selected participants.

ChemistryLEAD(Learning through Experiments And Demonstrations) at UVa *September 2012-Current*

Helped bring inquiry-based science experiments and demonstrations into elementary and middle school classrooms throughout Albemarle County, as well as developed and lead activities for the BLAST (Building Leaders for Advancing Science and Technology) summer camp. Graduate student-led activities taught K-12 students fundamental concepts in chemistry through hands on experimentation.

LEADERSHIP EXPERIENCE

University of Virginia **Charlottesville, VA**

Chemistry Graduate Student Council *2011-2012*

Acted as webmaster for the GSC website.

First Harmonics Social Chair *2009-2012*

Created posters, designs, and shirts, helped plan events for, and performed in UVa's only all science and engineering graduate student acapella group.

Franklin and Marshall College

Lancaster, PA

House Adviser, Thomas Hall Office of College House Administration *Fall 2008-Spring 2009*

Performed residential assistant duties to assist 30 upperclassmen residents

House Adviser, Marshall Hall Office of College House Administration *Fall 2007-Spring 2008*

Performed residential assistant duties to assist 20 first-year residents in their transition to campus life.

PROFESSIONAL AFFILIATIONS

Biology Faculty of 1000 (F1000), Contributing Associate Member	2012-present
Biophysical Society, Member	2011-2012
American Chemical Society, Member	2009-2010

ACADEMIC HONORS AND ACHIEVEMENTS

Robert J Huskey Travel Fellowship November 2013	
NSF Travel Award Recipient June 2012	
BTP Honorable Mention	May 2011
Hackman Research Fellowship	2006-2008
Dean's List	2005-2006

Table of Contents

Copyright page II

Dedication III

Abstract IV

Curriculum Vitae V

Table of Contents IX

List of Figures XIV

List of Tables XVI

Chapter 1: Membrane protein structure and function 1

1.1 Protein structure and function 1

1.2 The membrane environment 7

1.3 The pitfalls obstructing membrane protein progress 18

1.3.1 Pitfall 1: Protein expression is a difficult challenge 18

1.3.2 Pitfall 2: Membrane protein environments *in situ* are not conducive to biophysical characterization 22

1.3.3 Pitfall 3: Membrane mimics are necessary for *in vitro* studies 23

1.3.3.1 Detergent micelles 24

1.3.3.2 Liposomes 27

1.3.3.3 Mixed micelles and bicelles 28

1.3.3.4 Immobilized lipodisks 29

1.3.3.5 Nanodiscs 29

1.4 **Opacity-associated (Opa) outer membrane proteins of *Neisseria* 36**

1.4.1 *Neisseria meningitidis* and *Neisseria gonorrhoeae* 36

1.4.2 Opa protein structure, function, and diversity 37

1.5 Dissertation overview..... 41

1.6 References..... 42

Chapter 2: NMR spectroscopy 53

2.1 Introduction to the theory and principles of NMR..... 55

2.1.1 Nuclear spin 55

2.1.2 Nuclear shielding and chemical shift..... 58

2.1.3 Nuclear spin relaxation 60

2.1.4 Dynamics 61

2.2 Biomolecular NMR spectroscopy 65

2.2.1 The single pulse experiment 65

2.2.2 Heteronuclear NMR experiments 66

2.2.3 NMR 3D assignment experiments..... 67

2.2.4 Nuclear Overhauser Effect Spectroscopy (NOESY) 68

2.3 Overcoming the limitations of NMR 71

2.3.1 Transverse relaxation-optimized spectroscopy (TROSY)..... 71

2.3.2 Perdeuteration and selective amino acid labeling..... 73

2.3.3 Protein expression..... 74

2.4 Membrane protein structures solved by NMR 75

2.4.1 Beta barrel structures solved by NMR..... 76

2.4.2 Structures are solved in myriad conditions..... 82

2.5 References..... 85

3 **Chapter 3: Optimization of NMR-based membrane protein environments** **92**

3.1 Introduction **93**

3.2 Materials and Methods **97**

3.2.1 Cloning, expression, and purification of Opa proteins 97

3.2.2 Preparation of protein-detergent complexes 98

3.2.3 Detergent exchange of Opa solubilized complexes 99

3.2.4 Trypsin treatment of Opa_{50} 100

3.2.5 Circular dichroism and NMR spectroscopy 100

3.2.6 Cloning, expression, and purification of MSP1D1 proteins 101

3.2.7 Reconstitution of Opa proteins into nanodiscs 103

3.3 Results and Discussion **104**

3.3.1 Detergents affect the Opa_{50} fold 105

3.3.2 Temperature affects β-barrel and loop dynamics 107

3.3.3 Ionic strength affects Opa_{50} spectral quality 108

3.3.4 Proteolysis facilitates investigation of ionic effects 108

3.3.5 Relaxation data excludes Opa_{50} oligomerization 110

3.3.6 Comparison of detergent and lipid environments 111

3.4 Concluding Remarks **122**

3.5 Acknowledgements **123**

3.6 References **124**

Chapter 4: Assessing the binding of Opa_{50} and Opa_{60} with cognate receptors **129**

4.1 Introduction **130**

4.1.1 Opa proteins interact with two receptor families 130

4.1.2 Heparin sulfate proteoglycans interactions 130

4.1.3 CEACAMs 132

4.2 Materials and Methods 138

4.2.1 Preparation of Opa samples for binding studies 138

4.2.2 Preparation of cognate host receptors 138

4.2.3 ITC titration assays for Opa_{60}-receptor binding 139

4.2.4 NMR titration assays for Opa_{50} receptor binding 141

4.3 Results and Discussion 142

4.3.1 NCEACAM1-GST dimerization affects ITC results 142

4.3.2 Opa proteins in FC12 micelles do not bind to conjugate receptors 144

4.3.3 Preliminary evidence that Opa_{50} in DMPC nanodiscs binds to heparin 147

4.4 Concluding Remarks 153

4.5 Acknowledgements 155

4.6 References 156

5 Chapter 5: Dynamics of a model helical membrane protein TM0026 160

Introduction 162

5.1.1 Membrane protein dynamics by EPR 162

5.1.2 TM0026, a model polytopic membrane protein 168

5.2 Materials and Methods 171

5.2.1 Cloning, expression, and purification of TM0026 protein 171

5.2.2 Spin labeling of TM0026 171

5.2.3 EPR spectroscopy 172

5.2.4 NMR spectroscopy..... 172

5.3 Results and Discussion..... 174

5.3.1 SDSL mutations do not significantly perturb the fold of TM0026..... 174

5.3.2 The dynamics of model protein TM0026 by NMR and EPR 179

5.4 Conclusions..... 190

5.5 Acknowledgements 191

5.6 References..... 192

Figures and Tables

Figure 1.1 The permissible rotational angles in a peptide bond..... 5

Figure 1.2 A generalized Ramachandran plot adapted from Lovell *et al*. for structure validation by Cα geometry 6

Figure 1.3 The Fluid Mosaic Model, as proposed by Singer and Nicolson..... 10

Figure 1.4. Hydrophobicity scales determined for the 20 amino acid residues..... 16

Figure 1.5. The model of folding and membrane insertion of a β-barrel outer membrane protein, OmpA. 17

Figure 1.6. Models for the various lipid bilayer mimics used in membrane protein studies. 33

Figure 1.7. The proposed three stage model for detergent solubilization of lipid bilayers, from the bilayer perspective. 34

Figure 1.8. Optimization of nanodisc size for OmpX. 35

Figure 1.9. A simplified schematic for the phagocytosis of *Neisseria* bacteria. 39

Figure 1.10. The proposed topology structure for generic Opa proteins, with the sequences of Opa_{50} and Opa_{60}. 40

Figure 2.1. The graphical representation of slow, intermediate, and fast two site exchange on the chemical-shift time scale. 64

Figure 2.2. A simplified NOESY spectrum for two interacting protons of an Ala-Val peptide. 69

Figure 2.3. The INEPT experiment. 70

Figure 2.4. Double INEPT (HSQC) experiment for heteronuclear correlation with sensitive (S) nucleus detection. 71

Figure 2.5. The structures of the seven β-barrel membrane proteins determined by NMR. 84

Figure 3.1. The chemical structures of commonly used detergents and lipids in structural/functional studies. 96

Figure 3.2. Opa_{50} trypsin digestion. 115

Figure 3.3. Detergents that stabilize fold do not facilitate folding. 116

Figure 3.4. NMR spectral quality and structure of Opa_{50} in various detergents. 117

Figure 3.5. Ionic strength affects Opa_{50} in FC12. 118

Figure 3.6. Several protein-detergent complex (PDC) scenarios can lead to line broadening. 119

Figure 3.7. Loop interactions modulate spectral quality. 120

Figure 3.8. NMR ^{15}N, ^{1}H TROSY-HSQC spectral overlay of Opa_{60} in nanodiscs and detergent micelles. 121

Figure 4.1. A simplified figure for the structure of HSPG. 136

Figure 4.2. The crystal structure determined for the N-terminal D1 domain of human CEACAM1 (PDBid:2gk2). 137

Figure 4.3. N-CEACAM-GST oligomerizes at room temperature. 150

Figure 4.4. Opa_{50}–heparin binding interactions result in spectral intensity attenuation due to decrease in mobility. 151

Figure 4.5. A schematic for the proposed interaction of heparin binding to the Opa_{50} loops. 152

Figure 5.1. The chemical diagram of the spin labels used. 166

Figure 5.2. An example plot of ΔH_0^{-1} and $\langle H^{-2} \rangle$ for T4L, adapted from

Mchaourab *et al*..... 167

Figure 5.3. The NMR chemical shifts for TM0026, from Kronke *et al*..... 170

Figure 5.4. NMR spectra of select regions of interest with the predicted topology map of TM0026. 183

Figure 5.5. The EPR spectra of SDSL residues of TM0026. 184

Figure 5.6. The 15N NMR relaxation data for TM0026..... 187

Figure 5.7. The EPR M_S and $\langle H^2 \rangle$ data plotted against the residues of TM0026.. 188

Figure 5.8. ^{15}N relaxation data for TM0026 displayed as the R_1R_2 product. 189

Appendix Figure A1. The ^{15}N,^{1}H-HSQC spectral overlay of Opa_{50} and Opa_{60} in FC12. 197

Appendix Figure A2. The ^{15}N, ^{1}H HSQC for Opa_{50} in FC12 micelles. 198

Appendix Figure A3. Construction of truncated membrane scaffold protein variants (MSP). 199

Appendix Figure A4. A Western blot of the results from multiple N-CEACAM1-GST - Opa_{50} experiments. 200

Appendix Figure A5. The ^{15}N, ^{1}H HSQC NMR spectra for all TM0026 R1' mutants..... 201

Table 1.1: A table adapted from Opekarova *et al*. detailing the lipid composition of selected biological membranes..... 31

Table 1.2. The size and lipid content of optimized phospholipid nanodiscs, adapted from Hagn *et al*..... 32

Table 2.1 Gyromagnetic ratios, spin quantum numbers, and transition frequencies of some commonly observed NMR nuclei. 57

Table 2.2. The physical properties and buffer conditions of the seven β-barrel solution structures determined by NMR. 83

Table 3.1. Physical properties of pure detergents. 95

Table 3.2. The peak count from Figure 3.5. 113

Table 3.3. Correlation times estimated from T1/T2 ratios. 114

Table 4.1. The multiple sequence alignment of hypervariable regions of the extracellular loops for several Opa protein variants. 135

Chapter 1: Membrane protein structure and function

For membrane protein biophysicists, the opening words "Membrane proteins constitute about one third of all proteins in living organisms, continue to be prime drug targets, and yet are highly underrepresented in biochemical studies"(*1*, *2*) or some variant thereof have become mantra, finding a place in nearly every abstract and grant that has touched our eyes. It is easy, perhaps, to repeat, rather than consider the full implications of our motivations – despite the dearth in structural information present, over 60% of approved drugs target membrane proteins, including all 10 of the top 10 pharmaceutical drugs driving a billion dollar market.(*3*) These statistics only serve to complement the central role that membrane proteins play in basically all physiological processes. Indeed, the Royal Swedish Academy of Sciences recently awarded the 2012 Nobel Prize in Chemistry to Robert Lefkowitz and Brian Kobilka, for their work on G protein-linked receptors, a family of transmembrane receptors, in recognition of their outstanding contributions to this still growing field. This chapter will briefly outline protein biochemistry, elaborate on the importance of membrane proteins and discuss the pitfalls to progress in the field, and conclude with introduction of the opacity-associated outer membrane protein (Opa) from *Neisseria gonorrhoeae*, a model family of β-barrel membrane proteins used in these studies, as they will comprise the majority of this work.

1.1 Protein structure and function

If you have ever walked across a room, you have reaped the rewards from the hard work that your body's proteins have done. From ferrying oxygen from your lungs to the tissues of your body (hemoglobin) to helping you visually perceive light (rhodopsin), the proteins in your body are central towards every biological process that you perform. Proteins function as enzymes, catalyzing complex chemical reactions that drive our everyday processes.(*4*) Other proteins are involved in transporting and storing important biomolecules such as metal ions, oxygen, and glucose for later use, while still others generate the mechanical motion necessary for muscle contraction. It is no wonder that proteins have been called "biomolecular building blocks".(*5*) Ever since 1953, where Sanger determined the first complete amino acid sequence of bovine polypeptide hormone insulin, biochemists have been working to understand protein structure to gain a better knowledge of protein molecular mechanisms of action, as well as to tease apart evolutionary relationships among proteins and the organisms that produce them.(*6*) Protein structure, and subsequently function, can be best understood by physical properties correlating the chemical relationships between protein components. This organization is typically divided into four categories:

1. Primary structure, or 1° structure, is the amino acid sequence of a protein's polypeptide chains. The individual amino acids that make up the 1° structure contain a wide variety of functional groups that include H-bonding groups (alcohols, thiols, carboxylic acids, amines), groups capable of acid/base chemistry, and groups that drastically affect the flexibility and dynamics of a given protein. To form these biopolymers, the amino group

of one amino acid reacts with the carboxyl group of a second amino acid in a condensation reaction, resulting in the peptide bond of the polypeptide. This peptide bond has a rigid, planar, structure as a consequence of π character of the carbonyl-amide bond, which implies that the backbone of a protein is a linked sequence of rigid planar peptide groups. The polypeptide's backbone conformation is formed by its torsion angles about the C_α – N bond (φ) and the C_α – C bond (ψ), which are defined as 180° when the polypeptide chain is in its planar, fully extended conformation (Figure 1.1).

2. Secondary structure, or 2° structure, is the local spatial arrangement of a polypeptide backbone, without regard to the conformations of the amino acid side chains, that is largely defined by the hydrogen bonding network of the protein. In particular, the protein's φ and ψ dihedral angles are defined in their spatial orientations in this region and can be easily characterized, these secondary structure characteristics are typically referred to as α-helices or β-strands. This information can be summarized in a conformation map or Ramachandran diagram that represents the allowed conformations dictated by the van der Waals radii (Figure 1.2).(*4*, *7*)

3. Tertiary structure, or 3° structure, is often loosely referred to as a fold organizing 2° elements, which may define protein domains or assemblies of 2° structure, such as α-helical bundles or β-barrels. In general, soluble globular proteins consist of a core of hydrophobic amino acid residues that partition away from water, and a surface region of hydrophilic residues that are solute exposed; these arrangements stabilize interactions within the 3° structure.

4. Quaternary structure, or 4° structure, refers to the spatial arrangement of more than one polypeptide and details the non-covalent interactions and disulfide bonds between the protein's 3° polypeptide subunits.

X-ray crystallography and nuclear magnetic resonance (NMR) spectroscopy have been the primary methods of protein structure elucidation, leading to over 24,000 unique protein structures deposited in the Protein Data Bank. The vast majority of proteins that have been characterized thus far are soluble, globular proteins, a diverse group of proteins that, in their native states, exist in aqueous solvent. The growth rate for deposition of new soluble protein structures with sufficient backbone resolution has been exponential.(*8*) In contrast, there is a large discrepancy for deposition of membrane protein structures (proteins that are natively embedded in the lipid bilayer). Structure determination of membrane proteins lags far behind soluble proteins, and is hindered by many challenges associated with working with the membrane environment.

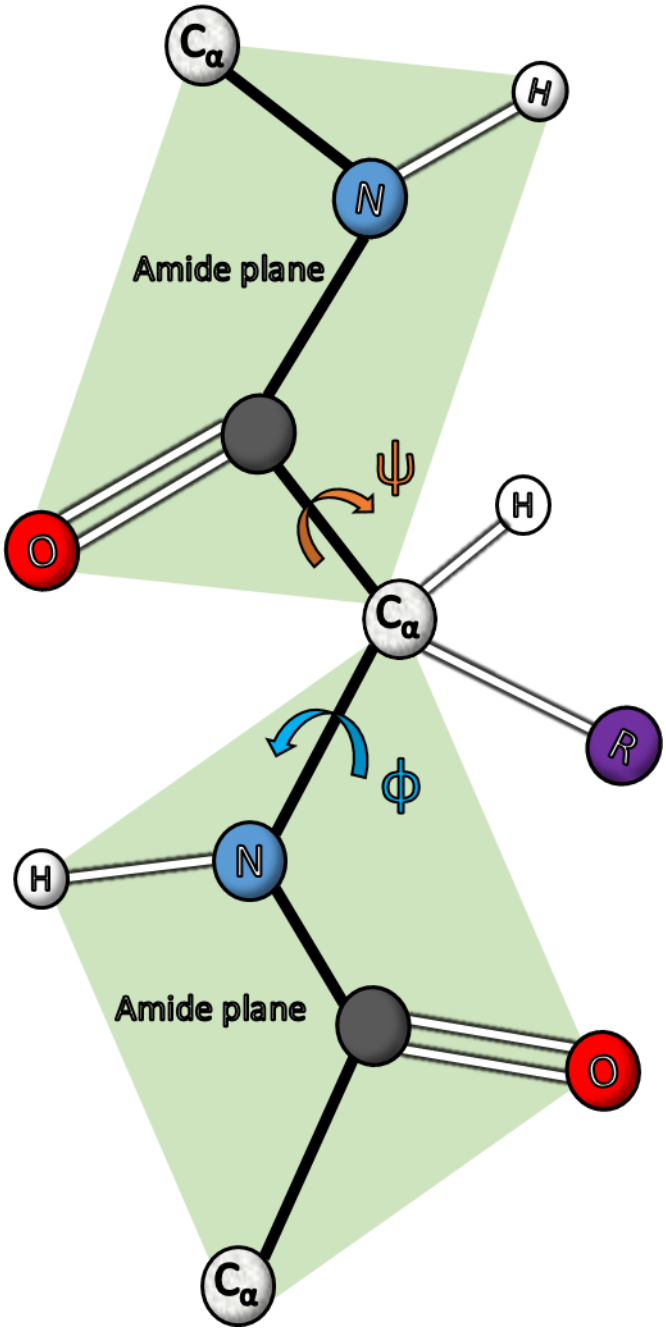


Figure 1.1. The permissible rotational angles in a peptide bond. The only reasonable free movements are rotations about the C_α-N bond (φ) and the C_α-C bond (ψ).

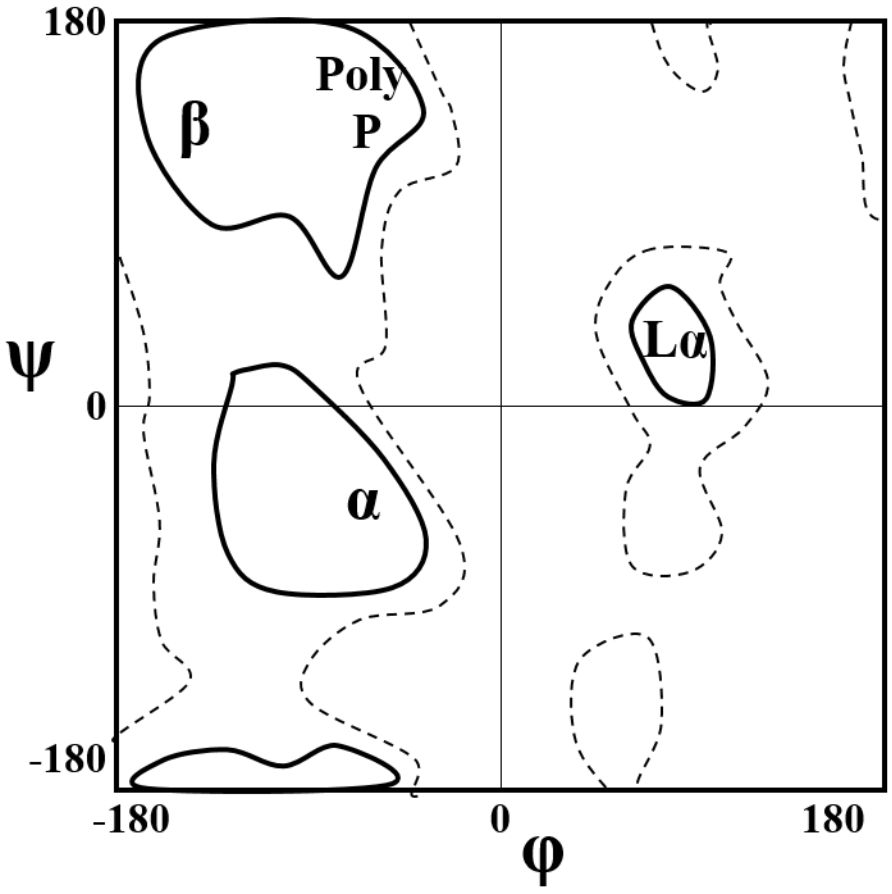


Figure 1.2. A generalized Ramachandran plot adapted from Lovell *et al*. for structure validation by Cα geometry.(*9*) Boundaries are defined by solid and dashed lines. Solid lines indicate "allowed" regions for φψ distributions, and dashed lines define "generously allowed" regions. Data to define regions were selected from 500 structures with 1.7 Å or better resolution.

1.2 The membrane environment

As opposed to soluble proteins, as the name implies, membrane proteins reside in the cell membrane, which composes the outer (in the case of Gram-negative bacteria and organelles such as mitochondria and chloroplasts, which are thought to have originated from bacteria according to the endosymbiotic theory) and inner membranes of cells, and surrounds various organelles including the nucleus, the endoplasmic reticulum, the Golgi apparatus, and other vesicles.(*10*) These natural semi-permeable barriers compartmentalize the cell and sequester the inner machinery, selecting for entry into the cytoplasm. The simplified model of the biological membrane features a lipid, sterol, and protein composition, with each component further classified based on the structural moieties.

The basic amphiphilic lipid of a lipid bilayer is generally characterized by a polar headgroup and nonpolar acyl tail groups. Lipids can be further divided into two classes: glycerophospholipids, which have glycerol-based head groups and two nonpolar acyl tails, and sphingolipids, which contain an O-linked backbone of long-chain bases, known as sphingoid bases and have a single acyl tail. Sterols are generally composed of fused cyclohexane (and 1 cyclopentane) rings known as steroids, with hydroxyl groups that overall affect membrane fluidity by altering lipid packing. Together, these amphiphilic molecules self-assemble into membranes and form hydrophobic and hydrophilic regions of the lipid bilayer, largely due to the hydrophobic effect. The driving force for assembly is an entropic effect; water molecules closest to the acyl chain of the lipid arrange to form clathrates, hydrogen bonded cages around the nonpolar chains.(*11*) These clathrates represent a drastic reduction in mobility and increase in order. To minimize water's

entropic cost, the nonpolar tails of the lipid aggregate and reduce the total surface area of nonpolar moiety exposed to water molecules, resulting in the two domain (hydrophilic headgroups and hydrophobic tail groups) system. The hydrophobic domain may be further enthalpically stabilized by van der Waals interactions between the acyl chains.(*12*)

To determine the energetics of amphiphile assembly, the hydrophobic domain was investigated by examining the solubility of hydrocarbons in water. The free energy of transfer of fatty acids from water to an alkyl solute can be calculated by the $\Delta G_{transfer}$ for the solute of interest.

$$K_p = K_{eq} = [solute]_{H2O}/[solute]_{alkane} \quad (1.1)$$

$$\Delta G_{transfer} = -RTln(K_{eq}) \quad (1.2)$$

where K_p is the partition coefficient, K_{eq} is the equilibrium constant, and $\Delta G_{transfer}$ is the free energy change for the transfer from alkane to water.

Based on experimental results by Tanford *et al.*, the energy cost is directly correlated with the chain length of the fatty acid and can be approximated, by the equation:

$$\Delta G_{transfer} = X - Y(n_C-1) \quad (1.3)$$

where X ranges from (0.2 - 2.9) kcal/mol and Y ranges from (0.6-0.8) kcal/mol depending on the moiety attached to the alkane (in N-alkyl betaines and N-alkyl trimethyl ammonium bromides), n_C is the chain length (the CH_2 group closest to the head group is not as hydrophobic and makes a negligible contribution). The sum of these small interactions stabilizes the overall lipid bilayer, maximizing water-water interactions while minimizing the entropically prohibitive acyl chain-clathrates effects.

The energetics of membrane formation are explained by the hydrophobic effect, but are not sufficient to describe the whole basic structure of biological membranes. The classic model is the Fluid Mosaic Model put forth by Singer and Nicolson in 1972 (*13*), which describes a dynamic membrane with components (lipids and randomly scattered patches of membrane proteins) that move laterally throughout the membrane (Figure 1.3). This model serves the biochemistry tyro, and is ubiquitous in introductory Biology textbooks as the foundation for understanding the structure of cellular membranes, but is largely an oversimplification and is currently evolving and being updated.

The primary challenge to the Fluid Mosaic Model is the assumption of stochastic distribution of proteins within the bilayer; rather than the "membrane protein icebergs in a sea of lipids" picture, experimental observations indicate that membranes have specialized microdomains ranging from 25-700 nm^2 sometimes referred to as lipid rafts.(*14*) These domains typically contain elevated cholesterol and highly saturated glycosphingolipid content which enables a tight rearrangement of the lipids.(*15*) Additional structures are observed as well. Caveolae form unique "cave-like" invaginations that are formation dependent on caveolin-1 protein. Planar (non-caveolar) lipid rafts have also been observed. These planar rafts are continuous with the plane of the plasma membrane and cannot be easily distinguished from the surrounding membrane.(*16, 17*) Both domains are exciting and currently being updated to advance the field of membrane protein biochemistry, but the breadth of information is too expansive to cover in this chapter alone, and the reader should seek references (*18, 19*) for additional information.

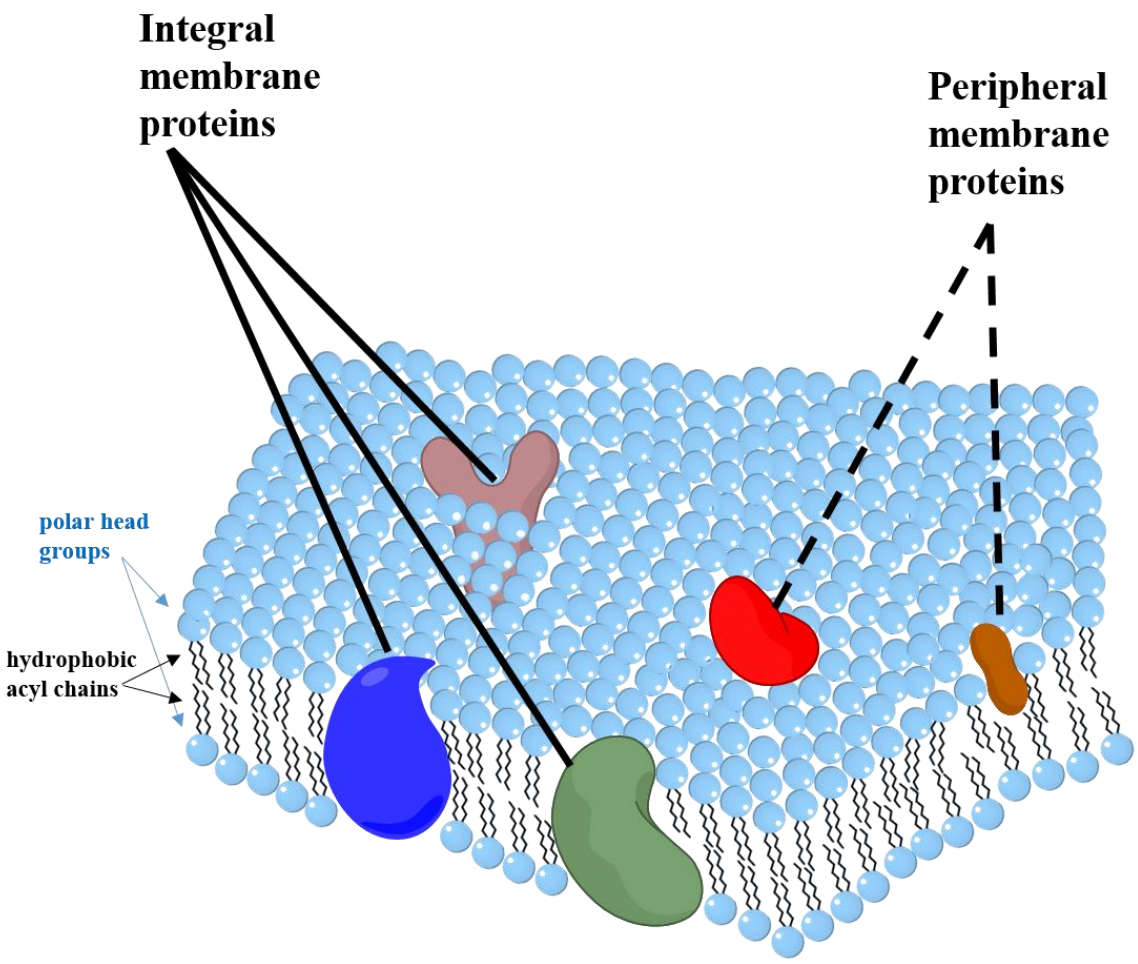


Figure 1.3. The Fluid Mosaic Model, as proposed by Singer and Nicolson. The basic structure of the membrane is a lipid bilayer, with acyl chains from each leaflet forming a nonpolar interior. Integral proteins span the length of the membrane, while peripheral proteins are partially embedded.

1.3 Membrane protein structure and function

Membrane proteins are the workhorses of the lipid bilayer that drive active and passive transport, cell communication by signal reception and transduction, biological energy conversion, and enzymatic functions. They are further classified into two groups based on membrane embedment. Integral (or intrinsic) membrane proteins span the length of the lipid bilayer and may have one or more transmembrane (TM) regions. peripheral (or extrinsic) membrane proteins are proteins that associate at the surface of the membrane and can be lipid anchored proteins that are attached covalently to fatty acids or lipids embedded within the membrane. Membrane proteins typically fall into three primary functional categories: receptors, enzymes, and transporters. Receptors, such as in the case of the G-protein coupled receptor family, bind to a variety of ligands and trigger a response that leads to cellular signaling. Membrane-bound enzymes carry out diverse covalent catalytic functions such as redox reactions in electron transport chain, and metabolizing membrane components such as phospholipids and sterols.(*10*) Transport membrane proteins fall into several different classes, depending on mechanism of action, stoichiometry of the transported molecule, and the energetics of the transport process. Membrane proteins involved in transport are divided into two categories. Active transport membrane proteins make use of either ATP hydrolysis or the cell's electrical and concentration gradients to pump solutes against a gradient. Passive transport membrane proteins direct the flow of solute through the membrane until their electrical or concentration gradients are dissipated.

Membrane proteins are further divided into two subclasses based on fold: α-helical bundles or β-barrel membrane proteins. The native environments of each are different; α-helical

proteins are typically found in cell membranes, and in the inner membranes of bacteria, whereas β-barrel membrane proteins are found only in Gram-negative bacterial outer membranes as well as in the outer membranes of mitochondria and chloroplasts, which are thought to be Gram-negative bacterial derivatives.(*20*) Membrane protein folding differs for each class of protein – in general, α-helical folding can be broadly thought of as a two stage process, as originally proposed by Popot and Engelman: insertion and folding.(*21*) The first stage typically consists of a translocon complex directing the insertion of the membrane protein across the lipid bilayer, establishing the topology of some of the membrane inserted fragments. (*22*) Many membrane proteins are inserted as they emerge from the ribosome through protein-conducting channels known as secretory complexes – SecY in bacteria and Sec61 in eukaryotes, though there are exceptions, such as diacylglycerol kinase (DAGK), which have been proposed to insert spontaneously and independently of the secreory complexes.(*23*) These secretory complexes are the "intelligent" gatekeepers to the membrane, selectively allowing passage of the polypeptide chain while halting travel of small ions and other molecules. SecY/Sec61 then determine whether or not to reverse the orientation of the emerging segment, while also deciding the fate of the segment, whether it should pass through to the periplasm, or if the SecY/Sec61 complex should open the channel laterally to allow for membrane insertion, all the while without membrane leakage. The topology of the membrane insertion is partially driven by the translocon, but also thought to be established by the positive-inside rule, which suggests that the cytoplasmic side of membrane proteins tends to be positively charged, due to the electrostatic potential differences between the outside and inside of the cell. In general, the non-translocated loops contain two to four times as many Lys and Arg residues as found

in the translocated domains, which makes it possible to predict the topological orientation of TM helices from amino acid sequences.(*24*) To establish the preference for amino acid partitioning into the membrane, Wimley and White introduced the hydrophobicity scale, which detailed the free energies of transfer for the membrane-active pentapeptide acetyl-WL-X-LL-OH (where X is one of the twenty natural amino acids) from octanol to water and observed that the amino acids with the greatest free energy for transition were most likely to be prevalent in the transmembrane region.(*25*) Moon and Fleming subsequently determined the energetics of membrane partitioning by guest amino acids by the [reversible] folding for outer membrane bacterial phospholipase OmpLa in phospholipids (Figure 1.4).(*26*) The second stage of protein folding is focused on building the tertiary and quaternary structures through assembly and reorientation of the TM segments, and is the result of a number of forces such as packing, electrostatic effects, and interactions among helical loops.

The folding of β-barrel is fundamentally different than α-helical bundle formation, as each α-helix can form independently with hydrogen bonds along the helix axis; β-barrels have hydrogen bonds between neighboring strands, including bonds between the N and C terminal strands. A single β-strand is therefore not feasible; rather, all strands of a β-barrel form concurrently, and are folded and inserted from the periplasmic side of Gram-negative bacteria.(*27*) *In vitro* folding studies of OmpA, an eight-stranded β-barrel that can be unfolded with urea and refolded by dilution (Figure 1.5) lends some insight to the spontaneous folding of β-barrels.(*28*) OmpA begins in the unfolded state (not shown) and hydrophobically collapses into a water soluble intermediate that associates with the lipid bilayer (1.5A) with disordered tryptophan residues. This bilayer-associated intermediate

then forms some β-character (in red), though tertiary contacts are not yet established; the intermediate is termed the (1.5B) "molten disk". Next, four tryptophan residues translocate to the center of the lipid bilayer; Trp7 does not translocate. This intermediate is more globular, but does not achieve the full tertiary fold, and is thus named the (1.5C) "molten globule". The complete translocation results in the (1.5D) native protein fold.

In vivo membrane insertion of integral β-barrel membrane proteins begins with co-synthesis of the membrane protein with N-terminal signal sequences that directs them to the secretion machinery in the inner membrane.(*29*) Inner membrane proteins are cotranslationally targeted to the Sec machinery, as previously indicated in α-helical insertion. To prevent co-translation of outer membrane proteins, an additional protein called trigger factor competes with the signal recognition particle for binding to the signal sequence.(*30*) Outer membrane proteins are instead post-translationally directed to the Sec machine by the chaperone protein, SecB, that binds to the proteins as they emerge from the ribosome.(*31*) Once in the periplasm, proteins with the propensity to form β-sheets are also prone to aggregation, given the stability of the multimeric structure. Chaperone proteins are therefore necessary for outer membranes to transit the periplasmic compartment in an unfolded state. SurA is a periplasmic peptidyl-prolyl isomerase in *E. coli* that has been shown to promote folding of several outer membrane proteins, including OmpA, OmpF, and LamB (*32*), and the bulk mass of outer membrane proteins have been shown to be transported by SurA.(*33*)

Identifying machinery for the insertion of Omp proteins in the outer membrane has been an arduous process. The relative populations of assembly machinery proteins are low,

which made fractionation, the classical approach to identifying outer membrane proteins, difficult.(*34*) The central component of the β-barrel assembly machine (Bam complex), BamA (originally known as Omp85) has been identified in *Neisseria meningitidis.* Voulhoux *et al.* found that a depletion of BamA led to both a reduction in lipopolysaccharide (which also stimulates Omp folding), as well as misassembly of various porin monomers, suggesting that the native BamA complex plays a role in outer membrane protein assembly.(*35*) The more recent availability of the bacterial genome sequences helped in the identification of BamA and BamD, two essential components of the Bam complex.(*34*) Though the mechanism by which the Bam complex folds outer membrane proteins is not known, BamA and BamD putatively participate directly in the assembly process of folding and inserting of outer membrane proteins. Other components of the Bam complex (BamB, BamC, and BamE) vary greatly, but are less crucial, as knock out mutants generally retain viability and are not essential.(*34*)

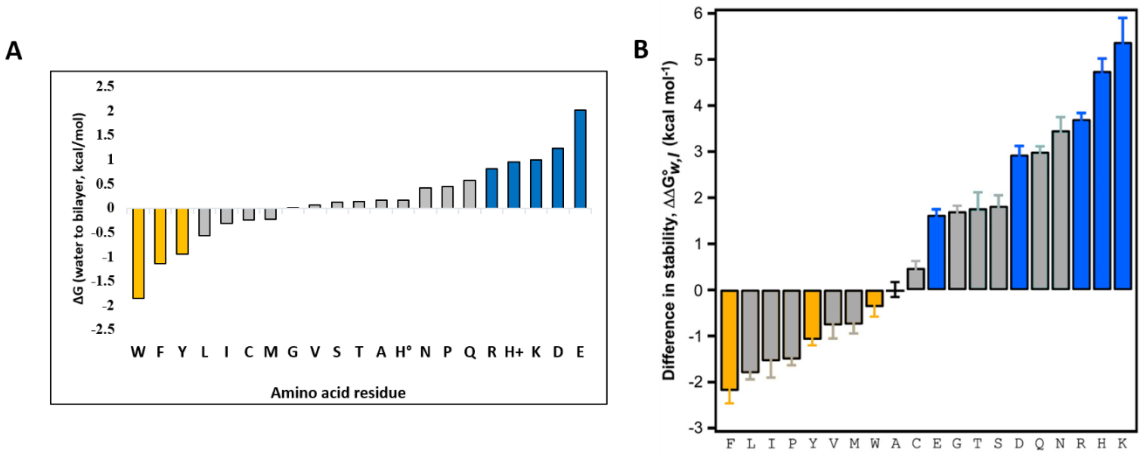


Figure 1.4. Hydrophobicity scales determined for the 20 amino acid residues.(*25*, *26*) Aromatic residues are denoted in yellow, and charged residues are denoted in blue. (A) The Wimley-White hydrophobicity scale of rank ordered experimental data for the free energy of solvation for the transfer of amino acid residues from an aqueous environment to the lipid bilayer hydrophobic region. H° and H^+ are unprotonated and protonated Histidines, respectively. (B) The whole-protein hydrophobicity scale determined for OmpLA by Moon and Fleming. The difference in the Gibbs free energy of unfolding ($\Delta\Delta G^\circ_{w;l}$) of each amino acid variant at A210, a lipid-facing exterior residue whose α–carbon is 0.2Å from the OmpLA membrane-water interface, is reported.

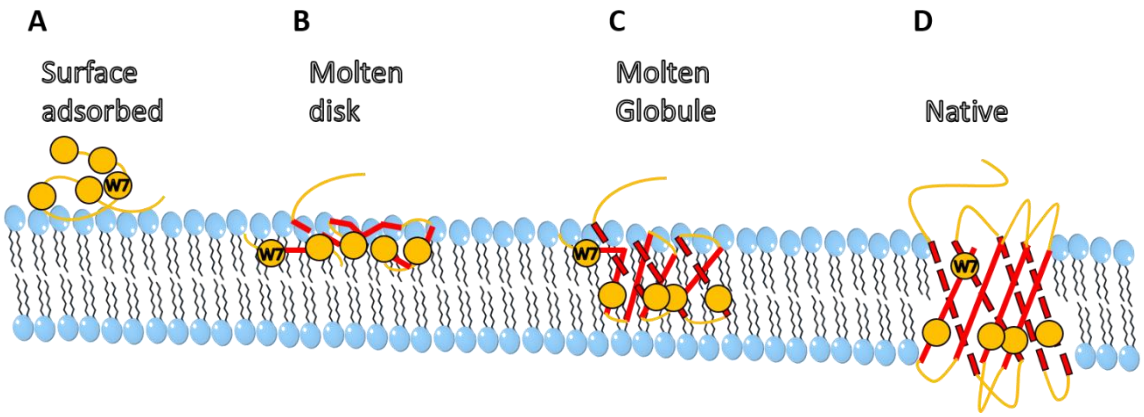


Figure 1.5. The model of folding and membrane insertion of a β-barrel outer membrane protein, OmpA. The folding mechanism was determined by Kleinschmidt *et al*. by time resolved tryptophan (Trp) fluorescence quenching, which permits direct observation of the translocation of the Trps of the protein (denoted as the yellow spheres).(*36*)

1.3 The pitfalls obstructing membrane protein progress

"*I worked on membrane proteins… once…*"
- Multiple tenured professors at various R1 institutions.

The study of membrane proteins is a particularly difficult challenge in protein biochemistry. As of February 2014, there are only 1,874 (448 unique structures) integral membrane protein structures of the 97,980 (24,101 unique structures) protein structures deposited in the Protein Data Bank (1.91%). Of the integral membrane protein structures, 1600 are integral α-helical proteins, and 274 are β-barrel structures. Traditionally, there are several major obstacles that hinder progress in both structural and functional studies. These can be divided into three pitfalls: protein expression yields are often inadequate, biophysical studies cannot be performed *in situ*, and membrane mimic selection is often empirically determined.

1.3.1 Pitfall 1: Protein expression is a difficult challenge.

The natural abundance of membrane proteins is typically too low to obtain sufficient concentrations (generally, milligram quantities are required) for structural and functional studies.(*37*) Overexpression systems are, therefore, prerequisite for structure and function determination. Overexpression of membrane proteins through accumulation in a membrane system avoids the inclusion body problem, but several problems are present. Chaperone proteins and cellular machinery necessary for protein folding are not overexpressed and consequently, macromolecular crowding of unfolded polypeptide can result in non-productive aggregation with other unfolded species.(*38*) Additionally, overexpression of membrane proteins is often toxic to the organism, severely reducing the net yield of protein.

Though the exact mechanism varies, there is a general assumption that the overexpressed membrane protein negatively affects the integrity of the membrane, which results in dampened cell viability from retarded growth and hampered division.(*37*) Despite the negative features of overexpression, β-barrel membrane proteins can be overexpressed in inclusion bodies, where they can be readily isolated, subsequently refolded *in vitro*, and used in further studies. However, there are very few examples of α-helical bundle membrane proteins that have been successfully refolded following denaturing isolation from inclusion bodies.(*39*) Additionally, in selecting the optimal *in situ* system for membrane protein expression, there are requirements necessary for correct folding of certain membrane proteins. Among these are lipid composition, the presence of molecular chaperones and secretory complexes for membrane protein targeting and insertion. Depending on the protein of interest, both prokaryotic and eukaryotic cell systems are employed; the most often used are the bacterial, insect, and yeast cell lines, though the lipid bilayer composition differs in each system (Table 1). Cellular machinery necessary for post translational modifications such as n-glycoslyations and disulfide bond formations must also be taken into account. In particular, glyscosylation for eukaryotic membrane proteins can be essential for proper folding, stability, and function. The composition and number of N-glycans can be crucial and is dependent on the overexpression host system used.(*40*)

Insect cell lines have the cellular machinery necessary for production of recombinant proteins that must be glycosylated, disulfide bonded, or membrane inserted for functional activity.(*41*) Expression of the proteins of interest are performed by the baculovirus-mediated expression system, which is one of the most efficient and popular systems among eukaryotic hosts. The baculovirus expression system is also conducive to incorporation of

stable isotopes in proteins for NMR study.(*42*) However, insect cell lines contain membranes with low levels of cholesterol, at ratio of 0.04 for sterol content:phospholipids (for mammalian and yeast membranes, it is >0.5) (*43*) which is a potential bottleneck for the overexpression of sterol-requiring proteins. Hua *et al*. have demonstrated that transcription can be repressed depending on a cell's sterol population.(*44*) Addition of cholesterol to the medium of infected baculovirus/Sf9 insect cells has been shown to increase expression and can be a critical factor for the function of membrane proteins expressed in a baculovirus/Sf9 cell system.(*45*) Insect cell lines do offer several benefits. However, insect cell lines are often expensive, and due to host cell incompatibilities, there is the potential that membrane proteins will be incorrectly folded, while certain post-translational modification machinery is not present, and membrane proteins may not be glycosylated.

The yeasts *S. cerevisiae*, *S. pombe*, and *P. pastorias* are an often used, relatively cheap eukaryotic system for expression. High yield protein expression levels coupled with simple minimal-defined growth media easily allows for isotopic labeling.(*46*) Certain post translational modifications, such as glycosylation have been shown to be possible in yeast systems. *P. pastoris* with reengineered glycosylation pathways were shown to successfully express glycosylated recombinant proteins.(*47*) The main detriments to using yeast systems include susceptibility to proteolysis, as well as the targeting of membrane proteins post expression to vacuoles. Yeast cells, like insect cells, also suffer from a shortage of specific sterols (sito-, stigma- and campesterol); the main fungal sterol, ergosterol takes the place of cholesterol, but is often not structurally similar enough to cholesterol to result in fully functional proteins.(*43*) This is evident in human P-glycoprotein, which had a

markedly decreased drug-binding activity in yeast, and did not function as a drug transporter in the yeast plasma membrane.(*48*)

E. coli is the most prevalent expression host for the production of recombinant proteins. These cells have been extensively characterized, have a short generation time, low cost, and are fairly easy to use, making them ideal candidates. Outer membrane proteins synthesized with amino-terminal leader sequences can be translocated across the cytoplasmic membrane.(*49*) Overexpressed proteins without such sequences in *E. coli* accumulate in the cytoplasmic (inner) membrane, or in cytoplasmic inclusion bodies. However, overexpression of many membrane proteins results in aggregate formation and subsequent partitioning to inclusion bodies; refolding from inclusion bodies into functional proteins is not often successful.(*50*) Polypeptide elongation and protein folding rates are also considerably higher in bacteria than in eukaryotes, which could result in mis-targeting and misfolding of heterologously expressed membrane proteins.(*37*) There have been several developments towards overcoming these limitations. Decreasing the temperature of cell culture during protein expression improves solubility and stability of proteins, and cold-shock expression vectors have been developed for this exact purpose, producing high yields of proteins.(*51*) Co-expressing molecular chaperones with the protein of interest has been shown to circumvent aggregation of the unfolded species.(*38*) Solubility enhancing fusion tags, such as the maltose binding protein (MBP) or glutathione-s-transferase (GST) tags have been used to enhance protein solubility as well as function as an affinity tag.(*52*)

Successful membrane protein expression does not necessarily result in functional protein. Function of the membrane protein itself may not be present in the chosen cell system, such

as in the case of ion channel potassium selective glutamate receptor in *E. coli*.(*53*) Overexpression of the membrane protein of interest followed by subsequent activity may also promote intracellular signaling, which in the case of β_2-adrenergic receptors in mammalian cells led to cell death.(*54*) Large quantities of membrane protein could potentially disrupt the membrane as well; overexpression of fumarate reductase in *E. coli* results in a lipid:protein ratio that is comparable to the lipid:protein ratio pre-induction of expression, indicating that the total amount of membrane doubles.(*55*) Finally, problems can occur at the transformation, transcription, and membrane insertion steps, which often results in cell death; Miroux *et al* found that expression of a variety of membrane proteins was toxic to BL21(DE3) host cells and that a double-mutant host, C43(DE3), was necessary to produce elevated levels of proteins. In these cells, transcription of the gene of interest was delayed after induction which ultimately resulted in incorporation of the over-produced protein in the membrane.(*56*) Overall, depending on the area of study, these systems are far from perfect and are still a highly studied area of research. These expression system hurdles have particularly contributed to the paucity of eukaryotic membrane protein structures. Of the thousands of proteins in the plasma membrane of eukaryotic cells, only two recombinant protein structures have been determined: rhodopsin and aquaporin.(*57, 58*)

1.3.2 Pitfall 2: Membrane protein environments *in situ* are not conducive to biophysical characterization.

Assuming membrane protein expression difficulties can be overcome, there is still the challenge of the membrane protein environment, which is complex, heterogeneous, and

dynamic. The mosaic lipid bilayer is not conducive to many standard biophysical techniques for structure/function determination, such as nuclear magnetic resonance (NMR) spectroscopy, X-ray crystallography, circular dichroism, and isothermal titration calorimetry (ITC). There are many difficulties in conducting any one of these experiments in the native environment due to the various contributions from the proteins and lipids that are not of interest. Instead, the membrane proteins of interest must be extracted and purified from the native environment, and because the majority are not soluble in the aqueous solution, they must be studied in an environment that mimics the lipid bilayer.

1.3.3 Pitfall 3: Membrane mimics are necessary for *in vitro* studies.

Membrane proteins necessitate a solvent that stabilizes both the hydrophobic nature of the transmembrane segments and the intracellular and extracellular soluble regions of the protein. This thesis is primarily concerned with an outer membrane protein, and in order to determine an optimal solvent condition, the lipid composition of the outer membrane of the native gram-negative bacteria should be considered. The outer membrane of Gram-negative bacteria is highly asymmetric and consists of phospholipids (70-80% phosphatidylethanolamine, 20-30% phosphatidylglycerol, and cardiolipin) in the inner leaflet, and lipopolysaccharide (LPS), which has multiple fatty acyl chains and complex polysaccharide structures linked to a glucosamine disaccharide backbone, in the outer leaflet.(*27*) LPS is negatively charged, and tends to have more acyl chain saturation, and is predicted to be more rigid than the inner leaflet, which is comprised of many phospholipids with unsaturated chains. In *Neisseria*, the bacteria of interest, LPS is replaced by lipooligosaccharide (LOS), which only possess the lipid A membrane anchor, and an

oligosaccharide core.(*59*) A major bottleneck in the field of membrane protein research is the difficulty in reconstituting a membrane protein in a model membrane system. Membrane mimic selection that stabilizes fold and function is often empirically determined. A number of systems that fulfill these requirements (though perhaps not as optimally as *in vivo*) have been developed to solubilize and reconstitute membrane proteins *in vitro*: detergent micelles, mixed lipid/detergent micelles and bicelles, liposomes, lipodisks, and nanodiscs, to name a few (Figure 1.6). Detergent micelles and nanodiscs were primarily used in this research, and detergent selection factors such as charge and tail length will be discussed in Chapter 3.

1.3.3a Detergent micelles

Detergents are amphipathic molecules consisting of a polar head group and (generally) a single hydrophobic tail. In solution, detergent molecules exist as monomer until the critical micelle concentration (CMC) is reached, upon which detergent monomers are in equilibrium with an aggregate, called a micelle. The number of detergent monomers in a micelle is called the micelle aggregation number (n). The CMC is determined by the combined effect of the head group repulsive forces and the hydrophobic interactions of the tails and decreases as the alkyl chain length increases.(*11*) Micelle size, and shape (oblate, disk shaped, or prolate, football shaped – *rarely* spherical) are determined by the monomer shape and size, and the ratio of these quantities (p, the packing parameter) p that is defined as:(*60*)

$$p = \frac{v}{la} \tag{1.4}$$

where *v* denotes the volume of the amphiphile hydrophobic tail, *l* denotes the length of the tail, and *a* denotes the surface area of the head group. In general, for a given detergent's p, when $0 < p < 1/3$, the resulting shape is oblate, prolate, or spherical, when $1/3 < p < 1/2$, the resulting geometry is a cylinder, and when $1/2 < p < 1$, formation of a bilayer or a disk like shape forms. Cylinders and bilayers are generally not conducive to structure determination, and may accommodate more than one protein per shape due to the larger volumes. Instead, micelles, which typically only solubilize one protein, are desirable membrane mimics that stabilize the fold of the membrane protein. The existence of detergent monomer at concentrations equivalent to the CMC may impose experimental difficulties and must be taken into account for any experiment.

Detergents are most often used to solubilize membrane proteins, especially for extracting the protein of interest from the native lipid bilayer. Non-ionic detergents are generally used for extraction and solubilization of the membrane protein to prevent denaturation of the target membrane protein; these are generally assumed to solubilize the protein without affecting important structural features.(*10*) Solubilization has been studied extensively, and is thought to occur in three stages. In Stage I, detergent monomer partitions into the phospholipid bilayer. This is followed by the intermediary Stage II, where the phospholipid membranes saturated with detergent coexist at equilibrium with mixed phospholipid-detergent micelles saturated with phospholipid. The process concludes with Stage III, where phospholipid is fully solubilized by detergent micelles. (*61*) (Figure 1.7) However, the detergent that solubilizes the membrane protein of interest often does not stabilize a functional fold, which is particularly the case for tetraethylene glycol detergents with short (C7-C10) hydrocarbon chains such as C_8E_4 and C_8E_5.(*62*)

While non-ionic detergents are often used for solubilization and extraction, other detergents are often used after lipid removal and purification. Among those used are zwitterionic and ionic detergents. Zwitterionic detergents describe a wide variety of detergents that possess positive and negative charges that have a net charge of zero. This group of compounds is highly heterogeneous depending on headgroup, and is comprised of foscholines, zwittergents, anzergents, and more curious cases such as lauryldimethylamine oxides (LDAO) that are zwitterionic at pH>7, and positively charged at pH<3. There is less data on steroid-based detergents such as bile salts and 1-propanesulfonate (CHAPS), but other studies have also recently emerged for their use as potential membrane mimics.(*63, 64*) Ionic detergents, which contain a head group with a net charge that is either cationic or anionic, such as sodium dodecyl sulfate (SDS) and Triton X-100 comprise the final category of detergents; these are often extremely efficient at solubilizing membranes, but are almost always denaturing. Detergent micelles are often useful for structural and functional due to their size and ease of use, though how the micelles affect native protein structure and function is still being examined. In particular, work is currently being done by a Columbus laboratory colleague, Ryan Oliver, by using small angle x-ray/neutron scattering (SAXS and SANS, respectively) to determine how detergent micelles perturb the structures of embedded membrane proteins, and if the membrane proteins also contribute to the overall structure of the protein detergent complex.

There are several drawbacks that preclude micelle use in biophysical characterization. In experimental design, free detergent monomer, as well as free detergent micelles can interfere with assay techniques. Experimental flaws are present in ITC, where measurements are affected by free detergent monomer in solution. For spectroscopic

experiments, certain detergents have non-ideal optical properties that interfere with light scattering and absorbance measurements. Detergent monomers and micelles can potential interact with the ligand of interest, affecting measurements. Detergent micelles do not offer a native lipidic environment, as they lack a "sided-ness" which makes transport experiments impossible to perform. Detergent micelles can induce curvature stress that results in a change in the structure of the membrane protein being studied.(*43*) These criticisms need to be addressed when planning experiments involving protein-detergent complexes.

1.3.3b Liposomes

Liposomes are lipid-bilayer vesicles that vary in curvature, size, and lipid composition. Depending on the method of liposome preparation, unilamellar (single bilayer) and multilamellar (multiple bilayer) vesicles can be formed. Small unilamellar vesicles (SUVs) of diameters 20 – 50 nm can be formed with extensive sonication of multilamellar vesicles.(*10*) Large unilamellar vesicles (LUVs) with diameters from 100 nm – 5 μm are formed by extruding or freeze thawing SUVs, and can encapsulate larger volumes, but often suffer from heterogeneous size distributions, and fragility in larger vesicles.(*10*) Giant unilamellar vesicles (GUVs) are 5 – 300 μm in diameter that are cell-sized, yet very fragile.(*10*) Selection for vesicle size must take into account the advantages and disadvantages of each system. Lipid selection is also not a clear cut case; Opekarova *et al.* present a case that membrane proteins may require specific lipids, as either cofactors for their functions or as "co-structures" for their correct folding and stability.(*43*) Many other proteins are reported to have altered activities dependent on the presence of specific lipids

and/or sterols.(*43*) Incorporation of protein into liposomes results in a proteoliposome, and generally can be prepared with two methods. A pre-incorporated protein-detergent complex can be diluted into a liposome solution (or the detergent can be dialyzed out of a detergent-lipid mixture), which causes the detergent concentration to fall well below the CMC, destabilizing the micelles which incorporate the membrane protein into the lipid bilayer. The second method involves direct refolding of protein into lipids from a denaturant, the conditions for two Opa proteins investigated in this thesis have been extensively screened by Dewald *et al*.(*65*)

1.3.3c Mixed micelles and bicelles

Membrane protein curvature stress can be relieved by matching membrane mimic dimensions to the membrane protein's hydrophobic surface. Mixed micelles and bicelles incorporate mixed populations of detergents and mixed populations of long-chain/short chain lipids, respectively, and can be used to design systems that match these hydrophobic regions. Mixed micelles are discoidal aggregates of either short chain detergents with long chain detergents, or long chain detergents with short chain detergents. Lipid bicelles contain populations of short-chain phospholipids and long-chain phospholipids. The size of the bicelles is dependent upon the ratio of long-chain to short-chain lipids, and the toal concentration of phospholipid.(*10*) The main drawbacks to using bicelles are the restrictions on the composition components, as well as determining the proper stoichiometry to use for bicelles formation (and subsequent characterization of the mixed micelle/bicelle ratios). This characterization can be fruitful; Columbus *et al*. demonstrated the feasibility of mixing and matching detergents to avoid hydrophobic mismatch that

could lead to rational design of mixed micelles to facilitate NMR structure determination.(*66*)

1.3.3d Immobilized lipodisks

Immobilized lipodisks are planar lipid bilayer structures that are stabilized by polyethylene glycol (PEG-ylated) attached lipids.(*67*) These systems are emerging as an alternative, detergent-free system that has potential for certain biophysical techniques. Lorigan and Saunders *et al.* have recently published pulsed EPR double electron-electron resonance (DEER) results for single transmembrane protein KCNE1 in lipodisks that suggest that the lipodisks system is conducive to longer, more accurate DEER measurements.(*68*) Lipodisks are not as extensively used as a membrane mimic, though recently functional cyclooxyngenase-1 has been shown to be successfully incorporated in immobilized lipodisks on a column by Meiby *et al*.(*69*)

1.3.3e Nanodiscs

Nanodiscs are non-covalent discoidal assemblies of phospholipids trapped within two membrane scaffold proteins (MSP) (Figure 1.8). Sligar *et al.* developed MSPs from a modified sequence of the human serum apolipoprotein A-I (A-I). A-I naturally binds to dietary lipid molecules in the liver and intestine, and is used by doctors to measure high density lipoprotein (HDL) and low density lipoprotein (LDL) content from blood tests. Native A-I is composed of an N-terminal globular domain followed by amphipathic α-helices (8 helices containing 20 amino acid residues and 2 helices containing 10 residues).(*70*) Sligar *et al.* introduced MSP1 which contains the A-I sequence, a His tag, a

Factor X cleavage site, and the N-terminal globular domain deleted. The MSP2 construct is a fusion of two MSP1 molecules.(*71*) These MSPs resulted in 10-nm diameter disks that contained approximately 160 saturated lipids. Experimental MSP N-terminal truncation mutants of 11 and 22 residue deletions resulted in nanodiscs of roughly similar size. This result indicates that the N-terminal residues are unnecessary for nanodisc formation, and N-terminal affinity tags and proteolytic cleavage site additions should not have adverse effects on nanodisc formation.

Hagn *et al.* constructed truncated membrane scaffold proteins (Table 2) that form smaller nanodiscs of 6.5-9.5 nm diameter.(*39*) These MSPs lack half of helix 4 (ΔH4/2), helix 5 (ΔH5), helices 4 & 5 (ΔH4H5), or helices 4-6 (ΔH4-6), and the resulting nanodiscs range in size from 52 – 160 kDa (Figure 1.8). Bacterial outer membrane protein OmpX was incorporated into each of the nanodiscs with DMPC, and somewhat surprisingly, only the wild type MSP1D1 and the ΔH5 construct represented stable systems (monitored via CD thermal melting curves) for OmpX. The OmpX-MSP1D1ΔH5 complex was best suited for NMR studies based on the long-range inter-β-strand NOEs observed with ^{15}N-edited 3D-TROSY-NOESY experiments, which enabled structure determination of OmpX in nanodiscs. The dynamics of OmpX in the DMPC nanodiscs were also probed and compared with FC12 micelles; the obtained data showed that dynamics in the nanosecond to picosecond range were not perturbed by the membrane mimic.

Table 1.1: A table adapted from Opekarova *et al*. detailing the lipid composition of selected biological membranes. CL denotes cardiolipin, PC denotes phosphatidyl choline, PE denotes phosphatidyl ethanolamine, PG denotes phosphatidyl glycerol, PI denotes phosphatidyl inositol, and PS denotes phosphatidyl serine.(*43*)

Organism	**Lipid Composition**
Prokaryotes	
Escherichia coli Gram-negative Inner membrane	PE 70–80%; PG 15–20%; CL 5%
Bacillus megaterium Gram-positive	PE—35%, PG—48%, CL—11%, glucosoaminyl PG—6%
Eukaryotes	
Yeast cell lines	
Saccharomyces cerevisiae	PC 17%; PE 20%; PI 18%; PS 34%; PA 4%; CL 0.2% (Sphingolipids~30%) Ergosterol/PL (mol/mol)~0.9
Pichia pastoris Whole cell extract	PL 48%; ceramides 2%; sterol (free) 31%; sterol derivatives 16%
Insect cell lines	
Spodoptera frugiperda Sf9 Whole cell extract	PC 35 (43)%; PI 23 (17)%, PE 36 (36)%; CL 4.6 (4.7)%
Drosophila melanogaster Embryonic membranes	PC, PE, PS, PI—not quantified Glycosphingolipids: two ceramides—not quantified Sterols: ergosterol 69%, cholesterol+dehydrocholesterol 14%, campesterol+ sitosterol 9%, others 8%
Animal cell lines	
Xenopus oocytes Whole cell extract	PE 19%; PC 65%; PI 10%; PS 2%; sphingomyelin 5% Cholesterol/PL (mol/mol) 0.6–0.7
BHK21 cell line Plasma membrane	PE 29%; PC 26%; sphingomyelin 24%; PS 18%; PI 3% Cholesterol/PL (mol/mol)~0.9
Plant cell lines	
Oat coleoptile and root Plasma membrane	Phospholipids 42-50% - PA 11–15%; PE 9–15%; PC 9–14%; PS 3–4%; PI 2%; PG 1–2% Sterols -19-25% : sitosterol 5–9%; campesterol 2%; stigmasterol 2–12%, others 4–8%

Table 1.2. The size and lipid content of optimized phospholipid nanodiscs, adapted from Hagn *et al*.(*39*) R_h denotes the hydrodynamic radius, and the calculated diameter was obtained using the formula, D = (*n*(aa) x 0.15 nm/ π) where n(aa) is the number of amino acids in contact with lipids.

MSP protein	MSP:DMPC ratio	SEC $2R_h$ (nm)	EM diameter (nm)	DLS $2R_h$ (nm)	Calculated diameter (nm)
MSP1D1	1:80	10.2	9.5 ± 1.1	9.4 ± 0.9	9.5
ΔH4/2	1:60	9.4	~	~	8.9
ΔH4	1:45	9.1	7.8 ± 0.8	8.2 ± 0.9	8.2
ΔH5	1:50	9.2	8.2 ± 0.6	8.4 ± 0.7	8.4
ΔH4/2 H5	1:30	8.4	~	~	7.9
ΔH4H5	1:20	7.8	6.9 ± 0.8	7.3 ± 0.3	7.1
ΔH4-H6	1:~10	6.8	6.3 ± 0.6	6.4 ± 0.5	6.3

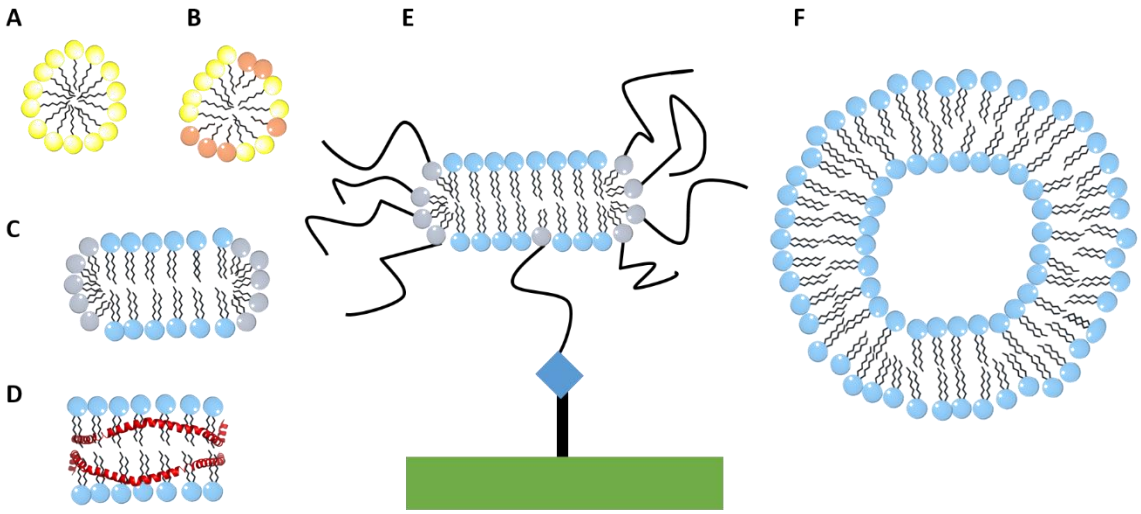


Figure 1.6. Models for the various lipid bilayer mimics used in membrane protein studies. Detergent based systems such as (A) pure detergent micelles (detergent monomers denoted in yellow), (B) mixed micelles (detergent monomers in yellow and orange) and (C) bicelles (long chain lipids in blue and short chain lipids/"detergents" in purple) are popular. More recent systems include (D) lipid nanodiscs, lipid patches (lipids denoted in blue) with two membrane scaffold proteins (denoted in red) wrapped around the patch, as well as (E) PEG-stabilized immobilized lipodisks, a system of mixed lipids (PEG-lipids in purple, form aggregates of high positive curvature and generally partition to the rim). (F) Liposomes, lamellar vesicles containing lipids (denoted in blue) are also commonly used.

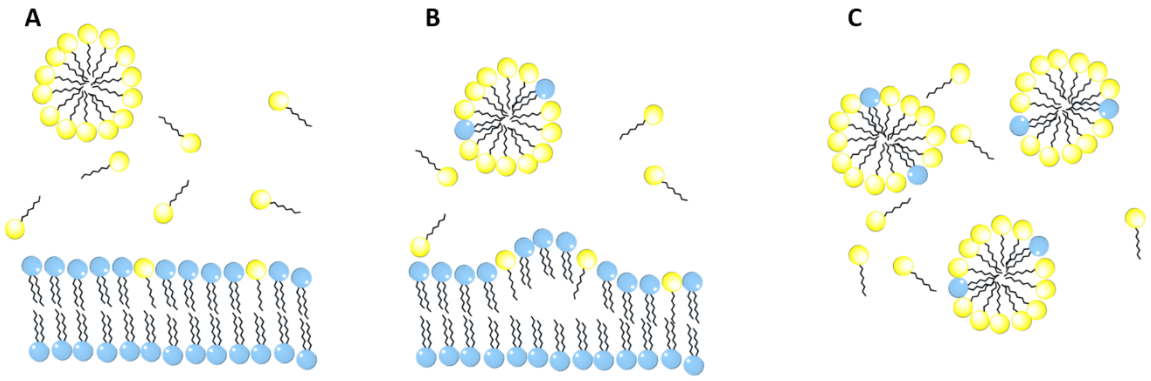


Figure 1.7. The proposed three stage model for detergent solubilization of lipid bilayers, from the bilayer perspective. (A) Stage 1 relates to detergent containing bilayers – interactions between detergents (in yellow) and lipid (in blue) that involves intercalation of detergent molecules within the bilayer. (B) Stage 2 is where a mixture of detergent saturated bilayers converts into lipid-detergent mixed micelles, and (C) Stage 3 is related to a reduction in the size of the mixed micelles as a result of their interaction with more detergent.

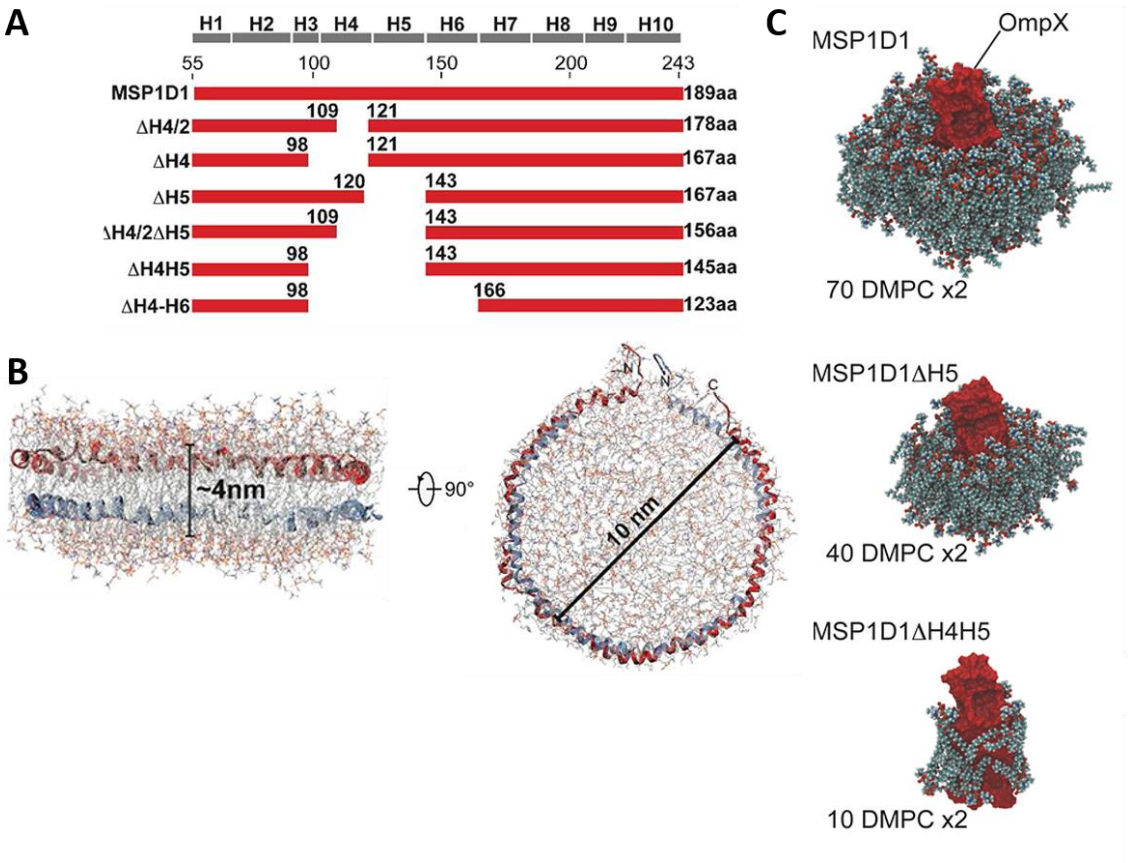


Figure 1.8. Optimization of nanodisc size for OmpX.(*39*) (A) The deletion constructs of MSP1D1 used by Hagn *et al*. The predicted secondary structure is shown on the top, with the length of each construct as indicated. (B) The proposed architecture of a phospholipid nanodisc, with two copies of MSP (denoted in red and blue) wrapped around a patch of phospholipid bilayer. (C) A cartoon of OmpX (denoted in red) and the number of DMPC lipids (denoted in blue) per bilayer leaflet, with MSP omitted for clarity.

1.4 Opacity-associated (Opa) outer membrane proteins of *Neisseria*

1.4.1 *Neisseria meningitidis* and *Neisseria gonorrhoeae*

Neisseria meningitidis and *Neisseria gonorrhoeae* are two species of obligate pathogenic Gram-negative bacteria responsible for human infection and mortality.(*72-74*) These obligate bacteria bind to the host receptors that trigger cell signaling pathways and result in engulfment. *N. gonorrhoeae* causes gonorrhea, and *N. meningitidis* is responsible for significant morbidity in children and young adults through epidemic or sporadic meningitis and/or septicemia.(*74*) Both pathogens initiate infection through colonization of mucosal surfaces – on the genitor-urinary tract in the case of *N. gonorrhoeae*, and nasopharynx in the case of *N. meningitidis*.(*75*) The strain of *N. gonorrhoeae* H041 obtained superbug status in 2011, initiating the first cases of untreatable, antibiotic resistant gonorrhoea.(*72*) These *Neisseria* species maintain a highly adapted pathogen-host relationship. The bacteria employ a strategy of continuously changing the functional characteristics and combination of prominent virulence factors by phase and antigenic variations which enables avoidance of the immune system. This strategy generates diverse subpopulations of bacteria specialized to the environmental niches within the human body, which is currently a difficult strategy for humans to counteract.(*76*)

1.4.2 Opa protein structure, function, and diversity

Neisseria begin invasion by adhering to host cells through pili and adhesion mediated interactions. Following this event, Opa proteins, which are integral membrane proteins, are promoters of bacterial engulfment, initiating the process by binding to the host cell's receptors. The opacity-associated nomenclatures is derived from the opaque phenotype visible on colonies that express the Opa protein.(*77*) Bacterial engulfment was shown in Opa-expressing *E. coli,* which were found to mediate adherence to and invasion of human cervical epithelial cells, similar to Opa-expressing *N. gonorrhoea* .(*78*) There are fifteen different known alleles that code for Opa proteins; eleven in *N. gonorrhoeae* and four in *N. meningitides.(79)* All Opa genes contain tandem repeats of $[CTCTT]_n$, which causes phase variable expression, resulting in bacterial expression of zero, one, or multiple different Opa variants.(*80*) Opa proteins are further classified into two sub-groups, Opa_{CEA} and Opa_{HS} based on host receptor specificity. The majority are Opa_{CEA} proteins, which bind to carcinoembryonic antigen-related cellular adhesion molecules (CEACAMs), while a small number are Opa_{HS}, which bind to heparin sulfate proteoglycans (HSPGs). Opa_{60}, an Opa_{CEA}, and Opa_{50}, an Opa_{HS}, are two Opa proteins investigated in these studies.

Opa_{50} and Opa_{60} are both eight-stranded β-barrel integral outer membrane proteins that bind to HSPGs and CEACAMs, respectively. The structure of Opa_{60} was determined by Daniel Fox of the Columbus laboratory in 2013 (pdbid: 2MLH). Opa proteins have a similar fold to other eight-stranded β-barrel such as OmpA and have a high sequence similarity in the β-barrel region. Opa_{50} and Opa_{60} are almost completely identical (97% identity) in the β-strand region; the differences occur in the first extracellular loop, termed

the semi-variable region (SV) and two extracellular loops with hypervariable regions (HV1 and HV2), and determine receptor specificity (Figure 1.10).

Opa_{50}'s HV1 and HV2 loops primarily interact with the heparin sulfate glycosaminoglycan chains and chondroitin sulfate side groups of the syndecan transmembrane proteins that together compose an HSPG. HSPG interactions are typically thought to be electrostatic in nature.(*81*) Deletion variants for the loops of the Opa_{50} protein indicate that the HV1 region is the most critical component to HSPG binding; binding is abolished without HV1, while the SV1 and HV2 regions were shown to only enhance binding to HSPG.(*82*)

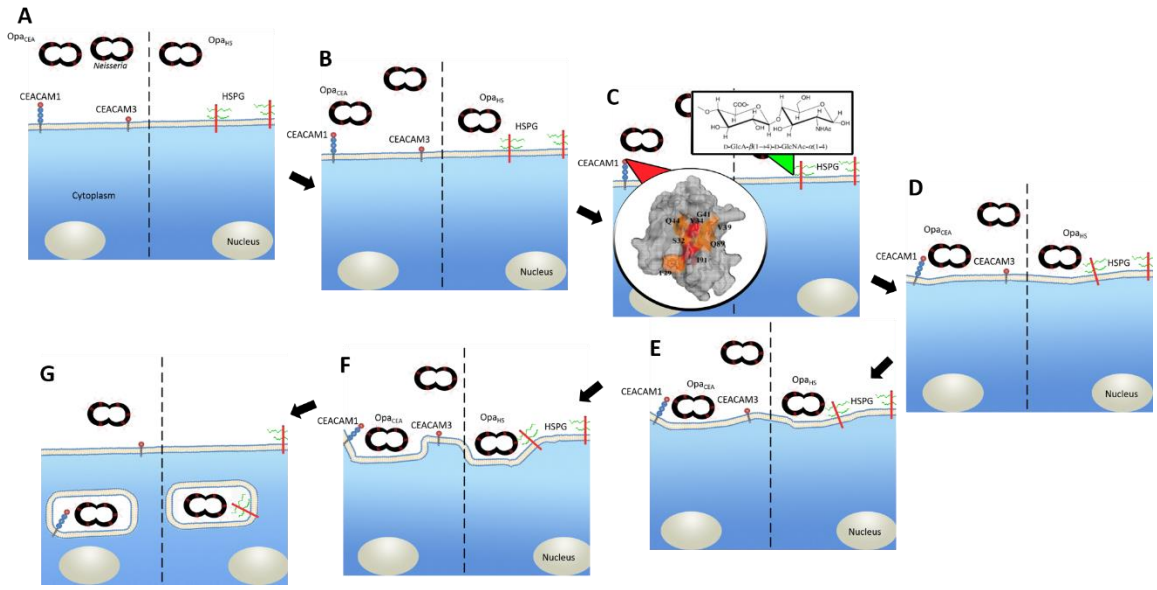


Figure 1.9. **A simplified schematic for the phagocytosis of *Neisseria* bacteria.** (A) *Neisseria* bacteria gain access to the host cell (and bind to the cell surface using pilli and adhesins, not shown), (B) approach the host cell's (C) receptors (CEACAM1 and HSPG), inducing (D,E,F) engulfment that results in complete entry (G) of the bacteria.

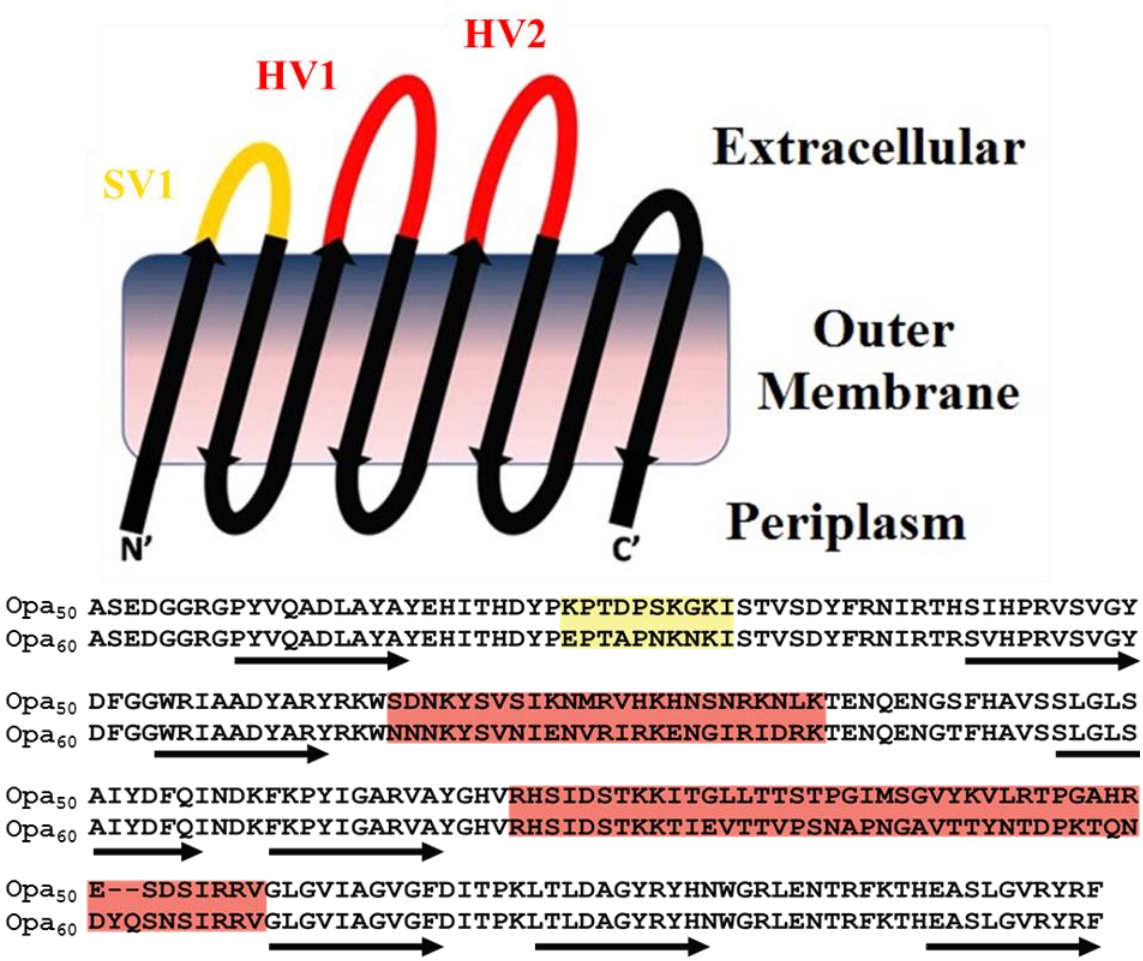


Figure 1.10. The proposed topology structure for generic Opa proteins, with the sequences of Opa_{50} and Opa_{60}. Arrows under the sequence represent β-strands, while semivariable regions in extracellular loop 1 (SV1) and hypervariable regions in extracellular loops 2 and 3 (HV1/HV2) are represented in yellow and red, respectively.

1.5 Dissertation overview

The structure of solution state structure of Opa_{60} in *n*-dodecylphosphocholine (FC12) micelles has been determined D. Fox of the Columbus laboratory. Chapter 2 presents a background on the NMR spectroscopy used for structural determination, as well as a brief history of the 7 β-barrel membrane protein structures determined by NMR. Determining the solution structure of Opa_{50} has been more difficult, and the investigation of the effects of membrane mimics as well as solution conditions on the protein-detergent complex is presented in Chapter 3. Chapter 4 describes preliminary investigations of Opa protein binding to soluble fragments and analogs of host receptors. The difficulties in Opa protein binding are discussed, as well as a potential membrane mimic system for future work. Chapter 5 describes an investigation of the dynamics of a model helical membrane protein, TM0026, using NMR as well as EPR methods.

1.4 References

1. Seddon, A., Curnow, P., and Booth, P. (2004) Membrane proteins, lipids and detergents: not just a soap opera, Biochim. Biohys. Acta 1666, 105-117.
2. Doerr, A. (2009) Membrane protein structures, Nat Meth 6, 35-35.
3. Herper, M. (2011) The best-selling drugs in America.
4. Voet, D. a. V., Judith (2004) Biochemistry, 3 ed., John Wiley and Sons, Inc, Danvers.
5. Bromley, E. H. C., Channon, K., Moutevelis, E., and Woolfson, D. N. (2008) Peptide and protein building blocks for synthetic biology: from programming biomolecules to self-organized biomolecular systems, ACS Chemical Biology 3, 38-50.
6. Stretton, A. O. W. (2002) The first sequence. Fred Sanger and insulin., Genetics 162, 527-532.
7. V., R. G. R. C. S. (1963) Stereochemistry of polypeptide chain configurations., J Mol Biol., 95-99.
8. White, S. H. (2004) The progress of membrane protein structure determination, Protein Science 13, 1948-1949.
9. Lovell, S. C., Davis, I. W., Arendall, W. B., de Bakker, P. I. W., Word, J. M., Prisant, M. G., Richardson, J. S., and Richardson, D. C. (2003) Structure validation by Cα geometry: ϕ,ψ and Cβ deviation, Proteins: Structure, Function, and Bioinformatics 50, 437-450.
10. Luckey, M. (2008) Membrane Structural Biology With Biochemical and Biophysical Foundations, 1st ed., Cambridge University Press, New York.

11. Tanford, C. (1973) The Hydrophobic Effect: Formation of Micelles and Biological Membranes, 1st ed., John Wiley & Sons, Inc

12. Makhatadze, G. I., and Privalov, P. L. (1993) Contribution of hydration to protein folding thermodynamics: I. The enthalpy of hydration, Journal of Molecular Biology 232, 639-659.

13. Singer, S. J., and Nicolson, G. L. (1972) The fluid mosaic model of the structure of cell membranes, Science 175, 720-731.

14. Pike, L. J. (2003) Lipid rafts: bringing order to chaos, Journal of Lipid Research 44, 655-667.

15. Lingwood, D., and Simons, K. (2010) Lipid rafts as a membrane-organizing principle, Science 327, 46-50.

16. Anderson, R. G. W. (1998) The cavelolae membrane system, Annual Review of Biochemistry 67, 199-225.

17. Yao, Y., Hong, S., Zhou, H., Yuan, T., Zeng, R., and Liao, K. (2009) The differential protein and lipid compositions of noncaveolar lipid microdomains and caveolae, Cell Res 19, 497-506.

18. Lee, J., and Glover, K. J. (2012) The transmembrane domain of caveolin-1 exhibits a helix–break–helix structure, Biochimica et Biophysica Acta - Biomembranes 1818, 1158-1164.

19. Rieth, M. D., Lee, J., and Glover, K. J. (2012) Probing the caveolin-1 P132L mutant: critical insights into its oligomeric behavior and structure, Biochemistry 51, 3911-3918.

20. Dyall, S. D., Brown, M. T., and Johnson, P. J. (2004) Ancient invasions: from endosymbionts to organelles, Science 304, 253-257.

21. Popot, J. L., and Engelman, D. M. (1990) Membrane protein folding and oligomerization: the two-stage model, Biochemistry 29, 4031-4037.

22. Bowie, J. U. (2005) Solving the membrane protein folding problem, Nature 438, 581-589.

23. Sanders, C. R., Czerski, L., Vinogradova, O., Badola, P., Song, D., and Smith, S. O. (1996) Escherichia coli diacylglycerol kinase is an α-helical polytopic membrane protein and can spontaneously insert into preformed lipid vesicles, Biochemistry 35, 8610-8618.

24. Heijne, G. v. (1986) The distribution of positively charged residues in bacterial inner membrane proteins correlates with the trans-membrane topology., The EMBO Journal 5, 3021-3027.

25. Wimley, W. C., and White, S. H. (1996) Experimentally determined hydrophobicity scale for proteins at membrane interfaces, Nat Struct Mol Biol 3, 842-848.

26. Moon, C. P., and Fleming, K. G. (2011) Side-chain hydrophobicity scale derived from transmembrane protein folding into lipid bilayers, Proceedings of the National Academy of Sciences 108, 10174-10177.

27. Tamm, L. K., Arora, A., and Kleinschmidt, J. H. (2001) Structure and Assembly of β-Barrel Membrane Proteins, Journal of Biological Chemistry 276, 32399-32402.

28. Tamm, L. K., Hong, H., and Liang, B. (2004) Folding and assembly of β-barrel membrane proteins, Biochimica et Biophysica Acta - Biomembranes 1666, 250-263.

29. Heijne, G. (1990) The signal peptide, J. Membrane Biol. 115, 195-201.

30. Crooke, E., and Wickner, W. (1987) Trigger factor: a soluble protein that folds pro-OmpA into a membrane-assembly-competent form, Proceedings of the National Academy of Sciences 84, 5216-5220.

31. Bechtluft, P., Nouwen, N., Tans, S. J., and Driessen, A. J. M. (2010) SecB-A chaperone dedicated to protein translocation, Molecular BioSystems 6, 620-627.

32. Lazar SW, K. R. (1996) SurA assists the folding of *Escherichia coli* outer membrane proteins., J Bacteriol 178, 1770-1773.

33. Vertommen, D., Ruiz, N., Leverrier, P., Silhavy, T. J., and Collet, J.-F. (2009) Characterization of the role of the *Escherichia coli* periplasmic chaperone SurA using differential proteomics, Proteomics 9, 2432-2443.

34. Hagan, C. L., Silhavy, T. J., and Kahne, D. (2011) β-barrel membrane protein assembly by the Bam complex, Annual Review of Biochemistry 80, 189-210.

35. Voulhoux, R., Bos, M. P., Geurtsen, J., Mols, M., and Tommassen, J. (2003) Role of a highly conserved bacterial protein in outer membrane protein assembly, Science 299, 262-265.

36. Kleinschmidt, J. H., den Blaauwen, T., Driessen, A. J. M., and Tamm, L. K. (1999) Outer membrane protein A of *Escherichia coli* inserts and folds into lipid bilayers by a concerted mechanism†, Biochemistry 38, 5006-5016.

37. Wagner, S., Bader, M. L., Drew, D., and de Gier, J.-W. (2006) Rationalizing membrane protein overexpression, Trends in Biotechnology 24, 364-371.

38. Young, J. C., Agashe, V. R., Siegers, K., and Hartl, F. U. (2004) Pathways of chaperone-mediated protein folding in the cytosol, Nat Rev Mol Cell Biol 5, 781-791.

39. Hagn, F., Etzkorn, M., Raschle, T., and Wagner, G. (2013) Optimized phospholipid bilayer nanodiscs facilitate high-resolution structure determination of membrane proteins, Journal of the American Chemical Society 135, 1919-1925.

40. Helenius, A., and Aebi, M. (2004) Roles of N-linked glycans in the endoplasmic reticulum, Annual Review of Biochemistry 73, 1019-1049.

41. Saxena, K., Dutta, A., Klein-Seetharaman, J., and Schwalbe, H. (2012) Isotope labeling in insect cells, Protein NMR Techniques (Shekhtman, A., and Burz, D. S., Eds.), pp 37-54, Humana Press.

42. Creemers, A. F. L., Klaassen, C. H. W., Bovee-Geurts, P. H. M., Kelle, R., Kragl, U., Raap, J., de Grip, W. J., Lugtenburg, J., and de Groot, H. J. M. (1999) Solid state ^{15}N NMR evidence for a complex Schiff base counterion in the visual G-protein-coupled receptor rhodopsin, Biochemistry 38, 7195-7199.

43. Opekarová, M., and Tanner, W. (2003) Specific lipid requirements of membrane proteins—a putative bottleneck in heterologous expression, Biochimica et Biophysica Acta - Biomembranes 1610, 11-22.

44. Hua, X., Sakai, J., Brown, M. S., and Goldstein, J. L. (1996) Regulated cleavage of sterol regulatory element binding proteins requires sequences on both sides of the endoplasmic reticulum membrane, Journal of Biological Chemistry 271, 10379-10384.

45. Gimpl, G., Klein, U., Reilaender, H., and Fahrenholz, F. (1995) Expression of the human oxytocin receptor in baculovirus-infected insect cells: high-affinity binding is induced by a cholesterol-cyclodextrin complex, Biochemistry 34, 13794-13801.

46. Pickford, A., and O'Leary, J. (2004) Isotopic labeling of recombinant proteins from the methylotrophic yeast *Pichia pastoris*, Protein NMR Techniques (Downing, A. K., Ed.), pp 17-33, Humana Press.

47. Hamilton, S. R., and Gerngross, T. U. (2007) Glycosylation engineering in yeast: the advent of fully humanized yeast, Current Opinion in Biotechnology 18, 387-392.

48. Saeki, T., Shimabuku, A. M., Azuma, Y., Shibano, Y., Komano, T., and Ueda, K. (1991) Expression of human P-glycoprotein in yeast cells-effects of membrane component sterols on the activity of P-glycoprotein, Agricultural and Biological Chemistry 55, 1859-1865.

49. Köster, W., Gudmundsdottir, A., Lundrigan, M. D., Seiffert, A., and Kadner, R. J. (1991) Deletions or duplications in the BtuB protein affect its level in the outer membrane of *Escherichia coli*, Journal of Bacteriology 173, 5639-5647.

50. Drew, D., Fröderberg, L., Baars, L., and de Gier, J.-W. L. (2003) Assembly and overexpression of membrane proteins in *Escherichia coli*, Biochimica et Biophysica Acta - Biomembranes 1610, 3-10.

51. Qing, G., Ma, L.-C., Khorchid, A., Swapna, G. V. T., Mal, T. K., Takayama, M. M., Xia, B., Phadtare, S., Ke, H., Acton, T., Montelione, G. T., Ikura, M., and Inouye, M. (2004) Cold-shock induced high-yield protein production in *Escherichia coli*, Nat Biotech 22, 877-882.

52. Esposito, D., and Chatterjee, D. K. (2006) Enhancement of soluble protein expression through the use of fusion tags, Current Opinion in Biotechnology 17, 353-358.

53. Chen, T., Belland, R. J., Wilson, J., and Swanson, J. (1995) Adherence of pilus- Opa+ gonococci to epithelial cells in vitro involves heparan sulfate, The Journal of Experimental Medicine 182, 511-517.

54. Lohse, M. (1992) Stable overexpression of human β2-adrenergic receptors in mammalian cells, Naunyn-Schmiedeberg's Arch Pharmacol 345, 444-451.

55. J.H. Weiner, B. D. L., M.L. Elmes, R.D. Bradley, and D.G. Scraba. (1984) Overproduction of fumarate reductase in *Escherichia coli* induces a novel intracellular lipid-protein organelle., J Bacteriol 158, 590-596.

56. Miroux, B., and Walker, J. E. (1996) Over-production of proteins in *Escherichia coli*: mutant hosts that allow synthesis of some membrane proteins and globular proteins at high levels, Journal of Molecular Biology 260, 289-298.

57. Sui, H., Han, B.-G., Lee, J. K., Walian, P., and Jap, B. K. (2001) Structural basis of water-specific transport through the AQP1 water channel, Nature 414, 872-878.

58. Wolfrum, U., and Schmitt, A. (2000) Rhodopsin transport in the membrane of the connecting cilium of mammalian photoreceptor cells, Cell Motility and the Cytoskeleton 46, 95-107.

59. Steeghs, L., de Cock, H., Evers, E., Zomer, B., Tommassen, J., and van der Ley, P. (2001) Outer membrane composition of a lipopolysaccharide-deficient N*eisseria meningitidis* mutant, The EMBO Journal 20, 6937-6945.

60. Israelachvili, J., Mitchell D.J., Ninham, B.W. (1976) Theory of self-assembly of hydrocarbon amphiphiles into micelles and bilayers, J. Chem. Soc., Faraday Trans. II 72, 1525.

61. Lichtenberg, D., Ahyayauch, H., and Goñi, Félix M. (2013) The mechanism of detergent solubilization of lipid bilayers, Biophysical Journal 105, 289-299.

62. le Maire, M., Champeil, P., and Møller, J. V. (2000) Interaction of membrane proteins and lipids with solubilizing detergents, Biochimica et Biophysica Acta - Biomembranes 1508, 86-111.

63. Cladera, J., Ricaud, J.-L., Verde, J. V., and DuÑach, M. (1997) Liposome solubilization and membrane protein reconstitution using CHAPS and CHAPSO, European Journal of Biochemistry 243, 798-804.

64. Lee, S. C., Bennett, B. C., Hong, W.-X., Fu, Y., Baker, K. A., Marcoux, J., Robinson, C. V., Ward, A. B., Halpert, J. R., Stevens, R. C., Stout, C. D., Yeager, M. J., and Zhang, Q. (2013) Steroid-based facial amphiphiles for stabilization and crystallization of membrane proteins, Proceedings of the National Academy of Sciences 110, E1203-E1211.

65. Dewald, Alison H., Hodges, Jacqueline C., and Columbus, L. (2011) Physical determinants of β-barrel membrane protein folding in lipid vesicles, Biophysical journal 100, 2131-2140.

66. Columbus, L., Lipfert, J., Jambunathan, K., Fox, D. A., Sim, A. Y. L., Doniach, S., and Lesley, S. A. (2009) Mixing and matching detergents for membrane protein NMR structure determination, Journal of the American Chemical Society 131, 7320-7326.

67. Agmo Hernández, V., Reijmar, K., and Edwards, K. (2013) Label-free characterization of peptide–lipid interactions using immobilized lipodisks, Analytical Chemistry 85, 7377-7384.

68. Sahu, I. D., McCarrick, R. M., Troxel, K. R., Zhang, R., Smith, H. J., Dunagan, M. M., Swartz, M. S., Rajan, P. V., Kroncke, B. M., Sanders, C. R., and Lorigan, G. A. (2013) DEER EPR measurements for membrane protein structures via bifunctional spin labels and Lipodisq nanoparticles, Biochemistry 52, 6627-6632.

69. Meiby, E., Morin Zetterberg, M., Ohlson, S., Agmo Hernández, V., and Edwards, K. (2013) Immobilized lipodisks as model membranes in high-throughput HPLC-MS analysis, Anal Bioanal Chem 405, 4859-4869.

70. McLachlan, A. D. (1977) Repeated helical pattern in apolipoprotein-A-I, Nature 267, 465-466.

71. Bayburt, T. H., Grinkova, Y. V., and Sligar, S. G. (2002) Self-assembly of discoidal phospholipid bilayer nanoparticles with membrane scaffold proteins, Nano Letters 2, 853-856.

72. Deguchi, T., Nakane, K., Yasuda, M., and Maeda, S.-i. (2010) Emergence and spread of drug resistant *Neisseria gonorrhoeae*, The Journal of Urology 184, 851-858.

73. Lo, H., Tang, C. M., and Exley, R. M. (2009) Mechanisms of avoidance of host immunity by *Neisseria meningitidis* and its effect on vaccine development, The Lancet Infectious Diseases 9, 418-427.

74. Rouphael, N., and Stephens, D. (2012) *Neisseria meningitidis*: Biology, Microbiology, and Epidemiology, In *Neisseria meningitidis* (Christodoulides, M., Ed.), pp 1-20, Humana Press.

75. Meyer, T. F., Pohlner, J., and Putten, J. P. M. (1994) Biology of the pathogenic *Neisseriae*, in bacterial pathogenesis of plants and animals (Dangl, J., Ed.), pp 283-317, Springer Berlin Heidelberg.

76. Gorvel, J.-P. (2004) Intracellular pathogens in membrane interactions and vacuole biogenesis, Kluwer Academic/Plenum Publishers.

77. Bos, M. P., Grunert, F., and Belland, R. J. (1997) Differential recognition of members of the carcinoembryonic antigen family by Opa variants of Neisseria gonorrhoeae, Infection and Immunity 65, 2353-2361.

78. Simon, D., and Rest, Richard. (1992) Escherichia coli expressing a Neisseria gonorrhoeae opacity-associated outer membrane protein invade human cervical and endometrial epithelial cell lines., Proc. Natl. Acad. Sci. 89, 5512-5516.

79. Robertson, B. D., and Meyer, T. F. (1992) Genetic variation in pathogenic bacteria, Trends in genetics : TIG 8, 422-427.

80. Merz, A. J., and So, M. (2000) Interactions of pathogeneic *Neisseriae* with epithelial cell membranes, Annual Review of Cell and Developmental Biology 16, 423-457.

81. Bishop, J. R., Schuksz, M., and Esko, J. D. (2007) Heparan sulphate proteoglycans fine-tune mammalian physiology, Nature 446, 1030-1037.

82. Grant, C. C. R., Bos, M. P., and Belland, R. J. (1999) Proteoglycan receptor binding by *Neisseria gonorrhoeae* MS11 is determined by the HV-1 region of OpaA, Molecular Microbiology 32, 233-242.

Chapter 2: NMR spectroscopy

Amongst membrane proteins of known structure, large membrane proteins or membrane protein complexes constitute the majority of high-resolution structural information of biomolecules deposited in the Protein Data Bank (PDB). The majority of these structures are also solved using X-ray crystallography – larger proteins have a larger ratio of protein volume to protein surface area in contact with lipid, thus providing more favorable conditions such as electrostatic contacts between adjacent units in a crystal lattice.[1] Conversely, small, integral membrane proteins are underrepresented, potentially due to the lack of exposed extra-membrane domains that could make neighboring crystal contacts necessary for 3D crystal formation.[2] For membrane proteins with determined crystal structures, nuclear magnetic resonance (NMR) spectroscopy can provide a complementary information to X-ray crystallography, such as elucidating the dynamics of membrane proteins embedded within biological membrane and examining conformational changes in membrane proteins across a wide range of time-scales (backbone oscillations in the ps-ns regime to large-scale domain movements in the μs-s time domain).[3] These dynamics are often related to the function of membrane proteins, and NMR spectroscopy is an excellent tool towards probing these fluctuations. Significant work has been performed in the biophysical membrane field towards recreating the native lipid bilayer environment that maintains native folding and function of the protein while preserving experimental sensitivity; recent advances have included the use of fast tumbling bicelles[4] as well as lipid nanodiscs[5, 6] to both bypass curvature strain present in micelles on solubilized transmembrane domains and provide a more bilayer-like environment,[2] as well as resolve

line broadened resonances. This chapter outlines the theory of NMR spectroscopy and highlight its applications to proteins, with an emphasis on overcoming some of the Pitfalls (Chapter 1, Section 1.1.4) of biophysical investigations of membrane proteins.

2.1 Introduction to the theory and principles of NMR

2.1.1 Nuclear spin

Elementary particles such as protons, electrons, and neutrons all exhibit behavior that can be explained in terms of a spinning motion, for which an angular momentum component is able to be quantized; this property is called spin.[7] These spins cause the particle to prefer a a distinct orientation when placed in a magnetic field; the magnetic field exerts a torque on the magnetic dipole, resulting in a rotation of the magnetic moment (called precession) at a frequency proportional to the applied magnetic field, termed the Larmor frequency.[7] The spin states of these particles are subject to Zeeman splitting – the splitting of degenerate nuclear or electronic spin states into energetically nonequivalent states in the presence of a magnetic field, resulting in distinct populations of orientations. The Zeeman splitting, or difference between the energies of the spin states is correlated with the magnitude of the magnetic field *B* and is given by the equation:

$$\Delta E_{\alpha\beta} = \hbar\gamma B \qquad (1)$$

where α and β are the lower and higher energy states, respectively, ħ is Planck's constant and γ is the gyromagnetic ratio of the nucleus (usually given in units of (rad s^{-1} T^{-1})). A variety of commonly used gyromagnetic ratios are present in Table 2.1.

Quantum mechanics dictate that a nucleus of spin *I* will have 2*I* + 1 allowed spin states. Nuclei with spin *I* = ½ are of significant importance to NMR, and will have 2 possible orientations: *I* = ½ and *I* = -½. If an external magnetic field is applied, the energy difference

between the two possible spin states at a given magnetic field strength will be proportional to their magnetic moments.[8] The splitting between nuclear spin states, however, is energetically very small, even at high magnetic fields; for example, the splitting between the lower energy (m = ½) and higher energy (m = -½) states for ^{1}H at 11.74 T (corresponding to 500 MHz) is only 0.239 J/ mol, which is much smaller than *RT* at ambient temperatures (*RT* = 2.48 kJ/mol at 298 K).[7] This energy difference corresponds to a very small population difference between the two states (1 in 10,000 for ^{1}H in a 500 MHz field), which results in a relatively low signal as compared to other forms of spectroscopy such as UV/visible and infrared, where almost all molecules are in the ground electronic state because the energy difference between the ground and excited states is large.[7] The entire ensemble of nuclear spins can undergo net absorption or emission as long as a population difference exists between the states. However, because the transition from α to β state and the reverse β to α state are almost equally likely, the re-establishment of the system to thermal equilibrium (called T_1 relaxation) tends to be quite slow. While the long T_1 relaxation times reduce the sensitivity of the NMR experiment, relaxation also enables multiple perturbations on a system of spins which gives rise to the basis for success of the NMR experiment.[7]

Table 2.1 Gyromagnetic ratios, spin quantum numbers, and transition frequencies of some commonly observed NMR nuclei. Values are from the Magnetic Resonance Periodic Table hosted by the Beckman Institute at the University of Illinois, Urbana.

Nucleus	Spin (I)	$\gamma/10^{-7}$ rad (s T)$^{-1}$	Frequency at 11.74 T (in MHz)
^{1}H	½	26.7510	500.0
^{2}H	1	4.1064	76.7
^{3}H	½	28.5335	533.3
^{13}C	½	6.7263	125.7
^{14}N	1	1.9331	36.1
^{15}N	½	-2.7116	50.7
^{19}F	½	25.1665	470.6
^{31}P	½	10.8289	202.6

2.1.2. Nuclear shielding and chemical shift

While the gyromagnetic ratio is constant among the same nuclei, the Larmor frequency differs for a given nuclear spin, depending on the location and environment of the spin. Electrons in atomic or molecular orbitals around every atom are constantly in motion, and respond to the presence of a magnetic field by moving in a circular path within the constraints established by the orbitals that they reside in. Moving electrons represent a movement of charge; thus, they generate their own effective magnetic field that interacts with the magnet's imposed field on the nucleus.[7] The result is a small, but detectable difference in the Zeeman splitting for a given nuclear spin, with the relationship between the observed resonance frequency of nucleus *I* and the applied field, defined by:

$$\omega_i = \gamma B_0(1-\sigma_i) \quad (2)$$

where σ_i is the shielding tensor that is dependent on the orientation of the molecule relative to B_0.

The major contribution to chemical shift usually comes from the influence of low-lying electronic excited states, and correlates well with electronegativity. Electronegative atoms tend to withdraw electron density from neighboring groups, increasing local fields at the nuclei of neighboring atoms, which leads to increased changes in Larmor frequency resulting in increased chemical shift values.[8] In proteins, chemical shifts are often highly perturbed and vary even amongst the same amino acid types. Effects such as hydrogen bonding,[9] backbone dihedral angles,[10] as well as ring current from aromatic moieties.[11] The chemical shifts of amide protons correlate with the hydrogen bond length, with

short hydrogen bonds correlating with low-field shifts, and larger bond lengths with shifts to the higher field.[9] The likely physical origins for these correlations were hypothesized to be electric field effects, local magnetic anisotropies, or potential polarization of the electron cloud near the hydrogen atom by neighboring atoms.[9] Amino acid residues displaying large variations in dihedral angles such as those located in dynamic loop regions tend to have chemical shifts different from those in non-flexible regions.[10] Comparison of resonance positions with reference chemical shift tables listing the random coil chemical shifts of all 20 amino acids is useful for carrying out spin identification.[12] Chemical shift predictions can be dependent upon two types of random coil chemical shift tables that exist: those derived stastically[13, 14], and those obtained experimentally.[15] Finally, ring current is induced by the π-electrons of aromatic moieties influences the magnetic field that adds to the applied B_0.[11]

2.1.3 Nuclear spin relaxation

NMR instruments expose the sample to radiofrequency (rf) waves at the nuclear Larmor frequencies of the specific nucleus observed in the experiment.[8] These pulses of rf waves disturb the equilibrium of the nuclear spin system and create transverse nuclear magnetization, which occurs when the net magnetic moment is perpendicular to the magnetic field. The longitudinal relaxation time, or T_1, is the time it takes for thermal equilibrium to be re-established after this perturbation. Since the processes that cause T_1 relaxation are those that restore the system to thermal equilibrium, they typically involve transfer of energy between the spin ensemble and the surroundings.[7] In solids, this implies a transfer of energy from the spins to the lattice, which is why T_1 relaxation is often also called spin-lattice relaxation. In practice, T_1 relaxation is typically slow and takes milliseconds to seconds to return to equilibrium and does not typically contribute significantly to the recorded signals of the nuclei.

Transverse relaxation time (T_2), or spin-spin relaxation, is the rate at which coherence is lost. The nuclear spins are shifted 90° following the radiofrequency pulse, and spins initially retain their net magnetization and precess along the perpendicular plane at their Larmor frequency. However, field influences cause a gradual loss of synchronization and contribute to T_2 relaxation resulting in the loss of coherence in the x,y (transverse) plane, thus T_2 is also called transverse relaxation time.[7] Spin interactions with other spins are important contributing factors to the loss of synchrony amongst other spins (that do not necessarily contribute to T_1, a change in state populations) – this loss of coherence from the spin-spin interactions is also why T_2 is referred to as spin-spin relaxation. Experimental

factors also contribute to the loss of coherence; inhomogeneity in the magnetic field (perhaps due to poor shimming) results in nuclei exhibiting slight differences in resonance frequency dependent upon the physical location of the NMR sample tube.

2.1.4 Dynamics

Broadly speaking, NMR can provide information concerning molecular dynamics on a variety of time scales, which include local motions in molecules by relaxation analysis (on the order of $10^{-15} – 10^{-10}$ s), chemical exchange and conformational averaging by saturation transfer and exchange contributions to relaxation (on the order of 10^{-3} s), and slow reactions by chemical shift and integration analysis ($10^{0} – 10^{3}$ s).[7]

The T_2 processes described in Section 2.1.3 represent a loss of phase memory. When a spin experience a change in electronic environment, the Larmor frequency of the spin is altered, and phase memory is lost. This loss of phase memory results in a decrease in observable coherence, and enhanced T_2 relaxation results in broadened resonances. Proteins are dynamic molecules constantly undergoing conformational exchange, and the local environment of one conformation can contribute differently to the shielding tensor than in another conformation. This exchange process can manifest in the NMR spectra in several ways. The NMR-active nucleus's sampling of two distinct environments, A and B, will give rise to two different resolved resonances. If the exchange is "slow", two different resonances are observed with relative integrations proportional to the fraction of time spent by the nucleus in each environment (Figure 2.1A). In intermediate exchange, the lifetime of a spin state in a given environment is shortened, and the rate of exchange is similar to

that of the difference in Larmor frequencies, which results in the nuclei switching between sites at random time intervals that are comparable to the time for one period of Larmor rotation (Figure 2.1C).[7] The fast exchange regime is reached as the exchange rate increases; the rate of exchange is much greater than the observed frequency of the states by the spectrometer, which gives rise to an average of both states (Figure 2.1E).

This phenomenon can also be used to quantitate protein-ligand interactions. The two molecules will be in equilibrium between the bound (A) and free (B) states, which is described by the dissociation constant K_d. These states are dependent on the exchange rate of the complex and the chemical shift difference between the free and bound states. In practice, for the observed slow exchange regime, two sets of signals are typically detected for each state (Figure 2.1B), while in the fast exchange regime, signals typically move in a continuous manner which gives rise to a single average line (Figure 2.1D).[7] Intermediate regimes time often give rise to detrimental peak broadening that obfuscates signal observation (Figure 2.1F).

Spin relaxation techniques are typically carried out with ^{15}N labeled proteins, and techniques have been developed to monitor reorientation of the ^{15}N-^{1}H bond vector of the amide protein backbone.[16] The pico- to nanosecond dynamics of each bond vector are derived from the Lipari-Szabo model-free approach, and the order parameter, S^2 describes the amount of local mobility for the ^{15}N-^{1}H vectors.[16, 17] Completely immobilized motions are described by $S^2 = 1$, and completely unrestricted local motions of the NH vectors are described by $S^2 = 0$. For typical relaxation studies, S^2 and the rotational correlation time are reported for various NH sites. This method allows for the detection of

extensive local dynamics, especially for mobile loops, linkers, and termini, in many proteins.

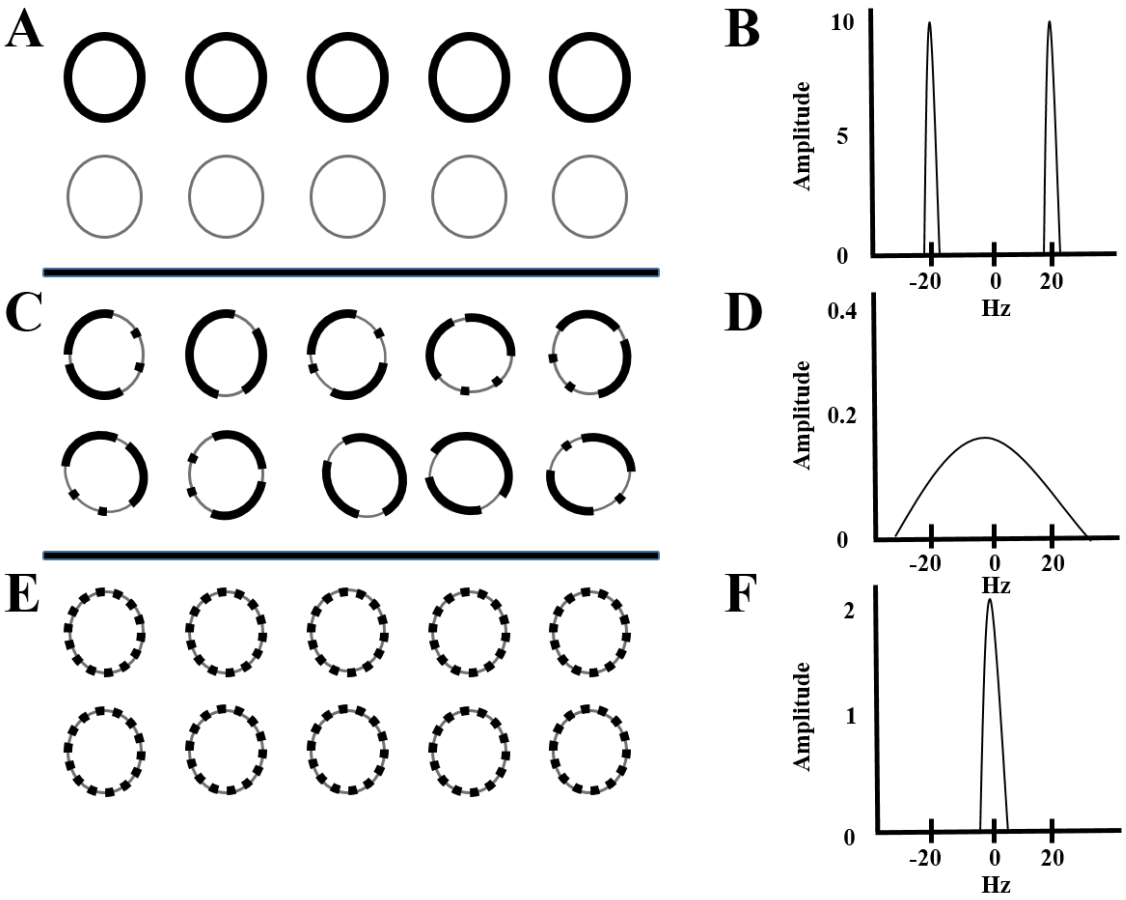


Figure 2.1. The graphical representation of slow, intermediate, and fast two site exchange on the chemical-shift time scale. Each circle represents the Larmor precession of a single nuclear spin, with precession in one state represented in black, and precession of the other state in gray. A) In slow exchange, precession is observed at each site and two FIDS are recorded, giving rise to B) two distinct populations observed in the NMR spectrum. C) In intermediate exchange, the possibility of site exchange becomes more likely, increasing the uncertainty of frequency measurement and D) broadening lineshapes. E) In fast exchange, intersite exchange occurs many times during a single cycle and is faster than can be distinguished by a spectrometer, which gives rise to F) a single average line.

2.2 Biomolecular NMR spectroscopy

2.2.1 The single pulse experiment

The simple one dimensional NMR experiment consists of a single rf pulse on a single rf channel, followed by signal detection. The experiment can be divided into four parts: initialization, excitation, detection, followed by processing and display. The initialization period occurs before the experiment begins, and the starting parameters such as pulse sequence, sampling frequency, and number of sampled points are downloaded to the pulse programmer and other spectrometer hardware, such as the synthesizer and the analog-digital converter. Post initialization, the pulse programmer executes a timed sequence of instructions to set the phase of the rf synthesizer and initiate pulse gate events. An rf pulse then travels into the probe from the amplifier and sets up resonant oscillations in the tuned circuit of the probe, irradiating the sample with an rf field close to the Larmor frequency of the chosen nuclear isotope. The rf pulse disturbs the equilibrium of the nuclear spin system and creates transverse nuclear magnetization. The pulse is then switched off, and after some microseconds, the pulse energy dissipates, initiating the detection stage. The precession of the nuclear spin magnetization in the x,y plane sets up oscillations in the tuned circuit, giving rise to an rf NMR signal known as the Free Induction Decay (FID) which is detected by the probe. This signal then travels from the probe to the detector, is amplified by a signal amplifier, and is converted down in frequency by the quadrature receiver. At the final stage, the FID is digitized, and the digital complex signal is Fourier transformed, which converts the NMR signal from the time domain to the frequency domain.

2.2.2 Heteronuclear NMR experiments

Heteronuclear polarization transfer forms the basis for an entire class of multidimensional NMR spectroscopy experiments. These experiments correlate the nuclear spins of two spin pairs from different nuclei, such as a single ^{1}H spin with a single ^{13}C or ^{15}N spin.[7] If the polarization, or equilibrium population difference across the spin transition, of a less sensitive (I) heteronucleus is transferred to a more sensitive (S) nucleus (such as the ^{1}H spin), the signal to noise ratio of the experiment is significantly improved. The Insensitive Nucleus Enhancement by Polarization Transfer (INEPT) experiment is based on these transfers of polarization from one system to another (Figure 2.3).[7] The sensitive (S) spin is excited and rotated into the plane perpendicular to the magnetic field [$P1 = \pi/2_x$ (S)]. The spins precess for time Δ, and both the S and I spins are excited with 180° radiofrequency pulses for each set of spins, simultaneously [$P2 = \pi_x$ (S), π_x (I)], which refocuses the chemical shift, but not the heteronuclear coupling. After a second mixing time (Δ), a 90° pulse is applied to both spins [$\pi/2_y$, (S), $\pi/2_y$ (I)], which inverts only one of the S components. By polarizing the population of the insensitive nucleus, it can be indirectly detected with enhanced sensitivity.

The double INEPT, or otherwise known as the Heteronuclear Single Quantum Coherence (HSQC) experiment has become the two-dimensional experiment of choice for detecting heteronuclear correlations in which the more sensitive nucleus (usually ^{1}H) is instead detected. The HSQC experiment has been called a “there and back again” experiment, due to the symmetric direction of magnetic polarization transfer during the experiment.[7] The protons (S) are used to provide polarization to the insensitive nucleus of interest to which

the S spins are coupled. This polarization evolves during t_1 on I (Figure 2.3), and is transferred back to S for detection. The detected proton signals evolve at their own Larmor frequencies, and report back on both these Larmor frequencies as well as the heteronuclear spins that they are attached to, resulting in a two dimensional spectrum correlating the two heteronuclei. This two dimensional spectrum, as expanded in the ^{13}C or ^{15}N dimension, resolves some of the overcrowded resonance issues for spectra that contain many Larmor frequencies (such as proteins), as well as overlapping frequencies.

2.2.3 NMR 3D assignment experiments

Large protein systems have a significant amount of spectral overlap in the 2-dimensional spectra, and assignment of resonances is difficult due to the convolution of degenerate peaks. The introduction of a third independent time variable into a multipulse experiment results in a third independent frequency domain that can represent a correlation between the sets of spins.[7] These experiments tend to use scalar coupling bonds to unambiguously identify heteronuclear connectivity, and are named after the direction of their magnetization transfer. The suite of experiments include the HNCO and HN(CO)CA, in which magnetization is transferred from the **H**ydrogen through a **N**itrogen, before correlating with the nearby **C**arbon of a carbonyl (the α-carbon, in the case of the latter). Other often used experiments are the HNCA and the HN(CO)CA, which correlate the α-carbons with the amide group, the HN(CA)CB and the HN(COCA)CB, which observe the β-carbons of the correlated with the amide group, and the HCACO and HCA(CO)N which correlate protons on the α-carbons with the rest of the heavy atoms of the protein backbone. Each type of amino acid has a unique chemical shift for the atoms observed in these spectra,

which helps in the identification of particular amino acids from these spectra. By observing identical carbon chemical shifts from different protons, pairs of systems can be sequentially connected and identified by amino acid type, which results in a sequential amino acid assignment.

2.2.4 Nuclear Overhauser Effect SpectroscopY (NOESY)

Direct through-space interactions between nuclear spins are called dipolar couplings (also known as dipole-dipole coupling), and are primarily observed in nuclear spin relaxation and the Nuclear Overhauser Effect SpectroscopY (NOESY). When two nuclei are close enough in space to detect each other's magnetic fields, the polarized spins stimulate relaxation in each other. This is observed experimentally as a change in intensity of a crosspeak. [7] These crosspeaks (Figure 2.2) have an r^{-6} distance dependence (where r is the internuclear distance) and are generally only effective to a few angstroms, where the upper limit of ^{1}H NOEs is approximately 5 Å.[7] This distance dependence, however, makes ^{1}H NOEs a useful tool for investigating the structures of multimolecular complexes, as well as molecular interactions; ^{1}H NOEs are primarily responsible for structural restraints because of the radius of interaction that they encompass.

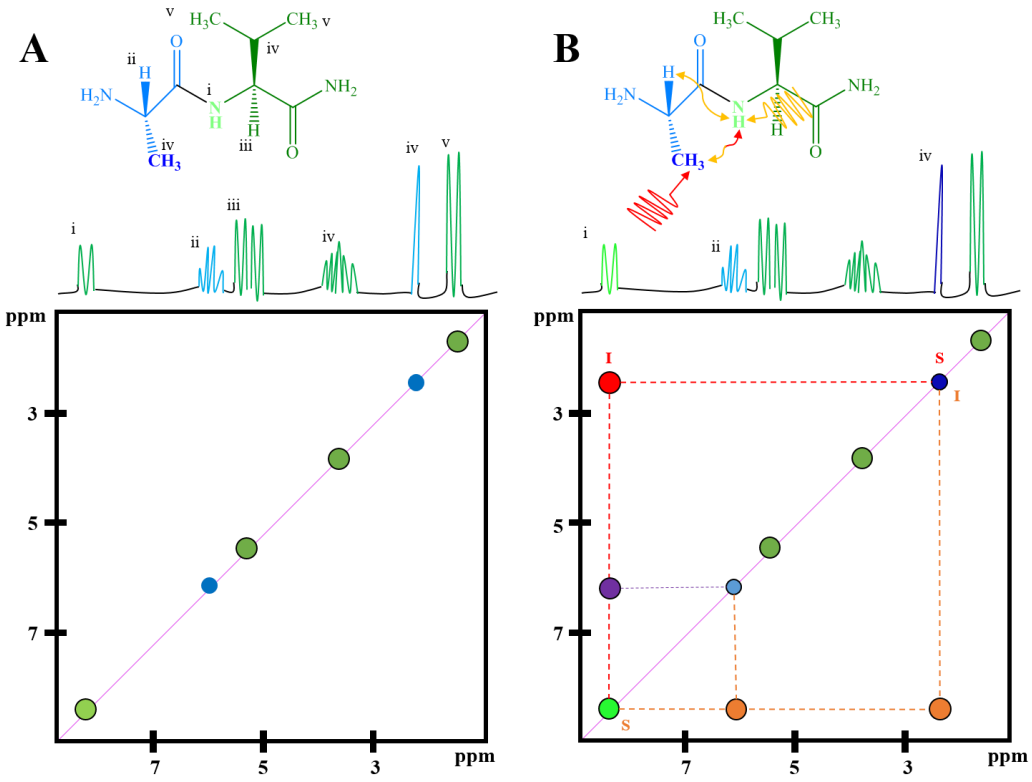


Figure 2.2. A simplified NOESY spectrum for two interacting protons of an Ala-Val peptide. A) The diagonal peaks not showing any NOE transfer. The coupled peaks of the peptide are shown on the x-axis. B) The diagonal peaks with NOE transfer of protons from systems i and iv displayed. The S spin of the Ala-methyl proton is irradiated (B, rf pulse in red) and the I spin detected through space is the amide proton (in green). The crosspeak (in red) should be aligned with the diagonal peak (i, in green). When the S spin of the amide proton is irradiated, (B, rf pulse in orange), the I spins are detected through space for neighbors within ~5, in this case the protons on ii and iv (in blue), which results in the corresponding crosspeaks (in orange). The proton of ii will also detect the same amide proton, resulting in a crosspeak (in purple). The intensity of the peaks in real spectra are modulated by the degree of saturation of the interacting spins.

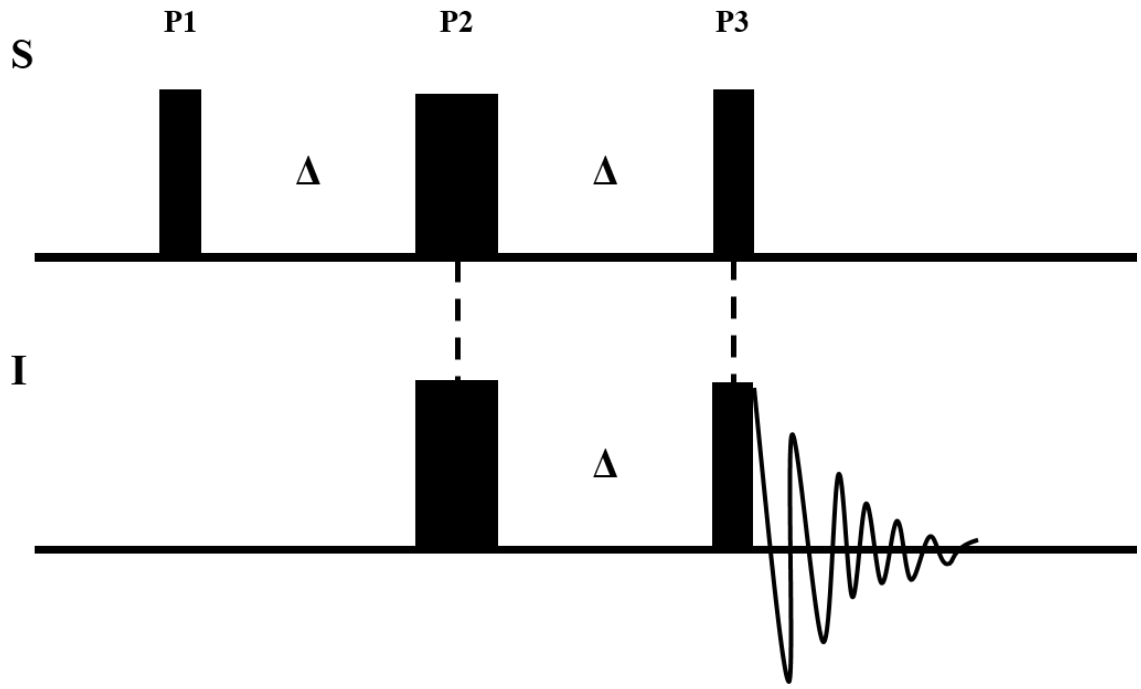


Figure 2.3. The INEPT experiment. This experiment transfers polarization from a sensitive nucleus, S (usually ^{1}H) to an insensitive nucleus, I (usually ^{13}C or ^{15}N). P1 = $\pi/2_x$ (S), P2 = simultaneous π_x (S), π_x (I), P3 = simultaneous $\pi/2_y$, (S), $\pi/2_y$ (I), $\Delta = ¼ J_{CH}$.

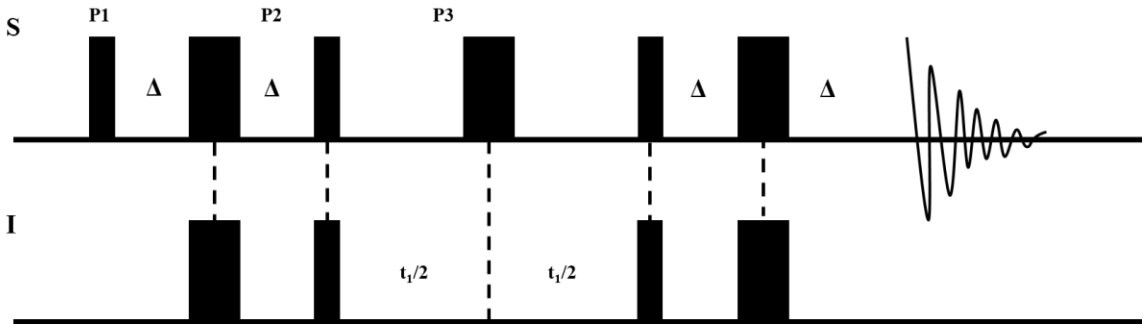


Figure 2.4. Double INEPT (HSQC) experiment for heteronuclear correlation with sensitive (S) nucleus detection. Polarization transfer from S to I: P1 = $\pi/2_x$ (S), P2 = simultaneous π_x (S), π_x (I), P3 = simultaneous $\pi/2_y$ (S), $\pi/2_y$ (I), $\Delta = ¼ J_{CH}$. P4 = π_x (S) for refocusing of SI coupling during I chemical shift evolution period t_1. Reverse INEPT step transfers polarziation from I back to S: P5 = simultaneous $\pi/2_y$ (S), $\pi/2_y$ (I), $\Delta = ¼ J_{CH.}$ P6 = simultaneous π_x (S), π_x (I).

2.3 Overcoming the limitations of NMR

2.3.1 Transverse relaxation-optimized spectroscopy (TROSY)

Both chemical shift anisotropy (CSA) and dipole-dipole interactions affect the fluctuating fields for a given spin, and can interact constructively or destructively. As previously addressed in Section 2.1.4., molecular reorientation results in relaxation.[7] Chemical shift anisotropy results from incomplete averaging of the chemical shift tensor as a function of molecular tumbling. The rate at which the molecule tumbles also modulates the efficiency of CSA relaxation. CSA has a larger effect on larger molecules as compared to smaller molecules, due to the slower tumbling of the larger molecule and the resulting incomplete averaging of the environment over time. Transverse relaxation-optimized spectroscopy (TROSY) selects transitions that experience mutual cancellation of dipole-dipole interactions as well as chemical shift anisotropy mechanisms.

Most HSQC experiments have decoupling of *I* during acquisition, which results in a singlet structure in both dimensions for the observed correlation between *I* and *S* in the transformed two-dimensional spectrum. However, if the HSQC is performed without refocusing or decoupling in either dimension, splitting in each dimension of the spectrum would result in a set of four peaks. Because the interference patterns of relaxation of CSA and dipole-dipole fields differ, the split peaks do not exhibit degenerate lineshapes. For one of the four peaks, the cross-terms in both dimensions reinforce the CSA and dipole-dipole relaxation, which originates from the rapidly relaxing components of both ^{1}H and ^{15}N, and results in a consequently broadened peak.[18] In two of the other peaks, the cross-terms

are opposite in sign during t_1 and t_2, resulting in an intermediate line width after transformation. For the fourth peak, the cross-terms would interfere destructively with the other relaxation terms, resulting in a narrower line width in both dimensions. The TROSY pulse sequence, through use of phase cycling, is designed to select only this fourth transition in which the cross-correlation terms suppress T_2 relaxation in both time domains. This selection is the basis for adapting the pulse sequence to multidimensional NMR methods using HSQC-type correlations for detection of larger size systems (such as proteins or protein-detergent complexes).

2.3.2 Perdeuteration and selective amino acid labeling

As protein systems become larger, slower tumbling of the system leads to faster relaxation time due to dipole-dipole interactions which limit the quality of spectra obtained; this is most notable in peak line-broadening. To tackle spin-spin relaxation, Venters *et al*. introduced perdeuteration to create isotopically labeled proteins on the protein human carbonic anhydrase II (HCA II) in 1995.[19] To increase the sensitivity of $^{1}H/^{13}C/^{15}N$ experiments by decreasing the rates of ^{13}C and ^{1}H-N T_2 relaxation, HCA II was grown in D_2O, essentially replacing all non-labile ^{1}H with ^{2}H. The only remaining ^{1}H were those protons solvent exposed and able to exchange with H_2O. The incorporated ^{2}H deuteron has a significantly smaller gyromagnetic ratio ($\gamma_D/\gamma_H \sim 1/6.5$) and is not as efficient at promoting dipole-dipole relaxation pathways in nearby nuclei. This inefficiency consequently dampens the dominant ^{1}H-^{13}C relaxation mechanism for ^{13}C nuclei.[20]

While perdeuteration is useful for mitigating rapid heteronuclear relaxation, the number of ^{1}H-^{1}H NOEs that can be measured and used for protein structure determination is drastically reduced. To acquire structurally informative NOEs in perdeuterated proteins, Kay *et al.* employed the use of specific α-keto acid precurors to specifically protonate methyl groups of the abundant Ile, Leu, and Val residues (often found within the hydrophobic regions of the proteins of interest).[21, 22] *E. coli* readily use these biosynthetic precursors to these amino acids ([3,3-$^{2}H_2$]^{13}C-2-ketobuterate and ^{13}C-α-ketoisovalerate preferentially over *de novo* synthesis. Simulations for an SH2 domain of phospholipase Cγ1 suggest that inclusion of the NOE correlations between methyl group within the hydrophobic core in conjunction with the backbone NOEs are sufficient to confidently characterize the global fold of the majority of proteins to within a few angstrom root mean square deviation (rmsd).[23]

Liang *et al*. reported on the use of paramagnetic relaxation enhancement (PREs) by means of nitroxide spin-labels, to provide valuable long distance information.[24] While spin labels are often reduced in highly nonpolar environments, restraints were obtained for 11 water-exposed and lipid-covered sites on OmpA, an integral 8 stranded β-barrel. These restraints covered a range of 15-24 Å, and addition of PREs resulted in significant improvement of the calculated backbone structure of OmpA.

2.3.3 Protein expression

The problems from Pitfall 1 (protein expression) translate over to NMR-based investigations, and in many cases cannot be overcome by simply growing more protein. It

is not financially viable to isotopically label all proteins of interest in *E. coli*, especially when yields are less than 1 mg L^{-1} of culture. Screening for better growth conditions while finding conditions that preserve proteins from proteolytic degradation may not necessarily lead to optimal conditions conducive to high-yield protein expression. One method to eschew this problem is the use of the recently developed *E. coli*-derived cell-free expression system, which uses a T7-based approach with an *E. coli* S30 cell extract in a coupled transcription/translation system. [25] During incubation, the reaction mixture, which contains all enzymes and high molecular mass compounds prerequisite for gene expression, is dialyzed against a low molecular mass substrate solution providing precursors to extend protein synthesis for greater than 10 hours. Integral membrane proteins produced in the expression system form precipitate that can subsequently be solubilized in mild detergents. [26] Keller *et al*. have demonstrated the functional preparation of resolubilized cell-free expressed rat organic cation transporters OCT1 and OCT2[27], while Klammt *et al*. reported nine out of fifteen human integral membrane cell-free expression preparations resulted in good ^{1}H-^{15}N, TROSY-HSQC spectra suitable for comprehensive NMR spectroscopy studies.[26] An additional 135 targets were selected in the 10-30 kDa range, and overall 111 of the 150 (74%) targets expressed at considerably high levels.

2.4 Membrane protein structures determined by NMR

Outer membrane proteins were among the first integral membrane proteins to have structures determined by X-ray crystallography. These outer membrane proteins have the β-barrel structural motif – an antiparallel β-sheet that closes on itself.[28] Additionally, the

vast majority of membrane protein structures have been determined by X-ray crystallography, and of the 1874 unique integral membrane protein structures deposited in the PDB, only 105 unique NMR structures have been deposited in the PDB. Progress has been relatively recent and localized to the past two decades; Bushweller and Tamm determined the first solution structure of a β-barrel, OmpA in 2001[29] while Wuthrich was simultaneously working on the structure of OmpX in 2001.[30] Zhou and Bushweller described the first NMR structure of a polytopic helical membrane protein, DsbB, in 2008.[31] To date, only seven β-barrel membrane protein structures determined by NMR have been deposited to the PDB: OmpA, OmpG, OmpX, VDAC-1, PagP, OprH and Opa_{60} (Figure 2.5).

2.4.1 Beta barrel structures determined by NMR

Outer membrane proteins – OmpA, OmpG, and OmpX

OmpA is a 35 kDa protein that consists of a transmembrane domain (19kDa) and globular periplasmic domain (16 kDa) with a multitude of putative functions, from providing physiological structural support for maintaining the shape of Gram-negative bacteria to forming ion channels and/or nonspecific pores for uncharged molecules in planar lipid bilayers.[28] OmpA is also one of the major surface antigens of Gram-negative bacteria; OmpA's extracellular loops adhere to human brain microvascular endothelial cells, facilitating passage through the blood brain barrier,[32] as well as contribute to pathogen evasion through serum resistance by binding to C4b binding proteins, which are complement fluid phase regulators.[33] The structure of the transmembrane domain of

OmpA was first determined by Schulz et al by X-ray crystallography, in C_8E_4 micelles.[34] Additionally, Bushweller and Tamm determined the solution structure of the transmembrane domain of OmpA (0-176 amino acid residues), which was refolded and studied in dodecylphosphocholine (FC12) micelles. Complete backbone assignments were obtained for 138 of the 177 residues, with an additional 18 partially assigned residues; the primary difficulties in assignments were the longer loop resonances on the extracellular side, due to peak broadening, likely as a result of conformational exchange. The conformational dynamics of the transmembrane domain of OmpA were also investigated using heteronuclear NOEs, which are sensitive to the mobility of individual amide N-H bond vectors on a ps-ns time scale. The average ratios were determined to be 0.73 in the β-barrel, 0.67 in the turns, and 0.46 in the loop regions; a static limit of 1.0 indicates a completely immobile residue, which suggests a fairly rigid barrel, with more mobile turns and flexible loops.

OmpX is a 148-residue outer membrane protein of *E. coli* that promotes bacterial adhesion and entry into mammalian cells. The first structure of OmpX was determined via crystallography in the detergent, *n*-octyltetraoxyethylene by Vogt *et al*.[35] The solution structure of OmpX was later determined in DHPC micelles by Fernandez *et al*. [36], and closely resembled the X-ray structure. The β-strands of OmpX were on average two residues shorter than that of the crystal structure, though poor resolution of loop residues and lack of long-range experimental constraints in these regions may have led to this result. Interestingly, Vogt *et al*. postulated that an Asn residue that was involved in a crystal contact point and protruded from the β-sheet could act as a hydrogen bonding partner; the

OmpX solution structure appears to confirm the conclusion that the protruding β-sheet is a structural feature of native OmpX and not a crystal artifact.

OmpG is a large integral membrane protein (33 kDa, 280 amino acid residues) that resides in the outer membrane of Gram-negative bacteria. OmpG functions as a monomeric porin that facilitates the uptake of large oligosaccharides,[37] and has been shown to form monomeric channels when reconstituted in lipid bilayers.[38] The crystal structure of OmpG has also been previously determined by two groups, and the determined structure forms a 14-stranded β-barrel with seven extracellular loops.[39, 40] Yildiz *et al*. also determined that OmpG adopts an open conformation at neutral pH (7.5), and a closed state at acidic pH (5.6) that is characterized by the folding of extracellular loop 6 into the barrel, catalyzed by the unzipping of hydrogen bonds by protonation of the solvent-exposed His231 and His261 at the ends of loop 6 and loop 7.[40] Liang and Tamm were able to determine the solution structure of OmpG by NMR using a variety of detergents.[41] Refolding of OmpG was attempted in short chain phospholipid 1,2-dihexanoyl-*sn*-glycero-3-phosphocholine (DHPC) and a variety of phosphocholines with different acyl chain lengths from 8-14, as well as two alkyl glucosides – β-octyl glucoside (OG) and *n*-dodecyl-β-maltopyranoside (DDM); ultimately, the best refolding condition was a rapid dilution into OG micelles over a span of 2 days. While initial ^{1}H-^{15}N TROSY spectra were promising, the stability of OmpG in OG micelles was an issue, and refolded OmpG in OG micelles were exchanged into FC12 micelles, allowing for weeks of measurements at 40°C. The resulting structure was in good agreement with the crystal structures in the β-region and had a majority of peaks assigned; the only unassigned peaks were in the entirety of the mobile loop 6, a few scattered residues in loops 1 and 2, and a variety of interfacial residues

on loop 7. The authors concluded, based on the flexibility of loop 6, that OmpG is likely present as a mixture of open and closed conformers that are in conformational exchange on the microsecond to millsecond time scale.

Voltage-dependent anion channel 1 (VDAC-1)

VDAC-1 is another large integral membrane protein that forms a 19-stranded β-barrel with the first and last strand parallel. VDAC-1 has been implicated in mitochondrial apoptosis.[42] Opening of VDAC-1 by pro-apoptoic protein Bax and Bak leads to the opening of the mitochondrial exit channel.[43] This event, in turn, allows the release of apoptogenic proteins such as catabolic hydrolases and their activators, that cause cell death through a variety of mechanisms.[44] Recombinant VDAC-1 has been shown to form voltage-gated channels in phospholipid bilayers in the presence of cholesterol.[45] Hiller *et al*. determined the solution structure of VDAC-1 in lauryldimethylamine oxide (LDAO) detergent micelles, and consistent with the function of VDAC-1 as a wide diffusion pore, observed no tertiary contacts between residues across the barrel diameter. The effects of adding cholesterol to the detergent micelle were investigated, but the overall structure of VDAC-1 is unchanged for 1:5:400 ratios of VDAC-1:cholesterol:LDAO. There are solely two distinct interaction sites with notable chemical shift changes greater than 0.05 ppm: at β-strands 7 and 8, and β-strand 11.

PhoPQ-activated outer membrane lipid A palmitoyltransferase enzyme (PagP)

PagP is a 20 kDa bacterial outer membrane protein composed of 162 residues that transfers a palmitate chain from a phospholipid to the glucosamine unit of lipid A. The palmitated lipid A is an antagonist of endotoxin signaling [46] and provides bacterial resistance against antimicrobial peptides by increasing outer membrane permeability.[47] To examine the effects of zwitterionic and non-ionic detergents on the fold and dynamics of PagP, Hwang *et al*. determined the solution structure of PagP in FC12 micelles and in OG micelles; precipitated PagP was rapidly diluted into FC12 for the former, and folded into SDS and dialyzed against OG in the latter. The refolded PagP specific activity has been shown to be indistinguishable from that of Native PagP purified from membranes, indicating a functional fold.[48] The rmsd between the structure determined for PagP in FC12 micelles and PagP in OG micelles was 0.91 Å in the β-barrel region, suggesting that the protein can adopt its native fold in the two different detergents, under different refolding conditions. All unassigned residues were localized to the extracellular loops, similar to the NMR studies of OmpA, which were hindered by broadened signals in the same regions. While the β-barrel is conserved between the two detergent conditions, exchange broadening occurred at the interfacial regions in the FC12 micelles, particularly with residues proximal to Pro28. This proline acts as a "hinge" by limiting H-bond formation between strands, and contributes to the mobility of the dynamic long extracellular loop.

Outer membrane protein H (OprH)

OprH is a 21-kDa, 200 residue protein native to the outer membrane of *Pseudomonas aeruginosa* that has been putatively thought to assist in establishing bacterial antibiotic resistance by preventing the uptake pathway of antibiotics across the outer membrane.[49]

OprH was proposed to bind lipopolysaccharide (LPS) sites which are normally occupied by divalent cations such as Mg^{2+}, thereby preventing access of polymyxin, gentamicin, and EDTA to these sites.[50] The only high resolution structure of OprH was determined by Edrington *et al*. via NMR spectroscopy in 2011, in 1,2-Dihexanoyl-*sn*-Glycero-3-Phosphocholine (DHPC) micelles. [51] Complete backbone chemical shift assignments were obtained for 156 of the 180 residues, with most of the unassigned residues occurring in loop and interfacial regions. Heteronuclear NOE data indicated dynamic extracellular loop regions on the ps-ns time scale that are far less ordered than the β-barrel region. These regions were proposed to tumble independently from the micelle-embedded body of the protein.

Opacity associated outer membrane protein (Opa_{60})

The opacity associated outer membrane proteins were discussed in Chapter 1, and will be the focus of Chapters 3 and 4. Briefly, Fox *et al*. (submitted manuscript) determined the first high resolution structure for Opa_{60}, an Opa_{CEA} binding protein, in FC12 micelles. Over 97% of the β-barrel and the periplasmic loops were assigned; portions of the extracellular loops were assigned using synthesized peptides, as well as specific amino acid labeling, though no NOE based distance restraints were detected for the extracellular loops. Two high variable (HV) regions on the extracellular loops determine receptor specificity. These regions are dynamic on the nanosecond timescale, and are predominantly disordered. Though specific long-lived intra-loop interactions are not observed, these loops are compact and weakly interact with each other. This diverse and dynamic nature is

potentially required for loop binding interactions, which enables the Opa protein to interact with a variety of host receptors.

2.4.2 Structures are solved in myriad conditions

None of the conditions leading up to the seven determined structures were identical and in many cases were not similar at all (Table 2.2). Apart from the variety of lipid bilayer mimics used, many of the buffer conditions also differed. No single optimal condition was found to be common for all membrane proteins. Rather, Pitfall 3 (membrane mimic selection) rears its ugly head, and exhaustive empirical screens are necessary for resolving spectra suitable for comprehensive structure determination. Some of these structures, such as PagP and OmpG, had multiple detergent environment, and progress towards optimizing detergent selection is still ongoing.

Table 2.2. The physical properties and buffer conditions of the seven β-barrel solution structures determined by NMR.

Protein	PDBID	Size (kDa)	Strands	Reconstitution Medium	Solution Conditions	τ_c (ns) MW (kDa)
OmpA	1G90 2GE4	19	8	600 mM FC12	10 mM KPO_4 50 mM NaCl 0.01% NaN_3	ND
OmpG	2JQY	33	14	70 mM OG +15 mM FC12	50 mM NaCl 0.05% NaN_3	ND
OmpX	1ORM 1Q9F	16	8	200 mM DHPC	20 mM $NaPO_4$ 100 mM NaCl 0.05% NaN_3	21 (60 kDa)
Opa_{60}	2MLH	29	8	110-150 mM FC12	20 mM $NaPO_4$ 150 mM NaCl	20
OprH	2LHF	20	8	150-175 mM DHPC	50 mM KCl 0.05% NaN_3	22
PagP	1MM4 1MM5	20	8	500 mM FC12 200 mM OG	50mM $NaPO_4$	20 (50-60 kDa)
VDAC-1	2K4T	32	19	300-500 mM LDAO	25 mM $NaPO_4$ 5 mM DTT	35 (75-90 kDa)

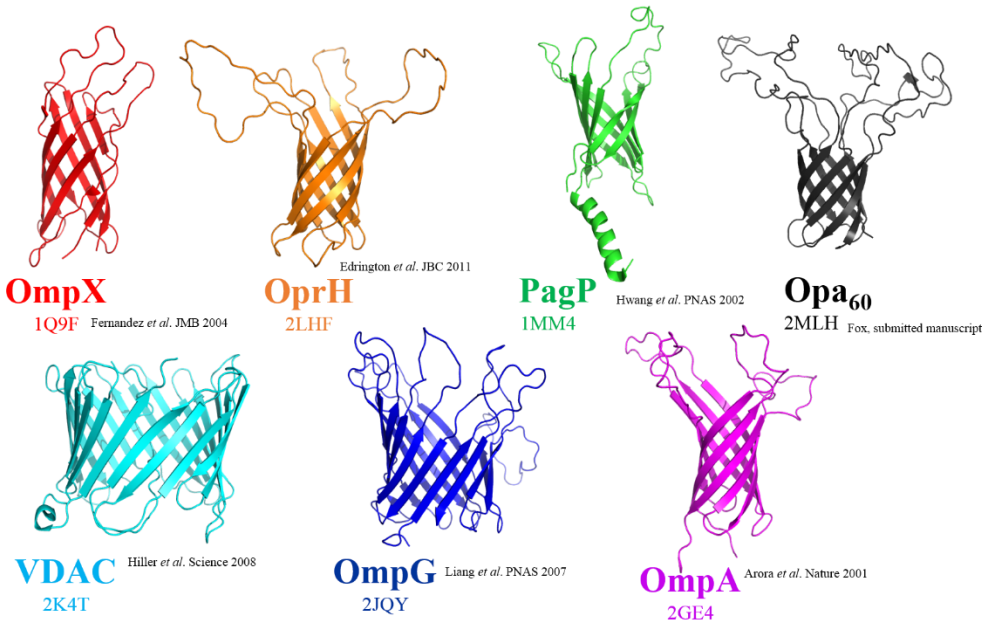


Figure 2.5. The structures of the seven β-barrel membrane proteins determined by NMR.

2.5 References

1. Gao, F. and T. Cross, Recent developments in membrane-protein structural genomics. Genome Biology, 2005. **6**(13): p. 244.
2. Poget, S.F. and M.E. Girvin, Solution NMR of membrane proteins in bilayer mimics: Small is beautiful, but sometimes bigger is better. Biochimica et Biophysica Acta (BBA) - Biomembranes, 2007. **1768**(12): p. 3098-3106.
3. Chill, J.H. and F. Naider, A solution NMR view of protein dynamics in the biological membrane. Current Opinion in Structural Biology, 2011. **21**(5): p. 627-633.
4. Prosser, R.S., et al., Current applications of bicelles in NMR studies of membrane-associated amphiphiles and proteins. Biochemistry, 2006. **45**(28): p. 8453-8465.
5. Luthra, A., et al., Nanodiscs in the studies of membrane-bound cytochrome P450 enzymes, in cytochrome P450 protocols, I.R. Phillips, E.A. Shephard, and P.R. Ortiz de Montellano, Editors. 2013, Humana Press. p. 115-127.
6. Hagn, F., et al., Optimized phospholipid bilayer nanodiscs facilitate high-resolution structure determination of membrane proteins. Journal of the American Chemical Society, 2013. **135**(5): p. 1919-1925.
7. Pochapsky, T., and Pochapsky, Susan., NMR for Physical and Biological Scientists, ed. G. Science 2007, New York.
8. Levitt, M., Spin Dynamics: Basics of Nuclear Magnetic Resonance, ed. J.W. Sons2008, Hoboken.

9. Wagner, G., A. Pardi, and K. Wuethrich, Hydrogen bond length and proton NMR chemical shifts in proteins. Journal of the American Chemical Society, 1983. **105**(18): p. 5948-5949.

10. Loth, K., et al., Determination of dihedral Ψ angles in large proteins by combining NHN/CαHα dipole/dipole cross-correlation and chemical shifts. Proteins: Structure, Function, and Bioinformatics, 2006. **64**(4): p. 931-939.

11. Marion, D., An introduction to biological NMR spectroscopy. Molecular & Cellular Proteomics, 2013. **12**(11): p. 3006-3025.

12. Wuthrich, K., NMR of Proteins and Nucleic Acids1991.

13. Wishart, D.S., B.D. Sykes, and F.M. Richards, The chemical shift index: a fast and simple method for the assignment of protein secondary structure through NMR spectroscopy. Biochemistry, 1992. **31**(6): p. 1647-1651.

14. Wishart, D. and B. Sykes, The 13C Chemical-Shift Index: A simple method for the identification of protein secondary structure using 13C chemical-shift data. Journal of Biomolecular NMR, 1994. **4**(2): p. 171-180.

15. Bundi, A. and K. Wüthrich, 1H-nmr parameters of the common amino acid residues measured in aqueous solutions of the linear tetrapeptides H-Gly-Gly-X-L-Ala-OH. Biopolymers, 1979. **18**(2): p. 285-297.

16. Fischer, M.W.F., et al., Characterizing semilocal motions in proteins by NMR relaxation studies. Proceedings of the National Academy of Sciences, 1998. **95**(14): p. 8016-8019.

17. Lipari, G. and A. Szabo, Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 1. Theory and range of validity. Journal of the American Chemical Society, 1982. **104**(17): p. 4546-4559.

18. Pervushin, K., et al., Attenuated T2 relaxation by mutual cancellation of dipole–dipole coupling and chemical shift anisotropy indicates an avenue to NMR structures of very large biological macromolecules in solution. Proceedings of the National Academy of Sciences, 1997. **94**(23): p. 12366-12371.

19. Venters RA, H.C., Famer BT, Trolard R, Spicer LD, Fierke CA, High-level 2H/13C/15N labeling of proteins for NMR studies. J Biomol NMR, 1995. **5**(4): p. 339-44.

20. Matthews, S., Perdeuteration/Site-Specific Protonation Approaches for High-Molecular-Weight Proteins, in Protein NMR Techniques, A.K. Downing, Editor 2004, Humana Press. p. 35-45.

21. Gardner, K.H. and L.E. Kay, The use of 2H, 13C, 15N multidimensional NMR to study the structure and dynamics of proteins. Annual Review of Biophysics and Biomolecular Structure, 1998. **27**(1): p. 357-406.

22. Goto, N., et al., A robust and cost-effective method for the production of Val, Leu, Ile (δ1) methyl-protonated 15N-, 13C-, 2H-labeled proteins. Journal of Biomolecular NMR, 1999. **13**(4): p. 369-374.

23. Gardner, K.H., M.K. Rosen, and L.E. Kay, Global folds of highly deuterated, methyl-protonated proteins by multidimensional NMR. Biochemistry, 1997. **36**(6): p. 1389-1401.

24. Liang, B., J.H. Bushweller, and L.K. Tamm, Site-directed parallel spin-labeling and paramagnetic relaxation enhancement in structure determination of membrane proteins by solution NMR spectroscopy. Journal of the American Chemical Society, 2006. **128**(13): p. 4389-4397.

25. Klammt, C., et al., High level cell-free expression and specific labeling of integral membrane proteins. European Journal of Biochemistry, 2004. **271**(3): p. 568-580.

26. Klammt, C., et al., Facile backbone structure determination of human membrane proteins by NMR spectroscopy. Nat Meth, 2012. **9**(8): p. 834-839.

27. Keller, T., et al., Cell free expression and functional reconstitution of eukaryotic drug transporters. Biochemistry, 2008. **47**(15): p. 4552-4564.

28. Arora, A., et al., Refolded outer membrane protein A of *Escherichia coli* forms ion channels with two conductance states in planar lipid bilayers. Journal of Biological Chemistry, 2000. **275**(3): p. 1594-1600.

29. Arora, A., et al., Structure of outer membrane protein A transmembrane domain by NMR spectroscopy. Nat Struct Mol Biol, 2001. **8**(4): p. 334-338.

30. Fernández, C., K. Adeishvili, and K. Wüthrich, Transverse relaxation-optimized NMR spectroscopy with the outer membrane protein OmpX in dihexanoyl phosphatidylcholine micelles. Proceedings of the National Academy of Sciences of the United States of America, 2001. **98**(5): p. 2358-2363.

31. Zhou, Y., et al., NMR solution structure of the integral membrane enzyme DsbB: functional insights into DsbB-catalyzed disulfide bond formation. Molecular cell, 2008. **31**(6): p. 896-908.

32. Shin, S., et al., *Escherichia coli* outer membrane protein A adheres to human brain microvascular endothelial cells. Biochemical and Biophysical Research Communications, 2005. **330**(4): p. 1199-1204.

33. Prasadarao, N.V., et al., A novel interaction of outer membrane protein A with C4b binding protein mediates serum resistance of *Escherichia coli* K1. The Journal of Immunology, 2002. **169**(11): p. 6352-6360.

34. Pautsch, A. and G.E. Schulz, Structure of the outer membrane protein A transmembrane domain. Nat Struct Mol Biol, 1998. **5**(11): p. 1013-1017.

35. Vogt, J. and G.E. Schulz, The structure of the outer membrane protein OmpX from Escherichia coli reveals possible mechanisms of virulence. Structure (London, England : 1993), 1999. **7**(10): p. 1301-1309.

36. Fernández, C., et al., NMR structure of the integral membrane protein OmpX. Journal of Molecular Biology, 2004. **336**(5): p. 1211-1221.

37. Misra, R.a.B.S., A novel mutation, cog, which results in production of a new porin protein (OmpG) of *Escherichia coli* K-12. J Bacteriol, 1989. **171**(8): p. 4105-4111.

38. Conlan, S., et al., Biochemical and biophysical characterization of OmpG: a monomeric porin. Biochemistry, 2000. **39**(39): p. 11845-11854.

39. Subbarao, G.V. and B. van den Berg, Crystal structure of the monomeric porin OmpG. Journal of Molecular Biology, 2006. **360**(4): p. 750-759.

40. Yildiz, Ö., et al., Structure of the monomeric outer-membrane porin OmpG in the open and closed conformation. The EMBO Journal, 2006. **25**(15): p. 3702-3713.

41. Liang, B. and L.K. Tamm, Structure of outer membrane protein G by solution NMR spectroscopy. Proceedings of the National Academy of Sciences, 2007. **104**(41): p. 16140-16145.

42. Vander Heiden, M.G. and C.B. Thompson, Bcl-2 proteins: regulators of apoptosis or of mitochondrial homeostasis. Nat Cell Biol, 1999. **1**(8): p. E209-E216.

43. Shimizu, S., M. Narita, and Y. Tsujimoto, Bcl-2 family proteins regulate the release of apoptogenic cytochrome c by the mitochondrial channel VDAC. Nature, 1999. **399**(6735): p. 483-487.

44. Kroemer, G., L. Galluzzi, and C. Brenner, Mitochondrial membrane permeabilization in cell death. Physiological Reviews, 2007. **87**(1): p. 99-163.

45. Koppel, D.A., et al., Bacterial expression and characterization of the mitochondrial outer membrane channel: effects of N-terminal modifications. Journal of Biological Chemistry, 1998. **273**(22): p. 13794-13800.

46. Muroi, M., et al., Lipopolysaccharide-mimetic activities of a toll-like receptor 2-stimulatory substance(s) in enterobacterial lipopolysaccharide preparations. Infection and Immunity, 2003. **71**(6): p. 3221-3226.

47. Guo, L., et al., Lipid A acylation and bacterial resistance against vertebrate antimicrobial peptides. Cell, 1998. **95**(2): p. 189-198.

48. Bishop, R.E., et al., Transfer of palmitate from phospholipids to lipid A in outer membranes of gram-negative bacteria. The EMBO Journal, 2000. **19**(19): p. 5071-5080.

49. Young M, B.M., Bell A, and Hancock R., Role of *Pseudomonas aeruginosa* outer membrane protein OprH in polymyxin and gentamicin resistance: isolation of an

OprH-deficient mutant by gene replacement techniques. Antimicrobial Agents and Chemotherapy, 1992. **36**(11): p. 2566-2568.

50. Hancock RE, B.A., Antibiotic uptake into gram-negative bacteria. Eur J Clin Micribiol Infect Dis, 1988. **7**(6): p. 713-20.

51. Edrington, T.C., et al., Structural basis for the interaction of lipopolysaccharide with outer membrane protein H (OprH) from *Pseudomonas aeruginosa.* Journal of Biological Chemistry, 2011. **286**(45): p. 39211-39223.

Chapter 3: Optimization of NMR-based opacity-associated protein environments

Of the 105 unique β-barrel membrane protein structures deposited in the PDB, seven of them were determined with solution NMR (with five of these having corresponding X-ray crystal structures). The underrepresentation of NMR structures is typically attributed to (1) the limitation in size that can be investigated (theoretically upper size limits of ~100 kDa), (2) the availability of expression systems that produce large yields of isotopically labelled folded proteins, and (3) the selection of solution conditions that yields quality NMR spectra. For β-barrel membrane proteins the size of the complex and the availability of large quantities of isotopically labeled folded protein do not appear to be the limiting factor. Recombinant expression to inclusion bodies, solubilization and refolding *in vitro* to produce large quantities for structural studies has been demonstrated for many β-barrel membrane proteins irrespective of the number of β-strands and size. Although the detergent micelle contributes to the overall molecular weight of the complex, the seven β-barrel membrane proteins determined thus far vary from 8–19 β-strands and 16–31 kDa (OmpX and VDAC, respectively) indicating that size is not the major bottleneck.(*1*, *2*) A significant hurdle is the optimization of solution conditions.

In this chapter, we systematically investigate detergents and ionic strengths that influence ^{15}N, ^{1}H-HSQC spectra of a model β-barrel membrane protein (Opa_{50} from *N. gonorrhoeae*) in order to gain an understanding of the physical forces that stabilize a protein fold for solution NMR structural investigations. Opa_{50} is an eight stranded β-barrel membrane

protein with four extracellular loops. The first three loops are long and bind to host receptors mediating phagocytosis of the bacterium by human cells (*3*, *4*). Opa_{50} binds to heparan sulfate proteoglycan receptors (HSPGs) as well as integrins via a heparan-mediated intermediate.(*5*) These interactions are predicted to be largely electrostatic in nature because the interaction *in vivo* can be disrupted with heparin.(*5*)

3.1 Introduction

Temperature, ionic strength, pH, and detergent all modulate the quality of NMR spectra of membrane proteins; specifically, chemical shift dispersion and line widths. The role of detergents in stabilizing folds for NMR structure is not well-understood. Once folded, β-barrel membrane proteins appear to be very stable in a variety of detergent micelles as determined by SDS-PAGE gel shift assays and circular dichroism; however, the corresponding NMR spectra are typically not conducive to NMR structure determination.(*6*) Although there is limited data on the effects solution conditions have on the NMR spectral quality of β-barrel membrane proteins, detergent type has been shown to influence membrane protein activity (*7-9*), folding (*8*), stability (*10*), structure (*11*), and side chain conformations (*12*). Opa_{50} was folded and/or detergent exchanged into FC12 (n-dodecylphosphocholine), FC10 (n-decylphosphocholine), OG (n-octyl-β-D-glucopyranoside), DM (n-decyl-β-D-maltopyranoside), and DDM (n-dodecyl-β-D-maltopyranoside). Spectra were recorded at elevated temperatures (30-50˚C), with ionic strengths varying from 0 mM to 200 mM NaCl, and pH values of 6.0 to 6.8.

Several β-barrel membrane protein structures have been determined in various detergents and detergent conditions, and different detergent conditions had been used for the refolding and solubilizing steps (e.g. PagP in OG and DM, OmpG in OG and FC12). This portion of the project systematically examined the Opa protein fold in five different detergents: FC12 FC10, OG, DM, and DDM. The chemical structures and differing headgroups between each are shown in Figure 1. Detergents were chosen based on detergents solubilizing previously determined structures, as well as systematically perturbing headgroups, ionic properties, and the average head group to head group distance (Table 3.1).

Table 3.1. Physical properties of pure detergents. The critical micelle concentration (cmc), shape of micelle, headgroup to headgroup distance (L), and aggregation number (N) are reported in this table for the detergents in this study.

detergent	cmc	shape	*L* (Å)	*N*
FC10 (59 mM)	11[a]	prolate	29.8	40-50
FC12 (77 mM)	1.5[a]	prolate	35.2	70-80
OG (50 mM)	18-20[b]	oblate	25.2	70-90
DM (80 mM)	1.8[c]	oblate	32.6	90-100
DDM (94 mM)	0.18[d]	oblate	37.4	135- 155

[a]Measurements performed by Anatrace (Affymetrix, Inc.), [b] Lorber *et al*. (*13*), [c] Alpes *et al.* (*14*), [d] Van Aken *et al*. (*15*), all other parameter are in Lipfert *et al*. (*16*)

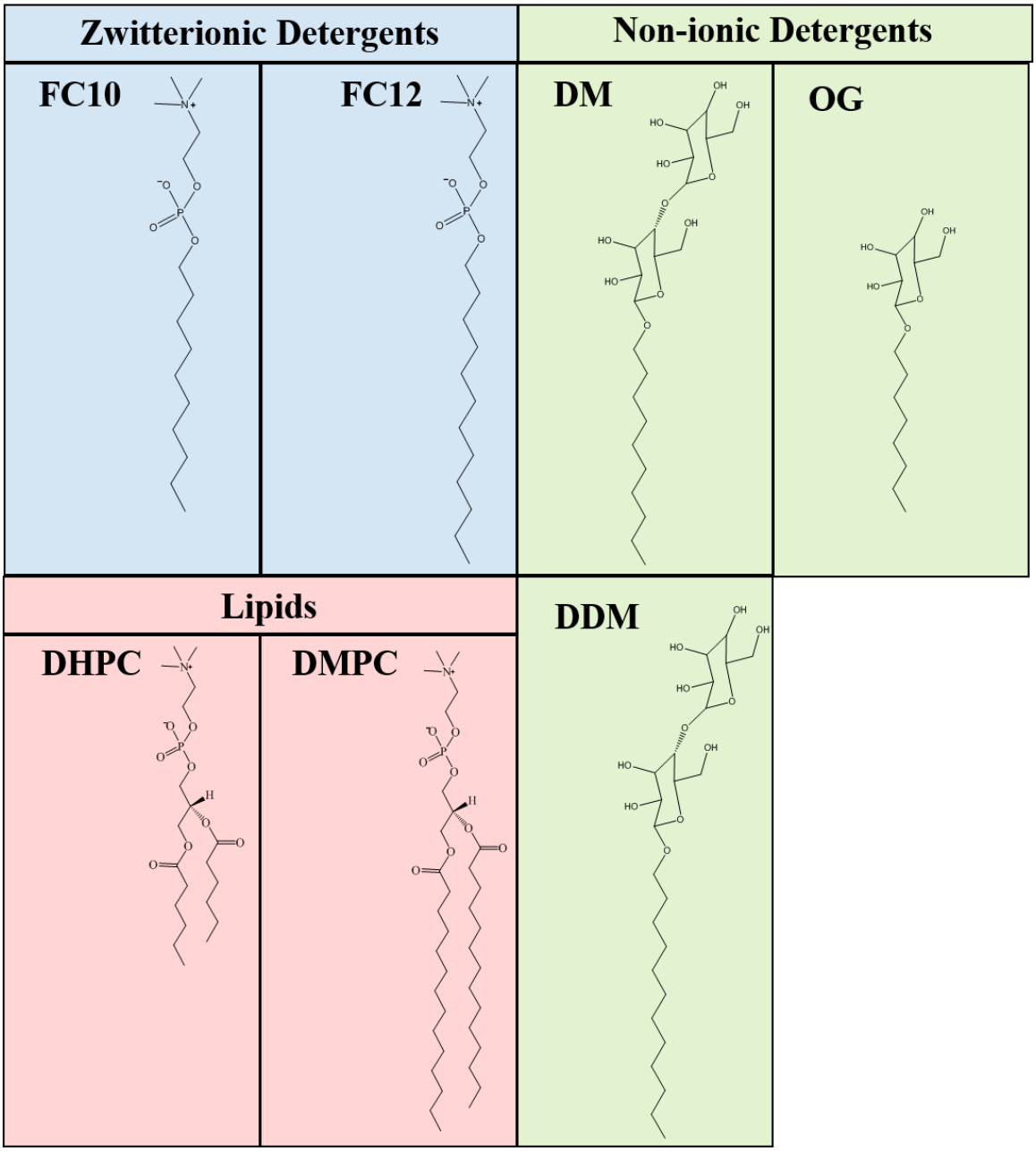


Figure 3.1. The chemical structures of commonly used detergents and lipids in structural/functional studies. The zwitterionic detergents (in blue) include *n*-decyl-phosphocholine (FC10) and *n*-dodecyl-phosphocholine (FC12). The non-ionic detergents (in green) include *n*-decyl-β-$_D$-maltoside (DM), *n*-dodecyl-β-$_D$-maltoside (DDM), and *n*-octyl-β-$_D$-glucoside (OG). The lipids (in red) include and 1,2-Dihexanoyl-*sn*-Glycero-3-Phosphocholine (DHPC) and 1,2-Dimyristoyl-*sn*-Glycero-3-Phosphocholine (DMPC).

3.2 Materials and Methods

3.2.1 Cloning, expression, and purification of Opa proteins

The *opa*$_{50}$ and *opa*$_{60}$ genes were sub-cloned from pEX vectors provided by Martine Bos (Utrecht University, The Netherlands) into the pET28b vector (EMD Chemicals, Gibbstown, NJ) between NdeI and HindIII restriction sites, introducing an N-terminal His$_6$-tag in the expressed construct. For expression, the Opa$_{50}$ plasmid was transformed into a BL21(DE3) *Escherichia coli* strain. A starting culture was prepared with 10 mL of Luria-Bertani (LB) media containing kanamycin (50 µg/mL) and inoculated with a single colony from a freshly transformed plate. The suspension was incubated at 37 °C with shaking at 250 rpm for 15-18 hours, and used to inoculate 1 L media. Inoculated cell cultures were grown in LB media or M9 minimal media media containing kanamycin (50 µg/mL) at 37 °C until the OD_{600} reached ~1, when protein expression was induced with 1 mM isopropyl-β-thio-D-galactoside (IPTG). After 4h (8 h for M9 MM), cells were harvested by centrifugation (5,000 x *g*, 20 min, 4 °C), the cell pellet was resuspended in 20 mL lysis buffer (150 mM NaCl, 50 mM Tris-HCl at pH 8.0) and lysed with several cycles of microfluidization (Microfluidics model 110L, Newton, Mass). The insoluble fraction (in which most of the Opa protein is observed) was pelleted (12,000 x *g*, 30 min, 4 °C), washed by resuspension in lysis buffer, and then centrifuged again. Opa$_{50}$ was solubilized from the insoluble fraction (pellet) in 25 mL of extraction buffer (lysis buffer with 8M urea) at room temperature with constant stirring overnight. Insoluble particulate was removed via centrifugation (12,000 x *g*, 30 min, 4 °C), and the supernatant was applied to 1.5 mL of cobalt charged chelating resin (GE Healthcare, Piscataway, NJ) previously equilibrated

with 15 column volumes (CV) of extraction buffer. The column was washed with 15 CV of wash buffer (20 mM sodium phosphate, pH 7.8, 150 mM NaCl, 20 mM imidazole, 8 M urea) and the protein was eluted with five CV of elution buffer (wash buffer with 680 mM imidazole). The protein was concentrated (MWCO=10kDa) to ~550 μM. Typical expression yields were ~15 mg of protein/L of cell culture. Protein concentration and purity were determined by A_{280} (ε = 40,340 M^{-1} cm^{-1}) and SDS-PAGE, respectively.

3.2.2 Preparation of protein-detergent complexes

Concentrated (~550 μM in 2 mLs) Opa_{50} in urea was diluted ≈100-fold by dropwise addition to 40 mL refolding buffer (20 mM sodium phosphate, pH 8.0, 3.8 mM n-dodecylphosphocholine, 500 mM NaCl) at room temperature. Opa_{50} was folded in each detergent (Affymetrix, Inc - Anatrace): n-dodecylphosphocholine, n-decylphosphocholine, n-decyl-β-D-maltoside, n-dodecyl-β-D-maltoside, and n-octyl-β-D-glucopyranoside in detergent concentrations yielding approximately 3:1 detergent micelle:protein ratios, where detergent micelles were calculated as follows:

$$[detergent\ micelle] = \frac{([detergent]-[CMC])}{aggregation\ \#} \tag{3.1}$$

where the CMC is the critical micelle concentration and the aggregation # is the number of detergent monomers that compose a detergent micelle. The [CMC] is subtracted from the overall detergent concentration to account for the free detergent monomer that does not contribute to micelles. Because Opa_{50} folds slowly, Opa-detergent solutions were left at room temperature over a minimum of four days. Folding was primarily assessed by SDS-

PAGE electrophoresis because folded and unfolded β-barrel proteins have different electrophoretic mobilities (*17, 18*). For NMR studies, folded Opa_{50} was concentrated to ≈600 μM and dialyzed into NMR buffer (150 mM NaCl, 20 mM phosphate buffer at pH=6.2). Detergent concentrations were measured by one-dimensional (1D) ^{1}H NMR spectroscopy by comparison of integrated methyl peaks with samples of known detergent concentrations.

3.2.3 Detergent exchange of Opa solubilized complexes

Folded Opa_{50} (in FC12) was applied to a cobalt chelating resin previously equilibrated with refolding buffer. The column was washed with 10 CV of wash buffer (150 mM NaCl, 20 mM imidazole, 20 mM phosphate buffer at pH=7.8) containing detergent ([FC12]=15 mM, [FC10] = 12 mM, [DM] = 10 mM, [DDM] = 21 mM, [OG] = 26 mM), and the protein was eluted with five CV of elution buffer (150 mM NaCl, 680 mM imidazole, 20 mM phosphate buffer at pH=7.0), yielding typical protein concentrations of 200 μM). The protein was concentrated (MWCO=10 kDa) to ~650 μM and folding was assessed by SDS-PAGE. The protein was then dialyzed into NMR buffer. Detergent concentrations were adjusted to yield 3:1 detergent micelle:protein molar ratios by dilution or concentration (MWCO=10 kDa). Detergent homogeneity and detergent concentrations were determined using 1D ^{1}H NMR spectroscopy. The detergent concentrations varied between 60–150 mM and protein concentrations were between 400–800 μM.

3.2.4 Trypsin treatment of Opa_{50}

Opa_{50} samples were dialyzed against refolding buffer without detergent for optimal trypsin activity. The molar ratio of trypsin to protein was 1:100 and the mixture was incubated at 37 °C. To assess the degree of cleavage and the extent of folding, aliquots were removed at 0, 1 min, 5 min, 10 min, 30 min, 1 h, 3 h, and 36 h time intervals and mixed with SDS-PAGE loading buffer. To remove trypsin, a column with 1 mL of benzamadine resin was equilibrated with 10 CV of refolding buffer, and the sample was applied. The flowthrough was collected, Opa_{50} was concentrated (MWCO=10-kDa) to ≈800 μM, and dialyzed against NMR buffer.

3.2.5 Circular dichroism and NMR spectroscopy

Isotopically ^{15}N, ^{2}H-labeled Opa_{50} or Opa_{60} for HSQC experiments were prepared as described above using M9 minimum medium containing $^{15}NH_4Cl$ (99%, Cambridge Isotope Laboratories, Inc) as a nitrogen source and D_2O (99%, Cambridge Isotope Laboratories, Inc). NMR samples included the addition of 10% D_2O for lock. All experiments were performed on a 600 or 800 MHz Bruker Avance spectrometer with a cryoprobe. Two-dimensional (2D) ^{15}N,^{1}H-TROSY (transverse relaxation optimized spectroscopy) spectra were recorded for each sample with a data matrix of 64 x 2048 complex points, with an experimental time of ~4 h. The backbone assignment of Opa_{60} was provided by Fox and Columbus (submitted). For ionic strength experiments, a 0 mM NaCl Opa_{50} sample was prepared as described above by, but dialyzed against NMR buffer with no salt. A 5 M NaCl stock was made and added to the NMR sample incrementally (to 25,

50, 75, 100, 125, 150, 175, 200 mM). After the addition of each aliquot the sample was equilibrate for 3h and the $^{15}N,^{1}H$-HSQC recorded. The following parameters were used to determine the ^{15}N R_1, and R_2 values - R_1 relaxation decays were set to 5, 10, 15, 25, 50, 100, 250, 500, and 1000 ms, and the R_2 experiments were recorded with 17, 34, 51, 68, 102, 140, 191, and 242 ms relaxation delays.

CD experiments were performed in an Aviv 410 circular dichroism spectropolarimeter using 0.5 nm bandwidth and 2 mm optical path cell. Opa concentrations varied from ~3 μM protein to ~110 μM protein.

3.2.6 Cloning, expression, and purification of MSP1D1 proteins

MSP1D1 wild type (wt) plasmid pMSP1D1 was obtained from Steven Sligar (Addgene). The truncated MSP protein variants were cloned from a pet28a expression vector (EMD) harboring the MSP1D1 gene sequence provided by Gerhard Wagner (Harvard University), introducing an N-terminal His_6 tag and a TEV cleavage site in the expressed constructs. Expression and purification of all MSP proteins was performed according to established protocols.(*19*) For expression, the MSP1D1 plasmid was transformed into a BL21(DE3) *E. coli* strain. A starting culture was prepared with 10 mL of LB media containing kanamycin (50 μg/mL) and inoculated with a single colony from a freshly streaked plate. The suspension was incubated at 37 °C with shaking at 250 rpm for 15-18 hours, and used to inoculate 1 L media. Inoculated cell cultures were grown in LB media containing kanamycin (50 μg/mL) at 37 °C until the OD_{600} reached ~0.7, when protein expression was induced with 1 mM IPTG. After 4h, cells were harvested by centrifugation (5,000 x *g*, 20

min, 4 °C) and the cell pellet was frozen. The frozen cell pellet was resuspended in lysis buffer (500 mM NaCl, 50 mM Tris-HCl, 1% TritonX-100, 1 mM EDTA, 6M GuHCl at pH 8.0) and lysed by pulse sonication for 30 seconds on, 30 seconds off at 20% amplitude for 15 minutes (Qsonica). Post lysis, 5 mM $MgCl_2$ and 100 U Benzoase was added to the lysate, and the solution was incubated on ice for 1 h. The lysate was then clarified by centrifugation (10,000 x *g*, 45 min, 4 °C). The supernantant was applied to a NiNTA column (~1 mL resin/1 L cell culture) equilibrated with Buffer A (50 mM Tris pH 8.0, 500 mM NaCL, 1% TX-100). The column was first washed with 5-10 CV of Buffer A, followed by 5-10 CV of Buffer B (Buffer A + 50 mM sodium cholate), 5-10 CV of Buffer C (50 mM Tris pH 8.0, 500 mM NaCL), then 5-10 CV of Buffer D (Buffer C + 20 mM imidazole), and eluted with Buffer E (Buffer C + 500 mM imidazole). The eluted fraction was buffer exchanged using into 50 mM Tris pH 8.0, 20 mM NaCl, 1 mM EDTA, 2 mM DTT an Amicon centrifugal filter unit of 10 kDa MWCO (Millipore). Next, TEV protease (1 A_{280} TEV for 100 A_{280} MSP protein) was added to the eluted fraction, and incubated overnight at 4 °C. This solution was buffer exchanged using an Amicon centrifugal filter unit of 10 kDa MWCO (Millipore) to 50 mM Tris pH 8.0, 500 mM NaCl, and applied onto a Buffer A equilibrated NiNTA column. The column was washed with Buffer D, and eluted with Buffer E. The eluted fraction was buffer exchanged into 20 mM Tris pH 7.5 100 mM NaCl, 0.5 mM EDTA using an Amicon centrifugal filter unit of 10 kDa MWCO (Millipore) and used immediately.

3.2.7 Reconstitution of Opa proteins into nanodiscs

Opa_{60} was reconstituted into nanodiscs according to established protocols.(*20*) MSP1D1ΔH5 construct (plasmid provided by Gerhard Wagner, Harvard University) was purified and assembled in 20 mM Tris/HCl pH 7.5, 100 mM NaCl, 5 mM EDTA buffer with the appropriate amount of dry lipid/detergent to obtain a mixture of MSP1D1ΔH5:DMPC:sodium cholate with the molar ratio of 1:50:100. Opa_{60} refolded in FC12 was added to the mixture, and the Opa_{60}: MSP1D1ΔH5 ratio was adjusted to 1:4. The mixture was incubated at 4 °C for one hour, and detergent was removed with ~0.5 g of washed Biobeads SM-2 (Biorad) per mL of assembly mixture. This suspension was gently agitated at 4 °C for 6–10 hrs. Biobeads were removed by centrifugation followed by decantation of the supernatant. The supernatant was then concentrated and purified on a Superdex 200 size exclusion column equilibrated with 20 mM sodium phosphate, pH 6.5, 50 mM NaCl, and 5 mM EDTA. The main peak of the chromatogram was pooled and concentrated using an Amicon centrifugal filter unit of 30 kDa MWCO (Millipore). The NMR sample consisted of ~0.5 mM $^{2}H,^{15}N$ Opa_{60} in MSP1D1ΔH5 nanodiscs with d_{54}-DMPC (Avanti Polar Lipids), in gel filtration buffer supplemented with 10% (v/v) D_2O.

3.3 Results and Discussion

Opa proteins play a key role in the invasion of human host cells by *Neisseria* bacteria, mediating the interaction of the bacteria with the host cell's receptors. All Opa proteins are eight-stranded β-barrels that have three periplasmic loops and four extracellular loops, three of which have regions of high sequence variability among the Opa proteins (semi-variable and hypervariable regions, labeled SV1, HV1, and HV2 - Figure 2A) and determine receptor specificity (*3*). The β-barrel amino acid sequence between Opa proteins is highly conserved (approximately ~70% identity) between all Opa proteins (*21*). Opa_{50} and Opa_{60} are two proteins which bind to heparin sulfate proteoglycans and carcinoembryonic antigen-related cellular adhesion molecules, respectively. These two proteins differ only by a single residue the β-barrel amino acid sequence, but have significantly different extracellular loop sequences. The NMR spectra of Opa_{50} displayed significantly more line broadening than that of Opa_{60}, particularly in the extracellular loop region. Compared to other β-barrel membrane proteins previously investigated with NMR, the Opa proteins have longer loops (loop 1 containing the SV1 sequence, 36 residues; loop 2 containing the HV1 sequence, 44 residues; and loop 3 containing the HV2 sequence, 49 residues) that comprise approximately half of the protein. As a result, the NMR spectrum has two distinct resonance populations corresponding to the highly dynamic loops in the aqueous environment and the globular ordered β-barrel regions embedded in the detergent micelle. In the case of Opa_{50}, the line broadening of both regions was more sensitive to buffer conditions (specifically ionic strength) than Opa_{60} and, therefore, required a systematic optimization of solution conditions for structure determination. A comparison

of the $^{15}N,^{1}H$-TROSY spectra of Opa_{50} and Opa_{60} in FC12 demonstrates the β-barrel fold is identical since the resonances overlap significantly (Appendix Figure A1); thus, the assignment of the β-barrel of Opa_{50} is complete by comparison. Using this assignment, the effects of each solution condition on regions of the barrel could be assessed. In addition, the loop and barrel can be spectroscopically dissected with trypsin cleavage. Trypsin treatment of the protein-detergent complex cleaves the exposed loop residues after arginine and lysine residues (Figure 2A and B); the detergent micelle sterically hinders the protease from cleaving the β-barrel region (Figure 2A and B). The trypsin cleavage is easily monitored via SDS-PAGE (Figure 3.2C). As was observed with Opa_{60}, treatment of Opa_{50} with trypsin yields a folded β-barrel (Figure 3.2C and 3.2D) in which the $^{15}N,^{1}H$-TROSY spectrum overlaps with the β-barrel region of the uncleaved full length Opa_{50} (Figure 3.2D). The assignment allows the effects of each solution condition on regions of the β-barrel to be assessed and the trypsin cleavage aids in determining whether the condition affects the loops, barrel, or both regions.

3.3.1 Detergents affect the Opa_{50} fold

For structural studies, many β-barrel membrane proteins are solubilized from inclusion bodies with a denaturant and folded into a membrane mimic by removing the denaturant in the presence of the detergents or lipids. To investigate the effects of different detergents, Opa_{50} was folded in FC12 and exchanged into four different detergents (FC10, DM, DDM, and OG) on an affinity column. SDS-PAGE migration indicated Opa_{50} remained folded in each condition, and CD spectroscopy indicated that Opa_{50} is ~50% β-strand and 50% random coil in each of the five detergents (Figure 3.3).

Detergent selection has been shown to dramatically affect the quality of NMR spectra of membrane proteins and Opa_{50} is no exception (Figure 3.4). The two main regions to consider are the β-barrel region, and the loop region. In FC12, all of the β-barrel region peaks were well-resolved. The observed β-barrel peaks overlap with Opa_{50} in FC12 allowing 84% of the β-barrel resonances to be assigned in FC10, 77% in DM, 75% in OG, and 44% in DDM (Figure 3.4B, D, F, and H). The resonances that are broadened beyond detection correspond to residues predominantly at the edges of the β-strands (the most in strands five and six, which flank the longest loop) near the headgroup region of the micelle. The phosphocholine detergents appear to be most conducive for preserving the embedded β-barrel fold, though the majority of residues are present for DM and OG as well. Only Opa_{50} in DDM resulted in large amounts of spectral broadening. DDM has the largest headgroup to headgroup distance of the five detergents, and hydrophobic mismatch of the embedded β-barrel is likely. Because the DDM hydrophobic length (the headgroup-headgroup volume) may be too large, Opa_{50} is likely in conformational exchange to expose hydrophilic residues to the solvent, or in is present in a heterogeneous population.

Although specific loop residues are not assigned in Opa_{50}, the Opa_{50} loop regions in each of the detergents experience a wide array of line broadening. The loop resonance quality appears to be highest for Opa_{50} in FC12, with the next best quality in OG and DDM, with the worst loop resonance quality in FC10 followed by DM. This trend does not appear to match any micelle attribute; the ionic character of the micelles varies in resonance quality, and the second (OG) and third (DDM) best conditions are on the opposite ends of headgroup-headgroup dimensions. The peaks with the greatest intensities are likely

residues near the SV and HV regions of the extracellular loops, which are furthest away from the barrel and fluctuating more rapidly than the β–barrel-micelle complex.

3.3.2 Temperature affects β-barrel and loop dynamics.

The line broadening observed in the $^{15}N,^{1}H$-TROSY spectrum of Opa_{50} in FC12 micelles indicated that regions of the protein are undergoing exchange between two or more conformations. Although protein aggregation is likely not contributing to the line broadening based on gel migration and spectral quality of the β-barrel region, there is a possibility that the loops are interacting with each other (within the Opa monomer), with empty micelles, or free detergent monomer. To investigate this exchange process, the $^{15}N,^{1}H$-TROSY spectrum of Opa_{50} was recorded at 20, 30, and 40 °C. In general, temperature has been observed to modulate the quality of Opa protein's loop and β-barrel regions.(*22*) However, no single temperature is optimal for both regions, as low temperatures yield the highest intensity for the loop region, while high temperatures yield the highest intensity in the barrel region. As the temperature decreases, the overall tumbling of the protein-detergent complex slows, which causes the β-barrel peaks, as well as many of the interfacial residues, to broaden beyond detection. Because of the globular nature of the β-barrel embedded in a detergent micelle, the tumbling ($\tau_c \approx 8.52$ns, Table 3.4 at 40 °C) of the protein-detergent complex increases with a decrease in temperature and yields broadened peaks. In particular, several of the residues that experience severe line broadening (S57, S120, G193, S231) are predicted to be at the detergent interface and, thus, thought to be most highly affected by lowered temperatures. The loop regions farthest from

the micelle fluctuate more rapidly, and are therefore more likely to be observed at lower temperatures.

3.3.3 Ionic strength affects Opa_{50} spectral quality

Peak broadening as a result of conformational exchange could potentially be modulated by (1) intramolecular loop interactions in the monomer, (2) intermolecular loop interactions between Opa proteins, or by (3) interactions between the Opa loops and detergent micelle or monomer. Native Opa_{50} loop binding interactions are predicted to be electrostatic in nature, and therefore ionic strength was explored as a systematic variation that could modulate interactions involving the charged loops. As was observed with temperature, the spectral quality of both the β-barrel and loop regions were affected by ionic strength (Figure 3.5). There is an ionic strength threshold close to 75 mM NaCl that must be surpassed to yield assignable spectra. Prior to the threshold, many cross peaks are not observed due to extensive line broadening that can be attributed to one of the outlined scenarios. Figure 3.5 shows the loop and barrel residues present at 25, 75, and 125 mM NaCl at different temperatures, and Table 3.2 details the number of peaks present under those conditions.

3.3.4 Proteolysis facilitates investigation of ionic effects

The effect of ionic strength on the line broadening of the β-barrel region could be due to an increase in molecular mass of the complex due to loop interactions between protein-detergent complexes and empty micelles (Figure 3.6A), protein-detergent complexes with other protein-detergent complexes (Figure 3.6B), changes in micelle shape and size that

modulate β-barrel dynamics (Figure 3.6C, Figure 3.6D), or modulation of the strength of the salt bridges within the β-barrel. Size increases (A and B) due to loop interactions could be investigated by comparing spectra of the full length and the trypsin cleaved Opa_{50}. The $^{15}N,^{1}H$-TROSY spectrum of trypsin cleaved Opa_{50} provides insight to the influence of solution conditions on the loops compared to the barrel. The spectrum of trypsin-treated Opa_{50} in FC12 is deconvoluted when compared to the full length spectra, and the overall fold of barrel is maintained in the trypsin-treated state (Figure 3.2). A comparison of the digested and full length 0 mM NaCl spectra reveals a large difference in the beta barrel peaks in the >8.5 ppm region; many of the beta barrel peaks are broadened in the full length spectrum as previously indicated in Figure 3.5. However, peaks that are not present in the full length are visible in the spectrum of the tryspin treated protein, indicating that loss of the loops removes line-broadening of the β-barrel residues. This is visually evident in the overlaid spectra of trypsin cleaved Opa_{50} in 0 mM NaCl (Figure 3.7A) and 150 mM NaCl (Figure 3.7B), which are almost identical (Figure 3.7C). Zwitterionic FC12 micelle size and shape changes are expected (Figure 3.6D) with changes in ionic strength due to an increase in ionic strength decreases repulsion between ionic head groups by screening of the anionic charge. This generally results in an increase of the overall aggregation number and size of the micelle with an increase in ionic strength. (*23*) However, because the trypsin cleaved micelles in both conditions result in similar β-barrel spectra, micelle perturbations are not the underlying factor resulting in line broadening. Similarly, modulation of the salt bridges within the barrel due to ionic strength changes are not contributing to the line broadening. Thus, loop interactions are modulating the line broadening and these interactions can be inter- or intra- molecular interactions.

3.3.5 Relaxation data excludes Opa_{50} oligomerization

To decipher between inter- and intramolecular interactions, relaxation studies were conducted. Considering that the Opa_{50} loop-HSPG interactions are electrostatic based on competitive assays with heparin, the Opa_{50}-detergent complex is predicted to undergo electrostatic interactions with other protein-detergent complexes or other empty charged detergent micelles. Such interactions would magnify the size of the complex resulting in significantly broadened peaks. Table 3.3 shows the estimate of the correlation times obtained from the T1/T2 ratios. A conservative estimate for the rotational correlation time of the micelle complex was 11 ns in 0 mM NaCl, whereas Opa_{50} in 175 mM NaCl was estimated to have a rotational correlation time of 14 ns. Rotational correlation times for the β-barrel in the trypsin-treated condition were similar – 12 ns at 0 mM NaCl, and 16 ns at 175 mM NaCl. The loops in each condition were estimated to have rotational correlation times of 7.0 ns and 6.5 ns in the 0 mM and 175 mM NaCl conditions, respectively.

Because the relaxation data for Opa_{50} in FC12 micelles is similar in both the 0 mM and 175 mM NaCl, as well as when trypsin treated, the overall size of the PDC (and therefore the rotational correlation time) is not changing between the two conditions, and therefore the PDC interacting with either itself or empty detergent micelles is unlikely. The PDC interaction with empty micelles was further addressed by investigating the detergent concentration effects on the $^{15}N,^{1}H$-TROSY spectral quality. The spectra recorded at a detergent micelle:Opa_{50} ratio of 1.2:1, 2:1, and 3:1, 4:1, 5:1, and 10:1 were nearly identical (Appendix Figure A2), indicating intramolecular interactions of the loops with themselves are possible. The line-broadening of the β-barrel cross peaks suggests an overall increase

in the size of the protein-detergent complex. Thus, the interactions leading to line broadening at low ionic strength appear to be due to intramolecular loop interactions, and localized charges might explain this interaction. Future site-directed mutagenesis of these charged residues will potentially define the interacting amino acid residues.

3.3.6 Comparison of detergent and lipid environments

One of the previous Pitfalls discussed in Chapter 1 (Section 1.1.4) addressed the drawbacks of using micelles in biophysical characterization. The primary criticism is that detergent micelle do not offer a native lipid environment, which could potentially perturb the structure of the embedded protein. To examine these results in a more native state, NMR and CD spectroscopy were performed on Opa_{60} in liposomes (CD) and nanodiscs (NMR)

CD spectra taken by D. Fox of Opa_{60} in FC12 and DMPC/DMPG lipid SUVs indicate that the protein structure is approximately 50% random coil and 50% β-strand in both conditions, indicating that the secondary structure is similar in the micelle and lipid conditions. (Fox *et al*., submitted manuscript)

Opa_{60} was also reconstituted into DMPC MSP1D1HΔ5 nanodiscs. The lipid composition and MSP construct were selected, based on the success that Fran Hagn demonstrated for OmpX (as introduced in Section 1.4.3e), which had a similarly sized embedded β-barrel region. The secondary structure similarity in micelles and lipids was supported by NMR spectra; Opa_{60} in nanodiscs with DMPC lipids have β-barrel chemical shifts that are superimposable with Opa_{60} in FC12 micelles (Figure 3.8B). However, there were several barrel resonances corresponding to residues on strands 3, 6, and 8 that are missing in the

nanodisc spectrum. Many of the loop resonances recorded at 10 °C are also superimposable (Figure 3.8A). These data indicate that Opa_{60} in DMPC nanodiscs has an overall similar fold in FC12 micelles and lipid environments. Overall, all residues are broadened compared to Opa_{60} in FC12 micelles, but this is likely due to size difference of nanodiscs compared to the smaller protein-detergent complex.

Table 3.2. The peak count from Figure 3.5. All peaks are present in ionic strength conditions above 125 mM NaCl.

	25 mM NaCl	75 mM NaCl	125 mM NaCl
40 °C (Total: 41)	9 (22%)	36 (88%)	41 (100%)
30 °C (Total: 38)	8 (21%)	22 (58%)	38 (100%)
20 °C (Total:31)	3 (10%)	8 (26%)	31 (100%)

Table 3.3. Correlation times estimated from T1/T2 ratios.

	Opa_{50} in FC12 micelles τ (ns)	
	0 mM NaCl	**175 mM NaCl**
Overall	7.4	8.5
Average Barrel residue	10.6	13.9
Average Loop residue	7.0	6.5
	0 mM NaCl trypsin treated	**175 mM NaCl trypsin treated**
Overall	-	-
Average Barrel residue	12.1	15.7
Average Loop residue	7.6	3.4

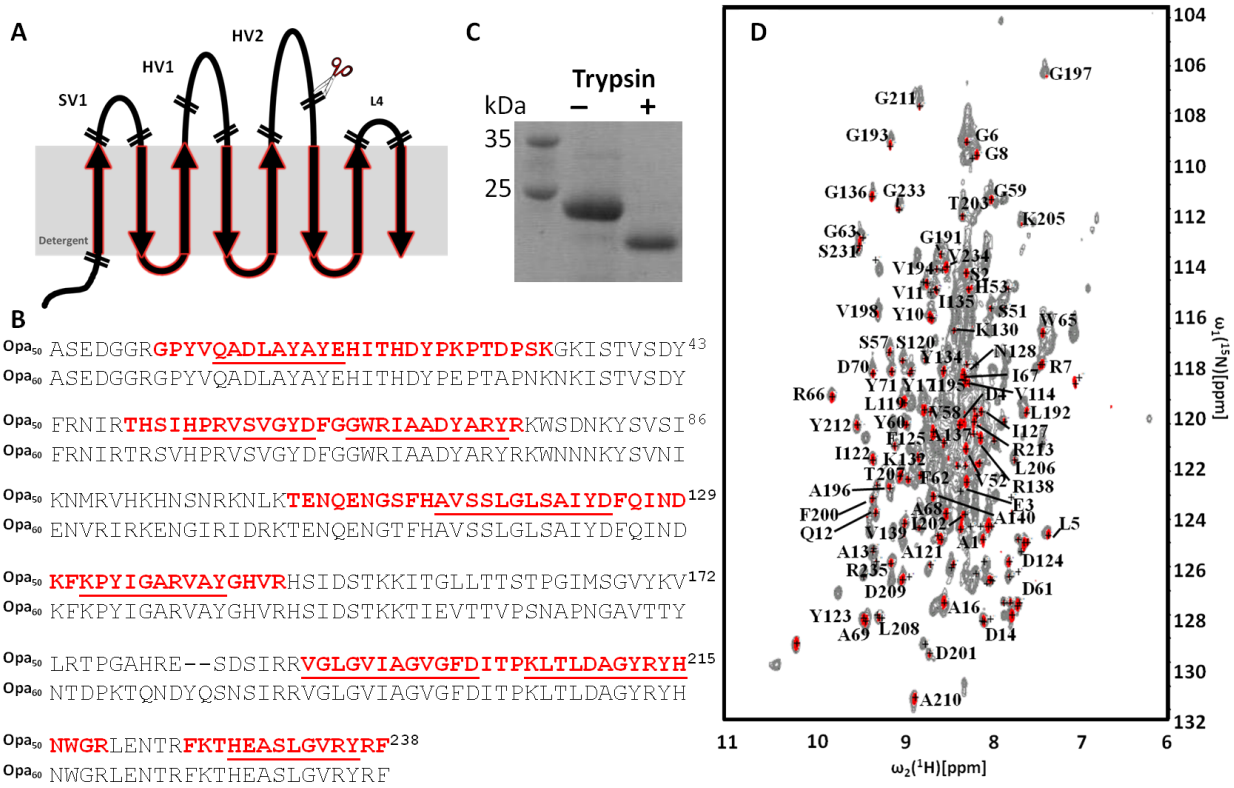


Figure 3.2. Opa$_{50}$ trypsin digestion. (A) A schematic of the Opa$_{50}$ topology with the β-barrel and loops preserved after cleavage outlined in red. (B) The primary amino acid sequence of Opa$_{50}$ and Opa$_{60}$, for comparison. The β-barrel strands are underlined and the expected residues post trypsin cleavage are in red. (C) SDS-PAGE of Opa$_{50}$ trypsin treatment. The lane labels Trypsin (-) and (+) identify the untreated Opa$_{50}$ and treated Opa$_{50}$ bands, respectively. (D) ^{15}N, ^{1}H-HSQC spectra of Opa$_{50}$ (black) and trypsin cleaved Opa$_{50}$ (red). Crosshairs denote the peaks remaining after trypsin digest; 130 peaks are observed from an expected 138 residues based on predicted cleavage patterns. The 800 μM Opa$_{50}$ was in 150 mM FC-12, 20 mM phosphate, 150 mM NaCl buffer at pH 6.2.

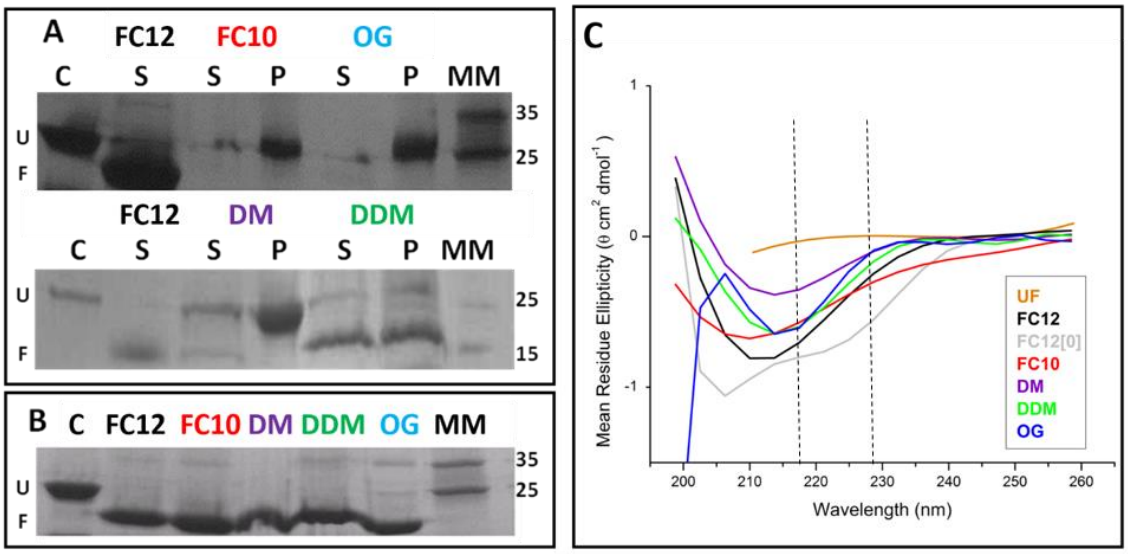


Figure 3.3. Detergents that stabilize fold do not facilitate folding. (A-B) SDS-PAGE of Opa_{50} in different detergents. Folded protein is differentiated by a gel migration shift (U identifies the unfolded protein band, and F indicates the folded protein band). The lane labels refer to C (the control Opa_{50}, unfolded in 8M urea without detergent), Opa_{50} in the detergents mentioned in Table 1: FC12, FC10, DM, DDM, OG, and MM (the molecular marker). (A) The gel shifts of Opa_{50} refolded directly into the denoted detergent. S denotes Opa_{50} in the supernatant of the solution, and P denotes Opa_{50} in the pellet from the precipitate. (B) The gel shifts of Opa_{50} first refolded into FC12, then exchanged into a separate detergent C) CD spectra of Opa_{50} in different detergents. FC12[0] denotes the spectra of Opa_{50} in FC12, with 0 mM NaCl present. Dashed lines indicate the representative minima (218 nm) and positive maxima (228 nm) typically reported for β-barrels.

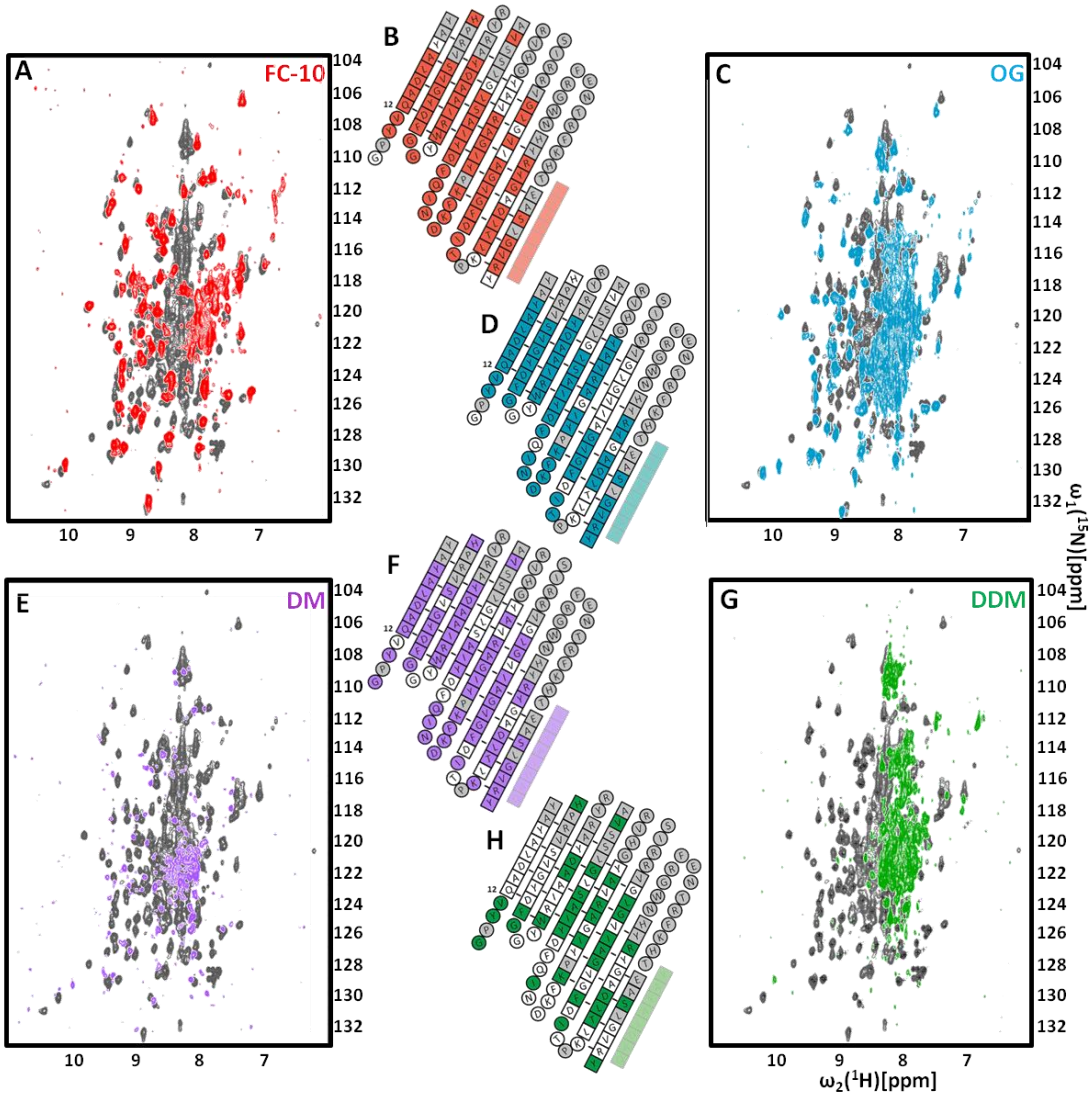


Figure 3.4. NMR spectral quality and structure of Opa_{50} in various detergents. (A-H) ^{15}N, ^{1}H-HSQC spectra of Opa_{50} in various detergents. The residues in the β-barrel that are observed in each are shown on the topology schematic. (A) The overlaid ^{15}N, ^{1}H-HSQC spectra of Opa_{50} in FC12 (gray) and Opa_{50} in FC10 (red) with (B) its corresponding topology map, (C) Opa_{50} in OG(blue) and (D) topology map, (E) DM(purple) and (F) topology map, (G) Opa_{50} in DDM (green) and (H) topology map. The observed β-barrel peaks overlap with Opa_{50} in FC12 allowing 84% of the β-barrel resonances to be assigned in FC10, 44% in DDM, 77% in DM, and 75% in OG.

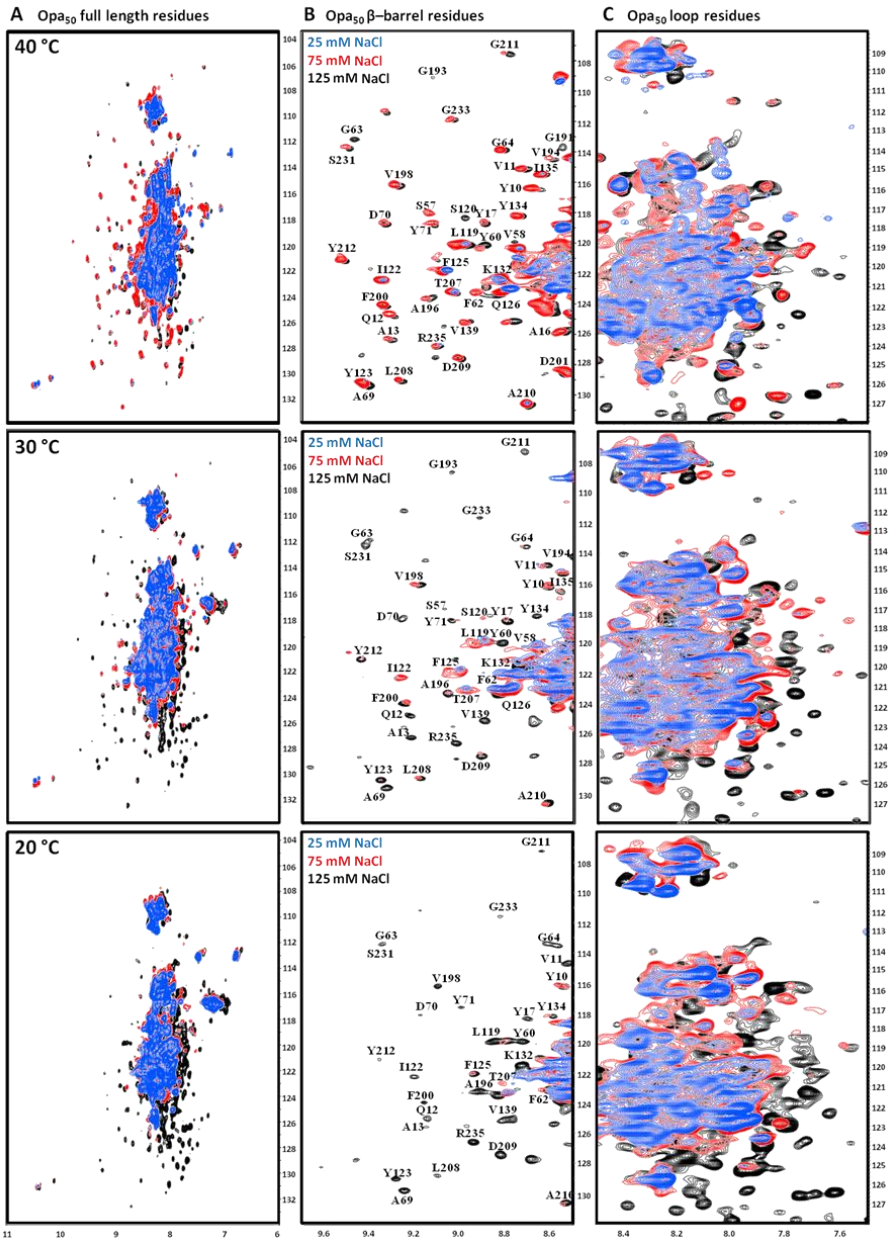


Figure 3.5. **Ionic strength affects Opa$_{50}$ in FC12**. The ^{15}N,^{1}H-HSQC spectra of Opa$_{50}$ in FC12 at 20, 30 and 40 °C, at three different representative ionic strengths - 25 mM NaCl (blue), 75 mM NaCl (red), and 125 mM (black). (A) The full length spectral overlay of Opa$_{50}$ at various ionic strengths and temperatures. (B) The spectral overlay of Opa$_{50}$ at various ionic strengths and temperatures, zoomed into the β-barrel region. (C) The spectral overlay of Opa$_{50}$ at various ionic strengths and temperatures, zoomed in on the loop region.

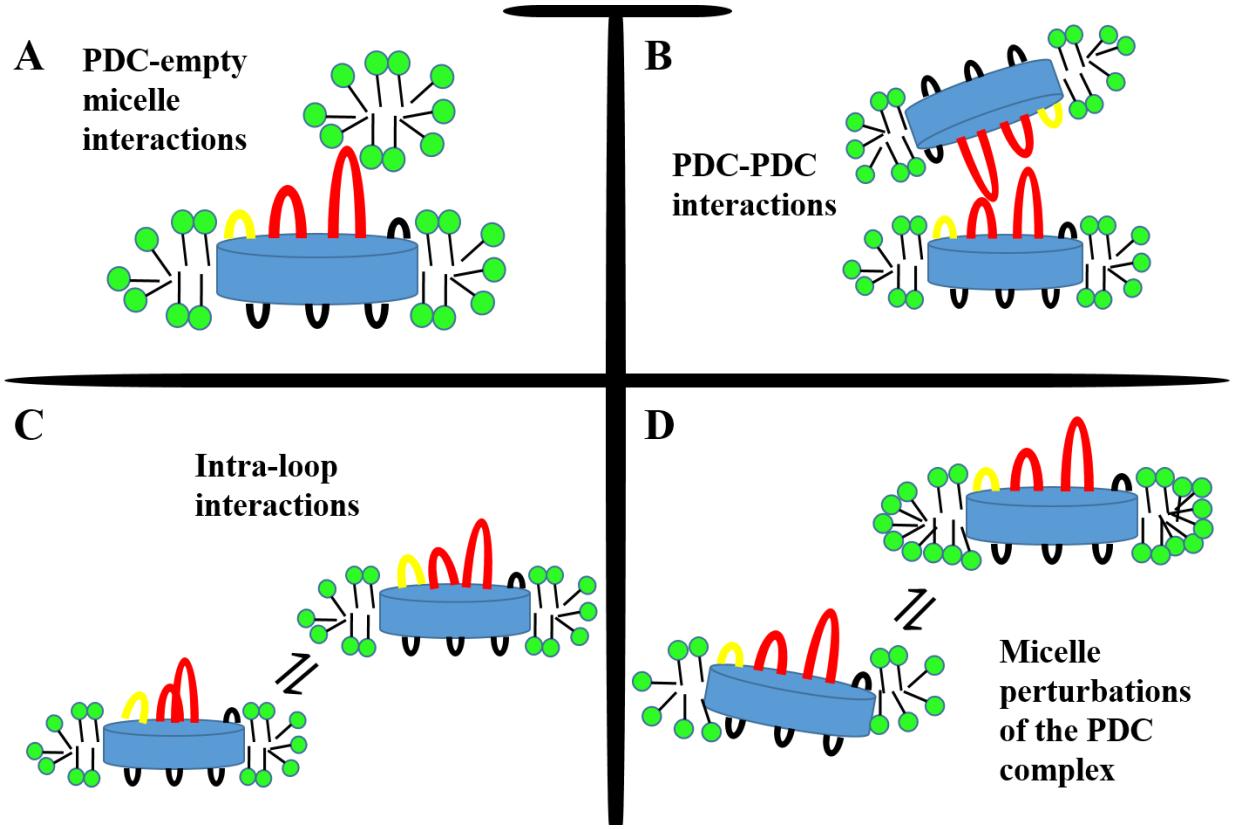


Figure 3.6. Several protein-detergent complex (PDC) scenarios can lead to line broadening. These could potentially be broken down into two categories. One is an overall increase in size of the PDC, which include (A) PDC-empty micelle interactions and (B) PDC-PDC interactions. The second is presence of conformational exchange or heterogeneous populations which include (C) intra-loop interactions, and (D) micelle perturbations of the PDC complex.

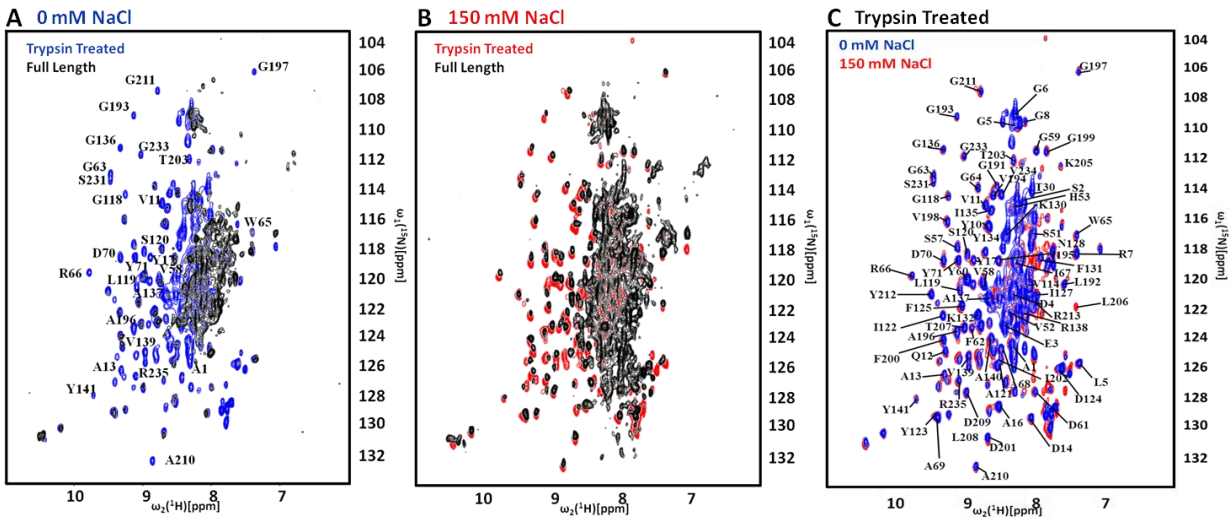


Figure 3.7. **Loop interactions modulate spectral quality.** (A) ^{15}N,^{1}H-HSQC spectra of Opa_{50} (black) and trypsin cleaved Opa_{50} (blue) in 0 mM NaCl. (B) ^{15}N,^{1}H-HSQC spectra of Opa_{50} (black) and trypsin cleaved Opa_{50} (red) in 150 mM NaCl. (C) ^{15}N,^{1}H-HSQC spectra overlay of Opa_{50} in 0 mM NaCl (blue) and Opa_{50} in 150 mM NaCl (red). Beyond the NaCl concentration the buffers were identical - 120 mM FC12 and 20 mM phosphate buffer at pH 6.2 The peaks labeled on each spectrum denote the assigned β-barrel peaks, as obtained by D. Fox for Opa_{60} (submitted manuscript).

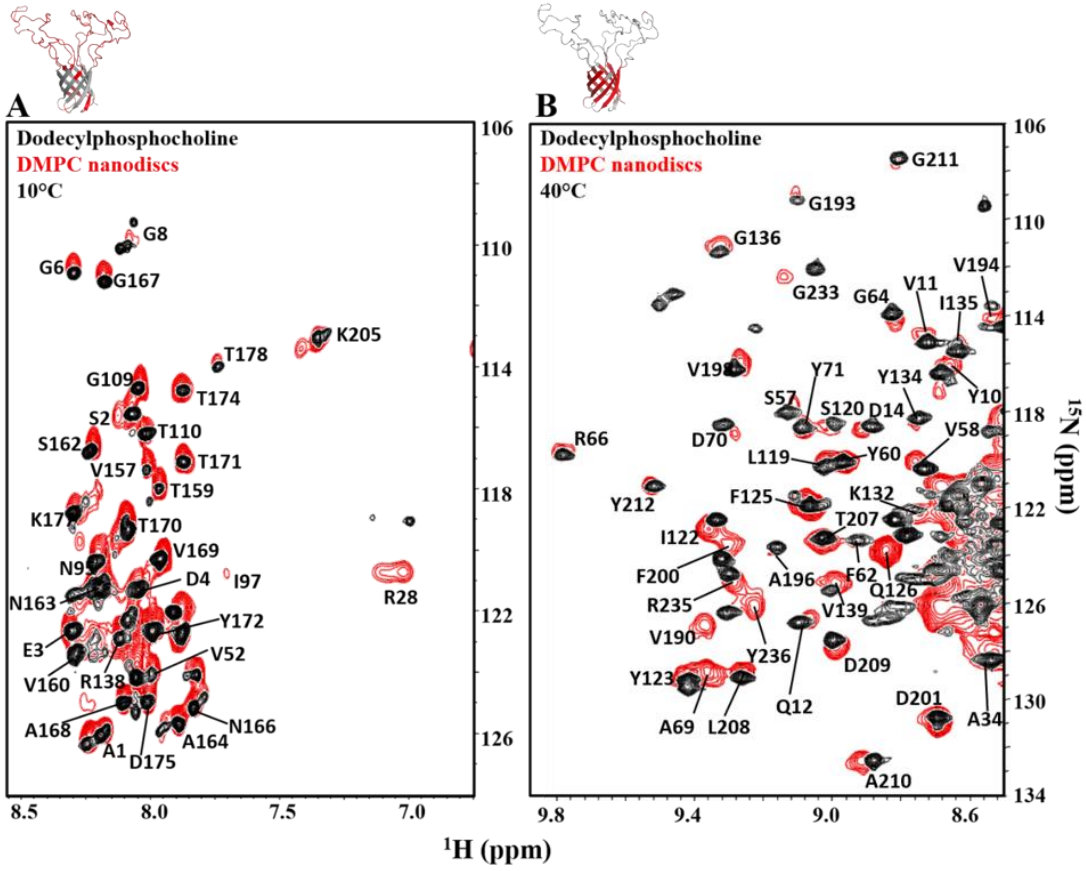


Figure 3.8. NMR ^{15}N, ^{1}H TROSY-HSQC spectral overlay of Opa$_{60}$ in nanodiscs and detergent micelles. Opa$_{60}$ in nanodiscs (in red) containing DMPC lipids with peaks labeled with the NMR backbone assignment and Opa$_{60}$ in FC12 (in black). (A) The spectral overlay of the loop region, recorded at 10 °C. (B) The spectral overlay of the β-barrel region recorded at 40 °C.

3.4 Concluding Remarks

Opa_{50}, an Opa_{HS} outer membrane protein that is predicted to electrostatically interact with its receptor, requires a critical ion concentration for high quality spectral resolution of the β-barrel and loop residues. At ionic strength conditions lower than 75 mM NaCl, over 70% of the assigned β-barrel residues are severely broadened, with significant loop residue broadening occurring as well. Trypsin treatment of Opa_{50} detergent-protein complexes at both 0 mM NaCl and 150 mM NaCl results in nearly identical spectra containing comparable β-barrel residue intensities in each condition, indicating that the Opa_{50} loops are the moieties modulated by ionic strength. Considering that the loop interactions with the receptor are proposed to be electrostatic, the ions present in solution may be screening the charges of the positive loop residues, likely preventing interactions with other protein-detergent complexes in solution. In the absence of ions in solution, these loop interactions with other protein-detergent complexes result in larger aggregates, resulting in slow overall tumbling and consequently short transverse relaxation times which accounts for the severe line broadening experienced at low ionic strength. Protein-detergent interactions of this order are important to consider in NMR structural determination, but are also critical in examining protein-ligand interactions.

Only seven integral transmembrane β-barrel protein structures have been determined in buffer conditions ranging from no salt to 100 mM NaCl.(*1*, *2*, *24-27*) Our study demonstrates the necessity of buffer condition optimization, particularly for membrane proteins that are known to have solvent exposed electrostatic interactions. Overall, the

results of this study provide new insight into the role of ionic strength on solvent-exposed moieties, while offering new evidence for the effects of charge on Opa_{50} extracellular loops.

3.5 Acknowledgements

Daniel Fox for obtaining the assignment of the Opa_{50} barrel, as well as general NMR tutelage.

Jeffrey Ellena for his technical support with the NMR spectrometers.

Franz Hagn and Binyong Liang for technical assistance in expressing, purifying, and reconstituting MSP1D1 nanodiscs.

3.6 References

1. Fernández, C., Hilty, C., Wider, G., Güntert, P., and Wüthrich, K. (2004) NMR Structure of the Integral Membrane Protein OmpX, *Journal of Molecular Biology 336*, 1211-1221.
2. Hiller, S., Garces, R. G., Malia, T. J., Orekhov, V. Y., Colombini, M., and Wagner, G. (2008) Solution Structure of the Integral Human Membrane Protein VDAC-1 in Detergent Micelles, *Science 321*, 1206-1210.
3. Sadarangani, M., Pollard, A. J., and Gray-Owen, S. D. (2011) Opa proteins and CEACAMs: pathways of immune engagement for pathogenic Neisseria, *FEMS Microbiology Reviews 35*, 498-514.
4. Belland, R. J., Chen, T., Swanson, J., and Fischer, S. H. (1992) Human neutrophil response to recombinant neisserial Opa proteins, *Molecular Microbiology 6*, 1729-1737.
5. van Putten, J., and Paul, S. (1995) Binding of syndecan-like cell surface proteoglycan receptors is required for Neisseria gonorrhoeae entry into human mucosal cells., *The EMBO Journal 14*, 2144-2154.
6. Tamm, L. K., Hong, H., and Liang, B. (2004) Folding and assembly of β-barrel membrane proteins, *Biochimica et Biophysica Acta (BBA) - Biomembranes 1666*, 250-263.
7. O'Malley, M. A., Helgeson, M. E., Wagner, N. J., and Robinson, A. S. Toward rational design of protein detergent complexes: determinants of mixed micelles that

are critical for the in vitro stabilization of a G-protein coupled receptor, *Biophys J 101*, 1938-1948.

8. Klammt, C., Schwarz, D., Fendler, K., Haase, W., Dotsch, V., and Bernhard, F. (2005) Evaluation of detergents for the soluble expression of alpha-helical and beta-barrel-type integral membrane proteins by a preparative scale individual cell-free expression system, *FEBS J 272*, 6024-6038.
9. Klammt, C., Schwarz, D., Eifler, N., Engel, A., Piehler, J., Haase, W., Hahn, S., Dotsch, V., and Bernhard, F. (2007) Cell-free production of G protein-coupled receptors for functional and structural studies, *J Struct Biol 158*, 482-493.
10. Galka, J. J., Baturin, S. J., Manley, D. M., Kehler, A. J., and O'Neil, J. D. (2008) Stability of the glycerol facilitator in detergent solutions, *Biochemistry 47*, 3513-3524.
11. Columbus, L., Lipfert, J., Jambunathan, K., Fox, D. A., Sim, A. Y. L., Doniach, S., and Lesley, S. A. (2009) Mixing and Matching Detergents for Membrane Protein NMR Structure Determination, *Journal of the American Chemical Society 131*, 7320-7326.
12. Jimenez, R. H., Freed, D. M., and Cafiso, D. S. Lipid and membrane mimetic environments modulate spin label side chain configuration in the outer membrane protein A, *J Phys Chem B 115*, 14822-14830.
13. Lorber, B., Bishop, J. B., and DeLucas, L. J. (1990) Purification of octyl β-d-glucopyranoside and re-estimation of its micellar size, *Biochimica et Biophysica Acta (BBA) - Biomembranes 1023*, 254-265.

14. Alpes, H., Apell, H. J., Knoll, G., Plattner, H., and Riek, R. (1988) Reconstitution of Na+/K+-ATPase into phosphatidylcholine vesicles by dialysis of nonionic alkyl maltoside detergents, *Biochimica et Biophysica Acta (BBA) - Biomembranes 946*, 379-388.

15. VanAken, T., Foxall-VanAken, S., Castleman, S., and Ferguson-Miller, S. (1986) Alkyl glycoside detergents: synthesis and applications to the study of membrane proteins, *Methods Enzymol 125*, 27-35.

16. Lipfert, J., Columbus, L., Chu, V. B., Lesley, S. A., and Doniach, S. (2007) Size and Shape of Detergent Micelles Determined by Small-Angle X-ray Scattering, *The Journal of Physical Chemistry B 111*, 12427-12438.

17. Surrey, T., and Jahnig, F. (1992) Refolding and oriented insertion of a membrane protein into a lipid bilayer., *PNAS 89*, 7457-7461.

18. Dewald, A. H., Hodges, J. C., and Columbus, L. (2011) Physical determinants of beta-barrel membrane protein folding in lipid vesicles, *Biophys J 100*, 2131-2140.

19. Ritchie, T. K., Grinkova, Y. V., Bayburt, T. H., Denisov, I. G., Zolnerciks, J. K., Atkins, W. M., and Sligar, S. G. (2009) Chapter 11 Reconstitution of Membrane Proteins in Phospholipid Bilayer Nanodiscs, In *Methods in Enzymology* (Nejat, D., Ed.), pp 211-231, Academic Press.

20. Hagn, F., Etzkorn, M., Raschle, T., and Wagner, G. (2013) Optimized Phospholipid Bilayer Nanodiscs Facilitate High-Resolution Structure Determination of Membrane Proteins, *Journal of the American Chemical Society 135*, 1919-1925.

21. Malorny, B., Morelli, G., Kusecek, B., Kolberg, J., and Achtman, M. (1998) Sequence diversity, predicted two-dimensional protein structure, and epitope mapping of neisserial Opa proteins, *J Bacteriol 180*, 1323-1330.

22. Fox, D. A., and Columbus, L. (2013) Solution NMR resonance assignment strategies for β-barrel membrane proteins, *Protein Science 22*, 1133-1140.

23. Tanford, C. (1973) *The Hydrophobic Effect: Formation of Micelles and Biological Membranes*, 1st ed., John Wiley & Sons, Inc

24. Arora, A., Abildgaard, F., Bushweller, J. H., and Tamm, L. K. (2001) Structure of outer membrane protein A transmembrane domain by NMR spectroscopy, *Nat Struct Mol Biol 8*, 334-338.

25. Liang, B., and Tamm, L. K. (2007) Structure of outer membrane protein G by solution NMR spectroscopy, *Proceedings of the National Academy of Sciences 104*, 16140-16145.

26. Edrington, T. C., Kintz, E., Goldberg, J. B., and Tamm, L. K. (2011) Structural Basis for the Interaction of Lipopolysaccharide with Outer Membrane Protein H (OprH) from Pseudomonas aeruginosa, *Journal of Biological Chemistry 286*, 39211-39223.

27. Hwang, P. M., Choy, W.-Y., Lo, E. I., Chen, L., Forman-Kay, J. D., Raetz, C. R. H., Privé, G. G., Bishop, R. E., and Kay, L. E. (2002) Solution structure and dynamics of the outer membrane enzyme PagP by NMR, *Proceedings of the National Academy of Sciences 99*, 13560-13565.

Chapter 4: Assessing the binding of Opa_{50} and Opa_{60} with cognate receptors

Neisserial Opa proteins induce engulfment of the bacteria by binding to specific host cell receptors. Opa proteins are classified into two families based on receptor selectivity. The majority of Opa proteins are Opa_{CEA}, which bind to carcinoembryonic antigen-related cellular adhesion molecules (CEACAMs). Less abundant are Opa_{HS}, which bind to heparin sulfate proteoglycan (HSPG) receptors, or to integrin receptors through an HSPG-mediated interaction with vitronectin or fibronectin. Opa receptor specificity is largely determined by two hypervariable regions in the 2^{nd} and 3^{rd} extracellular loops. There is large sequence diversity in these hypervariable regions, and sequence alignment does not readily identify a binding motif.

To understand the thermodynamic driving forces as well as the affinity of the interaction, binding of two Opa protein variants from *N. gonorrhoeae* MS11, Opa_{50} (an Opa_{HS}) and Opa_{60} (an Opa_{CEA}) with their cognate host receptors was investigated. Isothermal titration calorimetry (ITC) as well as NMR titration experiments were the two primary methods used in these experiments.

4.1 Introduction

4.1.1 Opa proteins interact with two receptor families

Opa proteins are eight-stranded β-barrels with four extracellular loops that dictate receptor specificity. A large sequence diversity is observed in certain regions of extracellular loops of all Opa variants (Table 1). There are three variable regions within the extracellular loops: the semi-variable region (SV) on extracellular loop 1, and the two hypervariable regions (HV1 and HV2) on extracellular loops 2 and 3. These regions have been shown to engage host receptors to induce engulfment of the bacteria, as well as to help the bacteria evade host immune responses.(*1*) The regions are variable in length, and do not constitute the entire length of the extracellular loop; SV is 3-10 amino acids, HV1 is 24-31 amino acids, and HV2 is 45-51 amino acids long. The HV loop regions have high sequence diversity, and have been shown to be necessary for receptor specificity to both HSPG and CEACAM receptors.

4.1.2 Heparin sulfate proteoglycan interactions

Heparin sulfate proteoglycans are cell surface receptors that are expressed to the plasma membrane of most mammalian cell types. HSPGs are comprised of a core protein (syndecan or glypican) embedded within the cell membrane, and are characterized by the attached glycoasaminoglycan sidechains composed of negatively charged heparin sulfates and chondroitin sulfates (Figure 4.1). The glycosaminoglycan sidechains are structurally heterogeneous and vary in length, charge, and monosaccharide composition, which contributes to the diversity of function and binding specificity of HSPGs.(*2*) HSPGs also contain two separate domains: a highly sulfated region of the chain forming blocks of 6-10

disaccharides that resemble heparin in structure, and 16-20 disaccharides that are largely unmodified. The majority of HSPG interactions with a variety of soluble and insoluble extracellular ligands are electrostatic in nature, though several proteins have been found to bind to the transmembrane domain.(*3*) These electrostatic interactions occur with the heparin sulfate chains, are a relatively high affinity interaction (K_d of 1-100 nM), and are generally resistant to physiological salt concentrations.(*2, 4*) Interacting proteins bind to HS via clusters of basic residues (K,R, and sometimes H) on their surfaces, mostly via protein domains.(*3*) Two HSPG-binding consensus sequences have been proposed: [X-B-B-X-B-X] or [X-B-B-B-X-X-B-X] where B is a basic amino acid, and X is a neutral or hydrophobic residue.(*5*) Opa_{HS} proteins primarily bind to HSPG syndecan 1 and syndecan 4.(*6, 7*) Bos *et al*. have indicated that a [B-X-B-B] sequence leads to observed HSPG recognition in Opa_{HS}, and that three Opa variants had this sequence in the HV1 region, while one Opa variant contained this sequence in the HV2 region.(*8*) Surprisingly, Bos *et al*. found that two chimeric Opa variants demonstrated HSPG binding activity while the Opa_{CEA} parents did not bind HSPG, providing support that potentially one HV loop or a combination of the HV loops is responsible for forming the HSPG binding domain.(*8*) Bos *et al*. also report (in unpublished data) that the HV1 region of OpaA (Opa_{50}) contains the sequence RVHK, and mutation of the arginine residue in this sequence abolishes HSPG binding by Opa_{50}.(*5*)

An intracellular signaling cascade is initiated upon Opa binding to HSPGs, which induces cytoskeletal rearrangement within the host cell, restructuring the membrane and engulfing the bacteria. The association between HSPG-stimulated intracellular signaling and microbial uptake is poorly understood, but is believed to occur via one of two potential

cell-dependent pathways. One mechanism focuses on the downstream activation of lipid modifying enzymes phosphatidylcholine-dependent phospholipase C (PC-PLC) and acidic sphingomyelinase (ASM) for the generation of secondary messengers diacylglycerol (DAG) and ceramide.(*9*) The second requires the presence of extracellular matrix proteins vitronectin and fibronectin, which are used as bridges for Opa_{HS} to indirectly engage human integrins, inducing integrin-mediated engulfment.(*9*)

Specific binding of Opa_{HS} expressing *Neisseria* to HSPG receptors is competitively inhibited by heparin and heparan sulfate. Heparin was found to be a more effective inhibitor, as it contains more negative charges in the form of sulfate ester groups and iduronic acid residues than heparin sulfate.(*7*) However, van Putten *et al*. found that two galactosaminoglycans (chondroitin sulfate and dermatan sulfate) did not prevent binding of gonococcal cells to host cell HSPGs, suggesting that geometry of the charged receptor moieties also plays a role in the electrostatic interactions between the charged molecules. (*7*)

4.1.3 CEACAMs

Cellular proliferation, migration, and differentiation are critically important processes for the development of all organisms. These cellular processes are modulated by interactions between cells with other cells, as well as the cell's microenvironment.(*10*) Cellular adhesion molecules (CAMs) facilitate these cell-cell interactions, and are comprised of cadherins, integrins, selectins, and the immunoglobulin superfamily (IgSF). The defining characteristic of the IgSF members is the presence of one or more immunoglobulin- (Ig-) like domains, which have a characteristic structure of two opposing antiparallel β-sheets

stabilized by a disulfide bridge (Figure 4.2).(*11*) The majority of IgSF molecules are transmembrane proteins that consist of an extracellular domain (which is generally composed of one or more Ig-like domains), a single transmembrane domain, and a cytoplasmic tail.(*12*) The N-terminal Ig-like domain commonly binds to other Ig-like domains of the same structure on an opposing cell surface (homophilic adhesion) but can also bind to integrins and carbohydrates (heterophilic adhesion).(*13*) The C-terminal intracellular domain often interacts with cytoskeletal or adaptor proteins, which can lead to cellular signaling.(*10*)

The carcinoembryonic antigen (CEA) family is a member of the IgSF involved in homophilic or heterophilic interactions with closely related molecules.(*14, 15*) These proteins are differentially expressed on the surfaces of leukocytic, epithelial, and endothelial cells, and have been implicated in a variety of physiological processes including cell proliferation, apoptosis, insulin regulation, and angiogenesis.(*14*) There are seven well-characterized CEACAM family members – CEACAM1, and CEACAM3-8. Of these, Opa proteins only bind to CEACAM1 (which is found in a variety of tissues, including epithelial cells),(*14*) CEACAM3 (localized to human neutrophil granulocyte cell surfaces),(*14, 16*) CEACAM5 (localized to epithelial cells),(*16*) and/or CEACAM6 (localized to organ epithelial tissue as well as neutrophils).(*17*) Binding of Opa proteins to the N-terminal Ig domain on the surface of cells induces bacterial engulfment, and is highly specific.(*16*) The structure of the N-terminal domain of CEACAM1 has been determined, and the residues that are critical for Opa binding are found on the non-glycosylated face (Figure 5.2).(*18*) The conserved residues Y34 and I91 were shown to be involved in binding of all Opa variants, and are conserved on both CEACAMs 1, 3,5 and 6. (*19*)

Bos *et al*. investigated the binding of both mutant and chimeric Opa proteins in *N. gonorrhoeae* MS11 with alterations in one or more of the extracellular loops.(*8*) While the SV1 region of the first extracellular loop was not critical for binding and only served to enhance binding of CEACAM1, deletion or mutation of either HV1 or HV2 resulted in a loss of receptor recognition.(*8*) Opa chimeras that included HV1 from one Opa_{CEA} and HV2 from another Opa_{CEA} did not bind CEACAM.(*8*) The results of the chimera studies, as well as with Opa deletion mutants, suggested that both HV loops are critical for CEACAM receptor binding.

Recent structure and dynamics data from D. Fox of the Columbus laboratory support the requisite HV1-HV2 interaction. Fox *et al*. indicate that Opa_{60} loops are intrinsically disordered, but sample a restricted volume such that frequent transient interactions occur between loops on the nanosecond time scale. (submitted manuscript) Fox *et al*. further suggest that the loops adopt an intermediate pre-molten globule state that may be a mechanism for the loops to remain disordered and tolerant to sequence variation, but still provide conformers that are able to interact with CEACAM.

Table 4.1. The multiple sequence alignment of hypervariable regions of the extracellular loops for several Opa protein variants. Opa_{60} is the first entry, presented with the full hypervariable sequence, and several other Opa proteins are aligned. Opa_{50} is the final entry, in white. The BxBB or BBxB sequences of the Opa_{HS} (Opa_{50}: RVHK, OpaK: HKSR) are boxed in red, as suggested by Bos *et al.(8)* The **.** indicates conservation of identical residues, **+** indicates conversation of charge, and **–** indicates a gap. The color gradient denotes receptor specificity, with CEACAM (shades of blue), HSPG (red), or no known interactions (green).

HV1 sequence	HV2 Sequence	CEACAM				HSPG
		1	3	5	6	
NNNKYSVNIENVRIRKENGI-RID------R	RHSIDSTKKT-IEVTTVPSNAPNGAVTTYNTDPKTQNDYQSNSI	√	√	√	√	X
.D.....TK..QVN.S..N-.Q.------L	+.QVH.VESKTTI..SK.TKGATQPGKLVSGPTPKPAYHE....	√	√	√	√	X
...Y....TKE..RNNNA.N-WR+------L	+.QVH.V..ETTTTFLA.TGDAKVPGKIVEGPFSKPAYHE.H..	√	X	√	√	X
H.......IKELE-..N.KTFGGNQLNIKYQ	.........ITGTL.AY..DADAAVTVYPDGH.QKNTYQK...S	√	X	√	X	X
KESNS.T--KK.TEEIN.N--YK+------T	+.QVH.VETKTTTI.SK.K.GSPQGGPIIQ...SKPPYHE.H..	√	X	X	X	X
........TK..QKNDNGNR--Q.------L	..........TNV..VAGAANTAPT--I.YAPETQNAYHE.H..	X	X	X	X	X
..S....S.KELGRNDNSTS-NSSHLNIKTQ	+.QVR.VESETTT...H--.GAP----VPQGPTPKPAYHK.R..	X	X	X	X	√
SD.....S.K.M.VH.H.SN-.KN------L	.........ITGLL..ST-PGIMSG--V.KVLRTPGAHRE.D..	X	X	X	X	√

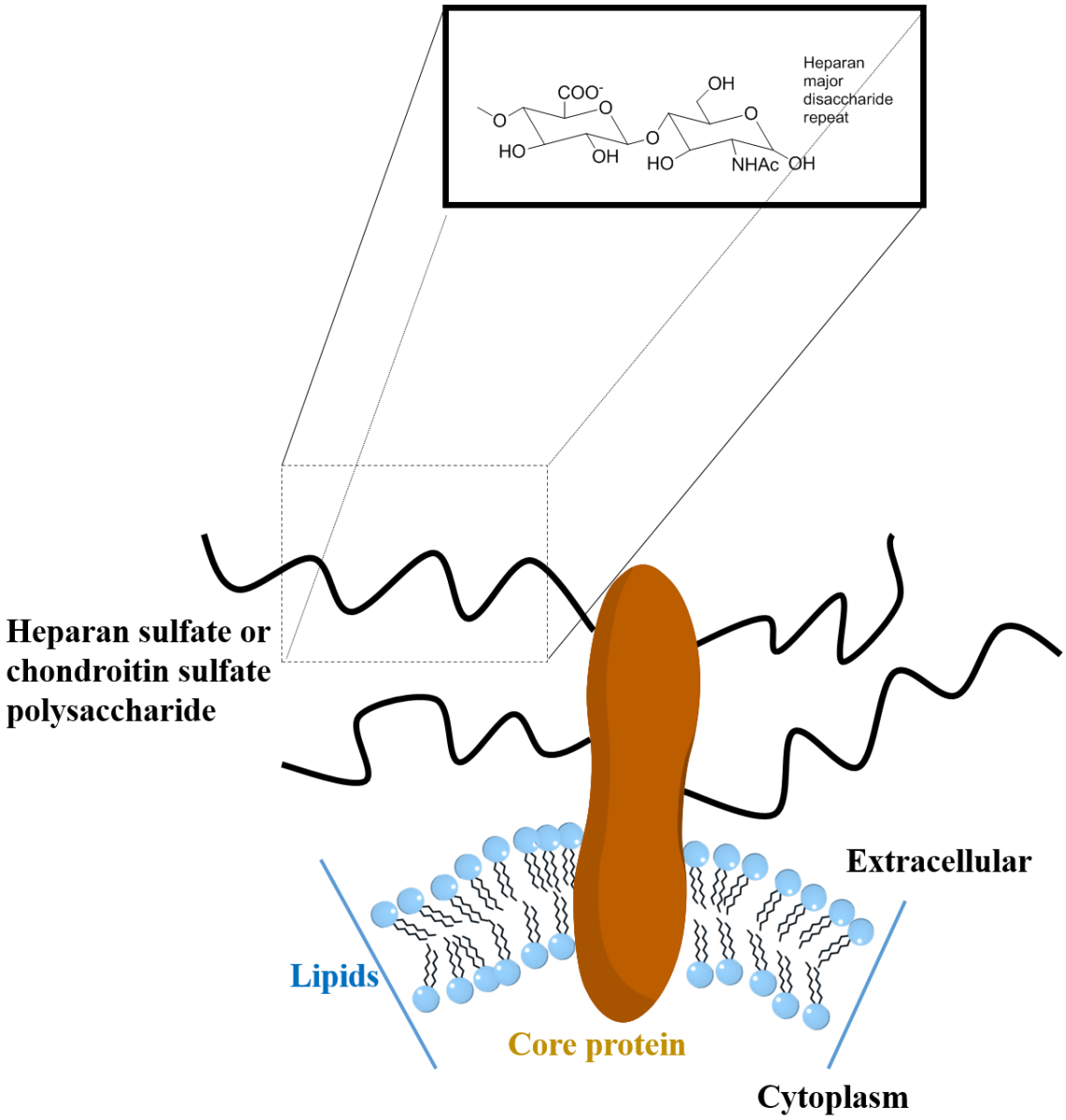


Figure 4.1. A simplified figure for the structure of heparin sulfate proteoglycans. The core protein (orange) is embedded within the lipid bilayer (blue), and has attached heparin sulfate or chondroitin sulfate side chains presented on the extracellular side. The major heparin disaccharide repeat is represented in the magnified region.

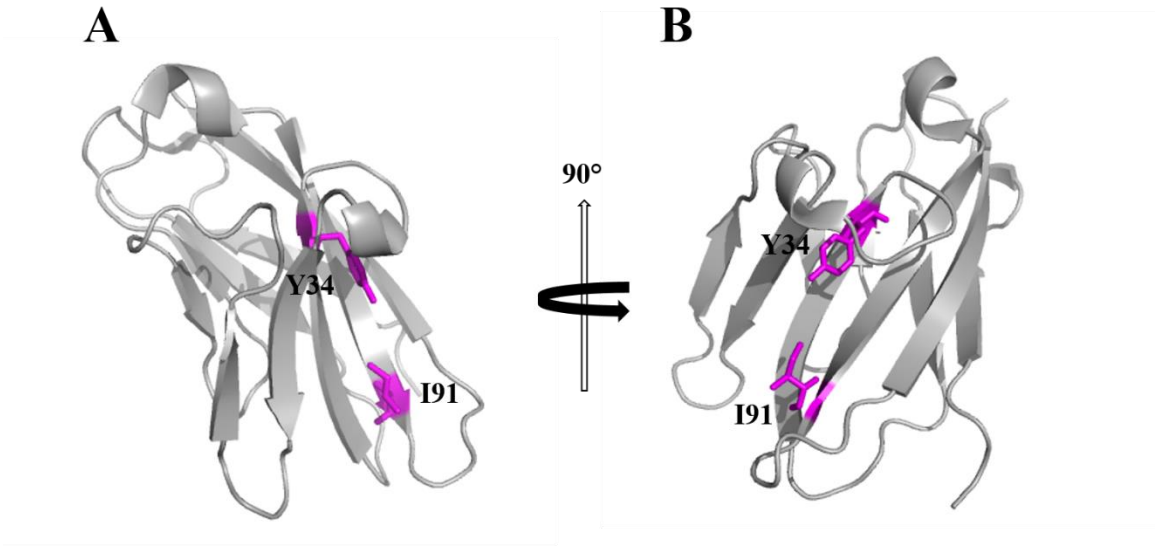


Figure 4.2. The crystal structure determined for the N-terminal D1 domain of human CEACAM1 (PDBid:2gk2). The (A) top and (B) side views of the D1 domain of human NCEACAM1. Critical residues for binding (Y34 and I91) are highlighted in magenta.

4.2 Materials and Methods

4.2.1 Preparation of Opa samples for binding studies

Opa_{50} and Opa_{60} samples were prepared as described in Section 3.2.2 for Opa-FC12 protein detergent complexes, and as described in Section 3.2.7 for Opa-DMPC nanodiscs.

4.2.2 Preparation of cognate host receptors

The N-terminal domain of human CEACAM1 (Addgene) was expressed and purified according to established protocols.(*18*) *E. coli* MC1061 cells transformed with a pGEX-2V plasmid encoding the N-terminal D1 domain of human CEACAM1 [amino acids 1-107 of the mature protein referred to as NCEACAM1; pdb 2gk2, r2gk2sf] were provided by Alena Fedarovich (Medical University of South Carolina). A linker region between GST and NCEACAM1 was designed to incorporate a TEV cleavage site (ENLYFQ) in the resulting fusion protein. A second uncleavable construct was also used and lacked the TEV cleavage site.

MC1061 *E. coli* cells with CEACAM plasmid were grown in LB media containing ampicillin (100 µg/mL) and streptomycin (50 µg/mL) at 37 °C until the OD_{600} reached ~0.6. Cell cultures were cooled to 25 °C and protein expression was induced with 1 mM IPTG overnight with shaking (250 rpm, 25 °C). Cells were harvested by centrifugation (5,000 x *g*, 20 min, 4°C) and the cell pellet was frozen. The frozen cell pellet was resuspended in lysis buffer (20 mM Tris, 150 mM NaCl, 2 mM EDTA, 2 mM DTT, 10% glycerol, with protease inhibitor tablet [Roche] at pH 8.0) and lysed with several cycle of microfluidization (Microfluidics model 110L, Newton, Mass). The lysate was then clarified by centrifugation (18,000 x *g*, 1 h, 4 °C) and protein was precipitated with the

addition of ammonium sulfate to 55% saturation with constant stirring (1 h, 4 °C). Precipitated protein was harvested using centrifugation (12,000 x *g*, 1h, 4 °C). Pellets were resuspended with 30 mL lysis buffer and added to a glutathione resin gravity feed column previously equilibrated with equilibration buffer (20 mM Tris, 150 mM NaCl, 2 mM DTT, 10% glycerol at pH 7.3). After loading the column ith protein, the column was washed with 10 CV of equilibration buffer, and eluted with 50 mL elution buffer (equilibration buffer + 10 mM reduced glutathione). In certain cases, TEV (~3.5 µM) was added to the purified GST-CEACAM fusion protein and dialyzed into equilibration buffer at pH 8.0 overnight at 4 °C (MWCO = 3,500 kDa). CEACAM was further purified using an HR Sephacrysl s-200 gel-filtration column (26/60 mm, GE Healthcare) previously equilibrated with 20 mM Tris, 500 mM NaCl, 2 mM DTT, 10% glycerol at pH 8.0. Fractions containing pure CEACAM (as assessed via SDS-PAGE) were pooled and concentrated using an Amicon centrifugal filter unit of 3.5 kDa MWCO (Millipore) and used immediately.

Heterogeneous sodium heparin was from porcine intestinal mucosa (Sigma) and had a molecular weight ranging from 3 – 30 kDa. For calculations, the midpoint MW of 13.5 kDa was used.

4.2.3 ITC titration assays for Opa_{60}-receptor binding

Purified Opa_{60} was reconstituted into MSP1D1ΔH5 DMPC nanodiscs as described in Section 3.2.7. Final concentrations of the Opa_{60} NMR sample ranged from ~100 – 250 µM. NCEACAM1 stock solutions were prepared as previously described, and both Opa_{60} and NCEACAM1 stock solutions were dialyzed in 20 mM phosphate, 150 mM NaCl buffer at pH 6.2. Binding experiments with Opa_{60} to CEACAM protein were performed with a VP-

ITC microcalorimeter (V_{cell} = 1.4 mL, $V_{syringe}$ = 290 µL) at 25 °C. For ITC, the optimal sample concentration is given by the equation:

$$C = K_d * [titrant] \quad (4.1)$$

where K_d is an estimated dissociation constant (or actual, if determined by a different method), and C is within the range of 1-1000. The optimal ligand concentration in the ITC syringe is dependent on the binding affinity and macromolecule concentration in the ITC cell, and is estimated by the equation:

$$M_i = 1400 * M * n\,(1.2 + 6/c^{1/2})/v_i \quad (4.2)$$

where M_i is concentration of ligand in syringe (mM), v_i is injection volume, 1400 (µL) is the reaction cell volume, M (mM) is macromolecule concentration in the ITC cell, n is the stoichiometry, and c is the parameter described in equation 4.1. The recorded ITC traces are typically obtained from multiple, baseline-subtracted injections. The area under each peak represents the enthalpy associated with the corresponding injection of ligand.

ITC experiments to monitor the Opa_{50}-heparin binding were not performed for two reasons. The first is that the heterogeneous populations of heparin oligosaccharide might have unforeseen adverse interactions that affect the measured enthalpy. The second reason is that Marissa Kieber, another graduate student in the Columbus laboratory, was working on Opa_{50}-heparin binding in proteoliposomes to investigate if binding could be monitored via ITC.

4.2.4 NMR titration assays for Opa_{50}-receptor binding

Purified $^{2}H,^{15}N$ Opa_{50} or Opa_{60} was reconstituted into MSP1D1ΔH5 d_{54}-DMPC nanodiscs as described in Section 3.2.7, and concentrated to a final volume of 300 – 400 µL supplemented with 10% D_2O in a Shigemi tube. Final concentrations of the Opa NMR sample ranged from ~80 – 200 µM. Heparin and NCEACAM1 stock solutions were prepared as previously described. All experiments were performed on a 600 or 800 MHz Bruker Avance spectrometer with a cryoprobe. Two-dimensional (2D) $^{15}N,^{1}H$-TROSY (transverse relaxation optimized spectroscopy) spectra were recorded for each sample with a data matrix of 64 x 2048 complex points, with an experimental time ranging from ~6-18 h.

4.3 Results and Discussion

4.3.1 NCEACAM1-GST dimerization affects ITC results

The thermodynamics of ligand binding to proteins can be measured with ITC, which quantifies the binding enthalpy (ΔH) and results in a binding isotherm. A fit of the isotherm provides the binding constant (K), which allows the free energy (ΔG) and indirect calculation of the entropy (ΔS) to be determined.(*20*) However, ITC measures all enthalpic contributions in solution which also includes undesired reactions such as precipitation, hydrolysis, redox, and competing equilibria with buffer and protons can compromise the measurement. To avoid other binding events, care must also be taken in matching buffers to avoid solvent interactions.

To measure the thermodynamics of the Opa-CEACAM interaction with ITC, several off target interactions posed setbacks. Use of detergent micelles as a membrane mimic presents one such difficulty, as titration of ligand can cause adverse interactions with free monomer in solution, dilute concentrations of detergent micelles below their CMCs, or result in micelle formation if detergent is in the titrant, which all affect the measured enthalpy. Therefore, the only systems available for ITC are liposomes and lipid nanodiscs. Marissa Kieber, a graduate student in the Columbus laboratory, and Sebastien Ortiz, an undergraduate in the Columbus laboratory are working on using liposomes to collect ITC binding data. I employed the use of DMPC nanodiscs to assess nanodisc viability.

Initial expression and purification of the N-terminal domain of CEACAM1 was successful, though at high enough concentrations, sample stability and solubility was an issue. An N-CEACAM1-GST construct was used for functional studies to reduce dimerization of N-

CEACAM1. Titration of N-CEACAM1-GST into Opa_{60}-MSP1D1ΔH5 DMPC nanodiscs produced a large change in heat (Figure 4.3A).

To investigate the absence of the unbound state, a negative control titrating N-CEACAM1-GST into buffer was performed. This experiment also resulted in a comparable isotherm (Figure 4.3B), indicating that the N-CEACAM1-GST titrant was responsible for the heat change of the interaction. This result was further supported by SDS-PAGE results; a band was present and correlated with the MW of the CEACAM1-GST dimer. To confirm that the band was from CEACAM1, anti-CEACAM Western blots (Appendix Figure A4) demonstrated the presence of a higher-order N-CEACAM1-GST oligomer. Bands were also present between 25-35 kDa that appeared to be GST degradation product, which might have resulted from proteolysis of the GST tag. Thus, it is likely that the enthalpic changes were due to either N-CEACAM1-GST oligomerization, or to N-CEACAM1 cleavage of the GST tag.

Another approach that reduces the GST cleavage contributions to the isotherm was the use of an NCEACAM1 construct lacking the TEV protease cleavage site. However, experiments with this construct also resulted in comparable isotherms, which indicates that the oligomerization of N-CEACAM1-GST is the primary contributing factor for enthalpic change.

Although not attempted, another approach would be to conduct the ITC experiments at 4 °C. Purification conditions of N-CEACAM1 is typically performed at 4 °C to avoid oligomerization, which is observed on SDS-PAGE. ITC experiments should be performed in native conditions, at 37 °C. However, ITC instrumentation typically limits the temperature of the experiment to room temperature. Temperature change can affect the

interactions between proteins and their ligands. These interactions are characterized by the free energy, enthalpy, and entropy changes associated with the binding reaction, which is related by the Gibb's free energy (ΔG) equation:

$$\Delta G = \Delta H - T\Delta S \tag{4.3}$$

where ΔH is the change in enthalpy of the reaction, T is the temperature of the reaction, and ΔS is the change in entropy.

The Gibb's free energy is related to the equilibrium constant *k* by the equation:

$$\Delta G = -RT \ln k \tag{4.4}$$

which combined with Equation 4.3, can be rearranged to give the equation:

$$\ln k = \frac{\Delta H - T\Delta S}{-RT} \tag{4.5}$$

Finally, the observed CEACAM dimer is concentration dependent. The concentration of the stock solution in the ITC syringe for injection may be too high. Thus, ITC may be too challenging to measure the Opa-NCEACAM1 interaction.

4.3.2 Opa proteins in FC12 micelles do not bind to conjugate receptors

As discussed in chapter 2, binding processes can be regarded as an equilibrium condition that can be monitored using NMR spectroscopy. One of the primary means of monitoring protein-ligand interactions is chemical shift perturbation mapping.(*21*) In chemical shift perturbation mapping, the ^{15}N-^{1}H HSQC spectrum of one protein is monitored while the unlabeled ligand is titrated in, and the perturbations of the chemical shift are observed.(*21*) These interactions cause environmental changes at the protein binding interfaces, which affects the chemical shifts of the nuclei in the area. However, in some cases, the entire protein may change conformations and all shifts may be affected, indicating allosteric

processes.(*22*) In general performing titrations with NMR allows for mapping of the protein binding interface, as well as provides a good estimation of the binding affinity and kinetics of binding.

The kinetics of binding determines how chemical shifts change during the titration. If complex dissociation is very fast, only a single set of resonances are observed, whose chemical shift is a fractionally weighted average of the free and bound chemical shifts.(*21*) This regime is referred to as fast chemical exchange, and is usually observed for weaker interactions, in general where K_d>10 μM.(*21*) The average amide chemical shift change can be represented by:

$$\Delta\delta_{avg}=\sqrt{\frac{(\frac{\Delta\delta N}{5})^2+\Delta\delta H^2}{2}} \qquad (4.6)$$

where ΔδN represents the change in the amide nitrogen's chemical shift, and ΔδH represents the change in the amide proton's chemical shift. These average shifts can be tracked to their fully bound state, and the binding constant can be extracted by fitting the fractional shift against a quadratic equation dependent on total protein and ligand concentrations.(*21*)

Two sets of resonances for the bound protein and unbound protein are observed for slow complex dissociation that are in the slow chemical exchange regime. As the titration continues, the unbound protein resonances will disappear and be replaced by the bound protein resonances. The majority of residues will have overlapping resonances, but larger differences will denote the interaction interface. This experiment is significantly more difficult to interpret, as the migration of bound state resonances are not readily identifiable, and may require independent assignment to determine the corresponding free-bound

resonances. The binding constant can still be determined by measuring the intensities of the disappearing or appear peaks as a function of titration.

In the intermediate chemical exchange regime, the frequencies of the changing resonances become poorly defined, and kinetic broadening of resonances occurs. Extensive broadening may result in disappearing resonances in the NMR spectrum. In intermediate exchange, the interaction interface is visible by progressively disappearing resonances. McAlister *et al*. reported on the binding of cell-cell recognition molecules in this time regime, and observed peaks that disappeared before any chemical shift change was observed.(*23*) The general broadening was attributed to the averaging of line widths of the resonances of unbound and bound cell recognition molecules.

Heparin titration experiments with Opa_{50}-FC12 micelles, as well as NCEACAM1 titrations with Opa_{60}-FC12 micelles (by D. Fox) were performed to saturating conditions. Spectra were recorded at 10 °C to obtain higher peak intensities for the hypervariable loop regions furthest from the β-barrel-micelle complex (described in detail in reference (*24*)). However, neither Opa protein-conjugate receptor experiment resulted in an observable change in spectra.

There are multiple scenarios that result in the absence of observable changes. The chosen ligand might not be a conjugate receptor, or some experimental characteristic is perturbing the ligand structure. For the NCEACAM1 interaction, NCEACAM1 could be oligomerizing as previously seen in the ITC experiments. Dimerization of the protein might result in the occlusion of the binding face, preventing any interactions with the Opa_{60} protein. This conclusion is unlikely, as the NMR experiments were recorded at 10 °C. Extremely slow kinetics for the binding interaction may also be affecting the experiment,

though this conclusion is also unlikely due to the Opa-proteoliposome:ligand interactions observed by A. Dewald and J. Martin (Data not shown).

Because binding data has been obtained for the Opa-liposome complex, there are three potential scenarios that may be occurring with the protein-detergent complex (PDC):ligand interaction. The detergent micelles (which are present at ~3:1 free micelle:PDC ratio) and monomers in solution could be interacting with (1) Opa loops,(2) the conjugate receptor, or (3)both. Rather than detergent micelles interacting with Opa loops (as discussed in Chapter 3), the free detergent monomer may be partitioning to the Opa protein's extracellular loops, sterically hindering ligand interactions. Another scenario that would affect Opa_{50}-heparin binding involves ligand-detergent effects; heparin interacts non-specifically with most charged molecules, and the positive choline moiety on the zwitterionic Foscholine-12 monomer or micelle could be electrostatically interacting with the heparin molecule. Also, shown in Chapter 3, the loop dynamics are dependent on detergent and ionic strength. The conditions that are optimal for NMR spectral quality may not be optimal for receptor binding. To investigate the effects of the lipid bilayer mimic, heparin binding experiments were performed with Opa_{50} in DMPC nanodiscs.

4.3.3 Preliminary evidence that Opa_{50} in DMPC nanodiscs binds to heparin.

Other members of the Columbus lab have used liposomes as membrane mimics to obtain preliminary binding data for Opa_{50}. Preliminary data from A. Dewald of the Columbus laboratory indicated that the dissociation constant obtained for pull-down assays with reconstituted Opa_{50} liposomes and fluorescently labeled heparin was approximately 600

nM (unpublished data). Data obtained from J. Martin and M. Kieber of the Columbus laboratory indicated that the dissociation constant obtained for fluorescence polarization assays with reconstituted Opa_{50} liposomes and fluorescently labeled heparin was ~500 nM. Considering the lack of data obtained for Opa:FC12 PDCs, lipids could possibly provide a better environment than detergent micelles for binding studies.

Opa_{50} was reconstituted into DMPC lipid nanodiscs, and heparin binding was examined using NMR spectroscopy. Because the β-barrel is not involved in ligand binding, and as indicated in Chapter 3, colder temperatures result in higher peak intensity in the HV loop regions, spectra were obtained at 10 °C. Initial results indicated that binding of heparin to Opa_{50} results in NMR spectral attenuation (Figure 4.4). The Opa_{50}:nanodisc:heparin complex is potentially in the intermediate chemical exchange regime; only the unbound protein population is observed. As ligand is added, the unbound protein population decreases, and the bound protein population should increase. However ligand binding decreases the mobility of Opa_{50}'s loops which results in the slow exchange, and the bound protein resonances are extremely broadened (Figure 4.5). The attenuation of peaks was fit to the Hill equation for sigmoidal slopes (Figure 4.4B,C, and D) which is given by the equation:

$$\theta = \frac{[L]^n}{K_d + [L]^n} \tag{4.7}$$

where θ is the fraction of occupied sites where ligand can bind to receptor, [L] is the free, unbound ligand concentration, K_d is the apparent dissociation constant, and n is the Hill coefficient which takes into account binding cooperativity. The sigmoidal shape of the binding curve with the linear scale could indicate cooperativity. Based on native HSPG interactions, between 2-6 disaccharides repeats are expected to bind to the Opa_{50} loops.(*25*)

However, the heterogenous nature of the heparin used could result in multiple Opa_{50} proteins binding to larger oligomeric heparin units. The K_d is difficult to estimate from these results alone, due to the extensive broadening from the intermediate chemical exchange, and techniques other than NMR might be better suited for measuring the Opa-receptor affinity.

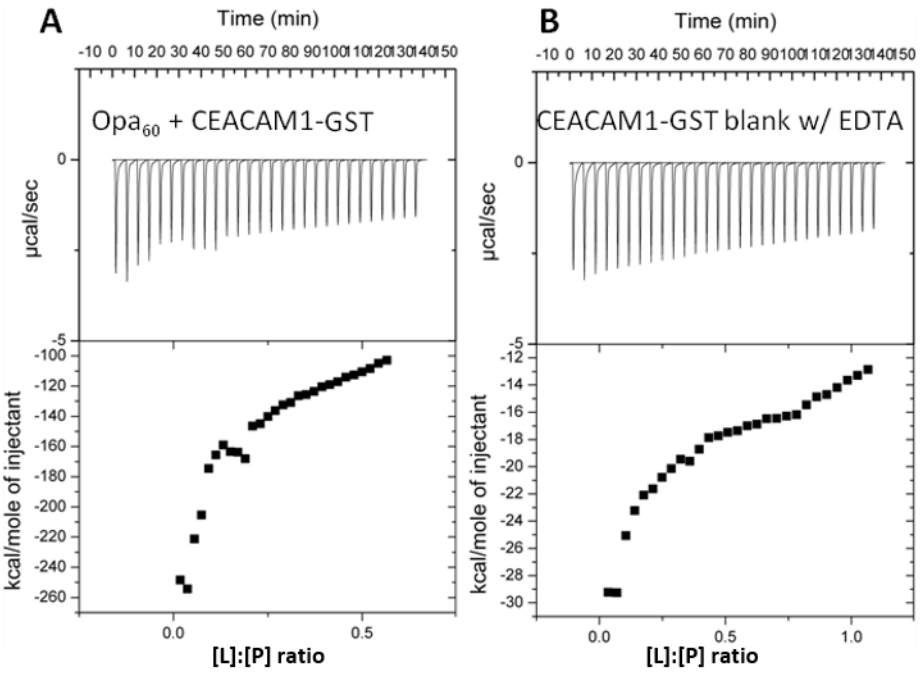


Figure 4.3. N-CEACAM-GST oligomerizes at room temperature. (A) The ITC data binding data for the N-CEACAM1-GST interaction with Opa_{50} in MSPΔH5 nanodiscs. (B) The ITC data binding data for the negative control for the experiment: the titration of N-CEACAM1-GST into buffer.

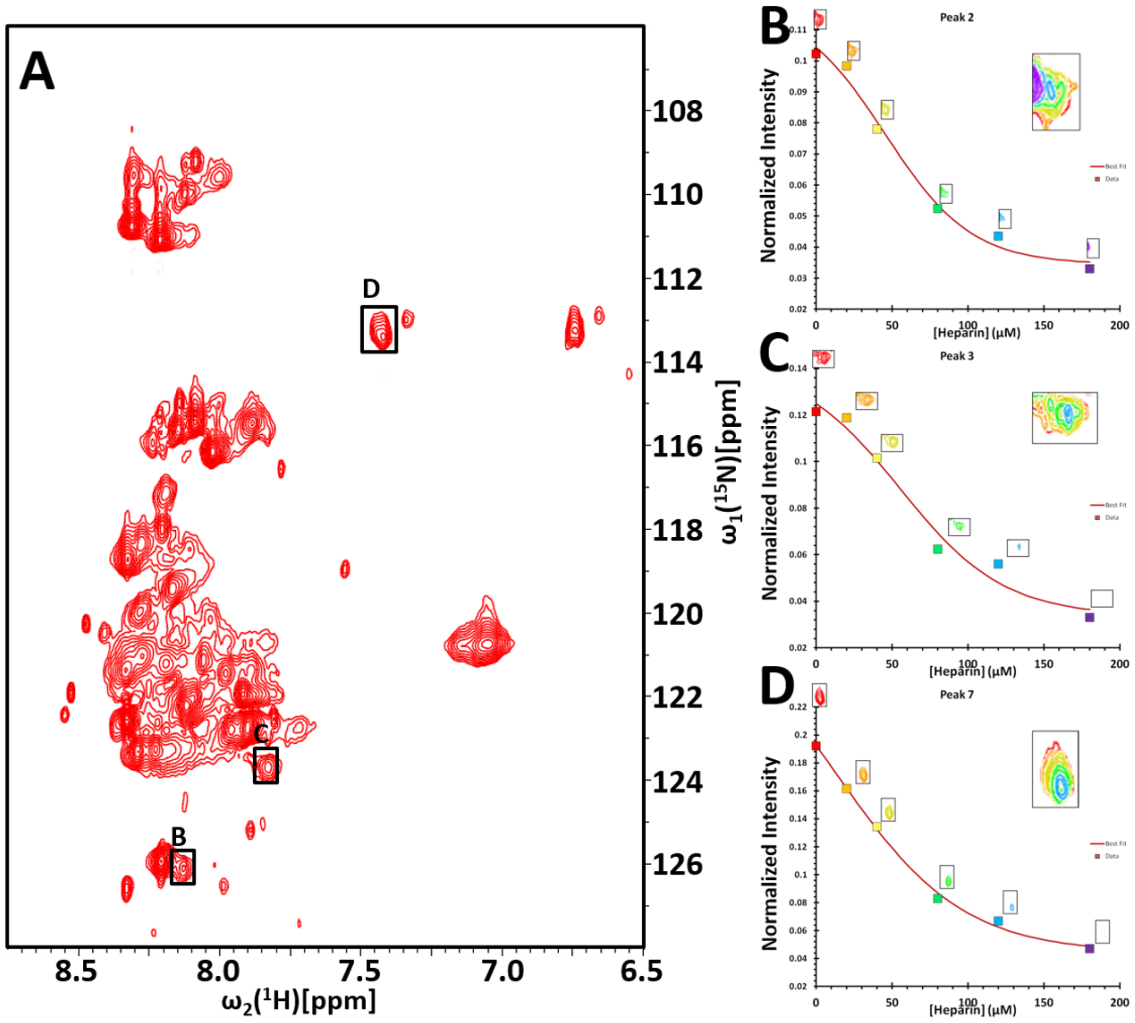


Figure 4.4. Opa$_{50}$–heparin binding interactions result in spectral intensity attenuation due to decrease in mobility. Spectra were recorded at 10 °C. (A) The ^{15}N, ^{1}H-HSQC spectra of the Opa$_{50}$ loop region in MSPΔH5 nanodiscs. (B-D) Plots for the peak intensity attenuation of select loop peaks, upon the addition of heparin.

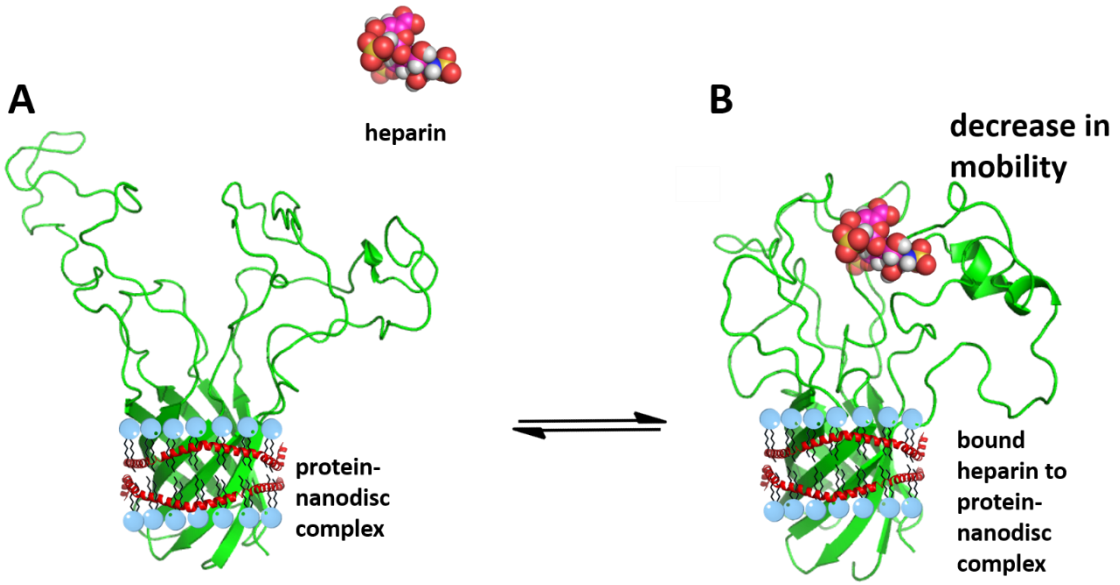


Figure 4.5. A schematic for the proposed interaction of heparin binding to the Opa$_{50}$ loops. A single heparin disaccharide repeat is shown for simplified purposes. (A) In the unbound state, the loops of the protein nanodisc are mobile, but upon binding (B) become immobilized.

4.4 Concluding Remarks

Significant difficulties have been presented in this chapter for binding of ligands to Opa proteins in FC12 detergent micelles, as well as in lipid nanodiscs. Opa protein detergent complexes have thus far been ineffective for obtaining binding data. ITC experiments are not conducive for use with PDCs, due to the free detergent monomer populations present. NMR investigations of both Opa_{50} and Opa_{60}:FC12 complexes have not yielded observable binding data.

The Opa-DMPC lipid nanodisc complex is a more promising system to work with for obtaining binding data. For ITC measurements, Opa_{60}-nanodisc systems are potentially useful for probing binding interactions, though problems resulting from NCEACAM1 dimerization affected the ITC measurements. Opa_{50}-DMPC nanodiscs were observed to interact with heparin via NMR, and the bound complex was predicted to undergo a loss in mobility, resulting in the broadened peaks. A complete assignment of Opa_{50} residues is required to gain a better understanding of the Opa_{50}-heparin complex's dynamics. Based on the difficulty of the Opa_{60} assignment, as well as the difficulties intrinsic to Opa_{50} that were addressed in Chapter 3, the assignment of Opa_{50} is going to be very challenging.

Subsequently, Opa liposomes (Jennifer Martin, Marissa Kieber, Alison Dewald) have provided better systems for measuring Opa-receptor affinities using pulldown assays and fluorescence polarization. For the fluorescence pulldown assays, A. Dewald incubated the Opa protein with the fluorescently labeled conjugate receptor for three days, centrifuged the reaction mixture at 12,000 x g to remove protein aggregate, and centrifuged the supernatant at 150,000 x g to pellet Opa proteoliposomes. Opa proteoliposomes, which pulled down the labeled receptor, reduced the overall fluorescence intensity of the

supernatant (free ligand). The results of the binding assay indicated that the K_d for the Opa50:heparin interaction was ~600 nM.

In the fluorescence polarization assays, the conjugate receptor (heparin or CEACAM) was fluorescently labeled and incubated with varying concentrations of Opa proteoliposomes. The unbound receptor tumbles and rotates rapidly in solution, and excitation by plane polarized light results in a low fluorescence polarization. The bound receptor results in a much larger Opa-receptor complex, which rotates slowly and results in high fluorescence polarization. The K_d was determined to be ~250 nM for the Opa_{60}:CEACAM1 binding interaction, and ~500 nM for the Opa_{50}:heparin binding interaction.

4.5 Acknowledgements

Jennifer Martin, for helpful discussion and advice in preparation of CEACAM1.

Daniel Fox, for helpful advice regarding NMR experiments and analysis.

Franz Hagn (Wagner lab, Harvard University) and Binyong Liang (Tamm lab, UVa) for guidance and discussion for preparation of nanodiscs.

4.6 References

1. Bilek, N., Ison, C. A., and Spratt, B. G. (2009) Relative contributions of recombination and mutation to the diversification of the opa gene repertoire of Neisseria gonorrhoeae, *Journal of Bacteriology 191*, 1878-1890.
2. Bernfield, M., Kokenyesi, R., Kato, M., Hinkes, M. T., Spring, J., Gallo, R. L., and Lose, E. J. (1992) Biology of the Syndecans: A Family of Transmembrane Heparan Sulfate Proteoglycans, *Annual Review of Cell Biology 8*, 365-393.
3. Bernfield, M., Götte, M., Park, P. W., Reizes, O., Fitzgerald, M. L., Lincecum, J., and Zako, M. (1999) FUNCTIONS OF CELL SURFACE HEPARAN SULFATE PROTEOGLYCANS, *Annual Review of Biochemistry 68*, 729-777.
4. H, C. (1997) *Heparin-binding proteins*.
5. Cardin, A. D., and Weintraub, H. J. (1989) Molecular modeling of protein-glycosaminoglycan interactions, *Arteriosclerosis, Thrombosis, and Vascular Biology 9*, 21-32.
6. Belland, R. J., Chen, T., Swanson, J., and Fischer, S. H. (1992) Human neutrophil response to recombinant neisserial Opa proteins, *Molecular Microbiology 6*, 1729-1737.
7. van Putten, J. P., and Paul, S. M. (1995) Binding of syndecan-like cell surface proteoglycan receptors is required for Neisseria gonorrhoeae entry into human mucosal cells, *The EMBO journal 14*, 2144-2154.
8. Bos, M. P., Kao, D., Hogan, D. M., Grant, C. C., and Belland, R. J. (2002) Carcinoembryonic antigen family receptor recognition by gonococcal Opa proteins

requires distinct combinations of hypervariable Opa protein domains, *Infection and immunity 70*, 1715-1723.

9. Popp, A., Dehio, C., Grunert, F., Meyer, T. F., and Gray-Owen, S. D. (1999) Molecular analysis of neisserial Opa protein interactions with the CEA family of receptors: identification of determinants contributing to the differential specificities of binding, *Cellular Microbiology 1*, 169-181.
10. Wai Wong, C., Dye, D. E., and Coombe, D. R. (2012) The Role of Immunoglobulin Superfamily Cell Adhesion Molecules in Cancer Metastasis, *International Journal of Cell Biology 2012*.
11. Dermody, T. S., Kirchner, E., Guglielmi, K. M., and Stehle, T. (2009) Immunoglobulin Superfamily Virus Receptors and the Evolution of Adaptive Immunity, *PLoS Pathog 5*, e1000481.
12. Juliano, R. L. (2002) SIGNAL TRANSDUCTION BY CELL ADHESION RECEPTORS AND THE CYTOSKELETON: Functions of Integrins, Cadherins, Selectins, and Immunoglobulin-Superfamily Members, *Annual Review of Pharmacology and Toxicology 42*, 283-323.
13. Barclay, A. N. (2003) Membrane proteins with immunoglobulin-like domains—a master superfamily of interaction molecules, *Seminars in Immunology 15*, 215-223.
14. Kuespert, K., Pils, S., and Hauck, C. R. (2006) CEACAMs: their role in physiology and pathophysiology, *Current Opinion in Cell Biology 18*, 565-571.
15. Beauchemin, N., and Arabzadeh, A. (2013) Carcinoembryonic antigen-related cell adhesion molecules (CEACAMs) in cancer progression and metastasis, *Cancer Metastasis Rev 32*, 643-671.

16. McCaw, S. E., Liao, E. H., and Gray-Owen, S. D. (2004) Engulfment of Neisseria gonorrhoeae: Revealing Distinct Processes of Bacterial Entry by Individual Carcinoembryonic Antigen-Related Cellular Adhesion Molecule Family Receptors, *Infection and immunity 72*, 2742-2752.

17. Bos, M. P., Grunert, F., and Belland, R. J. (1997) Differential recognition of members of the carcinoembryonic antigen family by Opa variants of Neisseria gonorrhoeae, *Infection and immunity 65*, 2353-2361.

18. Fedarovich, A., Tomberg, J., Nicholas, R. A., and Davies, C. (2006) Structure of the N-terminal domain of human CEACAM1: binding target of the opacity proteins during invasion of Neisseria meningitidis and N. gonorrhoeae, *Acta Crystallographica Section D 62*, 971-979.

19. Virji, M., Evans, D., Hadfield, A., Grunert, F., Teixeira, A. M., and Watt, S. M. (1999) Critical determinants of host receptor targeting by Neisseria meningitidis and Neisseria gonorrhoeae : identification of Opa adhesiotopes on the N-domain of CD66 molecules, *Molecular Microbiology 34*, 538-551.

20. Grossoehme, N., Spuches, A., and Wilcox, D. (2010) Application of isothermal titration calorimetry in bioinorganic chemistry, *J Biol Inorg Chem 15*, 1183-1191.

21. Zuiderweg, E. R. P. (2001) Mapping Protein−Protein Interactions in Solution by NMR Spectroscopy†, *Biochemistry 41*, 1-7.

22. Chen, Y., Reizer, J., Saier, M. H., Fairbrother, W. J., and Wright, P. E. (1993) Mapping of the binding interfaces of the proteins of the bacterial phosphotransferase system, HPr and IIAglc, *Biochemistry 32*, 32-37.

23. McAlister, M. S. B., Mott, H. R., van der Merwe, P. A., Campbell, I. D., Davis, S. J., and Driscoll, P. C. (1996) NMR Analysis of Interacting Soluble Forms of the Cell−Cell Recognition Molecules CD2 and CD48†, *Biochemistry 35*, 5982-5991.

24. Fox, D. A., and Columbus, L. (2013) Solution NMR resonance assignment strategies for β-barrel membrane proteins, *Protein Science 22*, 1133-1140.

25. Zhao, W., McCallum, S. A., Xiao, Z., Zhang, F., and Linhardt, R. J. (2012) Binding affinities of vascular endothelial growth factor (VEGF) for heparin-derived oligosaccharides, *Bioscience Reports 32*, 71-81.

Chapter 5: Dynamics of a model helical membrane protein TM0026

Electron paramagnetic resonance (EPR) spectroscopy is analogous to NMR, but relies on electron spins rather than nuclear spins. Site-directed spin labeling (SDSL) is a method to selectively introduce an unpaired electron into proteins, and has been a useful tool towards investigating the structure and dynamics of proteins. Within the past two decades, large strides have been made towards deconvoluting the complex dynamic contributions to spin label motion. Mchraourab *et al*. probed the motion of side chains in T4 lysozyme (T4L) in 1996.(*1*) Columbus *et al*. reported the model for motion of the (1-oxyl-2,2,5,5-tetramethyl-3-pyrroline-3-methyl)-methanethiosulfonate (R1) spin label in α-helices in 2004.(*2*) These investigation (among others) of soluble proteins characterized dynamics of solvent exposed helical sites that did not have observed contacts with surrounding residues. Studying membrane proteins is significantly more difficult, as spin labels act differently in membrane mimics, which makes interpretation of the lineshapes more difficult.(*3, 4*)

SDSL has been used to investigate conformational switching in membrane proteins, yet direct quantification of membrane protein dynamics is lacking.(*5*) Membrane proteins often have broad EPR lineshapes with multiple component regions, which are not thoroughly taken into account through central linewidth analysis. Fanucci *et al*. mapped the structure and dynamics of transmembrane β-barrel BtuB reconstituted in 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine (POPC) bilayers in 2002, and discovered a gradient in backbone fluctuations across β-sheets, with a higher mobility at the periplasmic face.(*6*)

In 2010, Kroncke *et al*. reported on the molecular motion of spin labels at solvent-exposed transmembrane α-helical sites on leucine transporter LeuT reconstituted in octyl glucoside detergent micelles.(*3*) In collaboration with Brett Kroncke and Tsega Solomon, I extended these investigations to mapping dynamics of a model polytopic membrane protein, TM0026 using SDSL and NMR.

5.1 Introduction

5.1.1 Membrane protein dynamics by EPR

Electron paramagnetic resonance is a powerful tool for investigating structure and dynamics of soluble as well as membrane proteins. The EPR spectrum of a probe can provide information characterizing the local environment of the probe.(*1*) The primary method for introducing a probe is site-directed spin labeling. SDSL incorporates the spin-labeled side chain into protein sequences, generally through a cysteine reaction with a sulfhydryl-specific nitroxide reagent. Cysteine substitution mutagenesis is used to selectively introduce nitroxide labels throughout the protein of interest.(*1*) The most commonly employed spin-label side chain is the methanethiosulfonate spin label I designated R1 when attached to cysteine residues (Figure 5.1A).(*7*) One of the primary goals in developing the SDSL method is to enable the quantification of protein dynamics directly from the EPR lineshape.(*7*)

The primary information of interest in the study of TM0026 is the detection of protein backbone dynamics using spin labeled side chains. The spectral lineshape of the EPR spectrum of R1 is determined by motion of the nitroxide ring on the ns timescale, and is related to three correlation times(*1*): (1) The rotational correlation time (τ_R) for the global tumbling of the entire protein. (2) The effective correlation time (τ_B) due to rotational isomerizations about bonds linking the nitroxide to the backbone, which depends on the primary, secondary, and tertiary structure of the protein as well as the structure of the spin-label.(*1, 3*) The effective correlation time (τ_S) for segmental motion of the backbone

relative to the protein structure. These contributions are broken down as the rotary diffusion of the protein, internal dynamic modes of the sidechain, and local backbone fluctuations, respectively.(*4*) Mchraourab *et al*. determined that contributions from rotary diffusion to the mobility of the R1 side chain in proteins with a τ_R > 18 ns are negligible, and EPR spectral lineshapes primarily reflect internal sidechain motions and internal backbone fluctuations.(*1*) Langen *et al*. demonstrated that the Sδ of the disulfide bond interacts with the C_α hydrogen atom in the backbone, likely restricting oscillations about the first two bond in the side chain.(*8*) Columbus *et al*. showed that this interaction, coupled with the high barrier of rotation around the disulfide bond, restricts the R1 motion around all but the terminal bonds of the linker (χ_4 and χ_5).(*9*) These data, cumulatively, indicate the label is sensitive to backbone dynamics when lacking interactions with surrounding neighbors.

The inverse central linewidth of the EPR lineshape, ΔH_0^{-1}, and the inverse second moment of the nitroxide CW spectrum, $\langle H^{-2} \rangle$, for the R1 side chain provide convenient experimental measures of the nitroxide mobility (Figure 5.2). The scaled mobility, M_s, is a measure of overall spin label mobility calculated from the inverse central linewidth, and is defined by the equation:

$$M_s = \frac{\delta_{exp}^{-1} - \delta_i^{-1}}{\delta_m^{-1} - \delta_i^{-1}} \tag{5.1}$$

where δ^{-1} is the inverse central linewidth, the subscript '*exp'* indicates the experimental spectrum, 'i', the spectrum corresponding to a fully immobilized spin, and 'm', the spectrum corresponding to a fully mobile spin label.(*10*)

High values of M_s reflect narrow central linewidth, which indicates fast motions of the nitroxide sidechain. The inverse second moment calculations are more sensitive to the outer regions of the spectra, where immobile components are observed.(*11*) The inverse second moment is defined by the equation:

$$< H^2 >= \int (B - < H >)^2 \frac{S(B)dB}{\int S(B)dB} \quad (5.2)$$

where <H> is the first moment (geometrical center of the spectrum), B is the magnetic field, and S(B) is the absorption spectrum. As indicated in Figure 5.2, R1 side chains in a given structural class are clustered in mobility, with buried sites located in the low-mobility region, and loop regions having the highest mobility. The low mobility of buried R1 side chains is due to steric interactions of the nitroxide within the densely packed core of the protein. High isotropic mobility of solvent exposed R1 side chains at loop sites are observed and may deviate from the χ_3/χ_4 mode, with increased rotational freedom about the C_α-C_β bond, compared to other sites of the protein.

This qualitative understanding was extended quantitatively by Columbus *et al*.(*9*) for a soluble protein, and the detected backbone dynamics was compared to NMR ^{15}N relaxation experiments (reviewed in Chapter 2).(*2*) The site-dependent variation of EPR spectra along the basic leucine zipper of GCN4 was experimentally evaluated. Columbus *et al*. demonstrated that M_s fluctuates along the sequence correlated to backbone fluctuations in the ps – ns time scale detected with NMR.(*2*) The spectra of the R1 side chains were fit with the microscopic order macroscopic disorder (MOMD) model to provide effective correlation times and order parameters. Based on simulated spectra, the authors determined that M_s is mostly dependent on the correlation time (τ) (and to a lesser extent, order parameter [S]), and that in multicomponent spectra, M_s is largely determined by the most

mobile component. A mobility gradient for M_s and τ^{-1} was present along the GCN4-58 sequence, and the periodic average of M_s over 3 – 4 residues (the putative turn of a helix) showed a linear decrease in mobility through the basic zipper region. This data corresponded well with NMR ^{15}N relaxation data, that also revealed a linear decrease in mobility from the N-terminus to the first residue of the leucine zipper indicating that M_s values can be used as a simple parameter to monitor backbone dynamics.(*12*)

However, the overall dearth of information on nitroxide EPR lineshapes of membrane proteins limits the interpretation of side chain and backbone dynamics, from the μs (interpretation of multiple spectral components) to the ns (directly from the lineshape) timescales. Whether or not EPR lineshapes of membrane proteins reflect backbone motion is still an unresolved question. Therefore, we investigated and compared the EPR lineshapes and the NMR ^{15}N relaxation data for the model membrane protein, TM0026.

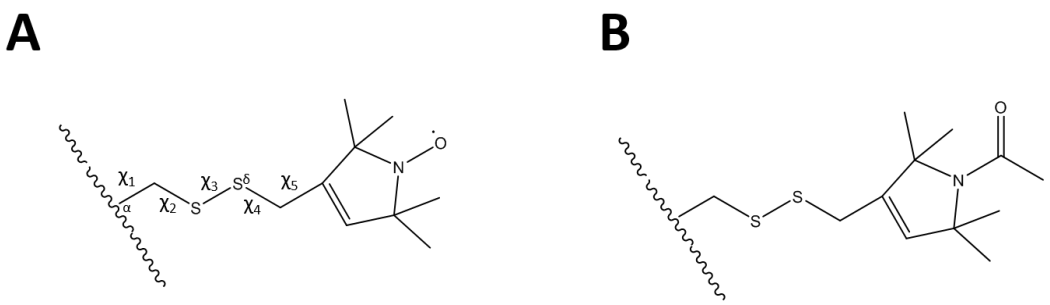


Figure 5.1. The chemical diagram of the spin labels used. (A) The structure of the nitroxide spin label (R1) showing side-chain dihedral angles (χ). χ_5 is defined by S-C-C=C. (B) The structure of the diamagnetic analog (R1') used for NMR spectral comparisons.

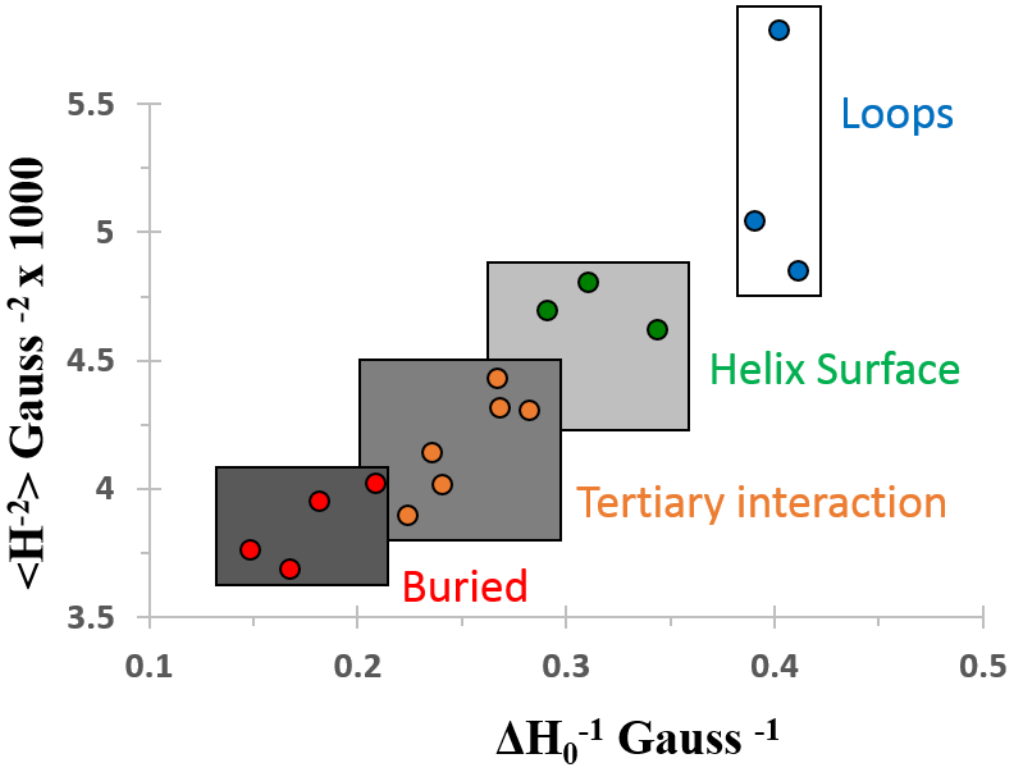


Figure 5.2. An example plot of ΔH_0^{-1} and $\langle H^{-2} \rangle$ for T4L, adapted from Mchaourab *et al.* (*1*) Reciprocal second moment versus reciprocal central linewidth are plotted from various EPR spectra of R1 for T4L.

5.1.2 TM0026, a model polytopic membrane protein

TM0026 is a model polytopic membrane protein from *Thermotoga maritima* that contains only two transmembrane helices (topology shown in Figure 5.3A). The function of TM0026 is not known, but the gene is located in the BgIR regulon along with TM0032 – the regulator, TM0024 – a lamarinase, TM0025 – a β-glucosidase, and TM0027–TM0031 – cellobiose and aminaribiose transporters.(*13*) Frock *et al*. established that in the presence of β-linked glucans (including laminarin, barley glucan, and pustulan), expression of all of these TM proteins was up-regulated which suggests that TM0026 may act as a regulating protein of sugar glucose transporters.(*13*)

The secondary structure,(*14*) low-resolution tertiary structure (unpublished data), and NMR resonance assignments have been previously established.(*15, 16*) Columbus *et al*. previously reported the effects of detergent on the structure of TM0026.(*16*) The results suggested that the matching of micelle hydrophobic thickness to the hydrophobic thickness of TM0026 was necessary to avoid potential exchange processes resulting from protein hydrophobic mismatch. Use of both DM and FC12 micelles resulted in a homogenous protein conformer, while TM0026 in DDM and FC10 micelles displayed significant conformational heterogeneity. Columbus *et al*. demonstrated that mixing DDM/FC10 micelles with a hydrophobic thickness parameter matched to FC12 resulted in nearly identical spectra, but with the benefit of additional cross peaks that facilitated assignment of 11 additional residues than in the pure detergent.(*16*) NMR structural data of a TM0026 construct in DDM/FC10 mixed micelles was collected by Brett Kroncke of the Columbus laboratory.(*17*) The construct contained all 68 native residues, as well as a 12 residue N-

terminal His_6 tag. Of the total 78 expected peaks of the ^{15}N, ^{1}H-HSQC spectrum, 71 resonances were observed. The NMR derived ^{13}C chemical shifts (accession number: 18494; Figure 5.3B-D) suggest that TM0026 consists of two transmembrane α-helices separated by an unstructured linker, which is consistent with the predicted secondary structure.(*15*)

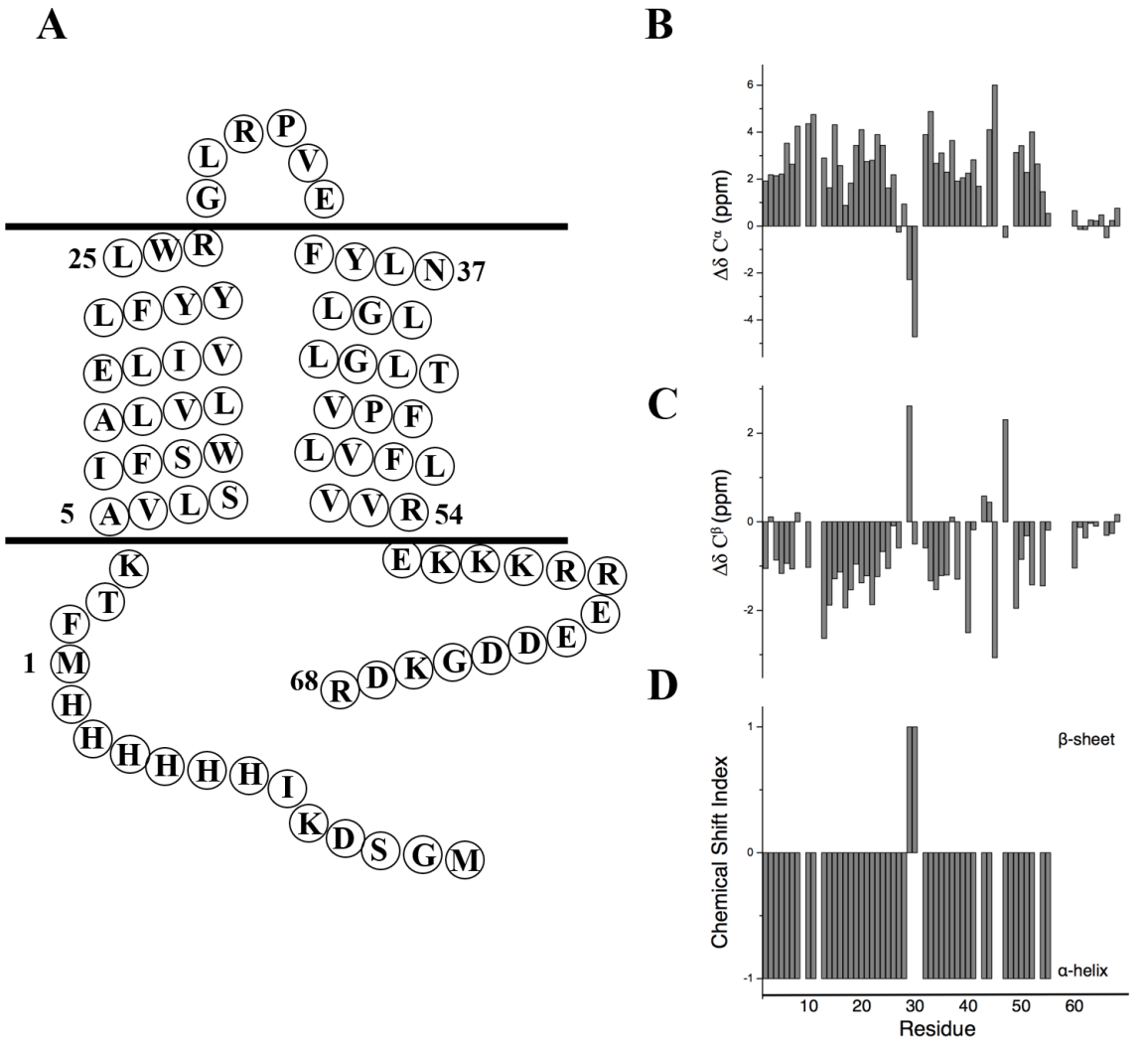


Figure 5.3. **The NMR chemical shifts for TM0026**, **from Kronke *et al*.(*17*)** (A) The topology predicted for a helical membrane protein containing two membrane-spanning helices. The NMR chemical shifts of (B) $^{13}C^{\alpha}$, (C) $^{13}C^{\beta}$ and (D) the derived chemical shift index (*18*) for TM0026 are shown. All three chemical shift plots are consistent with the topology.

5.2 Materials and Methods

5.2.1 Cloning, expression, and purification of TM0026 protein

Mutagenesis and protein expression of TM0026 were performed by Brett Kroncke and Tsega Solomon according to established protocols (*19*). All cysteine mutations were introduced using the PIPE polymerase chain reaction (*20*). Plasmid containing the TM0026 gene was transformed into BL21(DE3)RIL cells (Stratagene, Santa Clara, CA) for expression in either LB or minimal media containing 50 µg/mL ampicillin. Cells were grown until an OD_{600} of ~0.6–0.8 and induced with 1 mM IPTG for 4 h at 37 °C. ^{15}N-labeled samples were expressed in minimal media supplemented with $^{15}NH_4Cl$. The cells were lysed in a buffer containing 50 mM Tris-HCl (pH 8.0), 150 mM NaCl, 1 mM tris(2-carboxyethyl)phosphine (TCEP) and pelleted (15,000g, 60 min, 4 °C). After removing the cell debris, 10 mM *n*-decyl-β-D-maltopyranoside (DM; Anatrace, Inc., Maumee, OH) was added to the supernatant for 3 h at room temperature to solubilize TM0026. TM0026 was then purified into *n*-dodecyl-β-D-maltopyranoside (DDM ; Anatrace, Inc., Maumee, OH) and *n*- dodecylphosphocholine (FC10; Anatrace, Inc., Maumee, OH) by Co^{2+} immobilized metal affinity chromatography (IMAC) as previously described (*14*).

5.2.2 Spin labeling of TM0026

Spin labeling was performed by Brett Kroncke and Tsega Solomon. Spin labeling of TM0026 mutants was performed as previously described (*14*). TM0026 was concentrated to ~150-200 µM and passed through a PD-10 desalting column containing an elution buffer of 20 mM phosphate buffer (pH 6.2), 150 m*M* NaCl, 5 m*M* DDM, and 15 m*M* FC-10 to

remove TCEP and imidazole. The protein eluate was incubated with R1, (1-oxyl-2,2,5,5-tetramethyl-3-pyrroline-3-methyl)-methanethiosulfonate (MTSSL) or R1', the NO-acetylated diamagnetic equivalent, (1-acetoxy-2,2,5,5-tetramethyl-3-pyrroline-3-methyl) methanethiosulfonate (Figure 1B) at a 1:5 molar ratio of protein to spin label. After incubating the spin label with TM0026 overnight, excess MTSSL was partially removed by passing the sample through a PD-10 desalting column. The final excess of MTSSL was removed after a three day incubation at room temperature by Co^{2+} IMAC. The elution fraction was concentrated and dialyzed against 4 L of 20 mM phosphate buffer (pH 6.2), 150 m*M* NaCl to remove imidazole. The dialyzed protein was concentrated to 100 µM and spectra were recorded. (*14*)

5.2.3 EPR spectroscopy

EPR spectroscopy was performed by Brett Kroncke and Tsega Solomon. Protein samples of 5 µL (≈ 100 µ*M*) were loaded into Pyrex capillaries (0.60 mm id × 0.84 mm od; Fiber Optic Center, New Bedford, MA). X-band EPR spectra of TM0026 cysteine mutants were recorded on a Bruker EMX spectrometer with an ER4123D dielectric resonator (Bruker Biospin, Billerica, MA) at room temperature.

5.2.4 NMR spectroscopy

Isotopically ^{15}N, ^{1}H-labeled TM0026 for NMR experiments were prepared using M9 minimum medium containing $^{15}NH_4Cl$ (99%, Cambridge Isotope Laboratories, Inc) as a nitrogen source. NMR samples included the addition of 10% D_2O for lock. Chemical shifts were obtained from the published TM0026 assignment (BMRB 18494) (*17*). NMR experiments were performed on Bruker AVANCE spectrometers operating at proton

frequencies of 600 MHz and 800 MHz equipped with Bruker 5 mm TXI cryoprobes and recorded at 40°C. Spectra were processed with Topspin. Longitudinal ^{15}N relaxation (R_1), transverse ^{15}N relaxation experiments (R_2), and heteronuclear nuclear Overhauser effects were measured using two-dimensional ^{15}N-^{1}H TROSY-based experiments at both 600 MHz and 800 MHz. R_1 relaxation experiments employed longitudinal delay times of 50, 100, 250, 500, and 1000 ms and R_2 relaxation was measured with CPMG delays of 17, 51, 102, 204, and 492 ms. Relaxation measurements were performed at 40°C and data sets were processed and analyzed using NMRPipe (*21*).

5.3 Results and Discussion

5.3.1 SDSL mutations do not significantly perturb the fold of TM0026

The goals of SDSL studies are to investigate protein structure at the level of the backbone fold, study equilibrium dynamics, and resolve conformational changes in the protein fold.(*1*) Hence, in order to obtain and interpret meaningful SDSL data, the reporter group must not result in any significant perturbation of the secondary and tertiary organization of the protein. Matthews (1995) summarized the structural and thermodynamic characterization of over 200 T4 lysozyme mutants and determined that many amino acids in the sequence of the protein are noncritical for folding, stability, and activity.(*22*) In particular, surface site amino acid substitution has little effect on stability and structure, while buried sites have effects dependent on the nature of the residue introduced, as well as the local context of the introduced residue (for instance, relieving torsional strain for F153A).(*23*) Instead, mutations that affect the packing of the hydrophobic core result in the largest destabilization and structural alteration. For the majority of T4L mutations, the observed alterations involved side-chain rearrangement and main-chain shifts to repack the core, resulting in subtly altered structures.(*22*, *24*)

In membrane proteins, Faham *et al.* introduced alanine substitutions at 24 positions in the B helix of bacteriorhodopsin.(*25*) The authors found a large percentage of stabilizing mutants (17%), which indicated that membrane proteins are not highly optimized for stability. The four stabilizing mutations occurred in a fully exposed residue (L62), partially buried residues (L61 and M56), and a completely buried residue (V49). In particular, mutants of charged residues at the end of the helix contributed to destabilizing the protein.

These charged residues formed intra-helical salt bridges (K41, D38), while another a residue hydrogen bonded to backbone amide nitrogen atoms (D36), forming a helix cap.(*25*) Upon crystallizing the structure of the most stable mutant, M56A, and finding no significant alteration in the structure compared to the wild-type mutant, the authors concluded that the stability enhancement may be due to alterations in the unfolded state.(*26*) Interestingly, the authors found that the P50A mutation does not remove the kink, which indicates other residues could participate in bending of the helix. This is likely due to the energetic cost of bending the helix; the straightened helix could maintain a more regular hydrogen bonding pattern than the kink-induced helix.(*26*) To further examine the consequences of the P50A mutation, Cao and Bowie attempted to engineer transmembrane helix deformations in bacteriorhodopsin through amino acid mutations around the helix kink region.(*27*) To break the bend, Cao *et al*. introduced a second mutation, T46A, that allowed the helix to adopt several distinct conformations, that included noncanonical *i* -> *i* + 3 hydrogen bonds.(*27*) This indicated that the transmembrane helices were flexible, and that in the absence of proline, the helices could be shifted to distinct conformations by a single mutation. However, in both of these studies, helix B was investigated because it was not associated with bound retinal chromophore. The structural integrity of the mutant proteins was examined, rather than the function. Instead, the effects of R1 are often assessed using an activity assay, as well as by thermal stability.(*22, 24*) R1 substitutions should produce little to no change in either activity or thermal stability for solvent-exposed surface sites.(*24*) Substitutions at buried sites produced the greatest loss of activity, which was potentially due to rearrangement of the core to accommodate the increased volume of

R1.(*24*) TM0026 is not an enzyme and activity of mutants could not be assessed; therefore, overall fold was assessed with NMR spectroscopy

Of the TM0026's native 68 residues, 55 mutants were spin-labeled, spanning the N-terminus, transmembrane helices 1 and 2, the loop, and 2 residues of the C-terminus (at residues 1-27, 29-30, 31-38, 40-45, 47-54, and 57-61). The selected mutated residues were chosen to explore the various structural determinants of EPR spectral line shape to determine the structural and energetic consequences of residue substitution with a nitroxide spin label. Secondary structural elements were observed by the lineshapes. Narrow, sharp spectra were observed for residues 59-61, which indicates a disordered and mobile C-terminus. The N-terminus spectra have multiple components and are mobile, but are not completely disordered. These residues may likely interact with the micelle surface. The spin-labeled residues in the linker sequence between the two transmembrane helices have similar lineshapes to the N-terminus, which suggests that this loop is unstructured, but may interact with the transmembrane helices or the micelle. Lineshapes throughout the transmembrane regions vary in terms of dynamics and number of components. Typical lineshapes for detergent facing nitroxide side chains were observed (L7R1, S11R1, V20R1, T44R1, and F40R1). Certain spectra indicated nitroxides with highly restricted motions (A13R1, F34R1, L401, and L48R1), in which the spin label could be forming contacts between the transmembrane α-helices. The contrast in spin label dynamics between the solvent exposed and aliphatic exposed regions correlates well with the expected topology of TM0026. Based on the EPR lineshapes, A13 on TM1, and F47 and L48 on TM2 form the tertiary contacts between TM1 and TM2. Most transmembrane helices in polytopic membrane proteins are packed at an angle with respect to each other, and the EPR

lineshapes are consistent with a single cross point at a tertiary contact between these three residues. The EPR spectra for several sites in the transmembrane region as well as near the C-terminus and linker regions are indicative of a mobile spin label (I9R1, W12R1, I19R1, and V49R1 – Figure 5.5A-C). However, spin label sites that deviated from the expected topology and observed dynamic trends were investigated for structural perturbations. For instance, V45R1 indicates a mobile nitroxide spin label, but is embedded within TM2, next to a rigid proline residue. To investigate the effects of the R1 probe on the structure of TM0026, ^{15}N-^{1}H HSQC spectra of select mutated sites were examined and compared to the spectrum of the native protein. To avoid relaxation enhancement effects by the nitroxide spin label, an N-acetylated diamagnetic R1 probe (termed R1') was employed (Figure 5.1B). Deviations in trends of scaled mobility and lineshapes indicated that certain mutants might not retain the native fold. The R1' NMR spectra were therefore recorded for 15 sites based on the unexpected trends in lineshapes (F10, W12, V15, L16, E17, Y23), location (putative tertiary sites and termini - M1, F2, K4, A13) and precipitant present in the EPR tube (T3, A5, L7, S8, L16, E17, V45) (all NMR spectra are presented in Appendix Figure A5). For the majority of residues, the introduction of the R1' label did not significantly perturb the overall structure of TM0026 as the ^{15}N,^{1}H HSQC spectra were nearly identical to that of the wild type TM0026 (Figure 5.4B,C).

The primary exceptions were V45R1', K4R1', E17R1', and Y23R1' (Figure 5.4D-F). V45R1' is located in the second transmembrane helix. Significant spectral broadening was present in the NMR spectrum (Figure 5.4), and a large quantity of precipitate was present in the NMR sample solution, suggesting that the spin label at the V45 position could potentially be destabilizing. Of the 68 native resonances observed, only 43 peaks

overlapped from the V45R1' spectrum. The significant line broadening and missing backbone resonances indicated the overall fold and/or dynamics were significantly perturbed. Notable broadened resonances included residues from the immobile regions of both TM1 and TM2 (in particular: Y23, L25, V32, R27, L43, L48, V49, L51). V45 immediately precedes P46, and addition of the spin label at this residue could perturb the helix kink, which in turn disrupts the global fold of the protein.

K4 may be important for positioning the helix at the headgroup region of the micelle. E17R1' and Y23R1' are both located in the first transmembrane helix and had mild amounts of precipitate noticeably present in solution, prior to NMR spectroscopy. E17 is a likely protonated charged residue located in the hydrophobic region of the micelle that has no counter positive charge in proximity. In the E17R1' spectrum, 31 of the 68 peaks overlapped with the wild type spectrum. In the Y23R1' spectrum, 42 of the 68 peaks were observed and overlapping with the wild type resonances. In all three NMR spectra, the most broadened resonances were similar to those affected in the V45R1' mutant. The K4, E17, Y23, and V45 mutations appear to perturb both TM helices, which disrupts the overall fold of the protein. This interaction could potentially be due to displacement of a potentially necessary charge, or destabilization of the tertiary contacts. In addition, spectral broadening could be a result of conformational exchange of the protein. Subsequently, the spin label data for these four mutants were not included in the scaled mobility and second moment plots.

5.3.2 The dynamics of model protein TM0026 by NMR and EPR

The NMR R_1 relaxation data, which is highly sensitive to backbone nanosecond motions, corresponds well with the EPR data (Figure 5.7). The R_1 values are lower in the transmembrane regions of TM0026. The EPR and the NMR data suggest that the second transmembrane helix has two distinct regions; L38-V45 (TM2N) and F47-V53 (TM2C). The discontinuity in the motion of the helix may be due to a kink in the helix at P46. Prolines are recognized as inducing kinks in helices due to the disruption in hydrogen-bonding arising from lack of a backbone amide proton.(*26, 28*) Also present are two glycines (G39, G42) that are also typically assumed to be helix destabilizing,(*28*) and may contribute to the increased mobility of the TM2N compared to TM2C. The TM2C region contains two residues (F47 and L48) that form tertiary contacts between TM1 and TM2 (as assessed by the EPR lineshapes), and is likely stabilized by helix packing within the micelle. The R_2 and heteronuclear NOE data do not appear to support the observed decoupled dynamics in the second transmembrane helix, though these experiments are less sensitive to nanosecond dynamics compared to the R_1 experiments. The R_2 experiment is less sensitive for several reasons. The first is that the TM0026 – detergent complex has a relatively large overall correlation time, which decreases R_2 sensitivity due to the slow tumbling of the large protein-detergent complex. (*29*) The second is that the R_2 values are more difficult to interpret, since values decrease with nanosecond motions, and increase due to conformational changes in the μs – ms time regime. It is interesting that a gradient is observed such that R_2 increases from the termini towards the loop in both transmembrane helices. This trend is further observed in deuterium exchange rates and carbon chemical shifts for TM2 (data obtained by Brett Kroncke) but not in TM1. Because the gradient is

not observed in the R_1 or the heteronuclear NOE data, the observed gradient is likely due to μs – ms backbone dynamics.

The EPR spectra (Figure 5.5) for residues throughout TM0026 were recorded and compared to ^{15}N relaxation data. The ^{15}N relaxation data (Figure 5.6) suggests that the N-terminus (M1 – K4) and C terminus (E55 – R68) are extremely dynamic. The high R_1 values and the low R_2 values for the N- and C-termini, in contrast to the transmembrane helices, indicating that these regions are more flexible than the micelle embedded helices. Both the heteronuclear NOEs and the R_2 values also suggest that the termini are dynamic. The plotted EPR scaled mobility data is in agreement with the relaxation data for the C- (Figure 5.7) and N-. However, the range of M_s values observed for membrane proteins compared to soluble proteins is much smaller in scale and amplitude.(*3*) Kroncke *et al.* indicate that the different magnetic parameters and distinct motion of spin labels on membrane proteins reduces the sensitivity of M_s to backbone fluctuations.(*3*) These lower M_s values result from the restricted rate of nitroxide motion, which is due to weak interactions between the nitroxide ring and the protein surface.

There are several prominent features in the data for the transmembrane α-helices. According to the topology of the protein, there should be three unstructured regions of high mobility: the N- and C- termini, and the loop region. The loop and C-terminal regions have the highest scaled mobility indicated that they are much more dynamic than the transmembrane helices. However, as noted from the lineshapes, the loop is not as dynamic as the C-terminus. EPR lineshapes of the solvent exposed linker residues (G28 – E33) appear to mostly be mobile, with the exception of E33R1; however, only R30R1 is as

mobile as the typical random coil sequence such as R60R1, E61R1, and E62R1. Both R_1 relaxation data as well as the M_s and $\langle H^{-2} \rangle$ values suggest that the linker region is at least as mobile as the mobile region of the kinked helix (L38-V45). In contrast, the less sensitive R_2 and heteronuclear NOE relaxation data suggest that the linker region is as immobile as both transmembrane helices. The linker is short, with only six residues. G28 likely provides the backbone flexibility to break the α-helix and induce a turn. P31 provides rigidity to the linker backbone and induces the turn back towards the micelle to orient the second transmembrane α-helix. However, the remaining residues are bulky and may restrict the backbone motions.

In the transmembrane regions of the protein, the EPR and NMR R_1 relaxation data indicate that TM1 is more dynamic than TM2C. TM1 contains E17, which is likely protonated in the low dielectric, and may compete with hydrogen bonding within the helical backbone.(*30*) The dynamics data also indicate that TM2N is more dynamic than both TM1 and TM2C. This decoupling of backbone dynamics within TM2 is likely due to the helix kink caused by P46. TM2N has two glycine residues that form a GXXG motif, which could contribute to the overall higher mobility of the region. However, the two helices do not move independently of each other. A calculation of the overall correlation time using R_2/R_1 for TM1, TM1 and the loop, TM2, TM2 and the loop, and TM1, TM2, and the loop all estimate the correlation time to be approximately 13 ± 2 ns (MW of the PDC ≈ 22 ± 3 kD). This similarity in correlation time suggests that the differences in nanosecond dynamics are likely due to the local amino acid sequence. However, the R_2/R_1 approach does not distinguish between the effects of motional anisotropy and chemical exchange.(*31*) Instead, Kneller *et al.* proposed the utility of R_1R_2 analysis, which significantly attenuates the

effects of motional anisotropy.(*31*) The R_1R_2 product for the TM0026 relaxation data also suggest that the C- and N-termini are dynamic, compared to the transmembrane helices (Figure 5.8). The presence of the slower μs – ms motional processes increase the R_1R_2 values, which potentially obscures the mobile regions gradients observed in TM2.

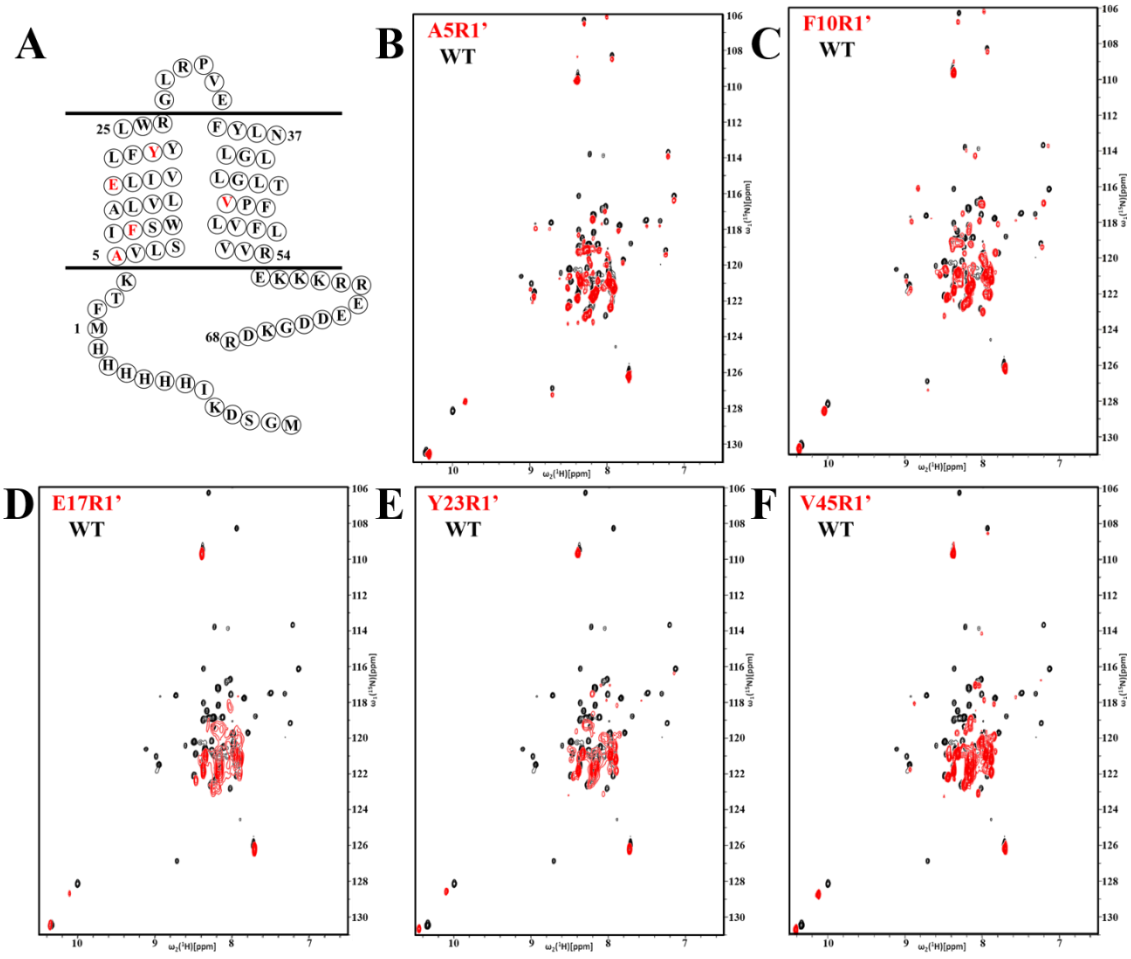


Figure 5.4. NMR spectra of select regions of interest with the predicted topology map of TM0026. (A) The topology map of the predicted structure for TM0026, with represented mutations highlighted in red. (B-C) Representative ^{15}N, ^{1}H HSQC spectra for SDSL sites (A5R1' and F10R1') that do not perturb the fold of TM0026. (D-F) The ^{15}N, ^{1}H HSQC spectra for SDSL sites (E17R1', Y23R1', and V45R1') that perturb TM0026's fold.

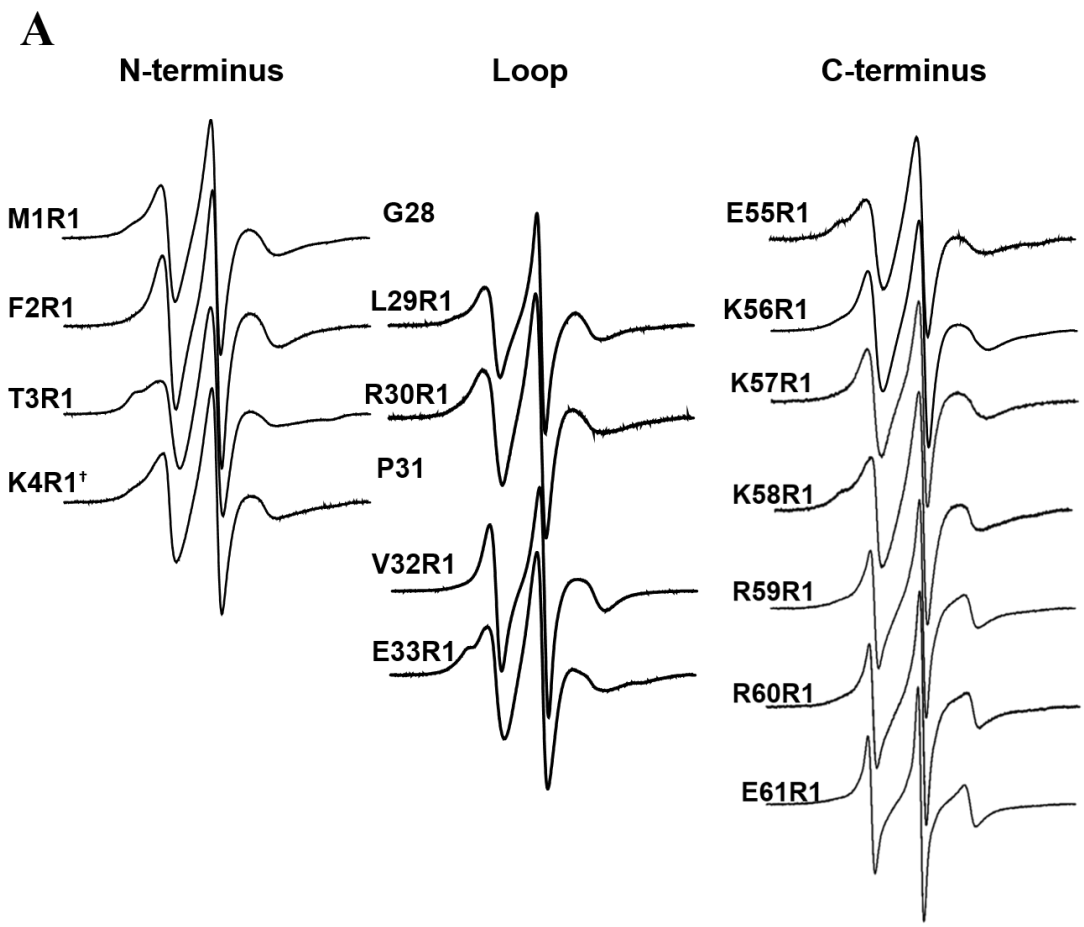


Figure 5.5. The EPR spectra of SDSL residues of TM0026. (A) The EPR spectra for the N-terminus, linker, and C-terminus regions.

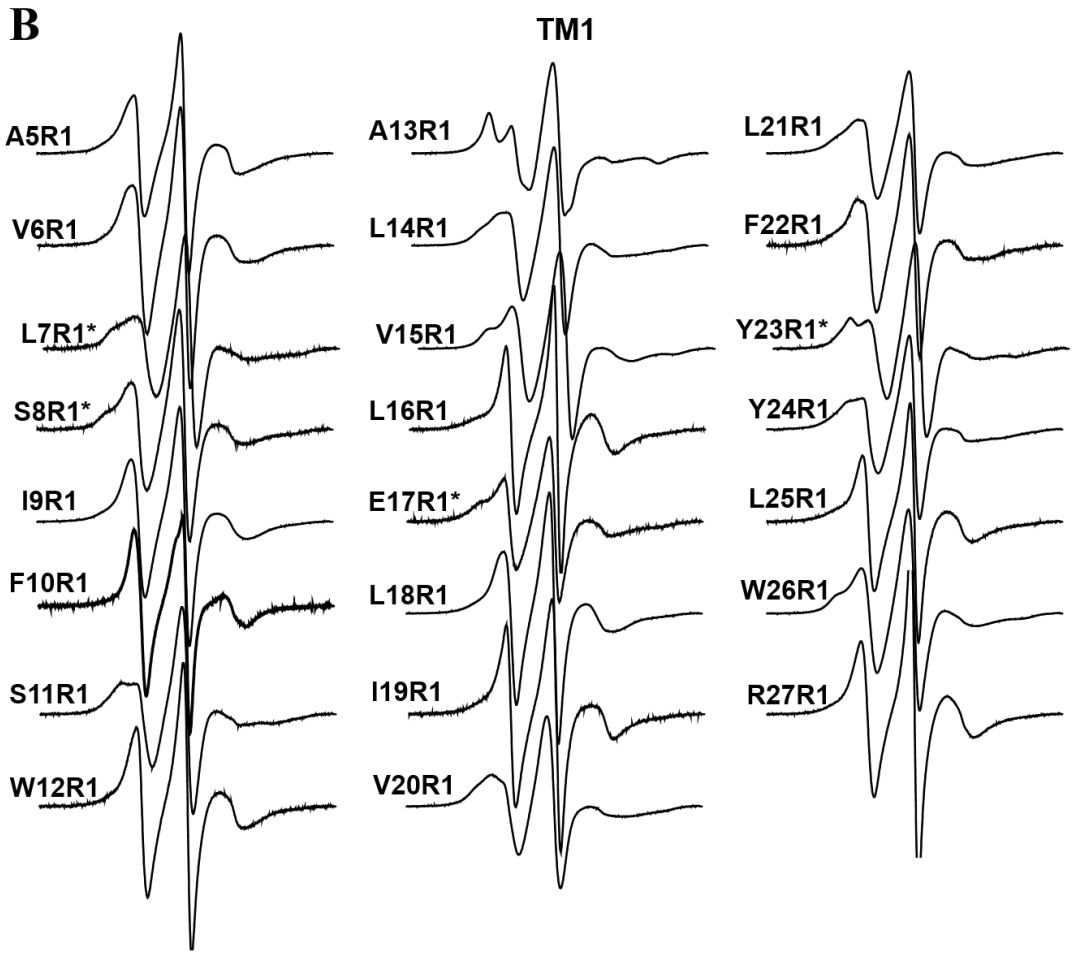


Figure 5.5. The EPR spectra of SDSL residues of TM0026. (B) The EPR spectra for the first transmembrane helix.

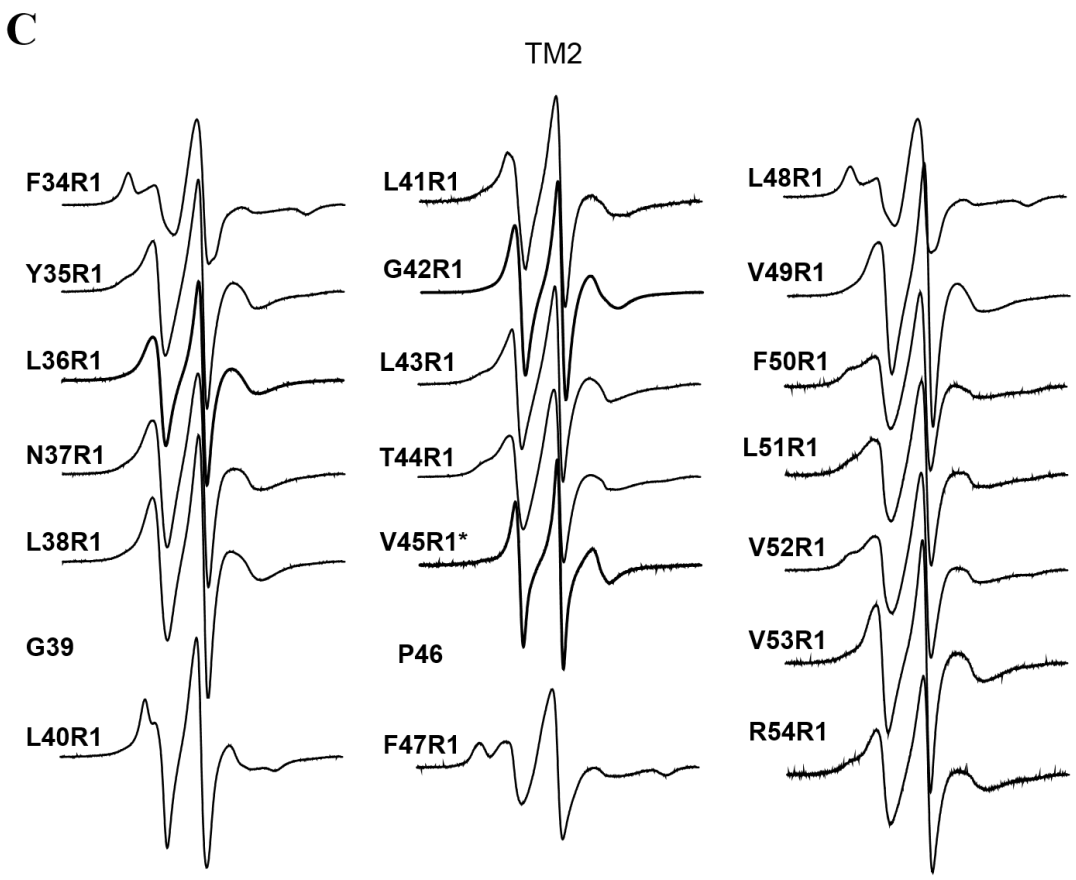


Figure 5.5. The EPR spectra of SDSL residues of TM0026. (C) The EPR spectra for the second transmembrane helix.

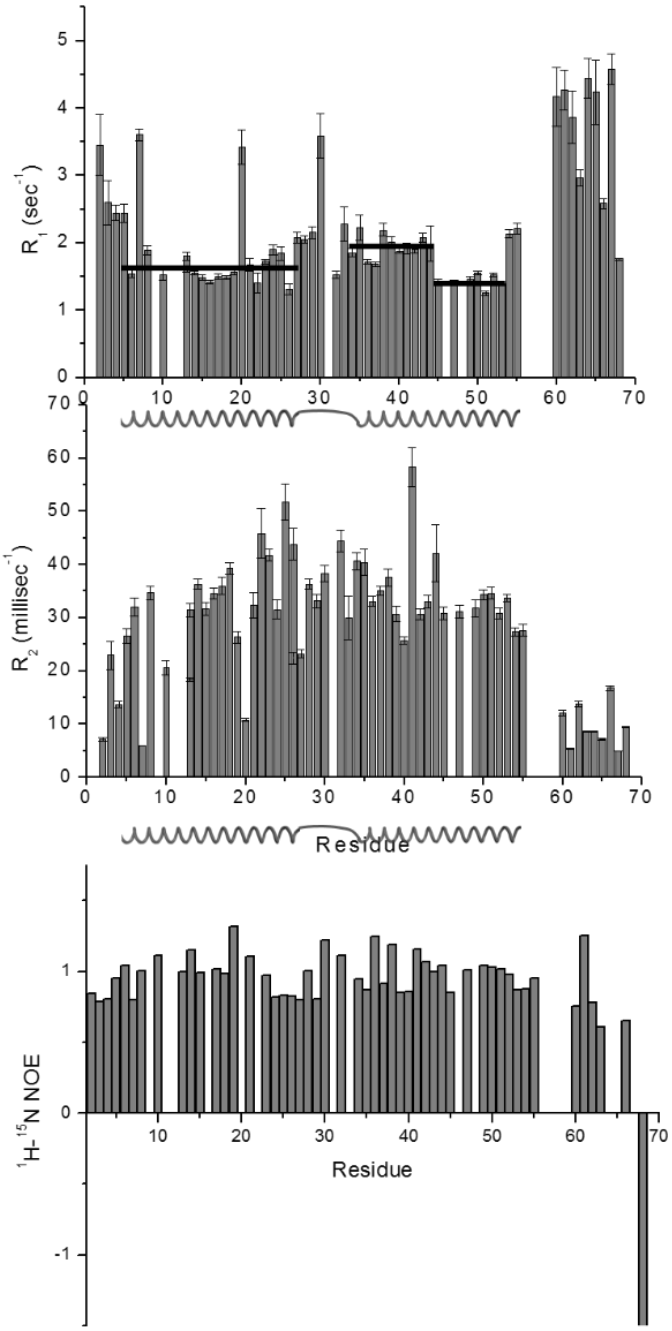


Figure 5.6. The ^{15}N NMR relaxation data for TM0026. The R1, R2, and heteronuclear NOEs are plotted, with the topology of TM0026 interspersed to show residue position and location.

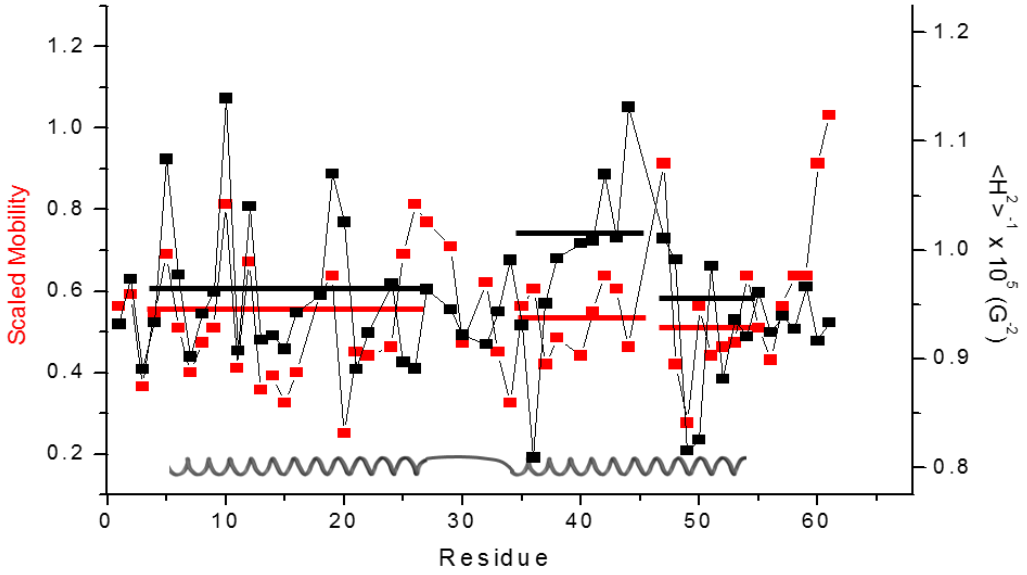


Figure 5.7. The EPR M_s and $\langle H^2 \rangle$ data plotted against the residues of TM0026.

Higher values of scaled mobility and of $\langle H^{-2} \rangle$ indicate greater mobility of the spin label.

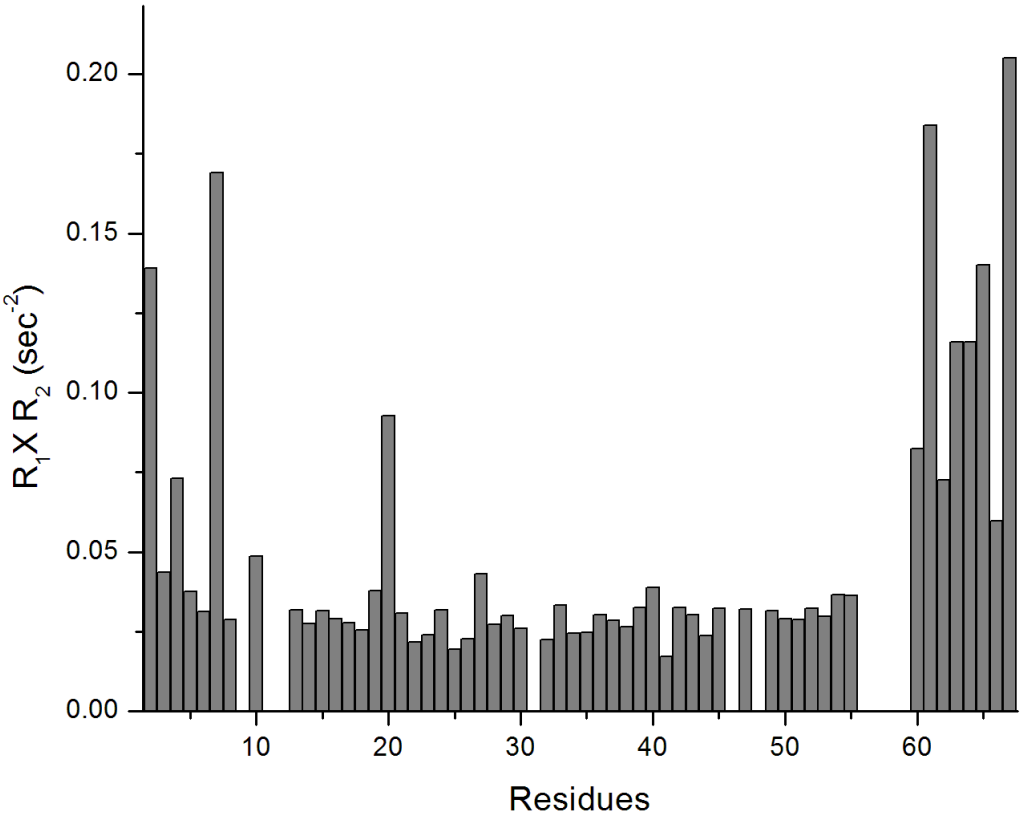


Figure 5.8. ^{15}N relaxation data for TM0026 displayed as the R_1R_2 product. The R_1R_2 product is plotted as a function of amino acid sequence.

5.4 Conclusions

The ability to detect nanosecond backbone dynamics with site-directed spin labeling in soluble proteins has been well-established. However, in membrane proteins, the nitroxide spin label has been previously shown to make contacts with hydrophobic residues on the protein's surface, which may interfere with the nitroxide dynamics coupling to the backbone.(*3*) Membrane protein backbone dynamics of model polytopic membrane protein, TM0026 were mapped with site-directed spin labeling and compared to NMR relaxation data. Nanosecond backbone dynamics were reflected in the EPR lineshapes of TM0026, and correlated with the ^{15}N R_1 relaxation data. Both the EPR and NMR data are in accord for describing the dynamics of the C- and N-termini, the transmembrane helices, and the flexible linker. The correlation between EPR lineshapes and the R_1 ^{15}N relaxation data suggests that nitroxide dynamics remain coupled to the backbone motions of membrane proteins. While certain labels were found to perturb TM0026's fold, the effects of labeling can also be investigated using NMR spectroscopy. Site-directed spin labeling methods can therefore, with some precaution, be employed to investigate the dynamics of membrane proteins in both detergent and lipid environments.

5.5 Acknowledgements

Analysis of the results is still ongoing, but many thanks are necessary for the hard work that Brett Kroncke and Tsega Solomon have put into this project.

Dan Fox for useful discussion regarding the heteronuclear NOEs.

5.6 References

1. Hubbell, W. L., Mchaourab HS, Altenbach C, Lietzow MA. (1996) Watching proteins move using site-directed spin labeling., *1996 4*, 779-783.
2. Columbus, L., and Hubbell, W. L. (2004) Mapping Backbone Dynamics in Solution with Site-Directed Spin Labeling: GCN4−58 bZip Free and Bound to DNA†, *Biochemistry 43*, 7273-7287.
3. Kroncke, B. M., Horanyi, P. S., and Columbus, L. (2010) Structural Origins of Nitroxide Side Chain Dynamics on Membrane Protein α-Helical Sites, *Biochemistry 49*, 10045-10060.
4. Columbus, L., and Hubbell, W. L. (2002) A new spin on protein dynamics, *Trends in Biochemical Sciences 27*, 288-295.
5. Wayne, L. H., David, S. C., and Christian, A. (2000) Identifying conformational changes with site-directed spin labeling, *Nature Structural & Molecular Biology 7*, 735-739.
6. Fanucci, G. E., Cadieux, N., Piedmont, C. A., Kadner, R. J., and Cafiso, D. S. (2002) Structure and Dynamics of the β-Barrel of the Membrane Transporter BtuB by Site-Directed Spin Labeling†, *Biochemistry 41*, 11543-11551.
7. Guo, Z., Cascio, D., Hideg, K., and Hubbell, W. L. (2008) Structural determinants of nitroxide motion in spin-labeled proteins: Solvent-exposed sites in helix B of T4 lysozyme, *Protein Science 17*, 228-239.
8. Langen, R., Oh, K. J., Cascio, D., and Hubbell, W. L. (2000) Crystal Structures of Spin Labeled T4 Lysozyme Mutants: Implications for the Interpretation of EPR Spectra in Terms of Structure†, *Biochemistry 39*, 8396-8405.

9. Columbus, L., Kálai, T., Jekö, J., Hideg, K., and Hubbell, W. L. (2001) Molecular Motion of Spin Labeled Side Chains in α-Helices: Analysis by Variation of Side Chain Structure†, *Biochemistry 40*, 3828-3846.

10. Huang, H., and Cafiso, D. S. (2008) Conformation and Membrane Position of the Region Linking the Two C2 Domains in Synaptotagmin 1 by Site-Directed Spin Labeling†, *Biochemistry 47*, 12380-12388.

11. (2006) *Methods in Enzymology*, Vol. 412, Academic Press.

12. Bracken, C., Carr, P. A., Cavanagh, J., and Palmer Iii, A. G. (1999) Temperature dependence of intramolecular dynamics of the basic leucine zipper of GCN4: implications for the entropy of association with DNA, *Journal of Molecular Biology 285*, 2133-2146.

13. Frock, A. D., Gray, S. R., and Kelly, R. M. (2012) Hyperthermophilic Thermotoga Species Differ with Respect to Specific Carbohydrate Transporters and Glycoside Hydrolases, *Applied and Environmental Microbiology 78*, 1978-1986.

14. Kroncke, B. M., and Columbus, L. (2012) Identification and removal of nitroxide spin label contaminant: impact on PRE studies of alpha-helical membrane proteins in detergent, *Protein Sci 21*, 589-595.

15. Kroncke, B., and Columbus, L. (2013) Backbone 1H, 13C and 15N resonance assignments of the α-helical membrane protein TM0026 from Thermotoga maritima, *Biomolecular NMR Assignments 7*, 203-206.

16. Columbus, L., Lipfert, J., Jambunathan, K., Fox, D. A., Sim, A. Y. L., Doniach, S., and Lesley, S. A. (2009) Mixing and Matching Detergents for Membrane Protein

NMR Structure Determination, *Journal of the American Chemical Society 131*, 7320-7326.

17. Kroncke, B. M., and Columbus, L. (2012) Backbone (1)H, (13)C and (15)N resonance assignments of the alpha-helical membrane protein TM0026 from Thermotoga maritima, *Biomol NMR Assign*.

18. Wishart, D. S., and Sykes, B. D. (1994) The 13C chemical-shift index: a simple method for the identification of protein secondary structure using 13C chemical-shift data, *J. Biomol. NMR 4*, 171-180.

19. Columbus, L., Lipfert, J., Klock, H., Millett, I., Doniach, S., and Lesley, S. A. (2006) Expression, purification, and characterization of *Thermotoga maritima* membrane proteins for structure determination, *Protein Sci 15*, 961-975.

20. Klock, H. E., and Lesley, S. A. (2009) The Polymerase Incomplete Primer Extension (PIPE) method applied to high-throughput cloning and site-directed mutagenesis, *Methods in Molecular Biology 498*, 91-103.

21. Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J., and Bax, A. (1995) NMRPipe: a multidimensional spectral processing system based on UNIX pipes, *J Biomol NMR 6*, 277-293.

22. Matthews, B. (1995) Studies on protein stability with T4 lysozyme, *Adv Protein Chem 46*, 249-278.

23. Eriksson, A. E., Baase, W. A., and Matthews, B. W. (1993) Similar Hydrophobic Replacements of Leu99 and Phe153 within the Core of T4 Lysozyme Have Different Structural and Thermodynamic Consequences, *Journal of Molecular Biology 229*, 747-769.

24. McHaourab, H. S., Lietzow, M. A., Hideg, K., and Hubbell, W. L. (1996) Motion of Spin-Labeled Side Chains in T4 Lysozyme. Correlation with Protein Structure and Dynamics†, *Biochemistry 35*, 7692-7704.

25. Faham, S., Yang, D., Bare, E., Yohannan, S., Whitelegge, J. P., and Bowie, J. U. (2004) Side-chain Contributions to Membrane Protein Structure and Stability, *Journal of Molecular Biology 335*, 297-305.

26. Yohannan, S., Faham, S., Yang, D., Whitelegge, J. P., and Bowie, J. U. (2004) The evolution of transmembrane helix kinks and the structural diversity of G protein-coupled receptors, *Proceedings of the National Academy of Sciences of the United States of America 101*, 959-963.

27. Cao, Z., and Bowie, J. U. (2012) Shifting hydrogen bonds may produce flexible transmembrane helices, *Proceedings of the National Academy of Sciences 109*, 8121-8126.

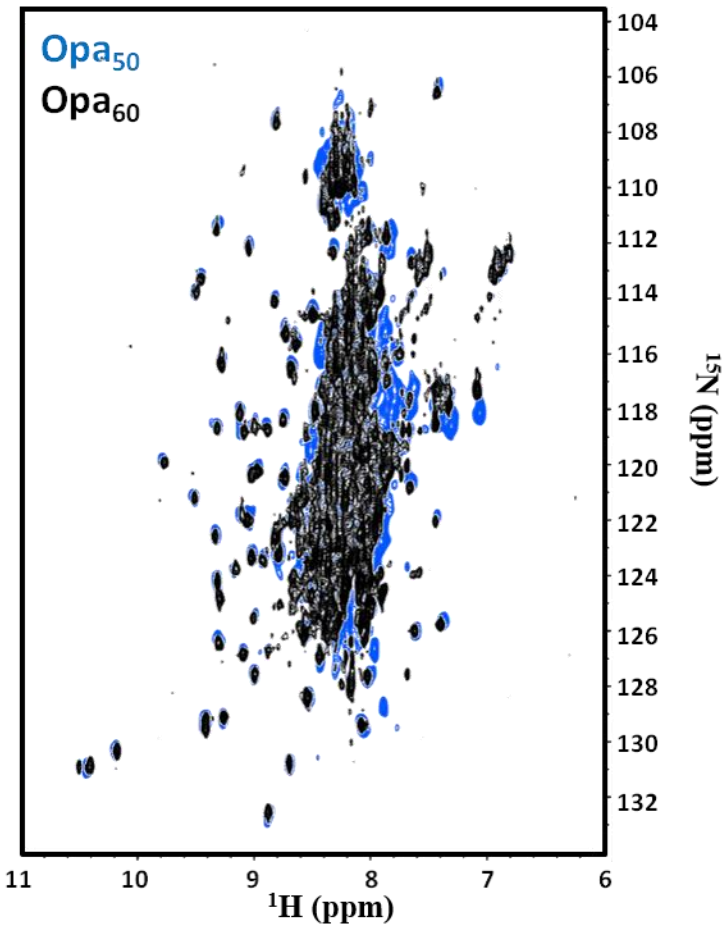
28. Riek, R. P., Rigoutsos, I., Novotny, J., and Graham, R. M. (2001) Non-α-helical elements modulate polytopic membrane protein architecture, *Journal of Molecular Biology 306*, 349-362.

29. Ishima, R., and Torchia, D. A. (2000) Protein dynamics from NMR, *Nat Struct Mol Biol 7*, 740-743.

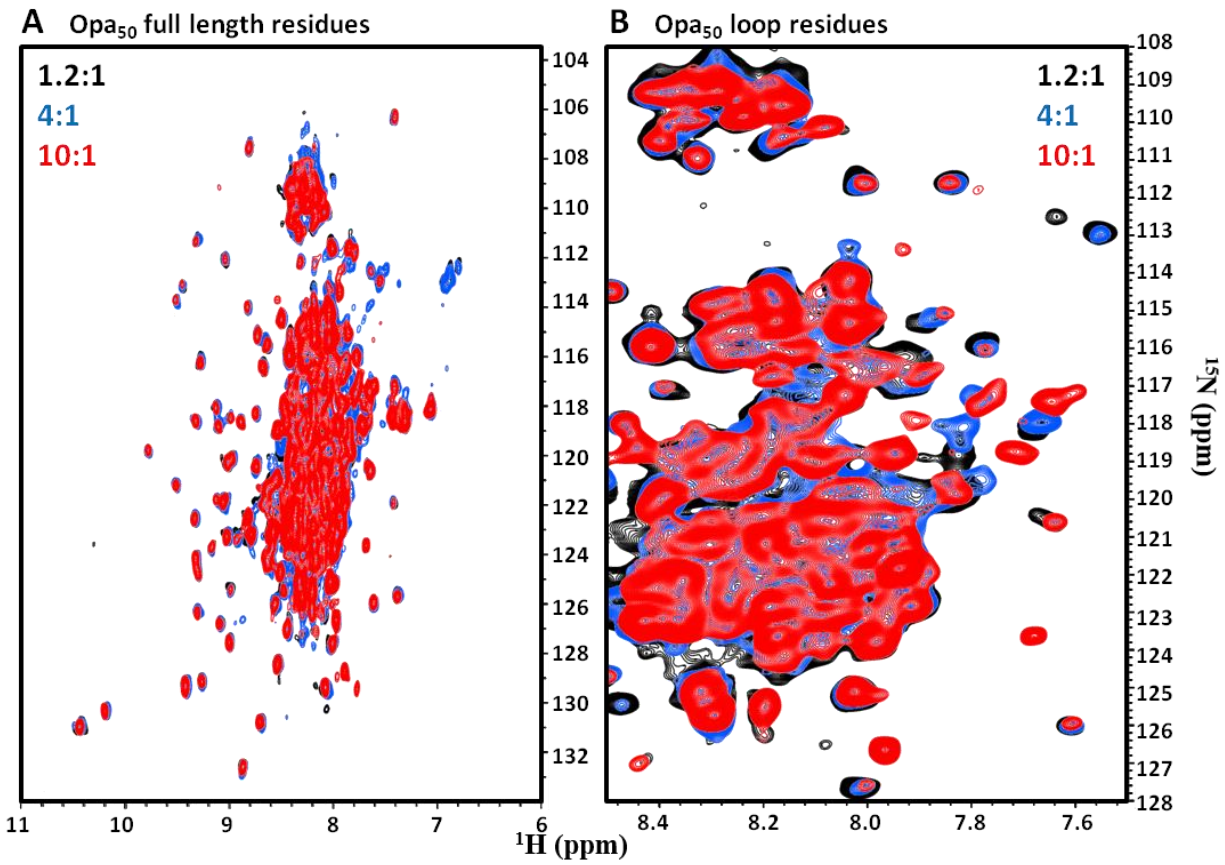
30. Liu, A., Hu, W., Majumdar, A., Rosen, M., and Patel, D. (2000) Detection of very weak side chain–main chain hydrogen bonding interactions in medium-size 13C/15N-labeled proteins by sensitivity-enhanced NMR spectroscopy, *Journal of Biomolecular NMR 17*, 79-82.

31. Kneller, J. M., Lu, M., and Bracken, C. (2002) An Effective Method for the Discrimination of Motional Anisotropy and Chemical Exchange, *Journal of the American Chemical Society 124*, 1852-1853.

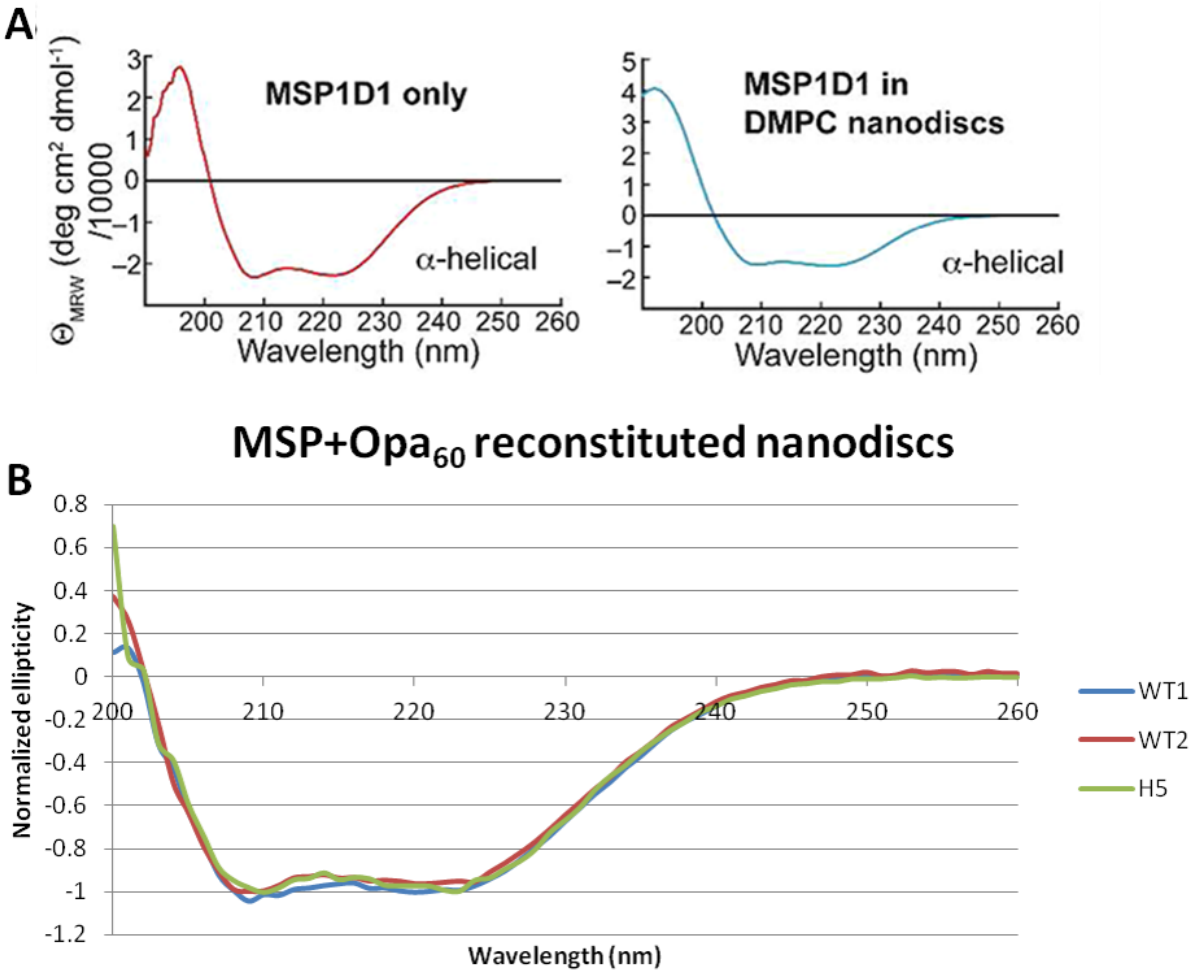
Appendix Figures



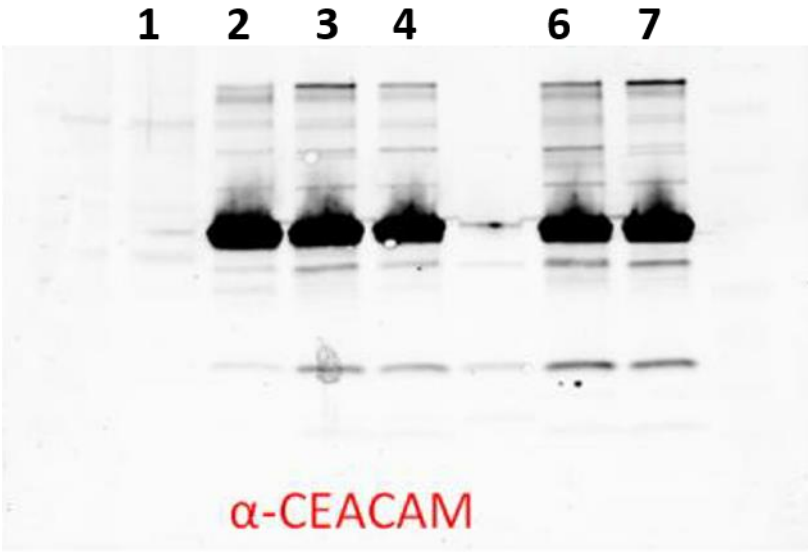
Appendix Figure A1. The $^{15}N,^{1}H$-HSQC spectral overlay of Opa_{50} and Opa_{60} in FC12. Opa_{50} (blue) and Opa_{60} (black) share significant spectral significant overlap.



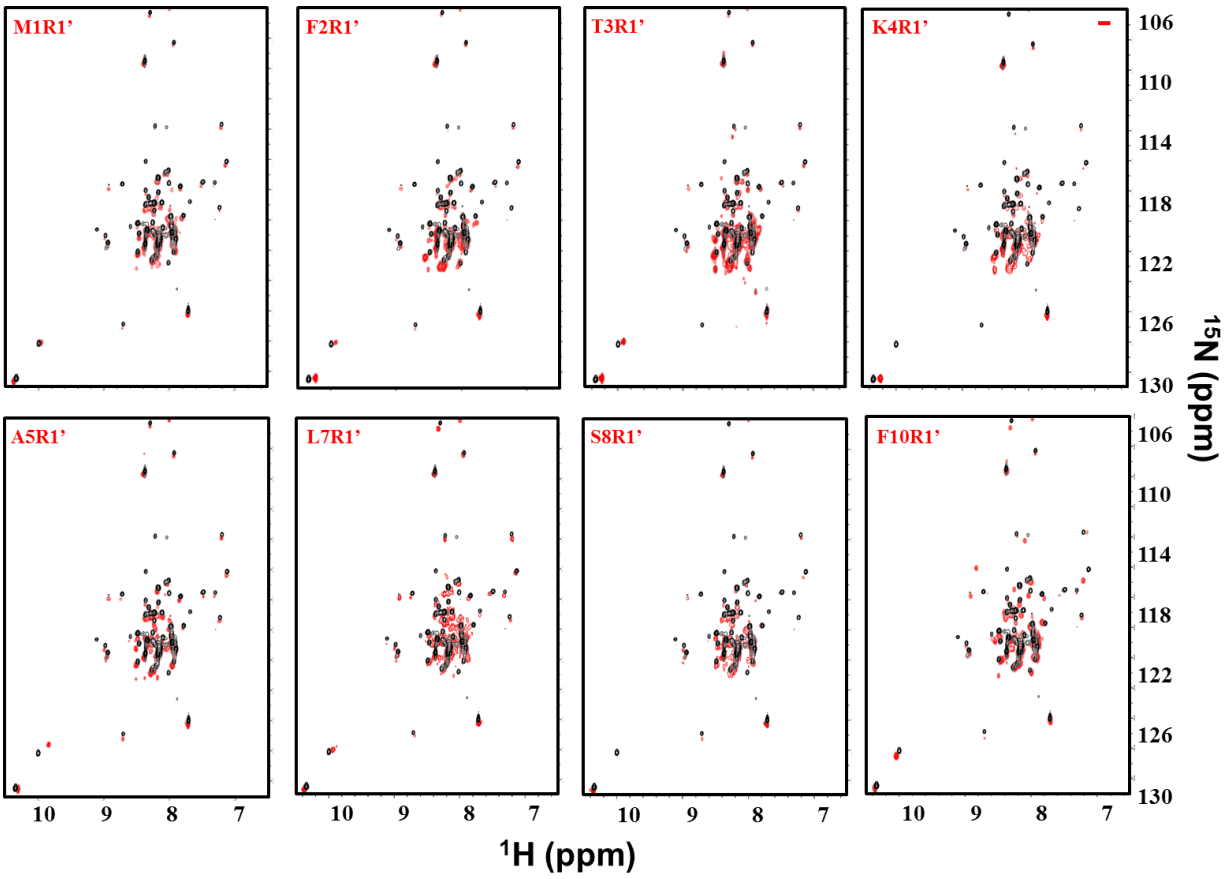
Appendix Figure A2. The ^{15}N, ^{1}H HSQC for Opa_{50} in FC12 micelles. The spectra recorded at a ratio of 1.2:1 (black), 2:1 (blue), and 3:1, 4:1, 5:1, and 10:1 (red) FC12 detergent micelle:Opa_{50} ratio were nearly identical.



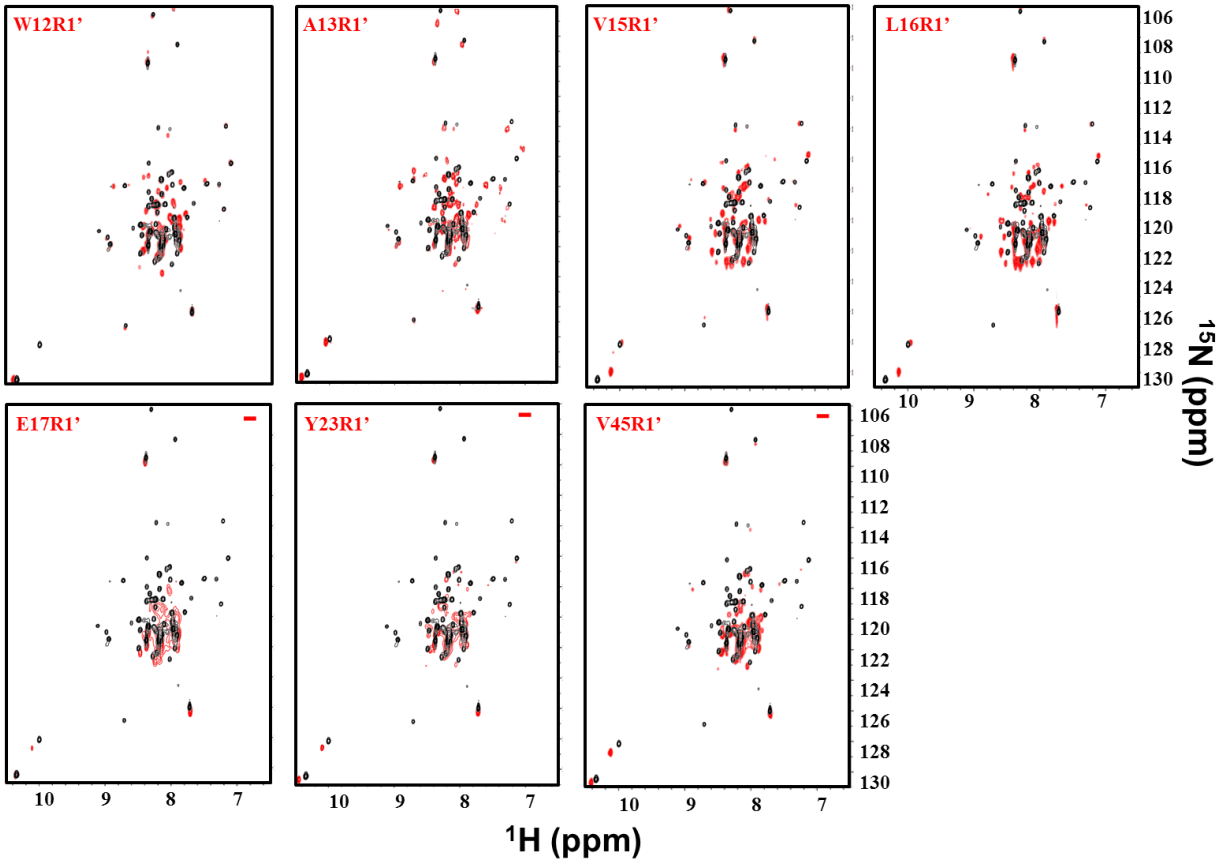
Appendix Figure A3. Construction of truncated membrane scaffold protein variants (MSP). (A) Far-UV spectra of membrane scaffold protein(MSP) variants both alone as well as in an assembled nanodisc, from Hagn *et al*. (B) Far-UV spectra of reconstituted MSP1D1 and Opa_{60} nanodisc variants – wild type MSP, as well as the truncated MSPΔH5. The predicted α-helical secondary structure matches the typical CD helical profile.



Appendix Figure A4. A Western blot of the results from multiple N-CEACAM1-GST - Opa_{50} experiments. Anti-CEACAM as the primary antibody. Lane 1 only has Opa_{60}, Lane 2 has NCEACAM1 only, Lane 3 has empty nanodiscs and N-CEACAM1-GST, and lanes 4, 5, and 6 have Opa_{60} with N-CEACAM1-GST.



Appendix Figure A5. The ^{15}N, ^{1}H HSQC NMR spectra for all TM0026 R1' mutants. (A) The spectra for M1R1', F2R1', T3R1', K4R1', A5R1', L7R1', S8R1', and F10R1'.



Appendix Figure A5. The ^{15}N, ^{1}H HSQC NMR spectra for all TM0026 R1' mutants. (B) The spectra for W12R1', A13R1', V15R1', L16R1', E17R1', Y23R1', and V45R1'.