

**NMR solution structure of Opa<sub>60</sub>: a *Neisserial* membrane protein that mediates host phagocytosis**


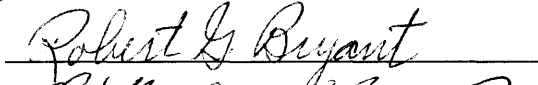
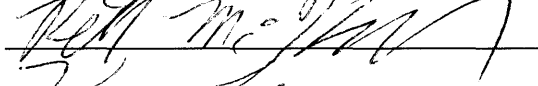


Daniel Andrew Fox  
Hatfield, PA

B.S. The University of Pittsburgh, 2007

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, 2013

© Copyright by  
Daniel Andrew Fox  
All rights reserved.  
May 2013

To my advisor, for seeing potential in me as a scientist and the ample opportunities for that potential to grow.

To my parents, for always encouraging me to better myself.

## Abstract

The family of Opa proteins from *Neisseria gonorrhoeae* and *N. meningitidis* are eight-stranded  $\beta$ -barrel outer membrane proteins that induce human cells to engulf the bacterium by engaging three different host receptors: carcinoembryonic antigen cellular adhesion molecules (CEACAM), heparansulfate proteoglycans (HSPG), or integrins via HSPG and fibronectin or vitronectin. The receptor engaged depends on the sequence of two of the extracellular loops (termed hypervariable (HV) loop 1 and 2), which are highly variable between isolates. The sequence variability is generated by multiple mechanisms including recombination among opa alleles (there are 11 opa alleles in *N. gonorrhoeae*), single point mutations, insertions and deletions, and insertion of opa genes from coinfecting species. There are multitudes of HV sequences identified; however, only approximately 25 Opa protein sequences have been characterized in terms of receptor engagement. Multiple sequence alignment of the HV loops does not reveal specificity motifs among the family of Opa proteins due to the extreme variability in the amino acid sequences. To investigate the determinants of Opa-receptor interactions, the NMR solution structure was determined.

In order to solve the structure, a suite of three dimension NMR experiments were performed to obtain an assignment, each optimized for the different domains of the protein. A variety of isotopic labeling techniques were also implemented to resolve crowded regions of the spectra. Opa<sub>60</sub> was cleaved using trypsin to isolate the micelle embedded  $\beta$ -barrel. NMR spectra for the dynamic extracellular loops were greatly improved at lower temperatures. Peptides corresponding to the most intrinsically dynamic regions of the hypervariable loops were synthesized to aid in assignment. The

structure of Opa<sub>60</sub> that was solved with these restraints was then refined using molecular dynamics simulations. With the aid of all these assignment strategies, the solution NMR structure was determined and reveals that Opa<sub>60</sub> is a canonical eight-stranded  $\beta$ -barrel and the HV loops are long, disordered, and highly dynamic that loosely associate with each other, displaying latent helical content in the hypervariable regions. Knowledge of the structure of Opa<sub>60</sub> will be integral in determining Opa<sub>60</sub>:host receptor interactions and may be of use for developing treatment for *Neisseria* related diseases

## Table of contents

Copyright page.....	i
Dedication .....	ii
Abstract .....	iii
Table of contents .....	v
List of figures .....	viii
List of tables.....	x
<b>Chapter 1: Membrane Protein Structure and Function .....</b>	<b>1</b>
Section 1.1 Protein Structure .....	1
Section 1.2 Membrane Protein Structure .....	5
Section 1.2.1 Physical properties of membrane proteins.....	5
Section 1.2.2 Functions of membrane proteins .....	7
Section 1.2.3 Outer membrane proteins.....	9
Section 1.3 Opacity-associated proteins.....	16
Section 1.3.1 <i>Neisseria meningitidis</i> and <i>N. gonorrhoeae</i> .....	16
Section 1.3.2 Opa protein function.....	17
Section 1.3.3 Carcinoembryonic antigen-like cellular adhesion molecules (CEACAMs).....	18
Section 1.3.4 Opa diversity .....	20
<b>Chapter 2: NMR Spectroscopy.....</b>	<b>35</b>
Section 2.1 Aspects of NMR Spectroscopy.....	36
Section 2.1.1 Nuclear spin .....	36
Section 2.1.2 Chemical shift.....	37
Section 2.1.3 Relaxation .....	39
Section 2.1.4 Dynamics .....	40
Section 2.2 Single Pulse Experiment .....	46
Section 2.3 Multidimensional NMR Spectroscopy.....	47
Section 2.3.1 Homonuclear NMR.....	47
Section 2.3.1.a COSY .....	47
Section 2.3.1.b TOCSY.....	48
Section 2.3.1.c NOESY .....	49
Section 2.3.1 Heteronuclear NMR.....	50
Section 2.3.2.a INEPT .....	50

Section 2.3.2.b HSQC.....	51
Section 2.3.1.c NOESY-HSQC, TOCSY-HSQC.....	52
Section 2.3.2.d Triple resonance spectra.....	52
<b>Chapter 3: NMR Protein Structure Determination.....</b>	<b>64</b>
Section 3.1 Protein Structure Determination Workflow .....	64
Section 3.1.1 Sequential assignment .....	64
Section 3.1.2 Sidechain assignment .....	65
Section 3.1.3 Torsion angles.....	66
Section 3.1.4 Through-space proton-proton interactions.....	67
Section 3.2 Techniques for larger systems .....	73
Section 3.2.1 Deuteration .....	73
Section 3.2.2 Methyl labeling .....	74
Section 3.2.3 Paramagnetic relaxation enhancement (PRE).....	75
Section 3.2.4 Transverse relaxation optimized spectroscopy (TROSY).....	77
Section 3.3 $\beta$ -barrel Structures Determined via NMR .....	83
Section 3.3.1 OmpX and OmpA.....	83
Section 3.3.2 OmpG and VDAC-1.....	85
Section 3.4 Overview.....	90
<b>Chapter 4: Backbone Assignment of Opa<sub>60</sub>.....</b>	<b>101</b>
Section 4.1 Stable Sample Preparation .....	102
Section 4.1.1 Refolding conditions .....	102
Section 4.1.2 Fold assessment .....	103
Section 4.2 NMR Assignment of Full Length Spectra .....	108
Section 4.2.1 Triple resonance experiments .....	108
Section 4.2.2 Specific amino acid isotope labeling.....	109
Section 4.3 Assignment of the $\beta$ -barrel Region.....	114
Section 4.4 Assignment of the Extracellular Loop Region.....	123
Section 4.4.1 Temperature effects on Opa <sub>60</sub> loops.....	123
Section 4.4.2 Synthetic peptide analysis .....	124
Section 4.5 Summary .....	132
<b>Chapter 5: Structure Calculations of Opa<sub>60</sub>.....</b>	<b>138</b>
Section 5.1 NMR Derived Structural Restraints .....	139
Section 5.1.1 NOE-based restraints of Opa <sub>60</sub> .....	139

Section 5.1.2 Dihedral angle restraints.....	140
Section 5.1.3 Membrane interface restraints.....	141
Section 5.2 NMR Determined Structure of Opa <sub>60</sub> .....	147
Section 5.3 Molecular Dynamics Refinement .....	151
Section 5.3.1 Molecular dynamics experiment .....	151
Section 5.3.2 Loop assessment .....	152
Section 5.3.2.a Loop position .....	152
Section 5.3.2.b Loop secondary structure .....	153
Section 5.3.3 Barrel assessment.....	153
Section 5.4 Function Significance of Structure .....	161
<b>Chapter 6: Prospects for Future Research .....</b>	<b>169</b>
Section 6.1 Overview .....	169
Section 6.2 Applications Towards Other Opa Protein Structures .....	170
Section 6.3 CEACAM Binding Studies.....	171
<b>Appendix: Materials and Methods .....</b>	<b>172</b>



## List of figures

Figure 1.1: Peptide bonds formed between amino acids .....	4
Figure 1.2: Representative integral membrane proteins .....	12
Figure 1.3: Wimley White hydrophobicity scale .....	13
Figure 1.4: Progress in membrane protein structure determination .....	14
Figure 1.5 Stabilizing interactions of outer membrane proteins .....	15
Figure 1.6: Micrograph of <i>Neisseria</i> invading cultured human cells .....	22
Figure 1.7: Proposed topology of Opa proteins .....	23
Figure 1.8: Crystal structure of N domain of CEACAM1 .....	24
Figure 2.1: Energy splitting of nuclear spin levels for a $^1\text{H}$ nucleus with respect to the magnetic field intensity. ....	42
Figure 2.2: Induced ring current .....	43
Figure 2.3 Effect of exchange rate on lineshape of NMR spectra .....	44
Figure 2.4 COSY pulse sequence and spectrum .....	54
Figure 2.5 Pulse sequence and spectrum of TOCSY experiment .....	55
Figure 2.6: Spin population of correlated heteroatoms and the INEPT pulse sequence .....	56
Figure 2.7: $^{15}\text{N}$ , $^1\text{H}$ -HSQC spectrum of OprH .....	57
Figure 3.1: Sequential connectivity used in small protein assignment .....	69
Figure 3.2: Magnetization transfer of HNCA experiment and Sequential assignment of HNCA strips .....	70
Figure 3.3 A. Magnetization transfer of H(CCO)NH, (H)C(CO)NH, and HCCH-TOCSY .....	71
Figure 3.4: Flowchart for TALOS+ database search procedure .....	72
Figure 3.5: Methyl-Methyl NOEs of OmpX .....	80
Figure 3.6 PRE distance restraints.....	81
Figure 3.7: Spectral quality improvements via TROSY .....	82
Figure 3.8: $\beta$ -barrel structures solved via solution NMR .....	87
Figure 3.9: Structural improvement of OmpX via methyl NOEs .....	88
Figure 3.10: Methyl contacts of N-terminus of VDAC-1 with barrel .....	89
Figure 4.1: Refolding of Opa <sub>60</sub> into different detergent micelles .....	105
Figure 4.2: Circular dichroism spectrum of Opa <sub>60</sub> .....	106
Figure 4.3: $^{15}\text{N}$ , $^1\text{H}$ TROSY-HSQC spectrum of Opa <sub>60</sub> .....	107
Figure 4.4: Spectral overlap of full length Opa <sub>60</sub> .....	111
Figure 4.5: Specific amino acid labeling of Opa <sub>60</sub> .....	112

Figure 4.6: Supplemented specific amino acid labeling of Opa <sub>60</sub> .....	113
Figure 4.7: SDS-PAGE of Opa <sub>60</sub> proteolysis .....	117
Figure 4.8: Proposed trypsin-cleaved region of Opa <sub>60</sub> .....	118
Figure 4.9: <sup>15</sup> N, <sup>1</sup> H HSQC of trypsin-treated Opa <sub>60</sub> .....	119
Figure 4.10: Overlay of trypsin-treated Opa <sub>60</sub> with trypsin-treated Opa <sub>50</sub> .....	120
Figure 4.11: Select assigned strips for amino acids G118-D129 of trypsin digested Opa <sub>60</sub> .....	121
Figure 4.12: Comparison of micellar interface amino acids of full-length and trypsin-cleaved Opa <sub>60</sub> .....	122
Figure 4.13: Overlay of Opa <sub>60</sub> <sup>15</sup> N, <sup>1</sup> H-HSQC spectra recorded at 10 and 40 °C .....	126
Figure 4.14: Select regions of <sup>15</sup> N, <sup>1</sup> H-HSQC spectra recorded over a series of temperatures .....	127
Figure 4.15: Overlay of HV2 peptide and Opa <sub>60</sub> <sup>15</sup> N, <sup>1</sup> H HSQC spectra recorded at 10 °C .....	128
Figure 4.16: <sup>1</sup> H, <sup>1</sup> H TOSCY spectrum of the 20 amino acid peptide .....	129
Figure 4.17: Backbone assignment of loop resonances using a synthetic peptide .....	130
Figure 5.1: Select strips from <sup>15</sup> N-edited NOESY spectrum of Opa <sub>60</sub> .....	143
Figure 5.2: Aromatic residues of Opa <sub>60</sub> at micellar interface .....	144
Figure 5.3: Environment NOE and water exchange peaks .....	145
Figure 5.4: Ensemble of structures generated in the absence of bilayer planar restraints .....	146
Figure 5.5: Ensemble of 20 lowest energy structures of Opa <sub>60</sub> derived from NMR restraints ..	149
Figure 5.6: Ensemble of Opa <sub>60</sub> structures after MD simulation .....	155
Figure 5.7: Experimental set up for MD simulations .....	156
Figure 5.8 Radius of gyration (R <sub>g</sub> ) of Opa <sub>60</sub> structures during MD simulation .....	157
Figure 5.9: Comparison of hypervariable regions of Opa <sub>60</sub> before and after MD simulation ....	158
Figure 5.10: Helical content of HV regions .....	159
Figure 5.11: Salt bridge formation in barrel .....	160

## List of tables

Table 1.1: Sequence alignment of HV receptors for a variety of CEACAM-binding Opa proteins .....	25
Table 2.1: Gyromagnetic ratio, spin quantum number and transition frequencies of commonly observed NMR .....	45
Table 5.1: NMR distance and dihedral angle constraints .....	150

## Chapter 1: Membrane Protein Structure and Function

### Section 1.1: Protein Structure

Proteins are diverse polymers made of various combinations of more than 20 different amino acids that perform many biological functions from structure and support to enzyme catalysis. The amino group of one amino acid reacts with the carboxyl group of another in a condensation reaction to form the peptide bond of a polypeptide. The  $\pi$ -character of this carbonyl-amide bond dictates that the nitrogen-hydrogen bond of the amide and the carbon-oxygen bond of the carboxyl will be restricted to the same plane (Figure 1.1). However, rotation around the central alpha-carbons is less restricted and is defined by dihedral angles phi and psi. These remaining dihedral angles are determined by a combination of steric and electrostatic restraints and can be grouped into three categories: 1.) The peptide structure may take an extended, or  $\beta$  sheet, conformation with a phi angle of approximately  $135^\circ$  and a psi angle of  $-135^\circ$ ; 2.) an  $\alpha$  helical state, in which phi and psi angles are clustered around  $-57^\circ$  and  $-67^\circ$ , respectively; or 3.) a random coil state which allows for a wider variety of phi and psi angles (Figure 1.1).

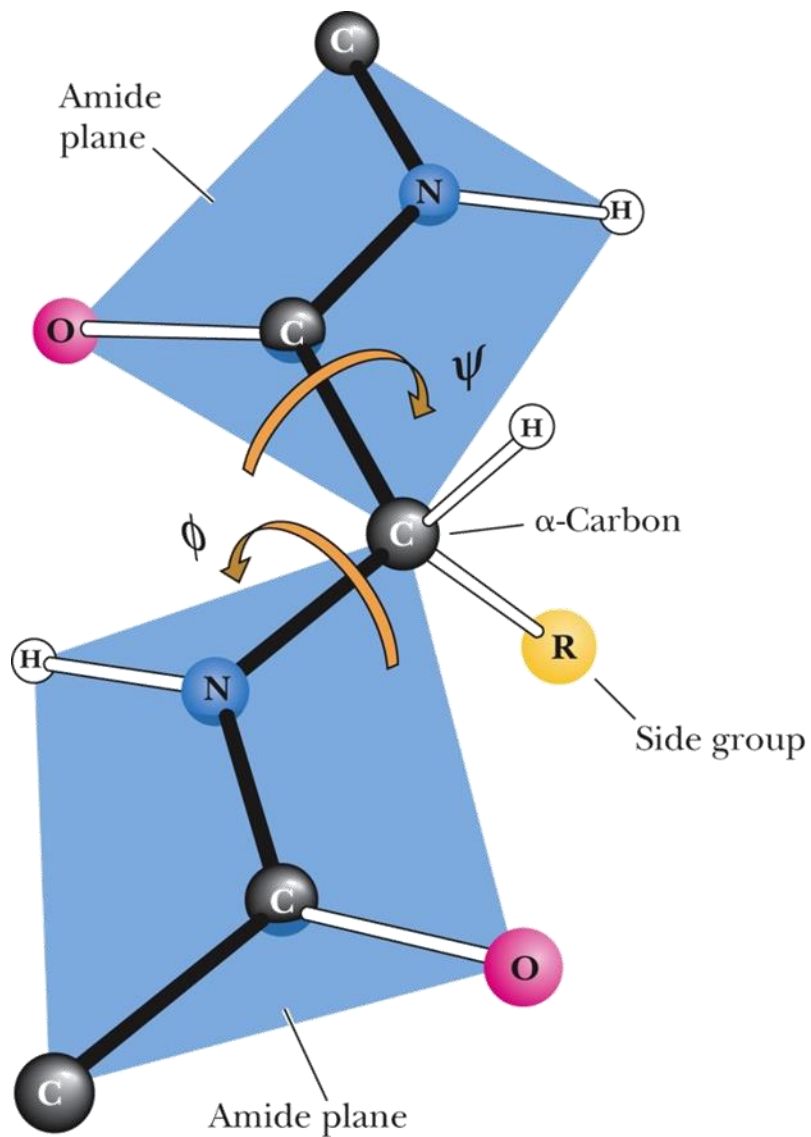
The conformation of the peptide backbone, and ultimately the overall tertiary structure of a protein results from the diversity, sequential neighbors and interactions of the side chains of each of the protein's amino acids. A wide variety of interactions may occur, including covalent, non-covalent monopole, dipole or induced dipole interactions. The protein's native conformation, or fold, is an

attempt to minimize the energies involved in these interactions, which may include hydrogen bonding ( $\Delta G$  of 1-10 kcal mol<sup>-1</sup>)(1), salt bridges (3-4 kcal mol<sup>-1</sup>)(2),  $\pi$ -cation (3 kcal mol<sup>-1</sup>)(3) and van der Waals interactions ( 50 cal mol<sup>-1</sup> Å<sup>-2</sup>)(4). Sterics, solvent, temperature and geometry also play large roles, each greatly affecting the stability, and structural preferences of each of these interactions.

While the weak force interactions listed above have favorable enthalpic contributions to the protein's structure, they carry an entropic loss associated with residue immobilization. This cost has been estimated to be between 4 and 19 cal mol<sup>-1</sup> K<sup>-1</sup>, or approximately 600 kcal mol<sup>-1</sup> for a moderately sized, 200 amino acid protein.(5, 6) The local environment of the sidechains greatly affects the total energy available. There is relatively little enthalpic gain (and a side-chain entropic cost) to the van der Waals interactions between hydrophobic sidechains. Yet, many hydrophobic residues are buried together in the core of globular protein structures. The burying of hydrophobic residues occurs because exposure of hydrophobic sidechains in an aqueous environment causes a reordering of the hydrogen bonding network of the solvent. Because the hydrophobic residues cannot form hydrogen bonds with the water molecules, the waters are forced to take on an entropically unfavorable, cage-like structure called a clathrate. The rearrangement of water molecules can result in an entropic penalty as high as 19 cal mol<sup>-1</sup> K<sup>-1</sup> per exposed non-polar residue.(6) Exposing a polar sidechain to water will result in less of a penalty, as water molecules can form hydrogen bonds with the sidechains, resulting in less

reordering, the penalty for an exposed polar residue varies from 0 to 10 cal mol<sup>-1</sup> K<sup>-1</sup>. (7) The hydrophobic effect plays a great role in protein folding.

For many proteins, the balance between these entropic and enthalpic costs results in a narrow temperature window in which the proteins form a stable fold. (8) In general the difference in free energy between a folded and unfolded soluble protein is only 5-15 kcal mol<sup>-1</sup>. (8)



$$\phi = 180^\circ, \psi = 180^\circ$$

**Figure 1.1: Peptide bonds formed between amino acids.** The planes that each amide nitrogen-hydrogen and carboxyl carbon-oxygen are restrained to are shown in green. The two dihedral angles that are free to rotate are highlighted with arrows and labeled as  $\psi$  and  $\phi$ , respectively. (Illustration: Irving Geis, reprinted with permission).

## **Section 1.2: Membrane Protein Structure**

A cell's inner machinery is sequestered from the extracellular environment by a cell membrane. All cells need to be able to interact with their extracellular environment, both to allow necessary ions and molecules to pass in and out and to interact with extracellular species. This unique subset of proteins, integral membrane proteins (IMPs), has evolved to be functionally stable embedded within and transversing the cellular membrane (Figure 1.2).

### **Section 1.2.1 Physical properties of membrane proteins**

For membrane proteins to be preferentially stable in the heterogeneous, largely apolar membrane environment, membrane proteins have unique amino acid sequences that allow for stable conformations in the lipid bilayer. Notably, the hydrophobic effect dictates the location of many of these residues, though the result is opposite of that in aqueous solvent. For protein regions spanning the apolar membrane, unfavorable water clathrate formation around exposed hydrophobic sidechains is not a possible protein-solution interaction. Instead, hydrophobic sidechains readily form van der Waals interactions with the apolar tails of the lipid bilayer, favoring their exposure.<sup>(9)</sup> Conversely, polar sidechains are rarely observed in the transmembrane region of the protein.

William Wimley and Stephen White investigated the preference of specific amino acids for partitioning into a membrane environment from an aqueous one.<sup>(10)</sup> The difference in free energy from water to bilayer is observed in Figure 1.3.



Unsurprisingly, the amino acids with favorable free energy in the transition into the bilayer are the most prevalent ones observed in the transmembrane region; More hydrophobic sidechains are commonly positioned to be exposed to the lipid environment. Lipid bilayers are approximately 30 Å in width from headgroup to headgroup, resulting in a minimum length of 20 amino acid residues for  $\alpha$ -helices (1.5 Å per residue) or 10 residues for  $\beta$ -strands (3.3 Å per residue) to span the membrane. Membrane proteins take on either  $\alpha$ -helical or  $\beta$ -strand secondary structure through transmembrane regions, and so plotting the hydropathy of the amino acid sequence is invaluable for predicting the topology of the transmembrane proteins. Specifically,  $\alpha$ -helices have a distinct 3.6 residues per turn.(11) while  $\beta$  strands follow a simpler repeating pattern of “exposed – buried” and are less reliably predicted based on hydropathy alone.(12-14)

The charged headgroups of the lipid bilayer are located at the interface of the membrane and the aqueous exterior. Membrane proteins are often rich with amphipathic residues, such as tyrosine, arginine, tryptophan and lysine at this interface.(15) These residues prefer to take a “snorkeling” conformation, winding aliphatic portions through the lipid tails and allowing their polar ends to surface with the polar headgroups.(16) The positioning of these amino acids (referred to as the “aromatic belt” due to the high abundance of tyrosine and tryptophan) with the headgroups is suggested to anchor the protein to the membrane, confining its orientation relative to the plane of the membrane.(17) Arginine and lysine residues are commonly located at the cytoplasmic interface for  $\alpha$ -helices and extracellular interface for  $\beta$ -barrels.(18) The reason for this

difference is attributed to the asymmetry of the outer membrane leaflets, which hosts  $\beta$ -barrel proteins. The outer leaflet contains many more charged headgroups from lipopolysaccharides (LPS) which contain negatively charged groups.(18)

### **Section 1.2.2 Functions of membrane proteins**

Membrane proteins are responsible for the vast majority of interactions that all cells have with their extracellular environment. As such, pharmaceuticals have made membrane proteins a principal target in developing drugs.(19) Approximately half of all drugs on the market target a membrane protein.(20) Thus, information gained about this class of proteins is in great demand. The most recent Noble Prize in Chemistry (2012) was awarded to Robert J. Lefkowitz and Brian K. Kobilka for their work on the membrane protein family of G-protein-coupled-receptors.(21)

Membrane proteins are not only important but are ubiquitous, comprising approximately 30% of the coding genome.(22) Despite this prevalence and a clear importance for both biological understanding and pharmaceutical research, membrane proteins are drastically under-represented compared to their soluble counterparts. Fewer than 1% of all unique structures deposited in the Protein Data Bank (PDB) are membrane proteins.(19) There was a 25 year gap from the time of the first soluble protein structure determined to that of the first membrane protein structure.(23) Despite the advances in technology over that time period, the rate of membrane protein structure determination still lags behind that of the

soluble proteins (Figure 1.4A).(24) Bottlenecks at expression, solubilization, and high resolution structural techniques have all contributed to this lack of data.

There are five main functional categories of membrane protein: receptor, enzyme, anchor, channel and transporter. Receptors function to bind a specific ligand (i.e., the G-Protein Coupled Receptor (GPCR) family can bind a variety of biomolecules, from hormones, glycoproteins, neurotransmitters to cytokines).(25) Upon ligand binding, the GPCR will undergo a specific conformational change which will then signal the cell's interior for a specific function to occur. GPCRs perform this action by promoting the exchange of bound guanine diphosphate (GDP) for guanine triphosphate (GTP) on the G protein, which results in downstream effects such as activating phospholipase or activating specific ion channels.(25)

Proteins which act as transporters or channels mediate the passage of a specific molecule or ion across the membrane. For example, LeuT, which is an amino acid transporter, undergoes a conformational rearrangement to transport amino acids across the membrane. LeuT, aliphatic and aromatic amino acids, such as leucine or tyrosine bind to a unique pocket.(26) Upon the addition of leucine and two sodium ions, two of the transmembrane helices partially unwind, allowing efficient hydrogen bonding with the amide and carboxyl groups of the leucine backbone as well as a hydrophobic pocket to stabilize the aliphatic sidechain, which allows for the amino acid to pass through.(27) This transport is effective with most aliphatic or aromatic amino acids, with the exception of tryptophan, which is suggested to be occluded due to a steric clash in which the amino acid

can bind but is too bulky to allow the protein to conformationally rearrange.(28) Transporters require an active energy source, as they are often transporting an ion against its electrochemical gradient, whereas channels allow the ions to flow down the electrochemical gradient and do not require added energy.

By contrast, channels such as the potassium ion channel KcsA has discrete pores which sterically and electrostatically allow only potassium ions to pass through. KcsA has a small pore lined with multiple carboxyl groups.(29) These carboxyls draw small charged ions, and can preferentially bind the positively charged ions, allowing them to shed their bound water molecules. The size of the pore dictates that only hydrated  $K^+$  ions can bind in these regions. These examples highlight just a few of the many ways in which membrane protein structure and conformation affect the roles that the membrane proteins partake in.

### **Section 1.2.3 Outer membrane proteins**

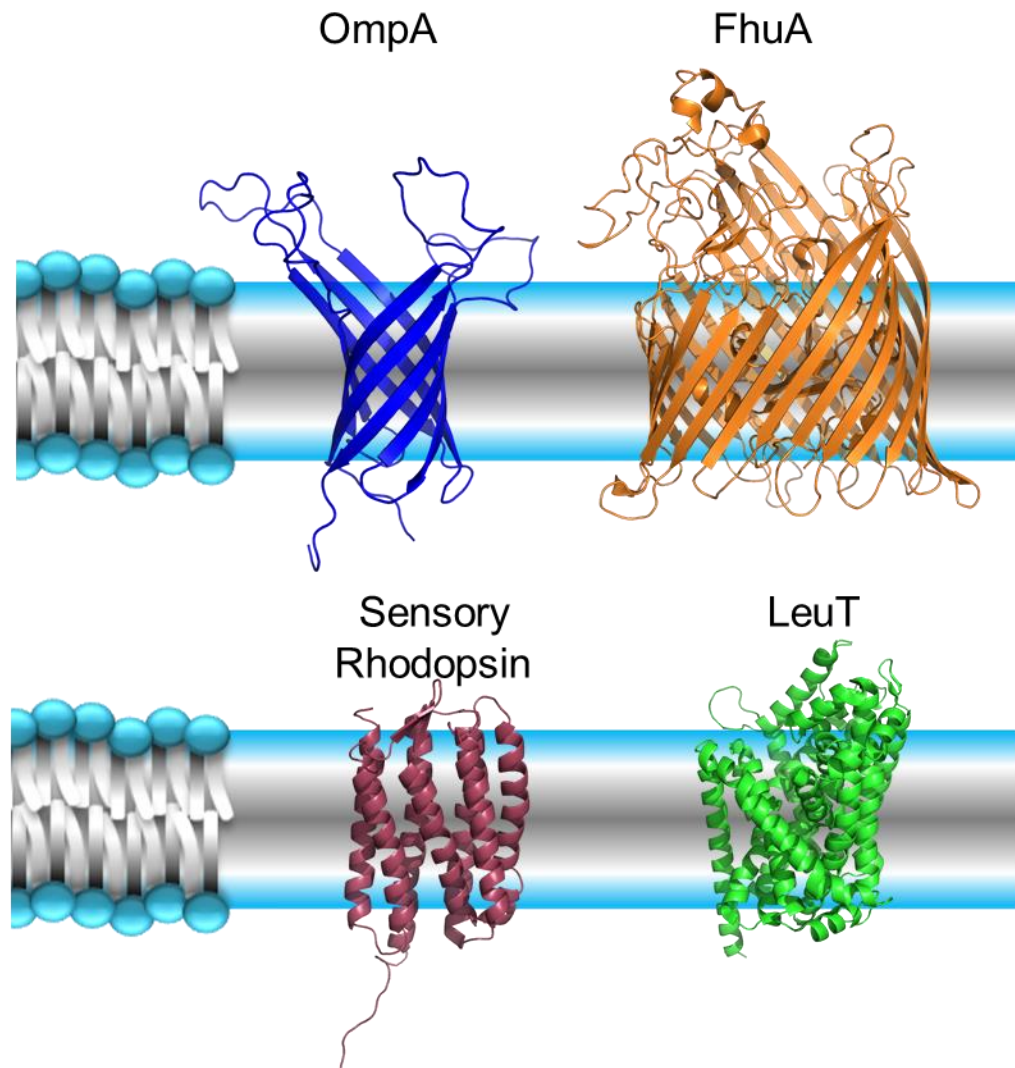
In order to satisfy the previously discussed folding energetics in the membrane,  $\beta$ -strands must expose predominantly nonpolar sidechains to the membrane and satisfy hydrogen bonding of the backbone polar groups. The only tertiary structure that meets these requirements is the anti-parallel  $\beta$ -barrel (Figure 1.5A). Anti-parallel  $\beta$ -barrel proteins are uniquely found in the outer membrane of gram negative bacteria and the outer membranes of organelles such as mitochondria and chloroplasts. These proteins, often called outer membrane proteins, or OMPs, are structurally unique compared to all other membrane protein because

their transmembrane regions are all defined by  $\beta$ -strands. Approximately 2-3% of all genes in Gram-negative bacteria code for outer membrane proteins, corresponding to anywhere from 40 to 120 unique OMP structures present per species.(30-32). To date, there are high resolution structures for 88 unique  $\beta$ -barrel proteins, ranging from relatively small eight-stranded barrels to massive barrels of 22 strands.(33, 34) Over 80 of these structures were determined with X-ray crystallography.(35)

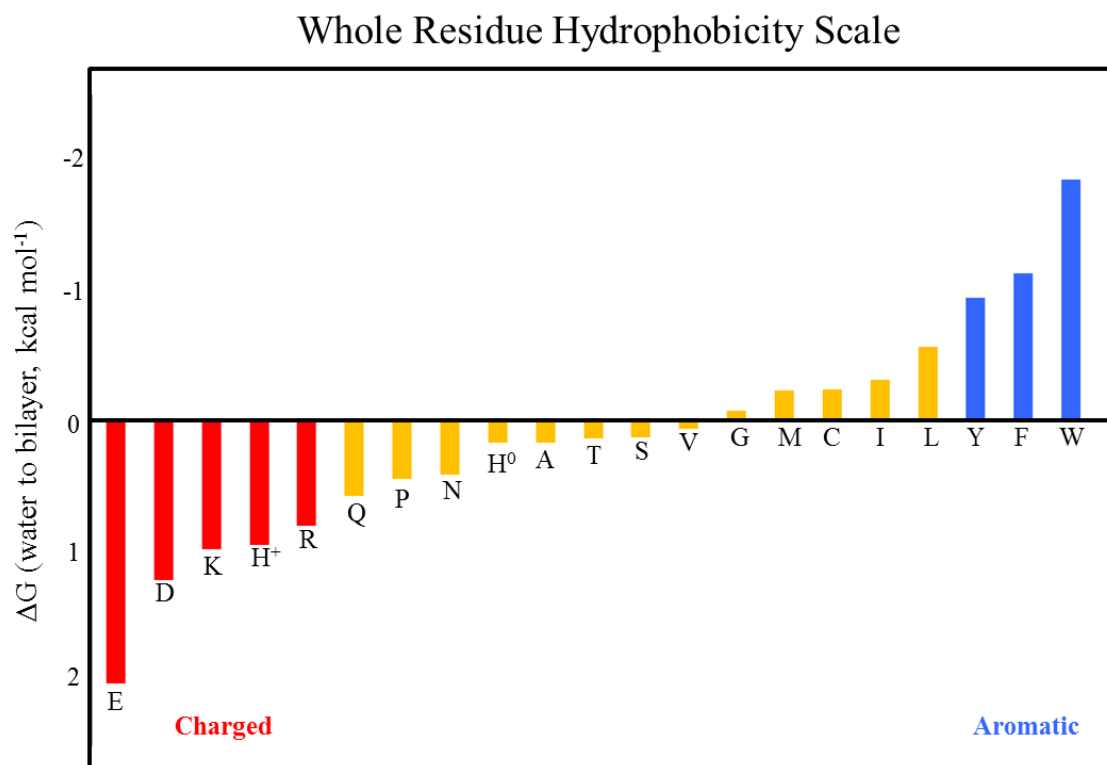
Functionally, outer membrane proteins are typically viewed as porins, channels, receptors or transporters.(36) The radius of the “pore” at the center of the barrel can range from small, proton exclusive channels, as observed in eight-stranded  $\beta$ -barrels, up to diameters greater than 30 Å, as calculated in crystal structures for barrels with 22 strands.(33, 37) The stability of the  $\beta$ -fold is remarkably high as nearly all polar backbone atoms are able to hydrogen bond with each other, creating a very strong network of approximately 30 kcal mol<sup>-1</sup> of hydrogen bonds formed between each strand (38). To stabilize such large structures,  $\beta$ -barrels have adopted four general folding motifs. For the larger barrels there is typically a separate domain, in-plug, that folds back into the interior of the barrel to create stabilizing contacts with the interior facing sidechains(Figure 1.5B).(39)  $\beta$ -barrels also utilize the out-clamp fold, in which an element will extend outside of the barrel and stabilize itself through external interactions, as seen with PagP (Figure 1.5C).(40) Additionally, many OMPs form oligomers which are stabilized by interactions with another OMP.(41) Additionally, OMP tertiary structure can be by specifically binding to lipids in the bilayer.(42) Finally, for the smaller eight-

stranded  $\beta$ -barrels, salt-bridges can form across the interior sidechains for stabilization.<sup>(43)</sup> These ionic interactions are strong forces, as there is rarely any water or other ions inside the barrel to compete with, resulting in very stable interactions.<sup>(44)</sup>

The work presented in this dissertation focuses on improving understanding of the structurally and functionally important features of Opacity Associated Protein (Opa), an OMP in *Neisseria*, a specific OMP.

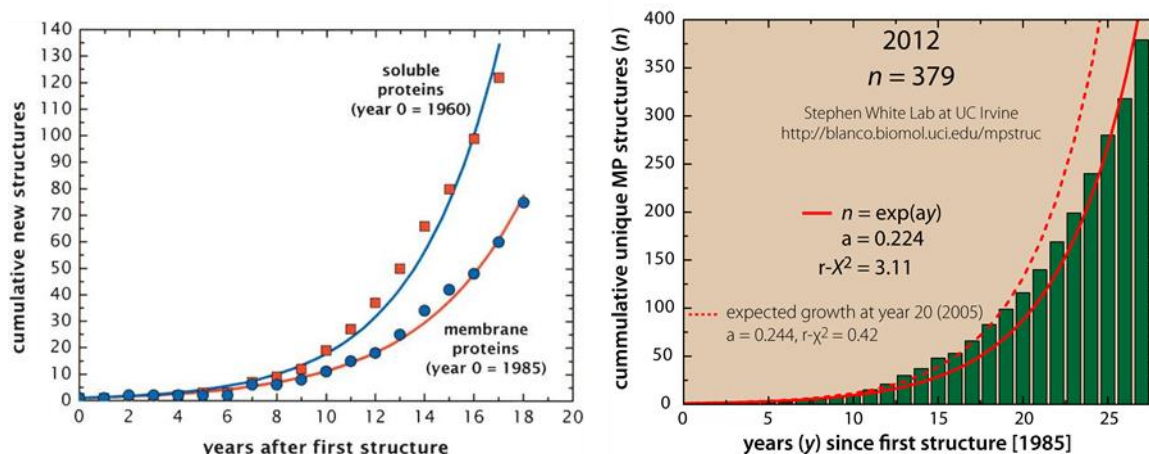


**Figure 1.2: Representative integral membrane proteins.** A sampling of both  $\beta$ -barrel (top) and  $\alpha$ -helical membrane proteins. The transmembrane segments are shown crossing the colored membrane bilayer with the extracellular region shown as the top part of the figures.



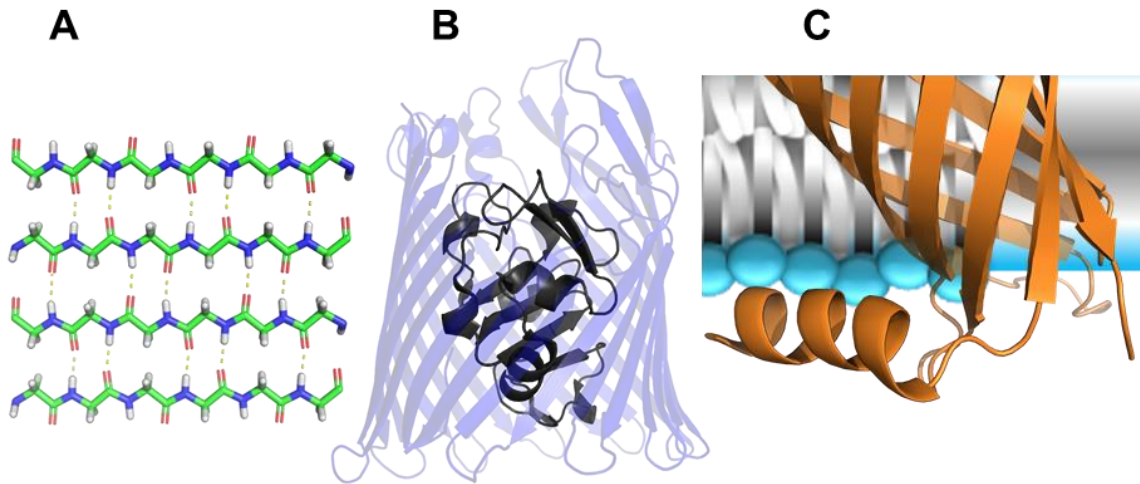
**Figure 1.3: A. Wimley White hydrophobicity scale.** Experimentally determined free energy of solvation from an aqueous environment to the aliphatic tails in lipid bilayers for each of the twenty amino acids.(10)





**Figure 1.4: A. Progress in membrane protein structure determination.**

Cumulative new structures of both soluble and membrane protein plotted over the 20 years after the first structure of each was determined. Both growth curves can be fitted with an exponential. As this trend continues, the slower growth rate of membrane proteins will continue to form a larger gap between soluble and membrane proteins.(24) **B. Updated growth curve.** The growth of membrane protein structure determination revisited from the past 7 years indicating that the exponential growth predicted in 2004 exceeded actual progress.(35)



**Figure 1.5 Stabilizing interactions of outer membrane proteins. A. Anti-parallel  $\beta$ -strand hydrogen bonding network.** In the anti-parallel  $\beta$ -barrel conformation, all polar residues on the backbone of the peptide are hydrogen bonded to each other, representing an enthalpically favorable condition in the presence of aliphatic lipid tails. **B. In-plug mechanism of BtuB.** For larger  $\beta$ -barrels, a separate domain (grey) folds back into the “pore” of the barrel (blue) to stabilize the fold as well as provide specific function for the protein. **C. Out-clamp mechanism of PagP.** The N-terminal helix of the eight-stranded  $\beta$ -barrel, PagP (orange) is shown interacting with the cartoon headgroups of the lipid bilayer, allowing the protein to stabilize its fold within the membrane.

### **Section 1.3: Opacity-associated proteins**

Opacity-associated (Opa) proteins are OMPs in the Gram-negative bacteria *Neisseria meningitidis* and *N. gonorrhoeae* that mediate bacterial engulfment by host cells. There are multiple Opa proteins with unique binding partners, determined by unique protein sequences in the receptor binding domain of the protein. These proteins are essential for the *Neisseria* pathogen to invade host cells.(45)

#### **Section 1.3.1 *Neisseria meningitidis* and *N. gonorrhoeae***

*Neisseria meningitidis* and *N. gonorrhoeae* are two species of pathogenic bacteria whose only known targets are humans (Figure 1.6).(46) *Neisseria meningitidis* is responsible for the most common form of bacterial meningitis. Whereas *N. gonorrhoeae* is responsible for gonorrhea, one of the most prevalent STDs, with over 700,000 new cases each year in the U.S. alone.(47) Both bacteria infect through colonization on mucosal surfaces; either the genitor-urinary tract in the case of gonorrhea or the nasopharynx in the case of bacterial meningitis.(48) These diseases can persist asymptotically for extended periods of time due to *Neisseria*'s ability to evade human immune response (47). Not only is *N. gonorrhoeae* capable of evading the immune response, the bacteria develops antibiotic resistance so quickly, that some strains have gained "super bug" status in the last decade.(49) An untreatable strain of *N. gonorrhoeae* was identified in Japan just two years ago with resistance to all known antibiotics.(49) Vaccines have failed to stimulate an immune response

and individuals who have previously been infected actually remain highly susceptible to re-infection.(48) *Neisseria meningitidis* causes approximately 500,000 cases of meningitis and septicaemia worldwide every year with a mortality rate of approximately 10%.(48)

### **Section 1.3.2 Opa protein function**

*Neisseria* are well adapted in invading human host cells. A key component of this invasion is a family of Opa proteins that are unique only to these two species. Opa proteins are OMPs that mediate interactions of *Neisseria* with various host cell types.(50) The *Neisseria* pathogen makes initial contact with either human immune or epithelial cells via adhesins and pili, anchoring the bacterium to the host cell.(51) Upon initial binding, Opa proteins interact with specific receptors on the human cell to induce the uptake of the bacterium into the host cells.(45, 51) This can occur even on host cells which do not normally undergo phagocytosis.(45)

Opa outer membrane proteins are named from the opaque phenotype of *Neisseria* colonies that are expressing Opa proteins.(52) The presence of Opa, in the absence of both pili and adhesins, is sufficient to induce engulfment, as *E. coli* engineered to express Opa proteins was phagocytosed by HeLa cells overexpressing the appropriate receptors.(51)

The family of Opa proteins is diverse, with the genomes of both *N. meningitidis* and *N. gonorrhoeae* containing at least 15 different *opa* loci (Four in *N. gonorrhoeae* and eleven in *N. meningitis*). (53, 54) Structure prediction indicated

that the Opa proteins are eight-stranded  $\beta$ -barrels (Figure 1.7A). The amino acid sequence of these proteins, specifically in the transmembrane barrel region, is highly conserved. Approximately 70% of these residues are identical across the entire family. One of the four extracellular loops (Loop 4) is well conserved and referred to as the conserved loop, (CL). There are three regions that have high sequence variability among the various Opa proteins. Two regions in extracellular loops 2 and 3 are the most variable, and so are named hypervariable loops 1 and 2 (HV1 and HV2), respectively. There is also a smaller semi-variable region (SV) in loop 1 (Figure 1.7B). The sequence in these loop regions determines the specific binding partners that lead to phagocytosis.<sup>(45)</sup> Opa is classified into two groups based on the receptor that the proteins interact with; either carcinoembryonic antigen-like cellular adhesion molecules (CEACAMs) or heparin sulfate proteoglycans (HSPGs; or integrin receptors through an HSPG-mediated interaction with fibronectin and vitronectin).<sup>(45)</sup> The vast majority of Opa proteins studied to date are classified as CEACAM binding proteins.

### **Section 1.3.3 Carcinoembryonic antigen-like cellular adhesion molecules (CEACAMs)**

CEACAM proteins serve a vital function as cell-cell adhesion, guiding cells to the proper locations during embryonic development and mediating the development of single cells into functional tissues.<sup>(55, 56)</sup> The CEACAM family is classified as an Ig cellular adhesion molecule (IgCAM). IgCAMs have been identified as modulators in apoptosis, cell proliferation, insulin regulation and

angiogenesis.(55, 56) All molecules in the Ig superfamily have at least one immunoglobulin (Ig-) domain. These domains are typically between 85-110 amino acids and are comprised of two  $\beta$ -sheets positioned in a “sandwich-like” fold (Figure 1.8).(56) This fold has been described as a “universal interface that can be fine-tuned for almost every binding task.” (57)

There are 7 members of CEACAM family, CEACAM1, and CEACAMs3-8. Each CEACAM consist of an N-terminal domain with an Ig fold and a various number of Ig repeat (0-6 repeats, depending on the CEACAM).(51) The CEACAM molecules are anchored to the cell membrane by either a transmembrane domain (CEACAM1, 3, and 4) or by a glycoposphatidylinositol (GPI) anchor. (55, 56)

Opa proteins are known to interact with CEACAM on the non-glycosylated face of the N-terminal Ig domain (Figure 1.8).(58) The N-terminal domain has a high sequence identity amongst the CEACAM proteins.(59) Point mutations of various CEACAMs revealed a pair of conserved residues in all CEACAMs, Tyr34 and Ile91 (of CEACAM1), that were required for binding to various strains of *Neisseria* (Figure 1.8).(59) Residues Gly41, Gln44, Val39, and Gln89 have shown to also have a role in various Opa specificity.(59-61) Mapping these residues on the crystal structure of CEACAM1 reveals that they are all located in a specific binding region which is concave and hydrophobic.(62) *Neisseria* has evolved several binding sequences to interact with these human binding targets.

### Section 1.3.4 Opa diversity

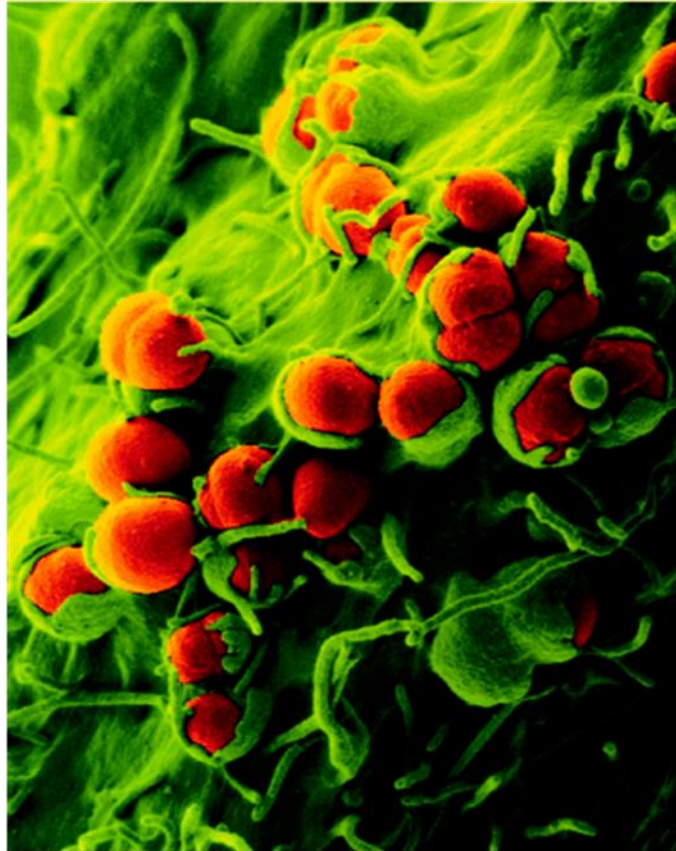
A single gonococcal strain can express anywhere from 0 to 11 *opa* genes (up to 4 for meningococcal) that can be turned on or off through translational phase variable expression.(63) In the *Neisserial* genome, there are pentameric sequence repeats in the Opa coding region. Strand slippage during translation results in a frequency phase shift which leads to different *opa* genes being read in frame. Each individual bacterium can express zero, one, or multiple Opa proteins at any given time and has the ability to switch which and how many Opa proteins are being expressed.(63)

There are only approximately 15 different *loci* for Opa proteins; however, there have been many more Opa protein sequences observed. Recombination and point mutations lead to the diversity observed within Opa proteins.(50) Full and partial *opa* genes are capable of recombining between loci of a single organism as well as with other organisms.(50) *Neisseria* are highly competent; thus, incorporating exogenous DNA sequences into their own chromosome leads to additional Opa protein production.(50) The majority of *opa* genetic variation has been attributed to this recombination, accounting for approximately three quarters of the mutations.(50) Despite a seemingly impossibly large variation for possible *opa* alleles based on these genetic variations (over 300,000), only 338 unique hypervariable sequences have been observed.(50) The reason for this relatively small sample may be a result of the other 99.9% of alleles producing either unstable proteins or perhaps more likely, functionally inactive proteins.

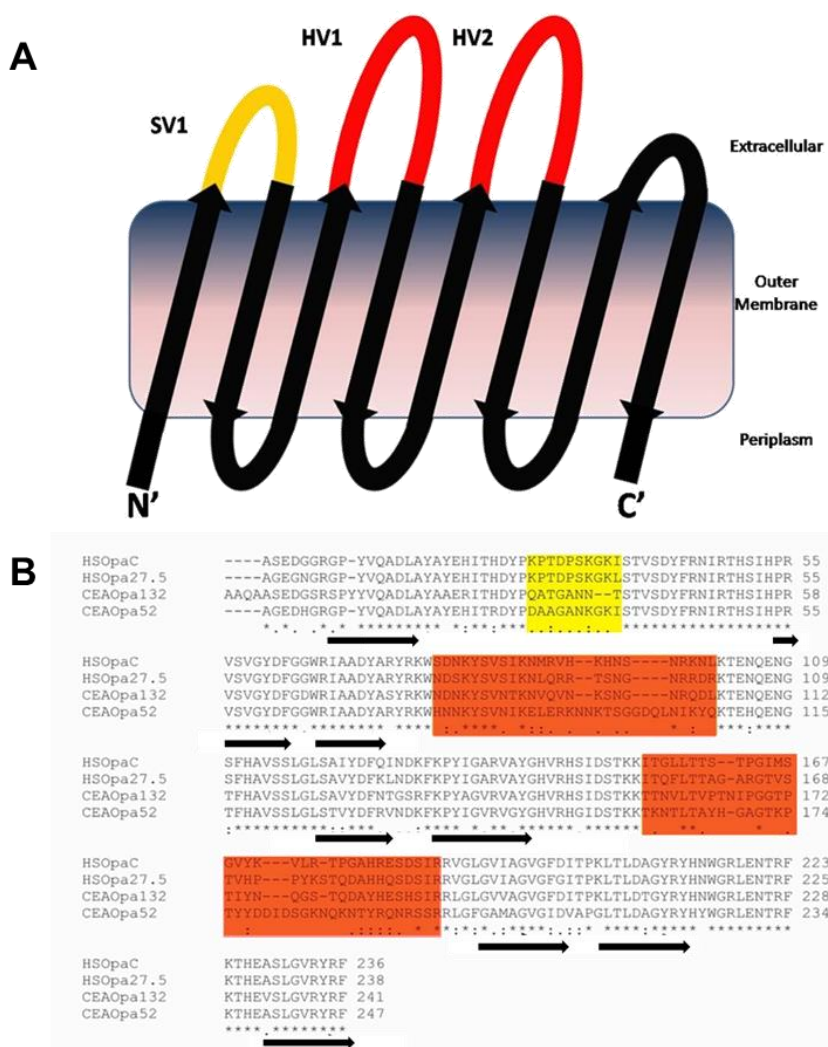
Evolutionarily, the bacteria producing these inactive proteins would be less likely to survive and promote these alleles.

There is still not a complete understanding as to what sequential motifs in the HV regions generate functionally active Opa proteins. Attempts have been made to determine how the Opa proteins are able to select for specific binding partners, despite their large sequential variability.<sup>(51)</sup> In observing the different sequences associated with the two hypervariable loops, both loop sequences were necessary for binding specificity.<sup>(51)</sup> In some cases there were identical HV1 or HV2 sequences between different Opa proteins, yet they both had distinctly different binding partners (Table 1.1). Furthermore, in chimeric studies in which CEACAM-binding Opa proteins were engineered to contain an HV1 loop from one Opa protein and an HV2 loop from a different sequence, binding and invasion decreased significantly for all chimeras generated.<sup>(51)</sup> However, HSPG binding Opa proteins had significant binding as long as the original HV2 remained intact.

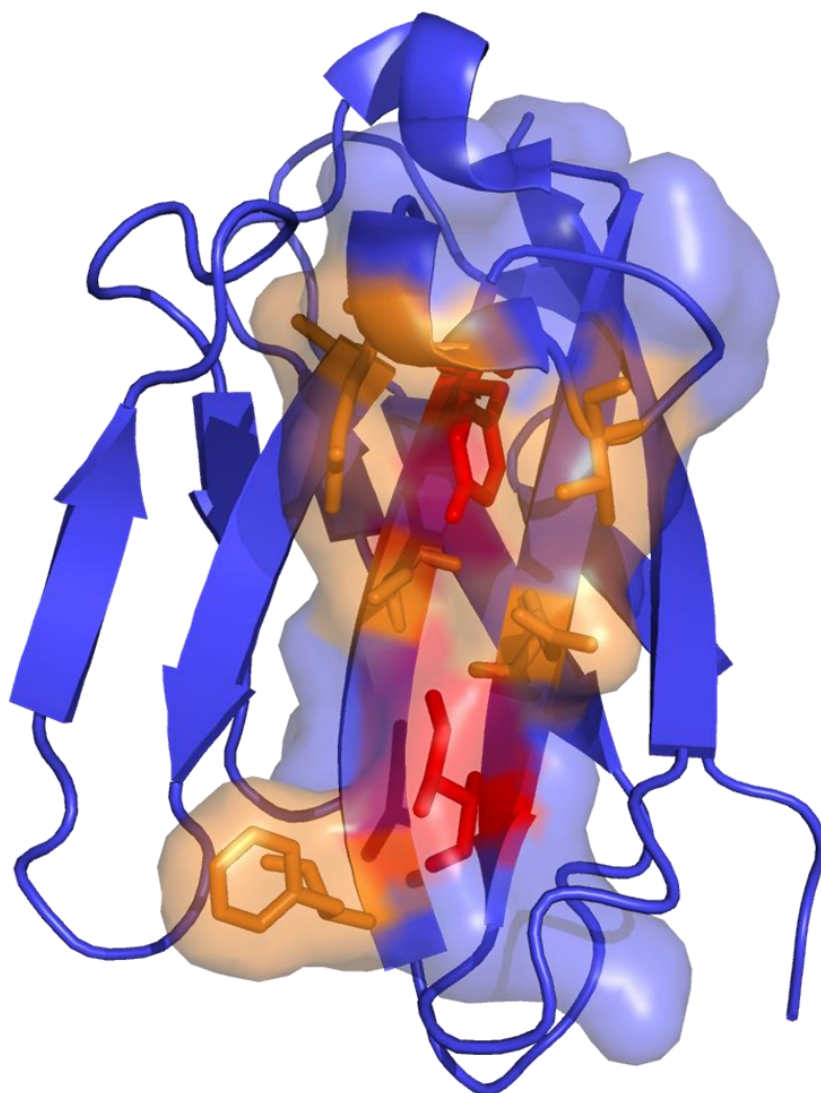




**Figure 1.6: Micrograph of *Neisseria* invading cultured human cells.** *Cover image of (58);* A false colored scanning electron micrograph is shown *Neisseria gonorrhoeae* (red) invading HeLa cells (green) expressing an Opa receptor, CEACAM1.



**Figure 1.7: A. Proposed topology of Opa proteins.** Opa proteins are predicted to be eight stranded  $\beta$ -barrels with four extracellular loops. The loops are classified based on the sequence variability amongst the Opa family with the semi-variable loops (SV) highlight in yellow, and the two hypervariable loops (HV1 and HV2) shown in red. **B. Sequential alignment of four Opa proteins.** An alignment of four different Opa (two HSPG-bind, two CEACAM binding) proteins with the semivariable (yellow) and hypervariable regions (red) highlighted. The rest of the protein alignment shows a high sequential conservation.



**Figure 1.8: Crystal structure of N domain of CEACAM1** (pdb: 2GK2).(62)

Conserved residues involved in all Opa binding, Tyr34 and Ile91, are colored in red. Residues implicated in different Opa protein interactions are colored orange. The binding region is rendered as a surface.

HV1 Sequence	HV2 Sequence	CEACAM			
		1	3	5	6
NDNKYSVNTKNVQVNKSNGN-RQDLK	KHQVHSVESKTTIVTSKPTKGATQPGKLVSGPTPKPAYHESNSISSL	√	√	√	√
NDNKY.VNT+N*QVNKSNGN-RQD*+	+ .S*D.*K-K..N*LTVPNTNIPGGTPT*YNQGSTQDA...H.*RRL	√	X	√	√
NDNKY.VNT+N*QVNKSNGN-RQD*+	+ .Q*H.*KKE..T*FLAPTGDADVPGK*VEGPFSKPA...H.*SSL	√	X	√	√
NNNYY.VNT+E*RRNNNAGN-WRE*+	+ .Q*H.*KKE..T*FLAPTGDADVPGK*VEGPFSKPA...H.*SSL	√	X	√	X
NNNKY.VNT+E*QRNNSNGTTWKE*+	+ .Q*H.*RKE..T*FSKPSGSTTKPGE*PS-LVTKPA...N.*SSL	√	X	√	X
NNNKY.VNT+E*ENKNNNK--RD*+	+ .S*D.*KKT.EV*TSSHA--PGTAPT*YNVPKTQNA...H.*RRL	√	X	√	X
KESNY.---+K*TEFKHQNGNKQED+	+ .Q*H.*ESK..I*TSKPTKGATQPGK*VSGPTPKPA...N.*SSL	√	X	√	X
KESNY.---+K*TEFKHQNGNKQED+	+ .Q*H.*ESK..T*TTN--NGGPVP---QGPTPKPA...H.*SSV	√	X	X	X
KESNS.T--K*TEEINNN--YKETQ	+ .Q*H.*ETK..T*TSKPKNGSPQGGP*IQTDPSKPP...H.*SSV	√	X	X	X
KESNF.T--K*TEEIKDN--YKET+	+ .Q*H.*ETK..T*TSKPKGGTPAGGP*IKTDPSKPP...H.*SSL	√	X	X	X

**Table 1.1: Sequence alignment of HV receptors for a variety of CEACAM-binding Opa proteins.** The HV1 and HV2 regions of several CEACAM binding Opa proteins are aligned. + indicates conservation of charge, \* conservation of hydrophobic residue, . conservation of polar residue and - indicates a gap. The variability and lack of specific motifs has hindered deducing the molecular determinants of the Opa-CEACAM interactions.

1. Wahl, M. C., and Sundaralingam, M. (1997) C-H...O hydrogen bonding in biology, *Trends Biochem Sci* 22, 97-102.
2. Kumar, S., and Nussinov, R. (1999) Salt bridge stability in monomeric proteins, *J Mol Biol* 293, 1241-1255.
3. Gallivan, J. P., and Dougherty, D. A. (1999) Cation-pi interactions in structural biology, *Proceedings of the National Academy of Sciences of the United States of America* 96, 9459-9464.
4. Makhatadze, G. I., and Privalov, P. L. (1993) Contribution of hydration to protein folding thermodynamics. I. The enthalpy of hydration, *J Mol Biol* 232, 639-659.
5. Lee, K. H., Xie, D., Freire, E., and Amzel, L. M. (1994) Estimation of changes in side chain configurational entropy in binding and folding: general methods and application to helix formation, *Proteins* 20, 68-84.
6. Privalov, P. L., and Makhatadze, G. I. (1993) Contribution of hydration to protein folding thermodynamics. II. The entropy and Gibbs energy of hydration, *J Mol Biol* 232, 660-679.
7. D'Aquino, J. A., Gomez, J., Hilser, V. J., Lee, K. H., Amzel, L. M., and Freire, E. (1996) The magnitude of the backbone conformational entropy change in protein folding, *Proteins* 25, 143-156.
8. Pace, C. N. (1990) Conformational stability of globular proteins, *Trends Biochem Sci* 15, 14-17.

9. White, S. H., and Wimley, W. C. (1999) Membrane protein folding and stability: physical principles, *Annual review of biophysics and biomolecular structure* 28, 319-365.
10. Wimley, W. C., and White, S. H. (1996) Experimentally determined hydrophobicity scale for proteins at membrane interfaces, *Nature structural biology* 3, 842-848.
11. Kyte, J., and Doolittle, R. F. (1982) A simple method for displaying the hydropathic character of a protein, *J Mol Biol* 157, 105-132.
12. Cserzo, M., Wallin, E., Simon, I., von Heijne, G., and Elofsson, A. (1997) Prediction of transmembrane alpha-helices in prokaryotic membrane proteins: the dense alignment surface method, *Protein engineering* 10, 673-676.
13. Claros, M. G., and von Heijne, G. (1994) TopPred II: an improved software for membrane protein structure predictions, *Computer applications in the biosciences : CABIOS* 10, 685-686.
14. Pasquier, C., and Hamodrakas, S. J. (1999) An hierarchical artificial neural network system for the classification of transmembrane proteins, *Protein engineering* 12, 631-634.
15. Chamberlain, A. K., Lee, Y., Kim, S., and Bowie, J. U. (2004) Snorkeling preferences foster an amino acid composition bias in transmembrane helices, *J Mol Biol* 339, 471-479.

16. Tanford, C., and Reynolds, J. A. (1976) Characterization of membrane proteins in detergent solutions, *Biochimica et biophysica acta* 457, 133-170.
17. Domene, C., Bond, P. J., Deol, S. S., and Sansom, M. S. (2003) Lipid/protein interactions and the membrane/water interfacial region, *Journal of the American Chemical Society* 125, 14966-14967.
18. Jackups, R., Jr., and Liang, J. (2005) Interstrand pairing patterns in beta-barrel membrane proteins: the positive-outside rule, aromatic rescue, and strand registration prediction, *J Mol Biol* 354, 979-993.
19. Patching, S. G. (2011) NMR structures of polytopic integral membrane proteins, *Molecular membrane biology* 28, 370-397.
20. Overington, J. P., Al-Lazikani, B., and Hopkins, A. L. (2006) How many drug targets are there?, *Nature reviews. Drug discovery* 5, 993-996.
21. Rasmussen, S. G., Choi, H. J., Rosenbaum, D. M., Kobilka, T. S., Thian, F. S., Edwards, P. C., Burghammer, M., Ratnala, V. R., Sanishvili, R., Fischetti, R. F., Schertler, G. F., Weis, W. I., and Kobilka, B. K. (2007) Crystal structure of the human beta2 adrenergic G-protein-coupled receptor, *Nature* 450, 383-387.
22. Liu, J., and Rost, B. (2001) Comparing function and structure between entire proteomes, *Protein science : a publication of the Protein Society* 10, 1970-1979.

23. Deisenhofer, J., Epp, O., Miki, K., Huber, R., and Michel, H. (1985) Structure of the protein subunits in the photosynthetic reaction centre of *Rhodospseudomonas viridis* at 3Å resolution, *Nature* 318, 618-624.
24. White, S. H. (2004) The progress of membrane protein structure determination, *Protein science : a publication of the Protein Society* 13, 1948-1949.
25. Dorsam, R. T., and Gutkind, J. S. (2007) G-protein-coupled receptors and cancer, *Nature reviews. Cancer* 7, 79-94.
26. Singh, S. K., Yamashita, A., and Gouaux, E. (2007) Antidepressant binding site in a bacterial homologue of neurotransmitter transporters, *Nature* 448, 952-956.
27. Yamashita, A., Singh, S. K., Kawate, T., Jin, Y., and Gouaux, E. (2005) Crystal structure of a bacterial homologue of Na<sup>+</sup>/Cl<sup>-</sup>-dependent neurotransmitter transporters, *Nature* 437, 215-223.
28. Singh, S. K., Piscitelli, C. L., Yamashita, A., and Gouaux, E. (2008) A competitive inhibitor traps LeuT in an open-to-out conformation, *Science* 322, 1655-1661.
29. Doyle, D. A., Morais Cabral, J., Pfuetzner, R. A., Kuo, A., Gulbis, J. M., Cohen, S. L., Chait, B. T., and MacKinnon, R. (1998) The structure of the potassium channel: molecular basis of K<sup>+</sup> conduction and selectivity, *Science* 280, 69-77.
30. Zhai, Y., and Saier, M. H., Jr. (2002) The beta-barrel finder (BBF) program, allowing identification of outer membrane beta-barrel proteins



encoded within prokaryotic genomes, *Protein science : a publication of the Protein Society* 11, 2196-2207.

31. Wimley, W. C. (2002) Toward genomic identification of beta-barrel membrane proteins: composition and architecture of known structures, *Protein science : a publication of the Protein Society* 11, 301-312.
32. Martelli, P. L., Fariselli, P., Krogh, A., and Casadio, R. (2002) A sequence-profile-based HMM for predicting and discriminating beta barrel membrane proteins, *Bioinformatics* 18 Suppl 1, S46-53.
33. Kurisu, G., Zakharov, S. D., Zhalnina, M. V., Bano, S., Eroukova, V. Y., Rokitskaya, T. I., Antonenko, Y. N., Wiener, M. C., and Cramer, W. A. (2003) The structure of BtuB with bound colicin E3 R-domain implies a translocon, *Nature structural biology* 10, 948-954.
34. Vogt, J., and Schulz, G. E. (1999) The structure of the outer membrane protein OmpX from Escherichia coli reveals possible mechanisms of virulence, *Structure* 7, 1301-1309.
35. White, S. Membrane Proteins of Known 3D Structure.
36. Wimley, W. C. (2003) The versatile beta-barrel membrane protein, *Current opinion in structural biology* 13, 404-411.
37. Arnold, T., Poynor, M., Nussberger, S., Lupas, A. N., and Linke, D. (2007) Gene duplication of the eight-stranded beta-barrel OmpX produces a functional pore: a scenario for the evolution of transmembrane beta-barrels, *J Mol Biol* 366, 1174-1184.

38. Chou, K. C., and Carlacci, L. (1991) Energetic approach to the folding of alpha/beta barrels, *Proteins* 9, 280-295.
39. van den Berg, B. (2005) The FadL family: unusual transporters for unusual substrates, *Current opinion in structural biology* 15, 401-407.
40. Huysmans, G. H., Radford, S. E., Brockwell, D. J., and Baldwin, S. A. (2007) The N-terminal helix is a post-assembly clamp in the bacterial outer membrane protein PagP, *J Mol Biol* 373, 529-540.
41. Naveed, H., Jackups, R., Jr., and Liang, J. (2009) Predicting weakly stable regions, oligomerization state, and protein-protein interfaces in transmembrane domains of outer membrane proteins, *Proceedings of the National Academy of Sciences of the United States of America* 106, 12735-12740.
42. Adamian, L., Naveed, H., and Liang, J. (2011) Lipid-binding surfaces of membrane proteins: evidence from evolutionary and structural analysis, *Biochimica et biophysica acta* 1808, 1092-1102.
43. Choutko, A., Glattli, A., Fernandez, C., Hilty, C., Wuthrich, K., and van Gunsteren, W. F. (2011) Membrane protein dynamics in different environments: simulation study of the outer membrane protein X in a lipid bilayer and in a micelle, *European biophysics journal : EBJ* 40, 39-58.
44. Donald, J. E., Kulp, D. W., and DeGrado, W. F. (2011) Salt bridges: geometrically specific, designable interactions, *Proteins* 79, 898-915.

45. Hauck, C. R., and Meyer, T. F. (2003) 'Small' talk: Opa proteins as mediators of Neisseria-host-cell communication, *Current opinion in microbiology* 6, 43-49.
46. Merz, A. J., and So, M. (2000) Interactions of pathogenic neisseriae with epithelial cell membranes, *Annual review of cell and developmental biology* 16, 423-457.
47. Caugant, D. A., and Maiden, M. C. (2009) Meningococcal carriage and disease--population biology and evolution, *Vaccine* 27 Suppl 2, B64-70.
48. Center for Disease Control and Prevention.
49. Ohnishi, M., Golparian, D., Shimuta, K., Saika, T., Hoshina, S., Iwasaku, K., Nakayama, S., Kitawaki, J., and Unemo, M. (2011) Is Neisseria gonorrhoeae initiating a future era of untreatable gonorrhea?: detailed characterization of the first strain with high-level resistance to ceftriaxone, *Antimicrobial agents and chemotherapy* 55, 3538-3545.
50. 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.
51. Muenzner, P., Dehio, C., Fujiwara, T., Achtman, M., Meyer, T. F., and Gray-Owen, S. D. (2000) Carcinoembryonic antigen family receptor specificity of Neisseria meningitidis Opa variants influences adherence to and invasion of proinflammatory cytokine-activated endothelial cells, *Infection and immunity* 68, 3601-3607.

52. 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.
53. Hobbs, M. M., Seiler, A., Achtman, M., and Cannon, J. G. (1994) Microevolution within a clonal population of pathogenic bacteria: recombination, gene duplication and horizontal genetic exchange in the opa gene family of *Neisseria meningitidis*, *Molecular microbiology* 12, 171-180.
54. Bhat, K. S., Gibbs, C. P., Barrera, O., Morrison, S. G., Jahnig, F., Stern, A., Kupsch, E. M., Meyer, T. F., and Swanson, J. (1991) The opacity proteins of *Neisseria gonorrhoeae* strain MS11 are encoded by a family of 11 complete genes, *Molecular microbiology* 5, 1889-1901.
55. Hammarstrom, S. (1999) The carcinoembryonic antigen (CEA) family: structures, suggested functions and expression in normal and malignant tissues, *Seminars in cancer biology* 9, 67-81.
56. Kuespert, K., Pils, S., and Hauck, C. R. (2006) CEACAMs: their role in physiology and pathophysiology, *Current opinion in cell biology* 18, 565-571.
57. Berg, J., J. Tymoczko, and L. Stryer (2002) *The Immunoglobulin Fold Consists of a Beta-Sandwich Framework with Hypervariable Loops*, W H Freeman, New York.

58. Billker, O., Popp, A., Gray-Owen, S. D., and Meyer, T. F. (2000) The structural basis of CEACAM-receptor targeting by neisserial Opa proteins, *Trends in microbiology* 8, 258-260; discussion 260-251.
59. 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.
60. 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.
61. Bos, M. P., Hogan, D., and Belland, R. J. (1999) Homologue scanning mutagenesis reveals CD66 receptor residues required for neisserial Opa protein binding, *The Journal of experimental medicine* 190, 331-340.
62. 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, Biological crystallography* 62, 971-979.
63. Stern, A., Brown, M., Nickel, P., and Meyer, T. F. (1986) Opacity genes in *Neisseria gonorrhoeae*: control of phase and antigenic variation, *Cell* 47, 61-71.

## **Chapter 2: Solution NMR Spectroscopy**

Nuclear magnetic resonance (NMR) is a spectroscopic technique useful in probing the local environment of nuclei that exhibit Zeeman splitting.(1) NMR techniques have found many uses in characterizing biomolecules and their functions, from elucidation of protein structure to determination (2) of binding affinities.(3) NMR can be used to gain insight into the bonding interactions of atoms within a molecule as well as provide information on the through-space interactions.(4) Other NMR experiments can be used to observe local changes of biomolecules indicative of conformational exchange (5), protonation state (6), and interactions with other molecules at a high resolution (7). Dynamics of the macromolecules that have shown to be of great importance for function can be measured through solution NMR(8, 9). In this chapter, I will outline the theory of NMR spectroscopy and highlight its applications to biomolecules – specifically proteins.

## Section 2.1 Aspects of NMR spectroscopy

### Section 2.1.1 Nuclear spin

NMR spectroscopy is able to take advantage of the Zeeman splitting of the energy levels of the nuclear spin. These states, which are energetically equivalent under typical conditions, have distinct orientation in the presence of a magnetic field.(1) The electronic component of the applied magnetic field causes the nuclear spins to precess orthogonally to the magnetic field at a frequency proportional to the strength of the applied magnetic field, defined as the Larmor frequency. The difference in energy of splitting of these states is also directly proportional to the applied magnetic field and (Figure 2.1) can be expressed as  $\Delta E = \hbar\gamma B$  in which  $B$  is the magnitude of the magnetic field and  $\gamma$  is the gyromagnetic ratio of the respective nucleus ( $\text{rad s}^{-1} \text{T}^{-1}$ ) (Table 2.1). The total angular momentum is defined with a spin quantum number  $I = \frac{1}{2}$ .(10) Although there are nuclei that exist which have a total spin greater than  $1/2$  and, therefore, more than two energy states ( $I = 1$  results in three states, 1, 0, and -1), the majority of work done with NMR spectroscopy focuses on the nuclei whose net nuclear spin is  $I = 1/2$ . Only  $2I + 1$  allowed spin states exist such that the observed transition state for the isolated nucleus is only between  $I = 1/2$  and  $-1/2$ .(10) One of the drawbacks of NMR spectroscopy is that this energy level splitting, even in the presence of a high magnetic field, is still orders of magnitude less than  $RT$  at ambient temperature ( $0.48 \text{ J/mol}$  for  $^1\text{H}$  in  $23.48 \text{ T}$ , compared to  $2.48 \text{ kJ/mol}$  for  $RT$  at  $298 \text{ K}$ ).(1) This energy splitting leads to a very small population difference between the two states which results in a relatively weak

signal compared to other spectroscopic techniques due to the ease at which the transition between the two states can be saturated. This transition between the two states that are at nearly saturated at room temperature also results in a relatively slow re-establishment of thermal equilibrium,  $T_1$ .<sup>(11)</sup> Because of the long relaxation time, NMR experiments can perturb the system during the relaxation, which is the basis for the longer and more complex NMR experiments.<sup>(10)</sup>

### **Section 2.1.2 Chemical Shift**

Even though the gyromagnetic ratio is a constant amongst all of the same nuclei, the Larmor frequency of each nucleus in a sample is rarely redundant. The reason for a variety of Larmor frequencies is a direct result of how the local environment of that nucleus affects the magnetic field that interacts with the nucleus, or to what degree the nucleus is shielded from the magnetic field. This shielding is a result of the magnetic contributions from the valence and bonding electrons<sup>(12)</sup>, typically indicative of specific functional groups. The unique frequencies that are determined from these electrons provide a fingerprint for small molecules as well as unique structural units of polymers <sup>(12)</sup>, such as amino acids or nucleotides. In proteins, the individual subunits exhibit unique shifts amongst even the same amino acid types. There are not just 20 specific shifts correlated for each of the  $\alpha$  protons of the 20 different amino acids, but rather an entire range typically falling anywhere from 3.2 to 5.8 ppm.<sup>(13)</sup> Changes in Larmor frequency due to shielding are typically reported in ppm; these differences are typically in the  $10^2$ - $10^3$  Hz range,<sup>(10)</sup> compared to the



typical  $10^8$  Hz proton precession frequency in an NMR spectrometer. The range of frequencies that chemically similar nuclei experience is a result of the surroundings. Previous work has shown that deviations of chemical shifts can be a result of the geometry of sidechains,(14) the backbone dihedral angle,(15) hydrogen bonding,(16) neighboring amino acids,(17) and aromatic ring currents(18, 19). These environmental factors are able to modulate the electron orbitals of the observed nuclei, by affecting bond geometry or by surrounding the observed nucleus with donating/withdrawing groups of different electronegativity. Ring current shift (Figure 2.2) is a result of aromatic rings which, when oriented perpendicular to the magnetic field will have an induced current due to the freely moving electrons of the ring which in turn will generate their own local magnetic field.(1)

As the ring current shift suggests, orientation of the molecule relative to the magnetic field can also have considerable effects. The precession of the electrons in their orbitals can be greatly affected by the applied magnetic field.(12) The orthogonal electric field associated from the applied magnetic field causes the electrons of aromatic rings to precess in a set direction.(10) The precession of electrons induces a magnetic field that constructively interferes with the applied magnetic field in the plane of the aromatic ring and destructively interferes above or below the ring (Figure 2.2). The local magnetic field variation leads to large Larmor frequency variations and as such, large chemical shift perturbations.

### Section 2.1.3 Relaxation

The tumbling rate of the proteins in solution also greatly affects the NMR spectra. In spectroscopic terms, relaxation is widely used to indicate re-establishment of thermal equilibrium. The simplest way to visualize this type of relaxation is to describe the time it takes for the two nuclear spin states (identically populated in the lack of a magnetic field) to reach their new equilibrium state in the presence of the magnetic field. This relaxation, commonly referred to as  $T_1$  or spin-lattice relaxation, is a slow process that typically takes milliseconds to seconds to return to equilibrium (1). In practice, this relaxation does not result in significant perturbations of the recorded signals of the nuclei. In fact,  $T_1$  is typically slow enough that the majority of the time recording a spectrum is spent waiting for the system to return to equilibrium after the system is perturbed and the signals are recorded(1).

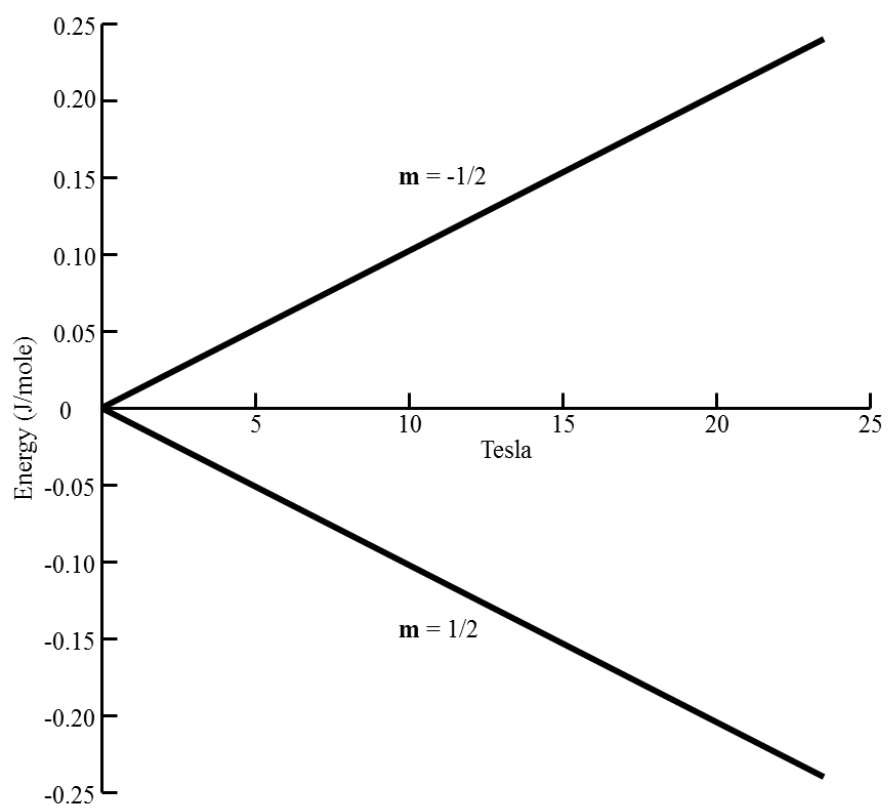
Transverse relaxation ( $T_2$ ), or spin-spin relaxation, is a faster type of relaxation that is unique to how NMR experiments are recorded; signal is observed in the plane perpendicular to the applied magnetic field(10). The nuclear spins, which were previously polarized to have a net alignment with the magnetic field, will be shifted to be in the plane perpendicular to the magnetic field after interacting with a radio frequency (RF) pulse. The spins will initially keep their net polarization and precess along the perpendicular plane at their Larmor frequency. However, transverse magnetization decays due to the spins' inability to maintain synchrony amongst other spins. This fluctuation is, in part, a result of molecular motion, in which the transverse relaxation of a molecule is approximated by the inverse of

the square of its correlation time(10). As systems become larger (systems over 100 kDa have been studied via NMR (20)) this faster relaxation time limits the quality of the spectra obtained, primarily, the linewidth of the signals obtained.

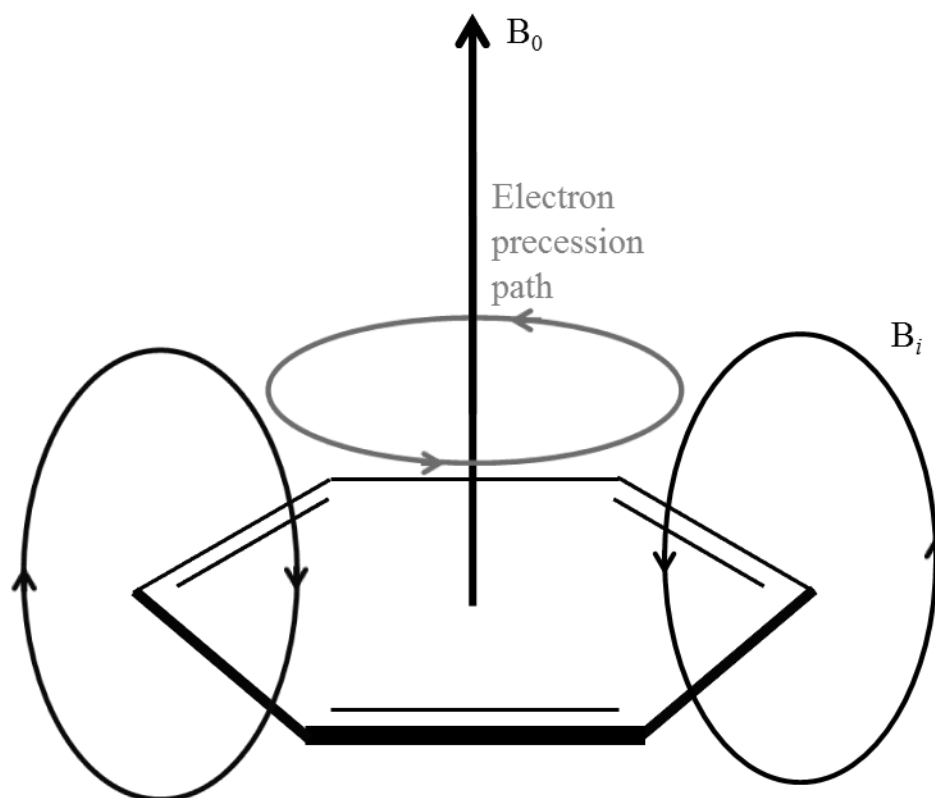
#### **Section 2.1.4 Dynamics**

$T_2$  processes represent a loss of phase memory. If a specific spin changes the electronic environment that it is experiencing, which results in a change in the Larmor frequency of the spin, phase memory can be lost. The loss of phase memory leads to a decrease in observable coherence and as such, enhanced  $T_2$  relaxation resulting in less intense, broader observed resonances. Proteins are not static molecules, and as such, the local environment of one conformation may shield the effective magnetization affecting a nucleus in a completely different way than another conformation.(5) This exchange process can affect NMR spectra in several ways. The first condition to consider is the exchange rate. In a simplified example in which the exchange rate is very slow, on the order of seconds, each individual scan would observe that nucleus in a unique environment, essentially recording two distinct free induction decays (FIDs). Fourier transforming these two distinct Larmor frequencies would result in the observation of two distinct peaks, corresponding to the two environments that the spin is experiencing (Figure 2.3, panel a). On the far extreme, if the nucleus is in exchange at a rate faster than what can be distinguished by the instrument, then the resulting FID will appear as one averaged frequency (Figure 2.3, panel h).

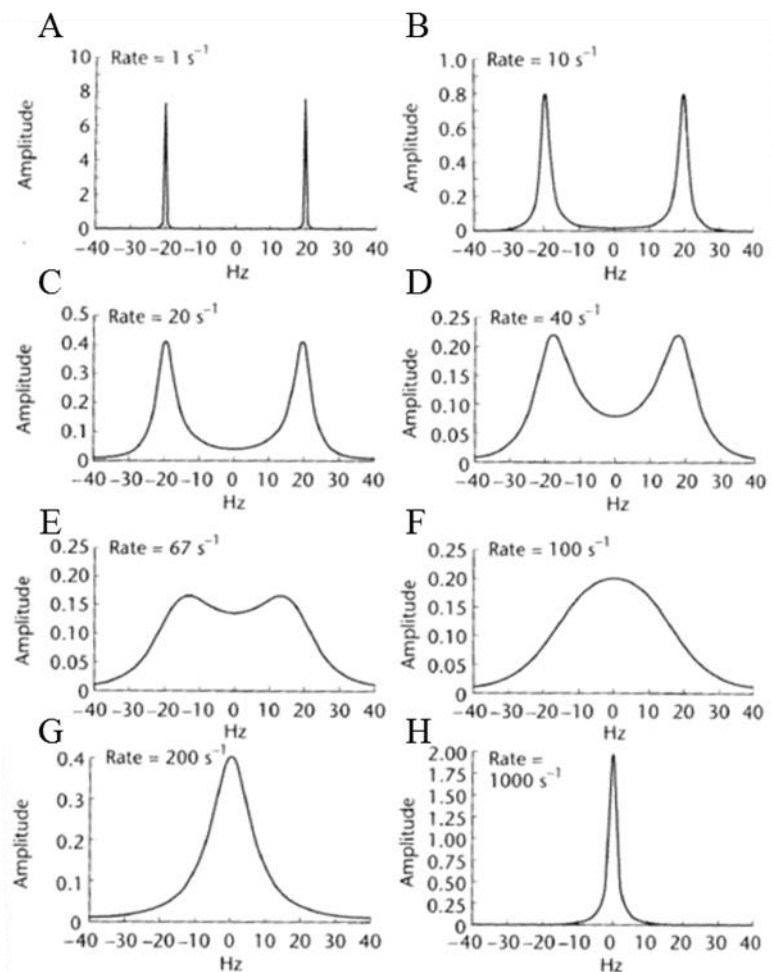
The time scale between these extremes, in which the lifetime of one spin state is shortened but not completely averaged with the other state, yields a broadening of the observed peak. As these exchange rates are plotted, there is an exchange rate in which the two peaks observed in slow exchange will coalesce into a broadened peak observed at the frequency of the averaged peak mentioned at the fast exchange condition (12). This broadening results in a signal with a much lower peak height (Figure 2.3, panel f, note signal intensity axis relative to other panels).(12)



**Figure 2.1: Energy splitting of nuclear spin levels for a  $^1\text{H}$  nucleus with respect to the magnetic field intensity.**



**Figure 2.2: Induced ring current.** In the presence of an applied magnetic field ( $B_0$ ), electrons in aromatic rings precess around the ring freely due to the electronic component of the magnetic field. This precession induces a smaller magnetic field ( $B_i$ ). Nuclei will experience different magnetic fields as the induced field adds constructively with the applied field outside of the ring, whereas the induced field destructively interferes with the applied field directly above or below the plane of the ring. These differences in fields result in greatly varied Larmor frequencies of nuclei induced by these induced ring currents.



**Figure 2.3 Effect of exchange rate on lineshape of NMR spectra.** (1) At very slow exchange rates, both populations are cleanly resolved as two distinct peaks with very narrow linewidths. As the exchange rate is increased and the frequencies are mixed in homogeneously in the recorded FID, the linewidths are broadened until they coalesce into a single averaged peak. This peak, over time, will more evenly averaged out, resulting in a narrow peak at the weighted average frequency of the different populations. Note the amplitude for the different graphs: the intermediate exchange peaks are far weaker due to their extensive broadening and are typically not observable over the noise.

Nucleus	Spin (I)	$\gamma/10^{-7} \text{ rad(s T)}^{-1}$	Frequency at 11.74 T (in MHz)
$^1\text{H}$	1/2	26.7510	500.0
$^2\text{H}$	1	4.1064	76.7
$^3\text{H}$	1/2	28.5335	533.3
$^7\text{Li}$	3/2	10.3964	194.3
$^{11}\text{B}$	3/2	8.5794	160.4
$^{13}\text{C}$	1/2	6.7263	125.7
$^{15}\text{N}$	1/2	-2.7116	50.7
$^{19}\text{F}$	1/2	25.1665	470.6
$^{31}\text{P}$	1/2	10.8289	202.6

**Table 2.1: Gyromagnetic ratio, spin quantum number and transition frequencies of commonly observed NMR.**



## Section 2.2 Single pulse experiment

The simple one dimensional NMR experiment is composed of two specific events. The first part of the experiment, known as the preparation period, is the excitation of spin systems by a radiofrequency (RF) pulse corresponding to the energy level splitting of the nuclear spins. Due to the unique gyromagnetic ratios of the different types of nuclei, the energy splitting between the two spin states is also unique to the different nuclei. As such, each nucleus will be excited by a RF frequency without affecting the other nuclei. A relatively short pulse of that frequency is used to excite a small range of frequencies, comprising the typical range of Larmor frequencies of that specific nucleus observed in the experiment. The RF pulse perturbs the nuclear spin system, rotating the spin system from parallel with the applied magnetic field by  $90^\circ$  to put the magnetization into the plane perpendicular to the magnetic field. The second part of the one pulse experiment is the detection period. Once the magnetization is in the perpendicular plane, the spins will precess at their respective Larmor frequencies in that plane. Due to the relaxation pathways discussed in section 2.1.3, the recorded signal decays over time as a free induction decay (FID). The FID is then Fourier transformed and the respective Larmor frequencies are reported.

## Section 2.3 Multidimensional NMR spectroscopy

Spectra of larger systems studied by NMR are often overcrowded with resonances that have similar Larmor frequencies overlapping. By expanding the spectra to an additional dimension or, in more recent experiments, as many as six dimensions, (21) these peaks can be more easily resolved. Multidimensional spectra correlate the resonances of multiple nuclei with each other in a variety of ways. In the previously discussed simple one dimension experiments, there were two stages to the experiment; the preparation period and the detection period. In multidimensional spectra, there are typically two additional periods. Following the preparation period, there is an evolution period, in which the spins precess a given time before the mixing period in which the spins interact with a set of RF pulses that can correlate the spins with each other in many different ways.(12)

### Section 2.3.1 Homonuclear NMR

#### *Section 2.3.1.a COSY*

The very first multidimensional spectrum was proposed by Jean Jeener in 1971.(22) Five years later, Richard Ernst (who would go on to earn a Nobel prize for this work) would first implement this work in a pulse sequence referred to as **CORrelation SpectroscopY** (COSY).(23) COSY experiments are able to identify spins that are coupled to each other by first exciting the spins with a 90° pulse, as discussed in single pulse, one-dimensional spectra. However, after this initial excitation, there is a mixing period (t) followed by a second 90° pulse before observation of the FID (Figure 2.4A). During the mixing time, the excited spins

that are coupled to each other are able to “mix” to allow some of the magnetization of one resonance to be transferred to the other. The majority of the magnetization will remain on the initial spin, which is represented on the Fourier transformed spectrum as the diagonal peaks (Figure 2.4B). The magnetization that was transferred to the coupled systems will show up as cross peaks, having two distinct frequencies in the x (initial spin) dimension and the y (coupled spin) dimension. This experiment is used to correlate systems unambiguously, which becomes necessary as systems grow in size and complexity.(24) COSY represents one of the simplest two-dimensional spectra, and many more experiments have been designed to gain even further insight.

#### *Section 2.3.1.b TOCSY*

Another homonuclear correlation spectroscopy technique that has seen substantial use in the field is the **T**otal **C**orrelation **S**pectroscop**Y** (TOCSY).(25, 26) In practice, this experiment can be thought of as an extended version of a COSY experiment. TOCSY spectra do not merely identify spins that are directly correlated with each other, but rather transfers the polarization step wise between spins that share correlated spins with each other (i.e. if spin systems 1 and 2 are correlated, and systems 2 and 3 are correlated, mixing of systems 1 and 3 will be observed in a TOCSY experiment) (Figure 2.5A). The original TOCSY pulse sequence is comparable to the COSY sequence, with the exception of the final 90° pulse, there is a train of longer, 180° pulses separated by short time intervals (Figure 2.5B).(26) The total duration of this pulse train should be long compared to the coupling of the systems (~10 Hz) to ensure

proper coupling amongst the spins, while the intervals within the train remain small, less than the inverse of the range of the of Larmor frequencies observed, (approximately 10 ppm, or 10 kHz in a 23.48 T, 1 GHz, magnet).(27) This experiment is very efficient at transferring magnetization amongst large multiple systems and is incredibly useful in protein assignment, as the magnetization can be transferred through the entire sidechain of amino acids to unambiguously assign a specific amino acid type.(28)

#### *Section 2.3.1.c NOESY*

COSY and TOCSY experiments observe J-coupling, or through-bond interactions. However, **N**uclear **O**verhauser **E**ffect **S**pectroscop**Y** (NOESY) highlights the dipole-dipole interactions that occur between spins.(29) Dipolar coupling between spins is a form of signal transfer in which polarized spins that are in the same spatial vicinity of each other induce relaxation of each other. By including an additional mixing time comparable but shorter than  $T_1$ , followed by another  $90^\circ$  pulse to the end of a COSY pulse sequence, nuclear Overhauser effect (NOE) magnetization transfer can be observed.(30, 31) These crosspeaks have a distance dependence which is related by a factor of  $r^{-6}$ , in which  $r$  is the distance between the two systems.(31) Practically, the upper limit range that  $^1\text{H}$  NOEs can be observed is approximately 5 Å.(32) This range of NOE intensities is useful for approximating the distance between two spin systems for which NOE crosspeaks are observed. NOEs have considerable value in protein structure determination as they are primarily responsible for structural restraints because they are not limited to bonding interactions but rather encompass all

spins within a radius. Structures of smaller molecules can be determined strictly using a variety of homonuclear experiments; however for larger biomolecules such as proteins, additional experiments are necessary.

### **Section 2.3.2 Heteronuclear NMR**

An entire separate class of multidimensional NMR spectroscopy is heteronuclear experiments. These experiments are able to correlate the nuclear spins of two different atoms (i.e.  $^1\text{H}$  with  $^{13}\text{C}$  or  $^{15}\text{N}$ ). There are a few issues in working with these multiple systems. Primarily,  $^1\text{H}$  is a far more abundant isotope (>99.9%) than other spin =  $\frac{1}{2}$  isotopes, including  $^{13}\text{C}$  or  $^{15}\text{N}$ , which appear at 1% and 0.1% natural abundance, respectively. The gyromagnetic ratios of these other isotopes are much lower than  $^1\text{H}$  (Table 2.1) resulting in a small energy level splitting, which translates to Fourier transformed resonances with lower peak heights. By overcoming these shortcomings, correlating heteronuclear spins, and multidimensional NMR spectroscopy, many complex systems can be investigated to answer structural and dynamic questions.

#### *Section 2.3.2.a INEPT*

One of the most important techniques in NMR spectroscopy is polarization transfer. By transferring the relative population states of a more sensitive nucleus (such as  $^1\text{H}$ ) to a less sensitive nucleus, the sensitivity of the experiment will improve by a factor of the ratio of the gyromagnetic ratios (Figure 2.6A). The **Ins**ensitive **N**ucleus **E**nhancement by **P**olarization **T**ransfer (INEPT) pulse sequence is able to transfer polarization from one system to another.(33) By

following the pulse sequence (Figure 2.6B), the proton spins are first excited and rotated into the plane perpendicular to the magnetic field. After letting the spins precess around this plane for a time  $\tau$ , both the proton and carbon spins are excited with  $180^\circ$  RF pulses for each set of spins simultaneously. This refocuses the chemical shift, but not the heteronuclear coupling. After a second mixing time, a  $90^\circ$  pulse is applied to both types of spin systems again which will selectively invert only one of the proton components, polarizing the population of the heavier atom which is now indirectly detected through the proton frequency with an enhanced sensitivity.(1)

#### *Section 2.3.2.b HSQC*

By indirectly detecting the insensitive nucleus, experiments that use the INEPT pulse sequence have become the preferred method of investigating heteronuclear correlations. One of the most popular heteronuclear experiments is the double INEPT or **H**eteronuclear **S**ingle **Q**uantum **C**oherence (HSQC) experiment.(34) In short, the polarization of the more sensitive protons is originally used to provide polarization of the less sensitive nuclei, during the first INEPT sequence of the experiment. The polarization can then precess during  $t$  between INEPT sequences before being transferred back to the protons. Each spin now has the ability to report back on both the Larmor frequencies of the proton and the heteronuclear spin that the polarization transferred to during the experiment, resulting in a two dimensional spectrum correlating the two heteronuclei. Common examples of highly overlapped proton regions in proteins include the amide region (8.0-9.5 ppm) (Figure 2.7), the aromatic region (6.0

ppm-8.0 ppm) and the aliphatic region (0.0-3.0 ppm).(13) By correlating these protons with their respective heavy atoms ( $^{15}\text{N}$  and  $^{13}\text{C}$ ), the resonances in these regions are more easily resolved while also adding environmental information about the correlated heteroatom.

#### *Section 2.3.2.c NOESY-HSQC, TOCSY-HSQC*

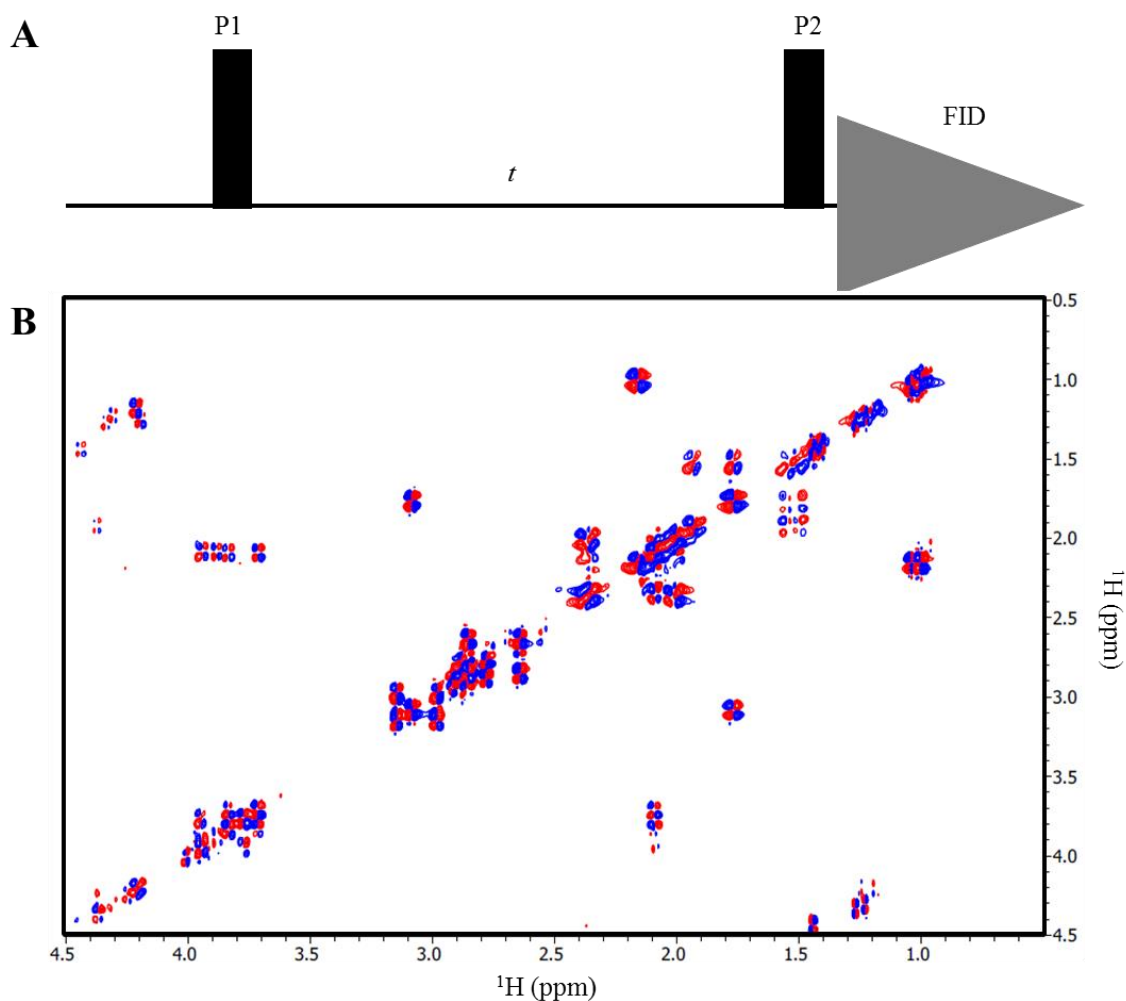
As was shown with the HSQC, which is made up of two INEPT sequences, many longer pulse sequences can be generated through a combination of previously discussed pulse sequences. One of the most straightforward examples of this is a three dimension NOESY-HSQC experiment.(35, 36) This pulse sequence is a combination of the two-dimension NOESY spectrum that leads directly into an HSQC. Various forms of this pulse sequence are used frequently throughout structural studies with large molecules. As discussed, NOESY spectra are the primary means of obtaining distance information in NMR experiments; however, as the systems being studied increase in complexity, the crowding of regions increases, reducing the data that can be obtained from each resonance, or ambiguous assignment of NOE peaks with correct diagonal peak.(37) Similarly, TOCSY spectra provide a wealth of information, specifically for sidechains of proteins, but can quickly become overly crowded, needing a third dimension to expand these crowded regions to resolve the peaks.(38)

#### *Section 2.3.2.d Triple resonance spectra*

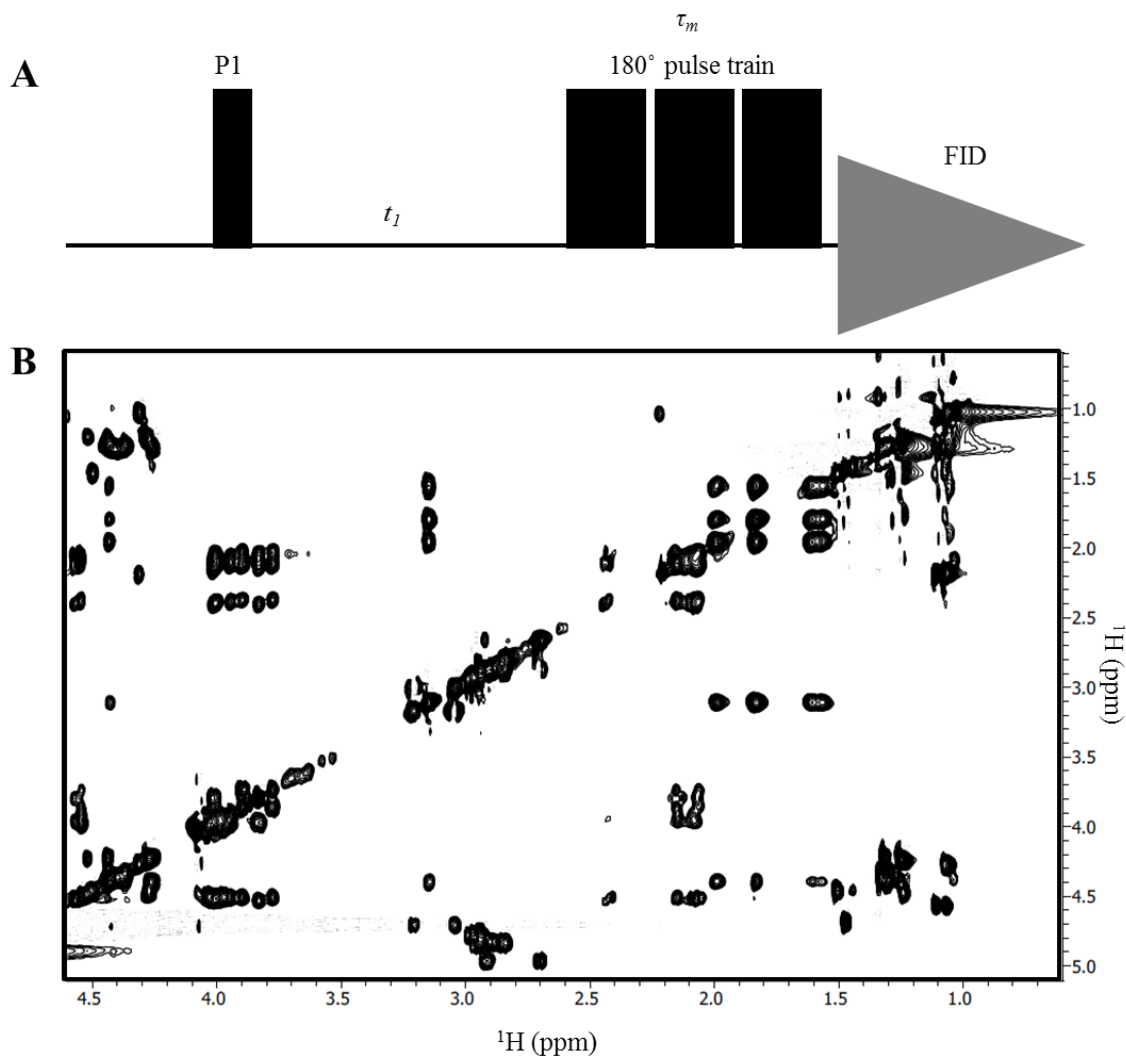
Assigning each resonance of the 2D TOCSY/NOESY-HSQC can be cumbersome for systems greater than 20 kDa.(37) An even more sensitive suite

of experiments, using only scalar coupling bonds, has been constructed and used with great success throughout the NMR community.(39, 40) These experiments are named after the direction of their magnetization transfer, i.e. in HNCO, the magnetization is transferred from the **H**ydrogen through a **N**itrogen before correlating with the nearby **C**arbonyl (**CO**). There is a basic suite of eight of these experiments used, all with very similar pulse schemes: HNCO and HN(CA)CO correlate the nearby carbonyl carbons on the backbone of a protein with the amide proton and nitrogen, HNCA and HN(CO)CA correlate the  $\alpha$ -carbons with the amide group, HN(CA)CB and HN(COCA)CB add a  $^{13}\text{C}$  COSY sequence to the HNCA sequences to observe the  $\beta$ -carbons,(41) and finally HCACO and HCA(CO)N correlate the proton on the  $\alpha$ -carbons with the rest of the heavy atoms on the backbone. Each type of amino acid has a range of unique chemical shifts for the respective atoms observed in these spectra, identifying the amino acid type specifically from these spectra. The main benefit of these spectra is that, unlike the TOCSY-HSQC, they are not strictly intra-residue. Sequential protons are capable of observing the same heavy atoms through these experiments. By observing identical carbon chemical shifts from different protons, the pair of systems can be connected sequentially. Combining the connection with known amino acid type, each peak can be assigned with the respective sequential amino acid. These experiments can be combined in a multitude of ways to generate high resolution protein structures.

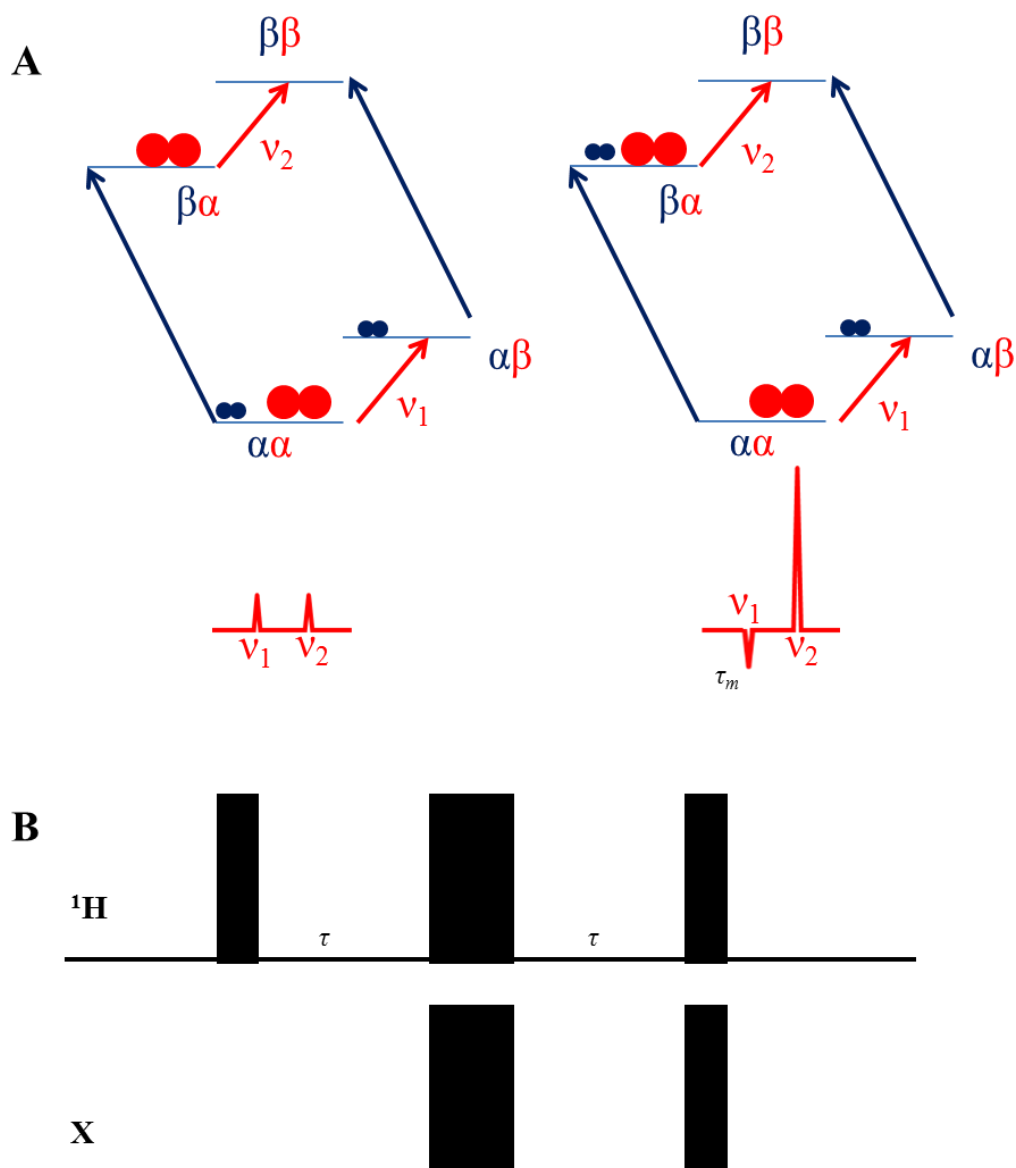




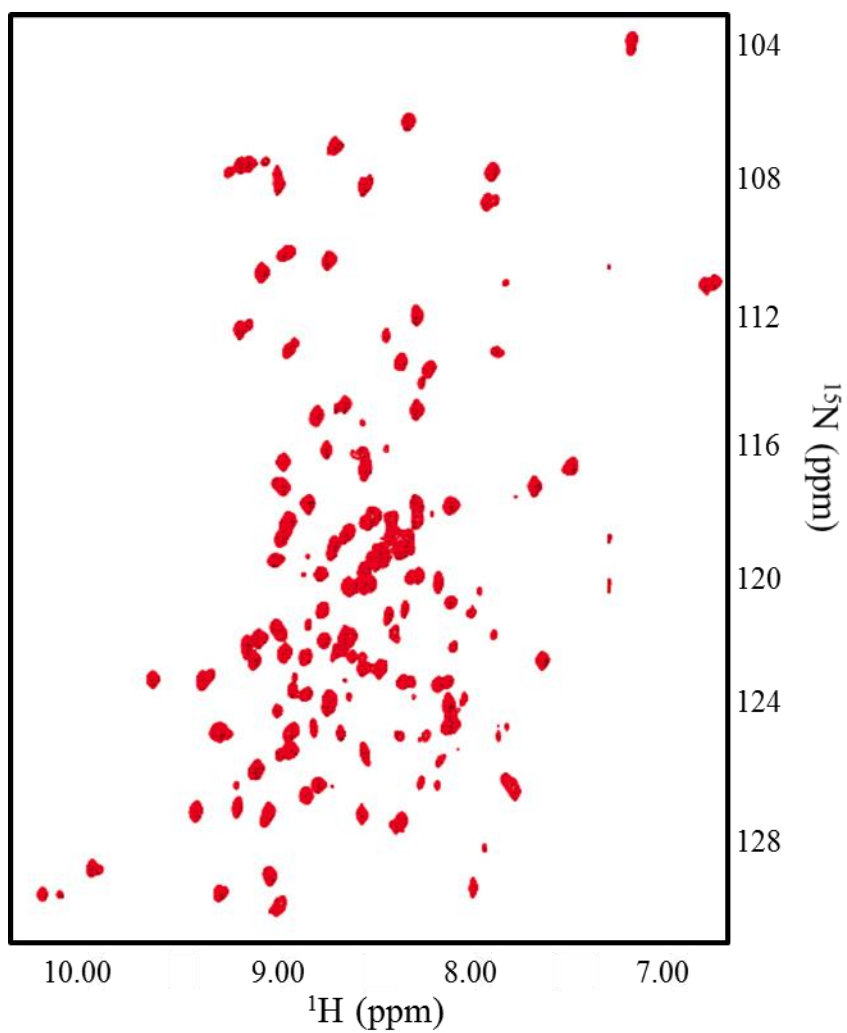
**Figure 2.4 COSY pulse sequence and spectrum.** **A.** COSY pulse schematic with a simple two  $90^\circ$  pulse and mixing time  $t$ . **B.** Recorded COSY spectrum of a 20 amino acid peptide. Peak splitting is observed due to the J-coupling of the spins. Positive peaks appear as red with negative peaks in blue.



**Figure 2.5 Pulse sequence and spectrum of TOCSY experiment. A.** The pulse sequence with mixing time  $t_1$  and total pulse train time of  $\tau_m$ . Note, not all  $180^\circ$  pulses are shown **B.** Recorded TOCSY spectrum of the same 20 amino acid peptide from Figure 2.4. The additional peaks observed in this spectrum are due to the total correlation of the sidechain protons, allowing the whole network of protons in an amino acid to correlate with each other.



**Figure 2.6: Spin population of correlated heteroatoms and the INEPT pulse sequence. A.** The relative population at equilibrium for protons (blue) and heavy atoms X (red) at equilibrium (left) and after INEPT selective inversion (right) with the Fourier transformed signal of X. **B.** INEPT pulse sequence.



**Figure 2.7:**  $^{15}\text{N}$ ,  $^1\text{H}$ -HSQC spectrum of OprH. Typical HSQC spectrum correlating the protein amide protons and nitrogens.

## References

1. Thomas C. Popchapsky, S. S. P. (2006) *NMR for Physical and Biological Scientists*, Taylor and Francis Group, New York.
2. Williamson, M. P., Havel, T. F., and Wuthrich, K. (1985) Solution conformation of proteinase inhibitor IIA from bull seminal plasma by  $^1\text{H}$  nuclear magnetic resonance and distance geometry, *J Mol Biol* 182, 295-315.
3. Fielding, L., Rutherford, S., and Fletcher, D. (2005) Determination of protein-ligand binding affinity by NMR: observations from serum albumin model systems, *Magnetic resonance in chemistry : MRC* 43, 463-470.
4. Kumar, A., Ernst, R. R., and Wuthrich, K. (1980) A two-dimensional nuclear Overhauser enhancement (2D NOE) experiment for the elucidation of complete proton-proton cross-relaxation networks in biological macromolecules, *Biochemical and biophysical research communications* 95, 1-6.
5. Ishima, R., and Torchia, D. A. (2000) Protein dynamics from NMR, *Nature structural biology* 7, 740-743.
6. Bachovchin, W. W. (1986)  $^{15}\text{N}$  NMR spectroscopy of hydrogen-bonding interactions in the active site of serine proteases: evidence for a moving histidine mechanism, *Biochemistry* 25, 7751-7759.
7. Shuker, S. B., Hajduk, P. J., Meadows, R. P., and Fesik, S. W. (1996) Discovering high-affinity ligands for proteins: SAR by NMR, *Science* 274, 1531-1534.

8. Mittermaier, A., and Kay, L. E. (2006) New tools provide new insights in NMR studies of protein dynamics, *Science* 312, 224-228.
9. Wand, A. J. (2001) Dynamic activation of protein function: a view emerging from NMR spectroscopy, *Nature structural biology* 8, 926-931.
10. Levitt, M. H. (2008) *Spin Dynamics*, John Wiley and Sons, Ltd, London.
11. Bernath, P. F. (2005) *Spectra of Atoms and Molecules*, Oxford University Press, Oxford.
12. Richard R. Ernst, G. B., Alexander Wokaun (1987) *Principles of Nuclear Magnetic Resonance in One and Two Dimensions*, Oxford University Press, Oxford.
13. Biological Magnetic Resonance Data Bank.
14. Spera, S., Ikura, M., and Bax, A. (1991) Measurement of the exchange rates of rapidly exchanging amide protons: application to the study of calmodulin and its complex with a myosin light chain kinase fragment, *Journal of biomolecular NMR* 1, 155-165.
15. Loth, K., Abergel, D., Pelupessy, P., Delarue, M., Lopes, P., Ouazzani, J., Duclert-Savatier, N., Nilges, M., Bodenhausen, G., and Stoven, V. (2006) Determination of dihedral Psi angles in large proteins by combining NH(N)/C(alpha)H(alpha) dipole/dipole cross-correlation and chemical shifts, *Proteins* 64, 931-939.
16. Wagner, G., Pardi, A., and Wuthrich, K. (1983) Hydrogen-Bond Length and H-1-Nmr Chemical-Shifts in Proteins, *Journal of the American Chemical Society* 105, 5948-5949.

17. Wishart, D. S., Bigam, C. G., Holm, A., Hodges, R. S., and Sykes, B. D. (1995) H-1, C-13 and N-15 Random Coil Nmr Chemical-Shifts of the Common Amino-Acids .1. Investigations of Nearest-Neighbor Effects (Vol 5, Pg 67, 1995), *Journal of biomolecular NMR* 5, 332-332.
18. Bergman, J. J., Chandler, W. D., and Moir, R. Y. (1971) Technique for Determining Proton Chemical Shifts Arising from Ring Current Effects - Conformations of Triply Ortho Substituted Diphenyl Ethers, *Can J Chemistry* 49, 225-&.
19. Christensen, A. S., Sauer, S. P. A., and Jensen, J. H. (2011) Definitive Benchmark Study of Ring Current Effects on Amide Proton Chemical Shifts, *J Chem Theory Comput* 7, 2078-2084.
20. Kreishman-Deitrick, M., Egile, C., Hoyt, D. W., Ford, J. J., Li, R., and Rosen, M. K. (2003) NMR analysis of methyl groups at 100-500 kDa: Model systems and Arp2/3 complex, *Biochemistry* 42, 8579-8586.
21. Zawadzka-Kazimierczuk, A., Kozminski, W., Sanderova, H., and Krasny, L. (2012) High dimensional and high resolution pulse sequences for backbone resonance assignment of intrinsically disordered proteins, *J Biomol Nmr* 52, 329-337.
22. Jeener, J. (1971), Amp'ere International Summer School II, Basko polje, Yugoslavia.
23. Ferretti, J. A., and Ernst, R. R. (1976) Interference Effects in Nmr Correlation Spectroscopy of Coupled Spin Systems, *J Chem Phys* 65, 4283-4293.

24. Wuthrich, K. (1983) Sequential Individual Resonance Assignments in the H-1-Nmr Spectra of Polypeptides and Proteins, *Biopolymers* 22, 131-138.
25. Bax, A., and Davis, D. G. (1985) MLEV-17 based two-dimensional homonuclear magnetization transfer spectroscopy, *J. Magn Reson.* 63, 355-360.
26. Braunschweiler, L., and Ernst, R. R. (1983) "Coherence Transfer by Isotropic Mixing: Application to Proton Correlation Spectroscopy, *J Magn Reson* 53, 521-528.
27. Inagaki, F., Shimada, I., Kohda, D., Suzuki, A., and Bax, A. (1989) Relayed HOHAHA: A useful method for extracting subspectra of individual components of sugar chains, *J Magn Reson* 81, 186-190.
28. Bax, A. (1989) Two-dimensional NMR and protein structure, *Annual review of biochemistry* 58, 223-256.
29. Jeener, J., Meier, B. H., Bachmann, P., and Ernst, R. R. (1979) Investigation of Exchange Processes by 2-Dimensional Nmr-Spectroscopy, *J Chem Phys* 71, 4546-4553.
30. Overhauser, A. W. (1953) Polarization of Nuclei in Metals, *Physical Review* 92, 411-415.
31. Anderson, W. A., and Freeman, R. (1962) Influence of a Second Radiofrequency Field on High-Resolution Nuclear Magnetic Resonance Spectra, *J Chem Phys* 37, 411-415.



32. Doreleijers, J. F., Raves, M. L., Rullmann, T., and Kaptein, R. (1999) Completeness of NOEs in protein structure: a statistical analysis of NMR, *Journal of biomolecular NMR* 14, 123-132.
33. Morris, G. A., and Freeman, R. (1979) Enhancement of Nuclear Magnetic Resonance Signals by Polarization Transfer, *Journal of the American Chemical Society* 101, 760-762.
34. Bodenhausen, G., and Ruben, D. J. (1980) Natural abundance nitrogen-15 NMR by enhanced heteronuclear spectroscopy, *Chem Phys Letters* 69, 185-189.
35. Zuiderweg, E. R. P., and Fesik, S. W. (1989) Heteronuclear three-dimensional NMR spectroscopy of the inflammatory protein C5a, *Biochemistry* 28, 2387-2391.
36. Marion, D., Driscoll, P. C., Kay, L. E., Wingfield, P. T., Bax, A., Gronenborn, A. M., and Clore, G. M. (1989) Overcoming the overlap problem in the assignment of proton NMR spectra of larger proteins by use of three-dimensional heteronuclear proton-nitrogen-15 Hartmann-Hahn-multiple quantum coherence and nuclear Overhauser-multiple quantum coherence spectroscopy: application to interleukin 1.beta., *Biochemistry* 28, 6150-6156.
37. Clore, G. M., and Gronenborn, A. M. (2000) NMR structure determination of proteins and protein complexes larger than 20 kDa, *Curr Opin Chem Biol* 2, 564-570.

38. Bax, A., Clore, G. M., and Gronenborn, A. M. (1990)  $^1\text{H}$ - $^1\text{H}$  correlation via isotropic mixing of  $^{13}\text{C}$  magnetization, a new three-dimensional approach for assigning  $^1\text{H}$  and  $^{13}\text{C}$  spectra of  $^{13}\text{C}$ -enriched proteins, *J Magn Reson* 88, 425-431.
39. Bax, A., Ikura, M., Kay, L. E., Barbato, G., and Spera, S. (1991) Multidimensional triple resonance NMR spectroscopy of isotopically uniformly enriched proteins: a powerful new strategy for structure determination, *Ciba Foundation symposium* 161, 108-119; discussion 119-135.
40. Kay, L. E., Ikura, M., Tschudin, R., and Bax, A. (1990) Three-dimensional triple-resonance NMR spectroscopy of isotopically enriched proteins, *J Magn Reson* 89, 496-514.
41. Grzesiek, S., and Bax, A. (1992) An efficient experiment for sequential backbone assignment of medium-sized isotopically enriched proteins, *J Magn Reson* 99, 201-207.

## **Chapter 3: NMR Protein Structure Determination**

### **Section 3.1 NMR Protein Structure Determination Workflow**

With solution NMR, an ensemble of structures is calculated using spectrally-derived atom-atom distances,(1) bond angles,(2) and orientation restraints.(3, 4) The resulting ensemble of structures is in contrast to the single three dimensional structure derived from electron density mapping in crystallography. The processes needed to determine structures via NMR spectroscopy will be detailed in this section.

#### **Section 3.1.1 Sequential assignment**

The first goal of an NMR spectroscopist is to empirically optimize conditions that lead to well dispersed resonances able to be individually resolved. Following this, the next goal is to assign each of the resonances to their respective atoms. For smaller proteins (<70 amino acids) the two dimensional homonuclear NOESY and TOCSY experiments can be enough to obtain a full assignment of all proton.(5) In these experiments three main types of NOE correlations between backbone atoms are resolved to sequentially assign each amino acid: the  $\alpha$ -proton from specific amino acid (i) correlated with the amide proton of amino acids up to four residues apart in the sequence ( $i + 1,2,3,4$ ); amide protons of residue (i) correlated with amide protons two residues away ( $i \pm 2$ ), or  $\alpha$ -protons associated with  $\beta$ -protons of amino acids three residues removed ( $i + 3$ ) (Figure 3.1A).(6)

The spectra of larger systems have crowded regions, due to peak overlap. The degenerate peaks result in too much ambiguity to assign the necessary resonances. Even adding an additional, heteronuclear dimension (typically  $^{15}\text{N}$  or  $^{13}\text{C}$ ) to NOESY and TOCSY experiments still leave excessive crowding and ambiguity in resonance assignment.(7) Triple resonance spectra, as introduced in the previous chapter have shown considerable success for the assignment of the backbone atoms of amino acids in large systems.(8) By performing the entire suite mentioned in 2.2.2 (HNCO, HN(CA)CO, HNCA, HN(CO)CA, HCACO, HCA(CO)N, HN(CA)CB and HN(CACO)CB), sequential amino acids are unambiguously assigned (Figure 3.2A).(9) These experiments are typically performed run in pairs, such that the two carbon resonances observed in the HNCA (Figure 3.2B) spectrum can be identified as the  $i$  or  $i-1$   $\alpha$ -carbon. Differentiation is possible because based on relative intensity ( $i-1$  is weaker) and by comparison to the HN(CO)CA spectrum in which only the  $i-1$   $\alpha$ -carbon is observed.(10) Because each unique spin system can report on observe both the  $i$  and  $i-1$  resonances, the same resonance is observed in two different systems as the  $i$  or  $i-1$  resonance, respectively (Figure 3.2A, red box). As a result, the systems can be assigned as sequential neighbors.(9) By correlating unique chemical shifts to specific amino acid identities,(11) these sequentially connected systems are identified as unique amino acid sequences present in the already known protein sequence. Thus, resonances are unambiguously assigned to the corresponding amino acid.

### **Section 3.1.2 Sidechain assignment**

Once a spin system has been identified as a specific amino acid in the protein sequence, many homo- and heteronuclear spins can be assigned to that amino acid. To accomplish this, spectra have been designed to record the resonances on the sidechains and to correlate them with the assigned backbone resonances.(12) Perhaps the most straightforward of these is the HCCH-TOCSY,(13) in which magnetization is transferred from hydrogen nuclei correlated to carbons before isotropic  $^{13}\text{C}$  mixing occurs. The results of mixing correlate each sidechain proton to its respective carbon, and, because a TOCSY sequence is performed throughout the entire side chain, will correlate all proton-carbon resonances with each other for that specific amino acid. Other experiments, such as the H(CCO)NH or (H)C(CO)NH,(14, 15) will correlate either the sidechain carbons or protons with the amide proton-nitrogen of the  $i+1$  residue (Figure 3.3A). While H(CCO)NH and (H)C(CO)NH spectra are less sensitive than those of the HCCH-TOCSY (Figure 3.3B),(16) they have the benefit of supplementing the triple resonance backbone experiments with specific amino acid identification, should the initial sequential assignment prove ambiguous.

### **Section 3.1.3 Torsion angles**

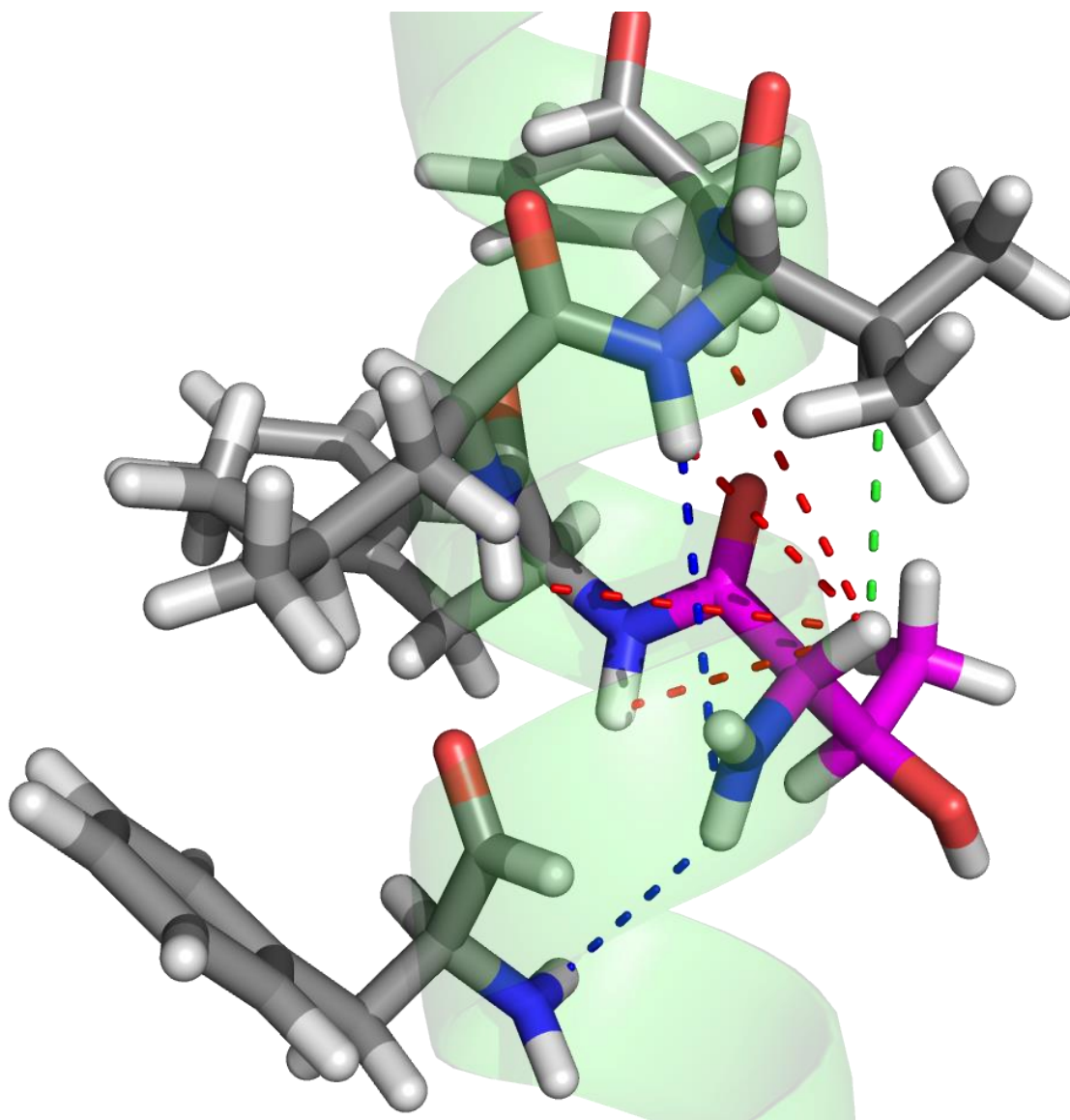
Torsion angles obtained from NMR provide the structural constraints necessary to give the assigned resonances physical meaning. In the early years of NMR protein structure determination, constraints were derived directly from the coupling constants.(17) However, decoupling pulses have become necessary to prevent spectral crowding, nullifying the practicality of directly measuring

coupling constants to obtain protein dihedral angles in large systems. As more protein structures were solved using NMR spectroscopy, a pattern of secondary chemical shifts emerged.(18) Secondary chemical shifts are deviations of chemical shifts of amino acids from their overall average, correlating with aspects of protein secondary structure. A program, TALOS, was designed to analyze the database of solved protein structures and compare the deposited chemical shifts of an amino acid and its two flanking neighbors (residues  $i-1$ ,  $i$ , and  $i+1$ ) with the chemical shifts obtained for the matching amino acid triplet of interest.(19) By averaging the dihedral angles of the ten closest chemical shifts in the database, TALOS calculates very accurate dihedral angles, when compared to direct coupling measurements (Figure 3.4). (20)

#### **Section 3.1.4 Through-space proton-proton interactions**

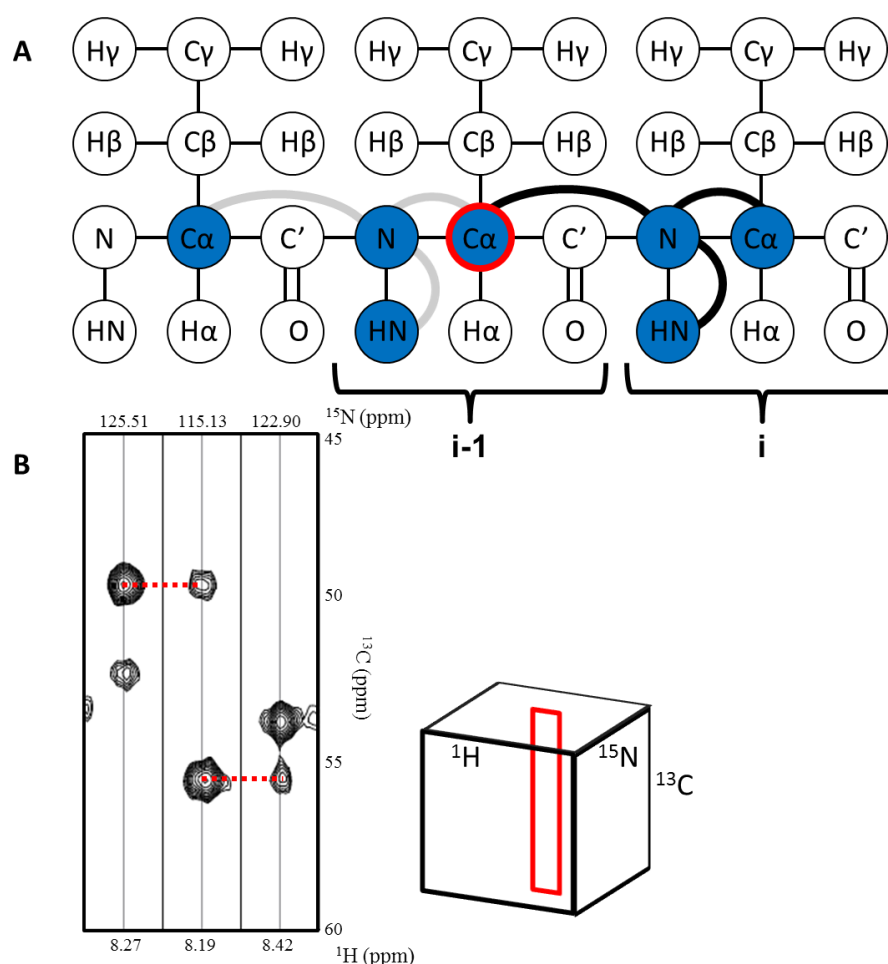
Assignment and dihedral restraints are important in solving protein structures; however, a structure cannot be determined by these techniques alone. All of the data obtained by these methods come via through-bond magnetization transfer, J-coupling. Through-bond magnetization transfer has the benefit of a relatively efficient magnetization transfer, but offers very little in terms of distance restraints or intra-protein interactions beyond the neighboring sequential amino acids. Thus, experiments that take advantage of the through-space properties of NOE interactions are commonly used to obtain spin-spin distance restraints.(21) These restraints are not limited just to spins that are closely connected sequentially, but can instead correlate spins of different domains of a protein that are spatially close. NOEs obtained through either  $^{15}\text{N}$  or  $^{13}\text{C}$  edited three

dimensional NOESY spectra are integral in determining the overall tertiary structure of proteins.(8) NOESY spectra have the added benefits of restricting sidechain conformations based on NOEs of protons with neighboring protons and intra-residue NOE interactions.(22) Due to the  $r^{-6}$  relationship of dipole-dipole interactions (section 2.2.1), these observed NOE crosspeaks can be binned into distinct distance ranges.(23) For proteins up to approximately 25 kDa, an assignment, dihedral angles and NOE distance restraints are sufficient to calculate a high resolution structure.(8)

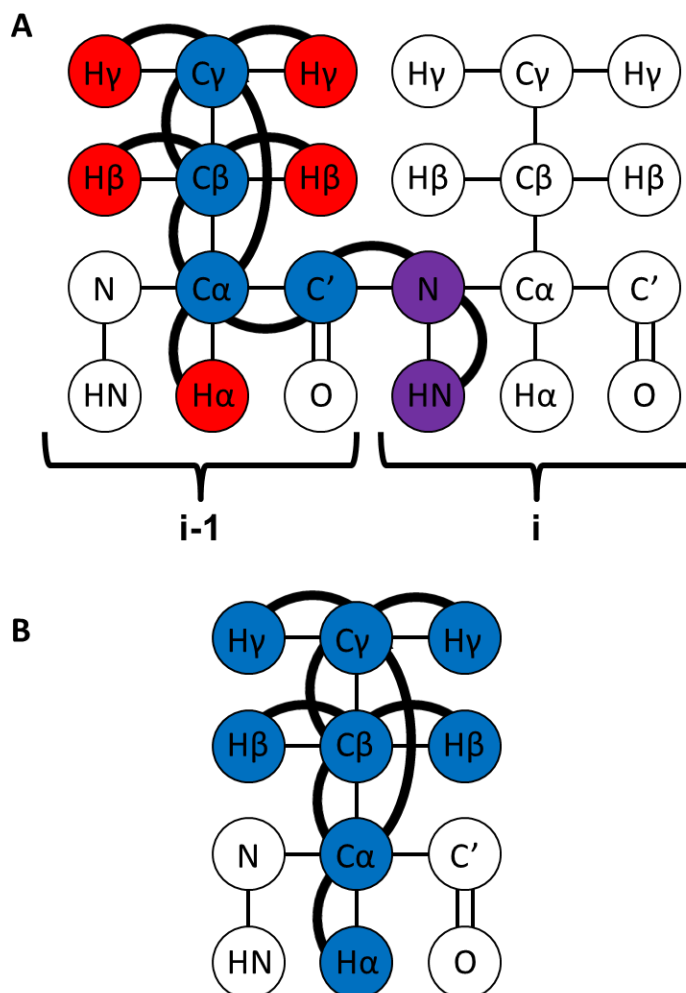


**Figure 3.1: Sequential connectivity used in small protein assignment.** Prior to the use of three dimensional NMR experiments, smaller proteins were assigned via a variety of two dimensional experiments. Three species of connective NOEs were crucial for assignment:  $\alpha\text{H}$  to  $\text{NH}$  of  $i + 1,2,3$  (red),  $\text{NH}$  to  $\text{NH}$  of  $i \pm 2$  (blue), and  $\alpha\text{H}$  to  $\beta\text{H}$  of  $i + 3$  (green). All of these connections are shown from one amino acid,  $i$ , (pink).



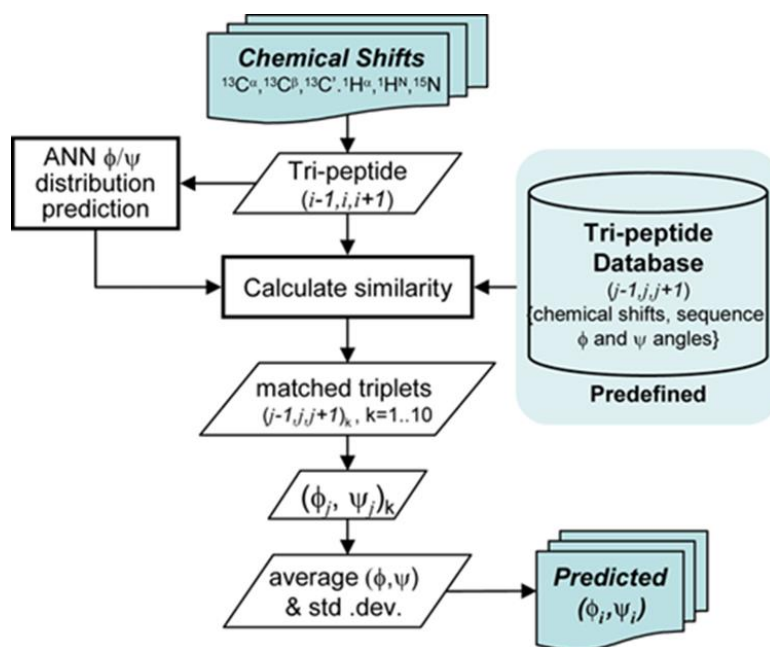


**Figure 3.2: A. Magnetization transfer of HNCA experiment.** Chemical shifts are obtained for atoms highlighted in blue. The magnetization begins on the respective amide protons and is transferred through the nitrogen and onto the  $\alpha$ -carbons of both the i and i-1 sequential amino acids as observed by following the black or grey paths. The neighboring amide protons both observe the same  $\alpha$ -carbon (outlined in red). **B. Sequential assignment of HNCA strips.** Each  $^{13}\text{C}$  strip (shown in red on 3D experiment space) represents all spins observed via each amide proton. As noted, where two amide protons observe the same  $^{13}\text{C}$  frequency (red dotted line), the two systems are sequentially connected.



**Figure 3.3 A. Magnetization transfer of both  $H(CCO)NH$  and  $(H)C(CO)NH$ .**

Magnetization transfer correlating the sidechain protons ( $H(CCO)NH$ , red) or side chain carbons ( $(H)C(CO)NH$ , blue) with the  $i + 1$  amide proton and nitrogen (purple). **B. Magnetization transfer of HCCH-TOCSY.** This experiment unambiguously assigns all sidechain protons with all sidechain atoms in each amino acid in the sequence. Sidechain assignments offer many spins to assign and obtain restraints for; however, fully assigning them is rarely accomplished in larger systems.



**Figure 3.4: Flowchart for TALOS+ database search procedure.(20)** Using a combination of predefined dihedral angles from previously solved structures and secondary structure prediction software, TALOS+, can generate accurate dihedral angles for a peptide backbone. Mining through previously published data, including chemical shift and sequentially neighboring amino acids, a very accurate dihedral restraint is estimated.

## Section 3.2 Techniques for larger systems

While the general workflow outlined so far is used in all NMR structural studies, the systems studied have continued to become more complex. Difficult systems may include larger proteins,(24) protein complexes,(25) or membrane proteins.(26) The primary limitation to NMR studies is the size of the system being studied. The larger a molecule is, the longer its correlation time, which will result in a faster T2 relaxation.(27) Long pulse sequences, such as the three dimensional backbone assignment experiments, will lose more coherence during the evolution and mixing periods, which results in dramatically reduced signals.(28) There have been a variety of techniques developed which have expanded NMR structural studies to systems greater than 100 kDa.(24) The techniques include deuteration, methyl labeling, transverse relaxation optimized spectroscopy (TROSY), and paramagnetic relaxation enhancement (PRE).

### Section 3.2.1 Deuteration

As a protein size increases, relaxation via dipole-dipole interactions becomes more prominent.(28) Because of the large gyromagnetic ratio of  $^1\text{H}$  spins, protons are the primary source for this relaxation mechanism.(29)  $^2\text{H}$  atoms have a gyromagnetic ratio over six times smaller than  $^1\text{H}$  atoms, and as a result are not nearly as efficient at promoting dipole-dipole relaxation pathways in nearby nuclei.  $^2\text{H}$  is a spin=1 nucleus, which when coupled to a spin =  $\frac{1}{2}$  nucleus like  $^{13}\text{C}$  or  $^{15}\text{N}$  should result in an expected triplet for those resonances. Instead, a collapsed singlet that is broadened due to scalar relaxation mechanisms is

observed. A high powered decoupling pulse in the  $^2\text{H}$  RF frequency is added to pulse sequences with perdeuterated proteins, removing this additional broadening, leaving narrower lineshapes. As an example, perdeuterating the 30 kDa amino-terminal domain of enzyme I resulted in the average  $T_2$  of amide protons doubling from 13 ms in the protonated sample to 28 ms in the perdeuterated protein.(30)

There is a drawback with perdeuteration as well. Most hydrogen atoms bonded to carbons are not labile and will remain attached, which is beneficial for decreasing relaxation pathways by keeping the  $^2\text{H}$  isotope bonded to the protein. However, hydrogen spin systems are not readily observable in most NMR experiments when perdeuterated. As a result, perdeuterated samples lack significant sidechain information.(4) Only the labile protons (most prominently the amide protons) freely exchange with water protons; thus, most experiments performed on deuterated samples only report on backbone proton information.(31) Deuterated samples are left with minimal distance restraints, almost exclusively between the labile backbone amide protons. (32)

### **Section 3.2.2 Methyl labeling**

Many larger protein studies have required more restraints than just those provided by amide protons. As a result, a variety of unique isotopic labeling techniques have been generated to aid in obtaining amino acid assignments and structural restraints.(33, 34) Perhaps the most common is the reintroduction of protons to the methyl carbons of isoleucine (Ile), leucine (Leu) and valine

(Val).(35) Protonated methyl groups are generated with a deuterated background by adding unique isotopically labeled amino acid precursors to growth media while protein expression is being induced.(36) Barring degradation, these precursors (Figure 3.5A) are used exclusively by the expressing bacteria (typically *E. coli*) to generate Ile, Leu and Val during protein synthesis. As a result, the expressed protein will have protons on the methyl groups from the precursors, as opposed to synthesizing these amino acids from the deuterated media. Methyl groups are popular for selective protonation for a number of reasons. First, the free rotation of methyl protons around the carbon axis and favorable relaxation properties result in large peak heights even in larger systems.(37) Second, because methyl groups are considerably hydrophobic, many Ile, Leu and Val groups lie in the hydrophobic core of the protein. By choosing a functional group that will be clustered in this manner, intra-methyl proton data, specifically NOE cross relaxation (Figure 3.5C), generate useful distance restraints with the lack of sidechain protons in deuterated proteins. (38)

### **Section 3.2.3 Paramagnetic relaxation enhancement (PRE)**

Paramagnetic relaxation enhancement (PRE) can be thought of a localization of enhanced relaxation: PRE involves adding a paramagnetic species with a very high gyromagnetic ratio to enhance relaxation of nuclei in the surrounding area resulting in decreased peak heights of nearby spin systems, as opposed to decreasing relaxation in the case of perdeuteration which acts to enhance the peak height and linewidth.(39)

There are many ways to practically introduce paramagnetic species to the system. Paramagnetic metals added to the solution will be in close proximity to any nuclei exposed to the solution.(40) This proximity will induce faster relaxation through electron-nuclear dipole-dipole interactions. Marked relaxation effects were previously mentioned between  $^1\text{H}$  and  $^2\text{H}$  nuclei (section 3.2.1), in which the gyromagnetic ratios differed by a factor of 6. The gyromagnetic ratio of a free electron is over 600 times larger than that of a  $^1\text{H}$  nucleus. The relaxation of protons near these paramagnetic species is enhanced so vastly that any NMR signal within the ranges proportional to this enhanced relaxation of paramagnetic species is broadened beyond detection.(39)

In protein structural studies, PRE is useful in supplementing NOE distance restraints.(39) Additional restraints are obtained by localizing a paramagnetic spin on the protein. First, a unique cysteine in the protein sequence (through mutation) is reacted with a spin label via a disulfide bond. These spin labels are designed to contain four methyl groups neighboring a nitrogen in a pyrrole or pyridine ring to stabilize and distribute radical density (Figure 3.6A).(41) Bonding of a paramagnetic species to a specific region on a protein localizes the relaxation enhancement. While all protons within 15 Å are spectroscopically invisible due to the enhanced relaxation of the radical, the rest of the peak intensities are modulated by the paramagnetic species. The peak heights ( $I$ ) of a paramagnetic spin label and a diamagnetic label can be approximated to be  $I_{\text{para}} = 1/R2^*$  and  $I_{\text{dia}} = 1/R2$  where  $R2^*$  is the summation of (spin-spin relaxation)  $R2$  with the spin contribution from the paramagnetic species ( $R2^{\text{sp}}$ ).(42, 43) The ratio

can be reworked to be  $I_{\text{para}}/I_{\text{dia}} = R_2 e^{(-R_2 \tau_c^*)} / (R_2 + R_2 \tau_c^*)$ .  $R_2 \tau_c^*$  can be converted into a specific distance using the following equation:

$$r = \left[ \frac{K}{R_2 \tau_c^*} \left( 4\tau_c + \frac{3\tau_c}{1 + \omega^2 \tau_c^2} \right) \right]^{1/6}$$

where  $r$  is the distance between the electron and the nucleus,  $\tau_c$  is the correlation time,  $\omega$  is the Larmor frequency of the nuclear spin and  $K$  is  $1.23 \times 10^{-32} \text{ cm}^6 \text{ s}^{-2}$ . The peak height modulation allows for specific distance restraints to be added into the structure calculation. Quantifiable peak height less can be assigned for distances between 15-23 Å (Figure 3.6B).<sup>(39)</sup> Long range distance restraints gained from PRE measurements can greatly supplement the relatively short range NOE-based restraints, especially in perdeuterated systems.<sup>(44)</sup> Tertiary structures such as  $\alpha$ -helical bundles lack sufficient short range interactions between domains, specifically using backbone atom-derived distance restraint, to allow a high resolution structure to be obtained by NOEs alone.<sup>(39)</sup>

### Section 3.2.5 Transverse relaxation optimized spectroscopy (TROSY)

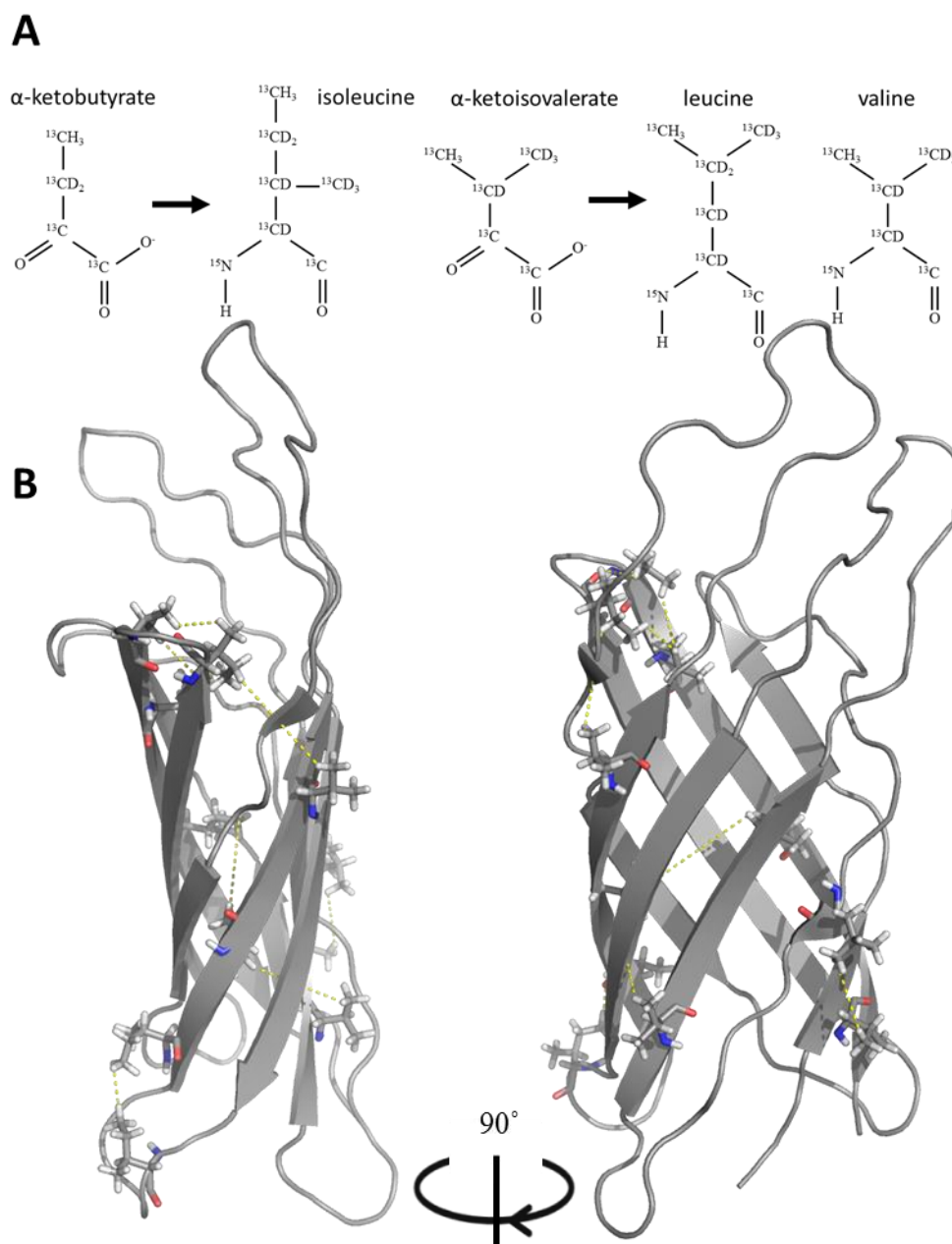
Thus far, the majority of the techniques introduced have manipulated dipole-dipole relaxation to obtain improved spectra or additional restraints. Transverse relaxation optimized spectroscopy (TROSY)<sup>(45)</sup> also manipulates dipole-dipole interactions to improve spectral quality, but unlike previous techniques, TROSY takes advantage of another form of relaxation, chemical shift anisotropy (CSA). CSA is the result of incomplete averaging of the chemical shift tensor as the molecule tumbles.<sup>(27)</sup> The specific orientation of a nucleus relative to the



magnetic field affects the local environment of that nucleus. This mechanism plays a more significant role in solid-state NMR, in which the molecules of interest are in an anisotropic medium, typically as crystals or dry powder in which molecular tumbling is limited.(28) In solution, as the molecule reorients, the chemical shift tensor of a spin is perturbed, and these time dependent fluctuations result in an enhanced relaxation.(46) Larger molecules are affected by CSA more than smaller molecules due to the slower tumbling of the larger molecule and resultant incomplete averaging of the environment over time.(46) Nuclei that are not in a symmetric electronic environment (i.e., larger contributions from p orbital characteristics in bonding electrons) such as  $^{15}\text{N}$  or  $^{13}\text{C}$ , can be dominated by CSA relaxation at higher magnetic fields or in slower tumbling molecules.

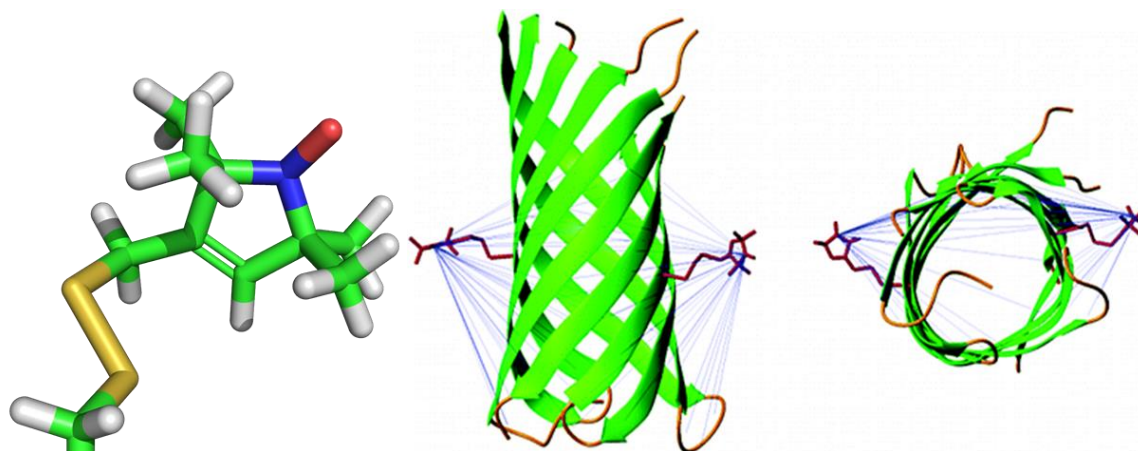
Both CSA and dipole-dipole interactions result in two different fluctuating fields for a given spin. These fields superimpose and can interfere in both constructive and destructive interference.(47) Most NMR spectra shown thus far have been decoupled, removing the splitting of the scalar coupling constant which, if included, would result in a splitting in each dimension of the spectrum (as observed in the COSY spectrum, Figure 2.4, from the previous chapter). Because of the interference patterns of relaxation of the CSA and dipole-dipole fields, the split peaks do not all exhibit identical lineshapes.(47) As observed for two dimensional HSQC type experiments, of the four peaks (splitting in both the x and y dimensions) relaxation of one resonance is enhanced by both mechanisms, resulting in a broader peak (Figure 3.7A). Two of the peaks have

components of constructive and destructive relaxation enhancement, resulting in both having sharper lineshapes and less efficient relaxation than the first peak. The final peak in this example will have destructive interference from both CSA and dipole-dipole interactions with other relaxation terms, resulting in a the narrowest linewidth. TROSY pulse sequences are able to select this specific narrower peak , resulting in an overall slower relaxation mechanism and narrower linewidths observed. TROSY allows for larger systems, such as large proteins or membrane-protein -micelle complexes), to be studied at greater resolution Figure 3.7).(45)

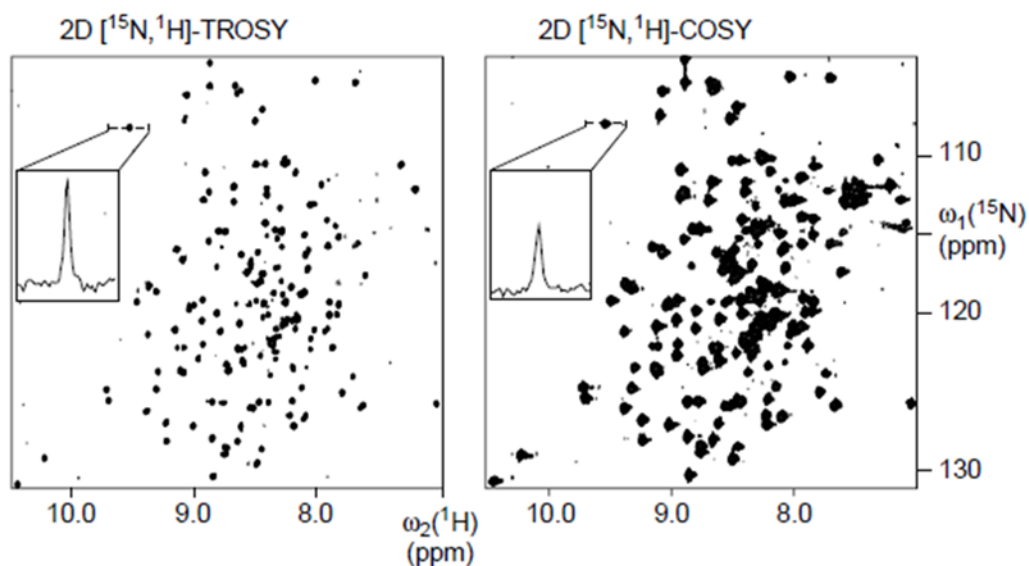


**Figure 3.5: A. Amino acid precursors to  $^1\text{H}$  methyl labeled Ile, Leu and Val.**

This unique labeling scheme, reintroduces  $^1\text{H}$  nuclei into a perdeuterated system in order to obtain extra spin systems to acquire data(35). **B. Methyl-Methyl NOEs of OmpX.** A sample of extra distance restraints gained via methyl-labeling. Nine of the 36 methyl-methyl NOEs are shown, whereas there were over 200 methyl-amide proton NOEs observed (PDB ID 1Q9F).(48)



**Figure 3.6 A. Structure of a commonly used spin label, S-(2,2,5,5-tetramethyl-2,5-dihydro-1H-pyrrol-3-yl)methyl methanesulfonothioate (MTSL).** The unique structure of the 2,2,5,5 methylation on the pyrrol ring stabilizes the unpaired, electron associated with the nitroxide group. Unpaired electrons have a gyromagnetic ratio over 600 times greater than any nucleus, which greatly enhances relaxation of any nucleus in the general area of the radical. **B. Long-range distance restraints.** When attached to a protein, distance restraints can be obtained for nuclei 15-23 Å away from the electron. Two different PRE experiments on OmpA with their observed restraints are presented. Overall, MTSL was placed in eleven different places, resulting in over 300 additional distance restraints.(44)



**Figure 3.7: A. Spectral quality improvement via TROSY:** Identically recorded and processed spectra of OmpX in DHPC micelles, a 60 kDa complex, with the exception that TROSY was implemented in the left figure. The inserts show cross sections taken in the  $^1\text{H}$  dimension to indicate the narrowing of the linewidths.

### Section 3.3 $\beta$ -barrel structures determined via NMR

Many of the techniques and processes outlined in this chapter have been used to solve  $\beta$ -barrel membrane protein structures. To date, there have been six unique  $\beta$ -barrel structures solved with NMR (Figure 3.8), though only one had not been previously solved via X-ray crystallography.(49-54) This section will highlight a sampling of these proteins with a brief overview of the biological and the spectroscopic techniques that were used in solving their structures.

#### Section 3.3.1 OmpX and OmpA

The first two  $\beta$ -barrel membrane proteins solved by NMR were outer membrane protein X (OmpX) (49) and outer membrane protein A (OmpA) (50) of *E. coli*. Both are 8 stranded  $\beta$ -barrels with 148 (OmpX) and 177 (OmpA) residues.

OmpA is fairly abundant in *E. coli* with a copy number of approximately 100,000 per cell.(55) With this large copy number for an outer membrane protein, OmpA serves a multitude of functions. The first two extracellular loops of OmpA are the two longest loops on the protein and are utilized for a variety of adhesion properties. *E. coli* is known to cause sepsis as a direct result of passing through the blood brain barrier (BBB).(56) In *E. coli*, OmpA has been shown to bind directly to brain microvascular endothelial cells (BMECs) on the BBB, facilitating passage.(56) OmpA is also involved an immune evasion response by binding to complement-binding protein 4 (C4bp).(57) When C4bp is bound, it acts as a cofactor in promoting the cleavage of C3b and C4b, leading to inhibition of cytokine induction.(57)

To obtain the sequential assignment of OmpA, TROSY-based triple resonance experiments were performed on the full length deuterated protein in which 138 of the 177 amino acids were assigned.(50) Multiple mutants were generated, replacing tryptophan residues with phenylalanine, to reduce exchange broadening associated with tryptophans.(58) Specific  $^{15}\text{N}$  labeling techniques were used to assign a variety of different amino acid types to aid in assignment. Using 49 HN-HN and 42-HN-H $\alpha$  NOEs with 142 dihedral angles obtained, the structure of the backbone of the barrel of OmpA was solved to a root mean squared deviation (rmsd) of 1.19 Å.(50)

Just a few months after the NMR structure of OmpA was published, the second NMR-determined  $\beta$ -barrel structure, that of the *E. coli* outer membrane protein X (OmpX), was released.(49) OmpX has been shown to promote adhesion to and entry into eukaryotic tissues.(59) The TROSY-based backbone assigned experiments were run for OmpX as well, resulting in 98% of the amino acids assigned.(49) A more complete assignment is observed compared to OmpA due to OmpX's shorter loops, which are where the majority of assignment ambiguity was observed for OmpA.(50) The structure obtained matched that of the previously solved crystal structure quite well. Uniquely, the barrel of OmpX is far more elliptical than most  $\beta$ -barrel proteins (axial ratio of 1.6, compared to, for example, 1.25 with OmpA).(60) An initial lack of NOEs led the group to revisit structural studies of OmpX several years later.(48) In this later study, the methyl labeling of valine, isoleucine and leucine were introduced to improve the total number of distance restraints. 220 NOEs were initially observed between amide

protons whereas 259 were observed between amide protons and  $^1\text{H}$  labeled methyl groups, with an additional 36 between methyl groups and 11 more between methyl groups and aromatic groups (Figure 3.5B).(48) These improved results led to an rmsd of the backbone of the barrel of OmpX to 0.93 Å (Figure 3.9).

### Section 3.3.2 OmpG and VDAC-1

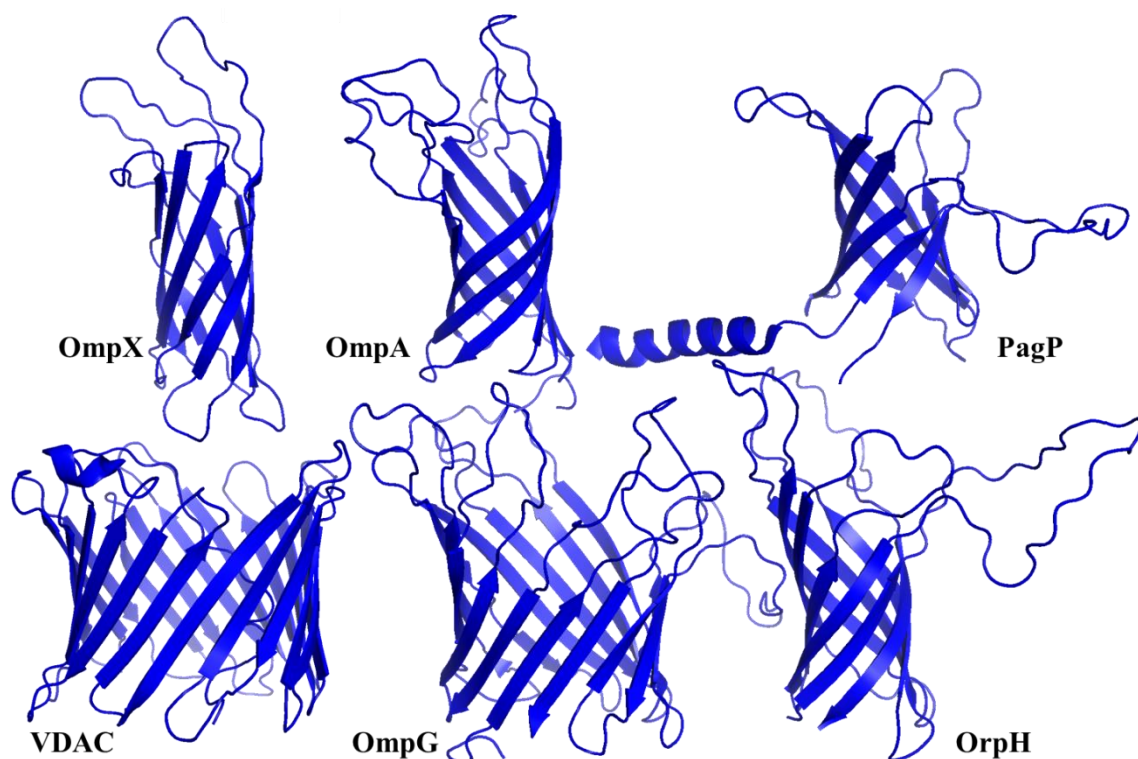
Outer member protein G (OmpG) and the voltage-dependent anion-selective 1 (VDAC-1) represent the two largest  $\beta$ -barrels solved by NMR spectroscopy. These two proteins are the only two barrel structures solved by NMR with strand numbers greater than eight; 14 for OmpG and 19 for VDAC.

OmpG functions as a porin in *E. coli*. The primary function of OmpG is to facilitate the uptake of large oligosaccharides.(61) Studies reveal that OmpG takes on a pH dependent open or closed state by a conformational change of extracellular loop 6, which results in blocking the central pore in acidic conditions.(62) For structural studies, OmpG was explicitly tested in different detergent micelles, optimizing the most stable environment to result in quality NMR spectra.(53) After initially attempting studies in octyl-glucoside micelles, more resolved spectra were obtained using dodecyl-phosphocholine micelles, which contain headgroups that more closely match those found *in vivo* lipid bilayers. TROSY-based experiments were performed, assigning 236 of the 280 amino acids. The majority of the unassigned residues were located in loop 6, the most mobile region of the protein.(53) The larger barrel size led to a total of 316

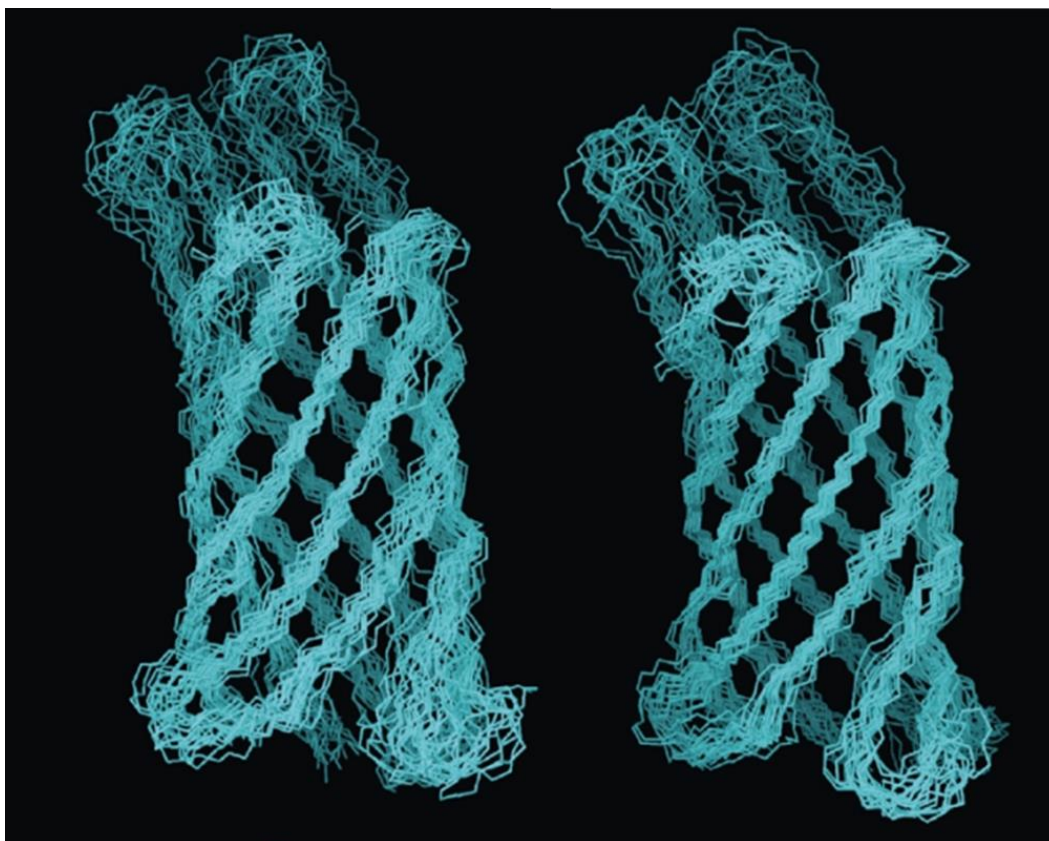


unique HN-HN NOE distances calculated. From these studies, the rmsd of the backbone of the barrel was determined to be 1.67 Å.(53)

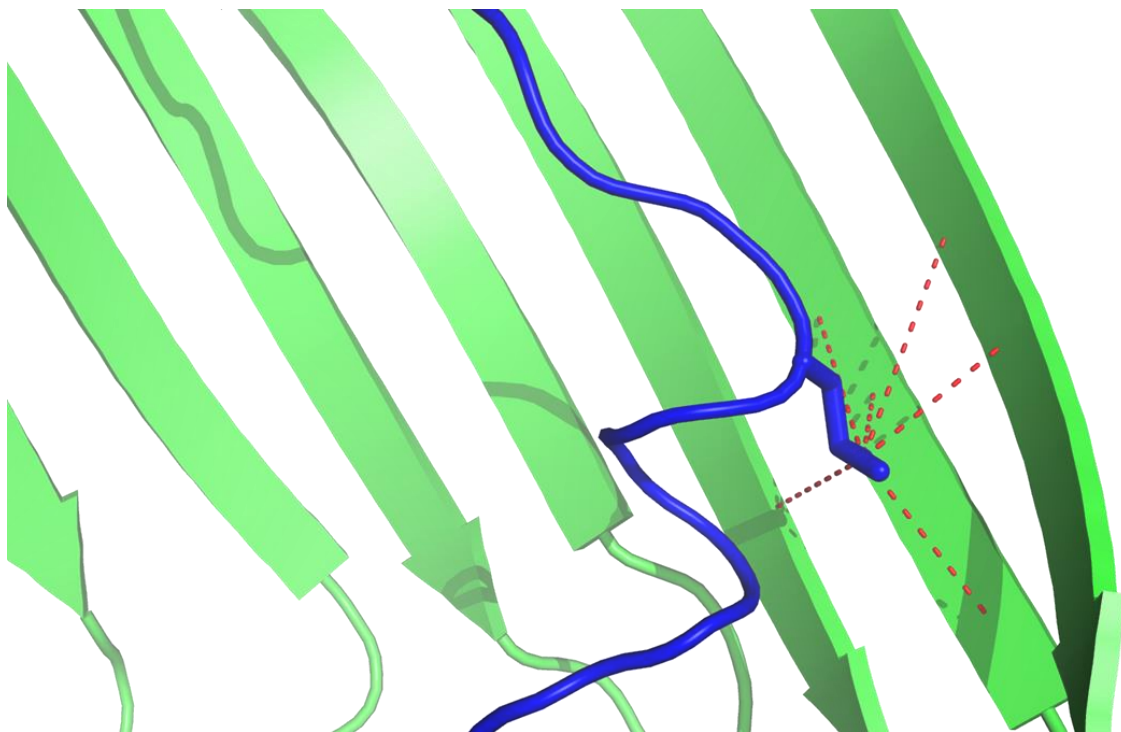
VDAC-1 represents the largest  $\beta$ -barrel structure solved by NMR.(52) The 19 stranded  $\beta$ -barrel contains 291 amino acids. Electron microscopy had previously shown that the transmembrane barrel contains a pore of approximately 20-30 Å in diameter.(63) When open, a variety of metabolites are able to pass through the channel.(64) The function of the VDAC family has been linked to mitochondrial apoptosis by forming a mitochondrial exit channel, allowing for the release of apoptogenic proteins that can then activate executioner caspase.(65) Using TROSY-based experiments, 80% of the backbone was able to be assigned. Methyl labeling was used to improve the distance restraints, most notably by obtaining contact information between the 23 amino acid “in-plug” N-terminus with the interior of the barrel (Figure 3.10).(52) Overall, 600 NOEs were obtained leading to structure determination.



**Figure 3.8:  $\beta$ -barrel structures solved via solution NMR.** There are six  $\beta$ -barrel structures that have been determined by solution NMR spectroscopy. They vary structurally from 18 kDa eight-stranded barrels to 33 kDa barrels comprised of nineteen  $\beta$ -strands.



**Figure 3.9: Structural improvement of OmpX via methyl NOEs.** With the incorporation of  $^1\text{H}$  labeled methyl groups of Ile, Leu, and Val, the rmsd of the ensemble of OmpX structures drastically improved by over 0.5 Å resolution. The ensemble determined without methyl NOEs (left) contained considerably less tight agreement amongst the strands than the structures solved with the methyl NOEs (right). The methyl restraints also showed considerable more ordering in the periplasmic loops.(48)



**Figure 3.10: Methyl contacts of N-terminus of VDAC-1 with barrel.** The location of the 25 amino acid N-terminus of VDAC-1 (blue) was determined to be stable within the barrel as several NOEs between a methyl on Leu10 contained several NOEs (red dashed lines) with several residues on a localized section of the interior of the barrel.

### 3.4 Overview

The research presented in this thesis focuses on the 238 amino acid protein, Opa<sub>60</sub>, from *Neisseria gonorrhoea*. Opa<sub>60</sub> is responsible for inducing phagocytosis of the bacterium by engaging human host cells. Many of the NMR techniques highlighted in this chapter were used to obtain residue assignments and the three dimensional structure of Opa<sub>60</sub>. Also presented are the development of new techniques necessary for the assignment and ultimately the structure determination of the protein. Following structure determination via NMR and in collaboration with Peter Kasson and Per Larrson, molecular dynamics of the NMR structures were obtained for further structure refinement, specifically for the large mobile loop regions which lacked restraints from NMR studies alone. The structural and dynamic insights of Opa<sub>60</sub> gained from these studies are crucial in determining how Opa proteins interact with human host cell receptors.

1. Williamson, M. P. (2009) Applications of the NOE in Molecular Biology, *Annu Rep Nmr Spectro* 65, 77-109.
2. Ottiger, M., and Bax, A. (1997) An empirical correlation between amide deuterium isotope effects on C-13(alpha) chemical shifts and protein backbone conformation, *J Am Chem Soc* 119, 8070-8075.
3. Clore, G. M., Gronenborn, A. M., Szabo, A., and Tjandra, N. (1998) Determining the magnitude of the fully asymmetric diffusion tensor from heteronuclear relaxation data in the absence of structural information, *J Am Chem Soc* 120, 4889-4890.
4. Billeter, M., Wagner, G., and Wuthrich, K. (2008) Solution NMR structure determination of proteins revisited, *J Biomol Nmr* 42, 155-158.
5. Wuthrich, K. (1990) Protein-Structure Determination in Solution by Nmr-Spectroscopy, *J Biol Chem* 265, 22059-22062.
6. Wuthrich, K., Billeter, M., and Braun, W. (1984) Polypeptide Secondary Structure Determination by Nuclear Magnetic-Resonance Observation of Short Proton Proton Distances, *J Mol Biol* 180, 715-740.
7. Kay, L. E., Ikura, M., Tschudin, R., and Bax, A. (1990) 3-Dimensional Triple-Resonance Nmr-Spectroscopy of Isotopically Enriched Proteins, *J Magn Reson* 89, 496-514.
8. Clore, G. M., and Gronenborn, A. M. (1998) NMR structure determination of proteins and protein complexes larger than 20 kDa, *Curr Opin Chem Biol* 2, 564-570.

9. Grzesiek, S., and Bax, A. (1992) An Efficient Experiment for Sequential Backbone Assignment of Medium-Sized Isotopically Enriched Proteins, *J Magn Reson* 99, 201-207.
10. Grzesiek, S., and Bax, A. (1992) Improved 3d Triple-Resonance Nmr Techniques Applied to a 31-Kda Protein, *J Magn Reson* 96, 432-440.
11. Wishart, D. S., and Nip, A. M. (1998) Protein chemical shift analysis: a practical guide, *Biochem Cell Biol* 76, 153-163.
12. Bax, A., Clore, G. M., and Gronenborn, A. M. (1990) H-1-H-1 Correlation Via Isotropic Mixing of C-13 Magnetization, a New 3-Dimensional Approach for Assigning H-1 and C-13 Spectra of C-13-Enriched Proteins, *J Magn Reson* 88, 425-431.
13. Olejniczak, E. T., Xu, R. X., and Fesik, S. W. (1992) A 4d-Hcch-Tocsy Experiment for Assigning the Side-Chain H-1-Resonance and C-13-Resonance of Proteins, *J Biomol Nmr* 2, 655-659.
14. Grzesiek, S., Anglister, J., and Bax, A. (1993) Correlation of Backbone Amide and Aliphatic Side-Chain Resonances in C-13/N-15-Enriched Proteins by Isotropic Mixing of C-13 Magnetization, *J Magn Reson Ser B* 101, 114-119.
15. Montelione, G. T., Lyons, B. A., Emerson, S. D., and Tashiro, M. (1992) An Efficient Triple Resonance Experiment Using C-13 Isotropic Mixing for Determining Sequence-Specific Resonance Assignments of Isotopically-Enriched Proteins, *J Am Chem Soc* 114, 10974-10975.

16. Sattler, M., Schleucher, J., and Griesinger, C. (1999) Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulsed field gradients, *Prog Nucl Mag Res Sp* 34, 93-158.
17. Griesinger, C., Sorensen, O. W., and Ernst, R. R. (1986) Correlation of Connected Transitions by Two-Dimensional Nmr-Spectroscopy, *J Chem Phys* 85, 6837-6852.
18. Williamson, M. P. (1990) Secondary-Structure Dependent Chemical-Shifts in Proteins, *Biopolymers* 29, 1428-1431.
19. Cornilescu, G., Delaglio, F., and Bax, A. (1999) Protein backbone angle restraints from searching a database for chemical shift and sequence homology, *J Biomol Nmr* 13, 289-302.
20. Shen, Y., Delaglio, F., Cornilescu, G., and Bax, A. (2009) TALOS plus : a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts, *J Biomol Nmr* 44, 213-223.
21. Clore, G. M., and Gronenborn, A. M. (1991) Structures of Larger Proteins in Solution - 3-Dimensional and 4-Dimensional Heteronuclear Nmr-Spectroscopy, *Science* 252, 1390-1399.
22. Lin, Y. X., and Wagner, G. (1999) Efficient side-chain and backbone assignment in large proteins: Application to tGCN5, *J Biomol Nmr* 15, 227-239.
23. Neuhaus, D., and Williamson, D. (2000) *The Nuclear Overhauser Effect in Structural and Conformational Analysis*, Wiley, New York.



24. Salzmänn, M., Pervushin, K., Wider, G., Senn, H., and Wüthrich, K. (2000) NMR assignment and secondary structure determination of an octameric 110 kDa protein using TROSY in triple resonance experiments, *J Am Chem Soc* 122, 7543-7548.
25. Montalvão, R. W., Cavalli, A., Salvatella, X., Blundell, T. L., and Vendruscolo, M. (2008) Structure Determination of Protein-Protein Complexes Using NMR Chemical Shifts: Case of an Endonuclease Colicin-Immunity Protein Complex, *J Am Chem Soc* 130, 15990-15996.
26. Hong, M., Zhang, Y., and Hu, F. H. (2012) Membrane Protein Structure and Dynamics from NMR Spectroscopy, *Annu Rev Phys Chem* 63, 1-24.
27. Levitt, M. H. (2008) *Spin Dynamics*, Wiley, Oxford.
28. Popchapsky, T. C., and Popchapsky, S. S. (2007) *NMR for Physical and Biological Scientists*, Taylor and Francis, New York.
29. Grzesiek, S., Anglister, J., Ren, H., and Bax, A. (1993) C-13 Line Narrowing by H-2 Decoupling in H-2/C-13/N-15-Enriched Proteins - Application to Triple-Resonance 4d J-Connectivity of Sequential Amides, *J Am Chem Soc* 115, 4369-4370.
30. Garrett, D. S., Seok, Y. J., Liao, D. I., Peterkofsky, A., Gronenborn, A. M., and Clore, G. M. (1997) Solution structure of the 30 kDa N-terminal domain of enzyme I of the Escherichia coli phosphoenolpyruvate:sugar phosphotransferase system by multidimensional NMR, *Biochemistry-US* 36, 2517-2530.

31. Clore, G. M., Wingfield, P. T., and Gronenborn, A. M. (1991) High-Resolution 3-Dimensional Structure of Interleukin-1-Beta in Solution by 3-Dimensional and 4-Dimensional Nuclear-Magnetic-Resonance Spectroscopy, *Biochemistry-Us* 30, 2315-2323.
32. Martin, J. R., Mulder, F. A. A., KarimiNejad, Y., vanderZwan, J., Mariani, M., Schipper, D., and Boelens, R. (1997) The solution structure of serine protease PB92 from *Bacillus alcalophilus* presents a rigid fold with a flexible substrate-binding site, *Structure* 5, 521-532.
33. Weigelt, J., van Dongen, M., Uppenberg, J., Schultz, J., and Wikstrom, M. (2002) Site-selective screening by NMR spectroscopy with labeled amino acid pairs, *J Am Chem Soc* 124, 2446-2447.
34. Parker, M. J., Aulton-Jones, M., Hounslow, A. M., and Craven, C. J. (2004) A combinatorial selective labeling method for the assignment of backbone amide NMR resonances, *J Am Chem Soc* 126, 5020-5021.
35. Tugarinov, V., and Kay, L. E. (2003) Ile, Leu, and Val methyl assignments of the 723-residue malate synthase G using a new labeling strategy and novel NMR methods, *J Am Chem Soc* 125, 13868-13878.
36. Tugarinov, V., and Kay, L. E. (2004) An isotope labeling strategy for methyl TROSY spectroscopy, *J Biomol Nmr* 28, 165-172.
37. Grishaev, A., Tugarinov, V., Kay, L. E., Trewella, J., and Bax, A. (2008) Refined solution structure of the 82-kDa enzyme malate synthase G from joint NMR and synchrotron SAXS restraints, *J Biomol Nmr* 40, 95-106.

38. Korzhnev, D. M., Kloiber, K., Kanelis, V., Tugarinov, V., and Kay, L. E. (2004) Probing slow dynamics in high molecular weight proteins by methyl-TROSY NMR spectroscopy: Application to a 723-residue enzyme, *J Am Chem Soc* 126, 3964-3973.
39. Battiste, J. L., and Wagner, G. (2000) Utilization of site-directed spin labeling and high-resolution heteronuclear nuclear magnetic resonance for global fold determination of large proteins with limited nuclear overhauser effect data, *Biochemistry-Us* 39, 5355-5365.
40. Fries, P. H., and Belorizky, E. (2005) Electronic relaxation of paramagnetic metal ions and NMR relaxivity in solution: Critical analysis of various approaches and application to a Gd(III)-based contrast agent, *J Chem Phys* 123.
41. Berliner, L. J., Grunwald, J., Hankovszky, H. O., and Hideg, K. (1982) A Novel Reversible Thiol-Specific Spin Label - Papain Active-Site Labeling and Inhibition, *Anal Biochem* 119, 450-455.
42. Gillespie, J. R., and Shortle, D. (1997) Characterization of long-range structure in the denatured state of staphylococcal nuclease .1. Paramagnetic relaxation enhancement by nitroxide spin labels, *J Mol Biol* 268, 158-169.
43. Gillespie, J. R., and Shortle, D. (1997) Characterization of long-range structure in the denatured state of staphylococcal nuclease .2. Distance restraints from paramagnetic relaxation and calculation of an ensemble of structures, *J Mol Biol* 268, 170-184.

44. Liang, B. Y., Bushweller, J. H., and Tamm, L. K. (2006) Site-directed parallel spin-labeling and paramagnetic relaxation enhancement in structure determination of membrane proteins by solution NMR spectroscopy, *J Am Chem Soc* 128, 4389-4397.
45. Pervushin, K., Riek, R., Wider, G., and Wuthrich, K. (1997) Attenuated T-2 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, *P Natl Acad Sci USA* 94, 12366-12371.
46. Tjandra, N., Szabo, A., and Bax, A. (1996) Protein backbone dynamics and N-15 chemical shift anisotropy from quantitative measurement of relaxation interference effects, *J Am Chem Soc* 118, 6986-6991.
47. Gueron, M., Leroy, J. L., and Griffey, R. H. (1983) Proton Nuclear Magnetic-Relaxation of N-15-Labeled Nucleic-Acids Via Dipolar Coupling and Chemical-Shift Anisotropy, *J Am Chem Soc* 105, 7262-7266.
48. Fernandez, C., Hilty, C., Wider, G., Guntert, P., and Wuthrich, K. (2004) NMR structure of the integral membrane protein OmpX, *J Mol Biol* 336, 1211-1221.
49. Fernandez, C., Adeishvili, K., and Wuthrich, K. (2001) Transverse relaxation-optimized NMR spectroscopy with the outer membrane protein OmpX in dihexanoyl phosphatidylcholine micelles, *P Natl Acad Sci USA* 98, 2358-2363.

50. 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 Biol* 8, 334-338.
51. Hwang, P. M., Choy, W. Y., Lo, E. I., Chen, L., Forman-Kay, J. D., Raetz, C. R. H., Prive, G. G., Bishop, R. E., and Kay, L. E. (2002) Solution structure and dynamics of the outer membrane enzyme PagP by NMR, *P Natl Acad Sci USA* 99, 13560-13565.
52. 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.
53. Liang, B. Y., and Tamm, L. K. (2007) Structure of outer membrane protein G by solution NMR spectroscopy, *P Natl Acad Sci USA* 104, 16140-16145.
54. 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*, *J Biol Chem* 286, 39211-39223.
55. Koebnik, R., Locher, K. P., and Van Gelder, P. (2000) Structure and function of bacterial outer membrane proteins: barrels in a nutshell, *Mol Microbiol* 37, 239-253.
56. Kim, K. S. (2002) Strategy of *Escherichia coli* for crossing the blood-brain barrier, *J Infect Dis* 186, S220-S224.

57. Prasadarao, N. V., Blom, A. M., Villoutreix, B. O., and Linsangan, L. C. (2002) A novel interaction of outer membrane protein A with C4b binding protein mediates serum resistance of *Escherichia coli* K1, *J Immunol* 169, 6352-6360.
58. Powers, R., Clore, G. M., Bax, A., Garrett, D. S., Stahl, S. J., Wingfield, P. T., and Gronenborn, A. M. (1991) Secondary Structure of the Ribonuclease-H Domain of the Human-Immunodeficiency-Virus Reverse-Transcriptase in Solution Using 3-Dimensional Double and Triple Resonance Heteronuclear Magnetic-Resonance Spectroscopy, *J Mol Biol* 221, 1081-1090.
59. Dekort, G., Bolton, A., Martin, G., Stephen, J., and Vandeklundert, J. A. M. (1994) Invasion of Rabbit Ileal Tissue by *Enterobacter-Cloacae* Varies with the Concentration of Ompx in the Outer-Membrane, *Infect Immun* 62, 4722-4726.
60. Vogt, J., and Schulz, G. E. (1999) The structure of the outer membrane protein OmpX from *Escherichia coli* reveals possible mechanisms of virulence, *Struct Fold Des* 7, 1301-1309.
61. Fajardo, D. A., Cheung, J., Ito, C., Sugawara, E., Nikaido, H., and Misra, R. (1998) Biochemistry and regulation of a novel *Escherichia coli* K-12 porin protein, OmpG, which produces unusually large channels, *J Bacteriol* 180, 4452-4459.

62. Yildiz, O., Vinothkumar, K. R., Goswami, P., and Kuhlbrandt, W. (2006) Structure of the monomeric outer-membrane porin OmpG in the open and closed conformation, *Embo J* 25, 3702-3713.
63. Guo, X. W., Smith, P. R., and Mannella, C. A. (1995) Molecular Design of the Voltage-Dependent, Anion-Selective Channel in the Mitochondrial Outer-Membrane, *J Struct Biol* 114, 41-59.
64. Hodge, T., and Colombini, M. (1997) Regulation of metabolite flux through voltage-gating of VDAC channels, *J Membrane Biol* 157, 271-279.
65. Desagher, S., and Martinou, J. C. (2000) Mitochondria as the central control point of apoptosis, *Trends Cell Biol* 10, 369-377.

#### **Section 4 Backbone assignment of Opa<sub>60</sub>**

NMR spectroscopy has shown to be an effective tool for studying the structure (1, 2) and dynamics (3, 4) of  $\beta$ -barrel membrane proteins. The field has developed over the twelve years since the first  $\beta$ -barrel structure was solved using NMR spectroscopy.(5) Despite a relatively low number of  $\beta$ -barrel structures determined over that time (only six total (5-10), or one every other year on average) each of the structures are unique and have helped push the envelope of structural NMR studies. The first step in determining the structure of Opa<sub>60</sub> is to assign the backbone resonances.(11) Tradition backbone assignment strategies prove to be insufficient to obtain a full assignment of Opa<sub>60</sub>. New assignment strategies are presented to circumvent the difficulties that the long unstructured loops present. (12) Assignment of the barrel was achieved by removing the loops with a protease, which leaves the barrel structure essentially unchanged basedon chemical shift perturbation. The most mobile regions of the loops were able to be assigned by adjusting the temperature such that only the mobile regions remain visible to NMR spectroscopy. Specific regions of the loops were synthesized and the NMR spectra of these smaller peptides overlapped well to the corresponding full length regions which aided in assignment and dynamic studies. These techniques were used to obtain sufficient assignment coverage to continue towards structural studies.



## **Section 4.1 Stable sample preparation**

### **Section 4.1.1 Refolding conditions**

After purifying denatured Opa<sub>60</sub> (see Appendix) in 8 M urea,(13) a proper membrane mimic needed to be determined to refold Opa<sub>60</sub>. Detergent micelles offer the best option for NMR studies as they form a hydrophobic region necessary to stabilize the transmembrane region of the protein (14) while adding a relatively small mass to the sample (~10-50 kDa per micelle, compared to > MDa masses for lipid vesicles(15)). The small mass is necessary for NMR to minimize the effect the mimic has on increasing the correlation time of the sample, which is directly correlated to the linewidth of the NMR signal. Detergents that have been shown useful in refolding other membrane proteins were screened,(7, 8, 10) in which a 200  $\mu$ L solution of unfolded Opa<sub>60</sub> was rapidly diluted twenty fold into a solutions containing detergent micelles with at least a 10:1 ratio of micelle to protein upon dilution. After incubating the sample at room temperature for four hours, each sample analyzed using sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) to analyze total folding of the protein. Due to the stability of the  $\beta$ -barrel fold, SDS does not completely denature the protein and as such, a folded  $\beta$ -barrel will migrate at a different rate than the unfolded state (Figure 4.1A).(16) By assessing the ratio of unfolded to folded protein in each lane on the gel, dodcyl-phosphcholine produced the most folded Opa<sub>60</sub>.

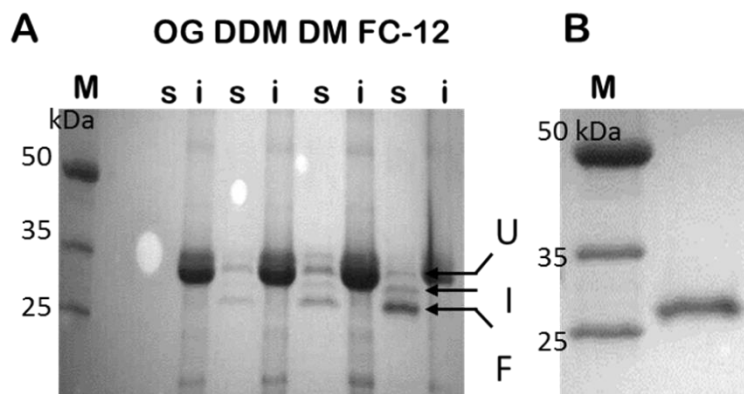
Other  $\beta$ -barrel proteins in literature indicated complete refolding in less than a day, however Opa<sub>60</sub> required a much longer time. The reason for the discrepancy is attributed to the relatively large extracellular loops. In order to refold, the  $\beta$ -strand regions of the protein must interact with each other to form the  $\beta$ -barrel, which is less probable when larger sequences, such as loops 1-3 in Opa<sub>60</sub>, separate the  $\beta$ -strand sequences.  $\beta$ -barrel folding studies have also demonstrated that folding occurs by the extracellular region of the protein traversing through the membrane mimetic.(17, 18) The longer loop regions of Opa<sub>60</sub> contain longer stretches of residues needed to transverse the micelle than other  $\beta$ -barrels previously studied, resulting in a longer refolding time. Complete protein refolding was confirmed by the lack of the unfolded band SDS-PAGE after a five day room temperature incubation time (Figure 4.1B).

#### **Section 4.1.2 Fold assessment**

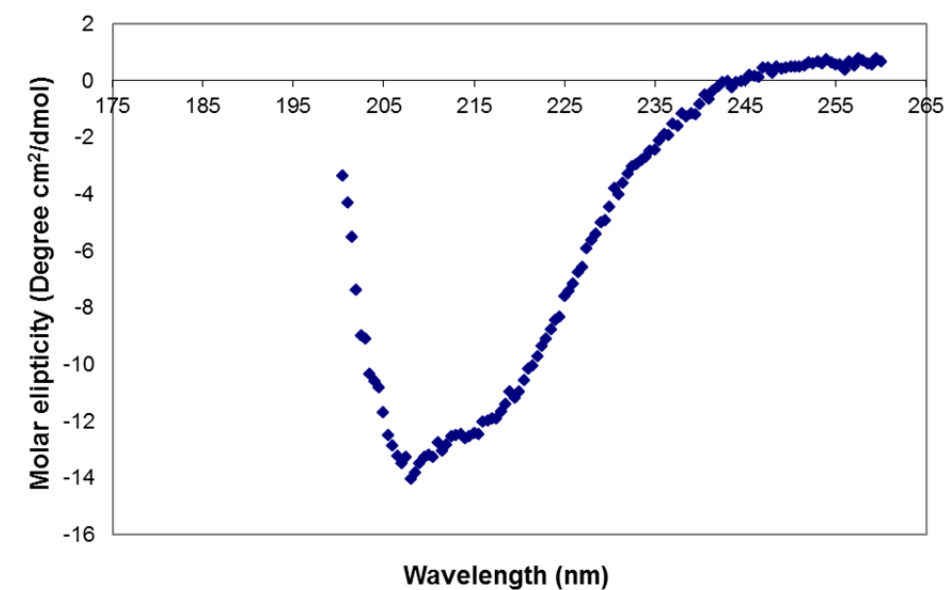
Once Opa<sub>60</sub> was refolded, the next step was to assess the secondary structure as well as a homogenously folded protein. To assess the secondary structure, a circular dichroism (CD) spectrum was recorded for the refolded of Opa<sub>60</sub>. CD spectroscopy records the differential absorption of left and right circularly polarized light, which is unique depending on the secondary structure of the protein.(19) Recorded CD spectra are then compared to a set of standard protein CD spectra to determine the approximate ratio of  $\alpha$ -helix,  $\beta$ -strand and random coil secondary structure present in the protein. The Opa<sub>60</sub> spectrum (Figure 4.2) was analyzed using an algorithm from CD-Pro, estimating that approximately

40% of the protein is  $\beta$ -strand and 60% is random-coil, in good agreement with secondary structure sequence prediction tools.

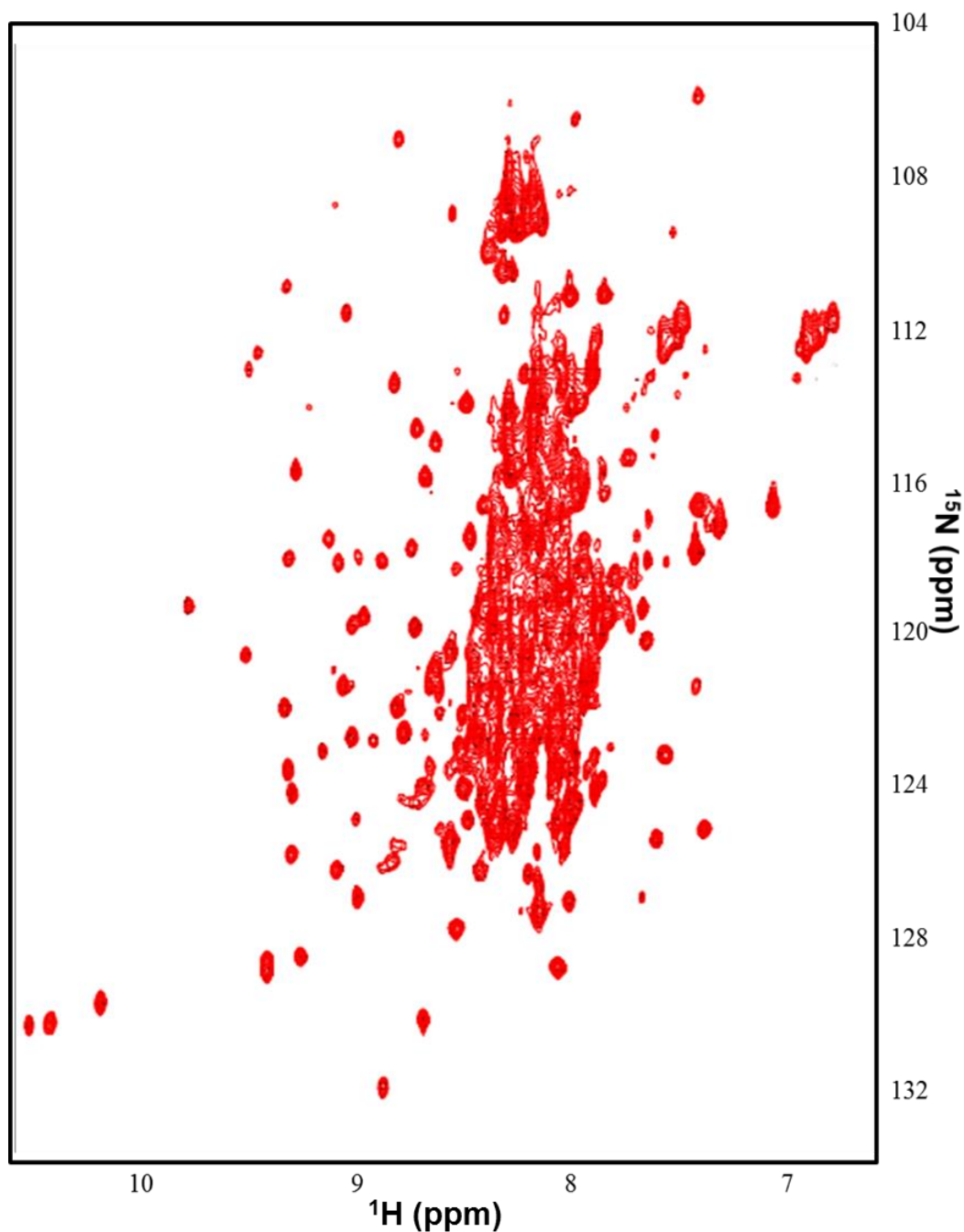
Homogenous fold was tested by recording a two dimension  $^{15}\text{N}$ ,  $^1\text{H}$  TROSY-HSQC spectrum of uniformly  $^{15}\text{N}$ ,  $^2\text{H}$ -labeled Opa<sub>60</sub> (Figure 4.3) recorded on a Bruker AVANCE spectrometer operating at proton frequencies of 600 MHz equipped with Bruker 5 mm TXI cryoprobe at 40 °C. The amide protons that are shifted greater than 8.5 ppm downfield are associated with protons in  $\beta$ -strands. The resonances observed are all approximately of uniform linewidth and peak height of these resonances, indicating a homogenous fold within the barrel region.



**Figure 4.1: A. Refolding of Opa<sub>60</sub> into different detergent micelles.** After an initial incubation period of four hours, Opa<sub>60</sub> was rapidly diluted into four different detergent micelles (OG – octylglucoside, DDM - dodecylmaltocide, DM – decylmaltocide, FC-12 – dodecyl phosphocholine), the solution was centrifuged and both soluble (s) and insoluble (i) fractions were observed via SDS-PAGE to determine amount of folded protein (F) unfolded (U) or intermediate (I). **B. Complete refolding of Opa<sub>60</sub> in FC-12 micelles.** After optimization of refolding conditions, Opa<sub>60</sub> was observed on SDS-PAGE to be completely folded in FC-12 micelles.



**Figure 4.2: Circular dichroism spectrum of Opa<sub>60</sub>.** Analysis of the spectrum by CD Pro indicated that the sample was approximately 40%  $\beta$  character and 60% random coil.



**Figure 4.3:**  $^{15}\text{N}$ ,  $^1\text{H}$  TROSY-HSQC spectrum of **Opa<sub>60</sub>**. The sample was uniformly  $^2\text{H}$ ,  $^{15}\text{N}$  labeled. Disperse downfield peaks indicate a homogenous  $\beta$ -fold whereas the random coil region ( $< 8.5$  ppm) reveals significant spectral crowding.

## Section 4.2 NMR assignment of full length protein spectra

### Section 4.2.1 Triple resonance experiments

There were 225 unique resonances observed in the  $^{15}\text{N}$ ,  $^1\text{H}$ -HSQC spectrum of Opa<sub>60</sub> (Figure 4.3). 71 resonances were observed shifted downfield in the well disperse region of the spectrum. The majority of the resonances were recorded between 7.5 and 8.5 ppm in the  $^1\text{H}$  dimension, typical chemical shifts for amide protons of  $\alpha$ -helix or random coil secondary structure.(20) The resonances in this region were also not of uniform intensity or linewidth. These lineshapes imply that the recorded amide protons are in different dynamic regions of the protein. The more intense peaks can indicate regions of the protein where the tumbling is enhanced. The broadening of these peaks also implies conformational exchange of specific regions.(21) Due to the spectral crowding of varying intensities, many of the less intense signals were unable to be observed in this region.

A suite of TROSY-based triple resonance experiments were executed, including HNCA, HN(CO)CA, HNCO, HNCACB, and HN(CO)CACB.(22) (23) A variant of the HNCACB experiment, i-HNCACB,(24) was also performed, which results in resonances of the aliphatic carbons only on the *i* amino acid, as opposed to the *i*-1 in the case of the HN(CO)CACB (Figure 4.4). These experiments were performed at 40 °C. The higher temperature allowed for the protein to tumble faster, resulting in sharper, more observable peaks for the less intense, broadened peaks of the  $\beta$ -barrel region. While over 50 amino acids were able to be assigned via these experiments, the vast majority of the protein sequence

remained unable to be assigned. There were multiple issues that led to this dearth of assignments from the full length spectra. One issue was sequential redundancy, in which resonances can be connected via triple resonance experiments, but there are multiple possible peptide sequences that fit that series of chemical shifts. The spectral crowding of many resonances resulted in multiple apparent carbon chemical shifts for a single spin system; or in some cases the less intense peaks that were in densely populated regions were lost in the noise of the more intense peaks surrounding them.

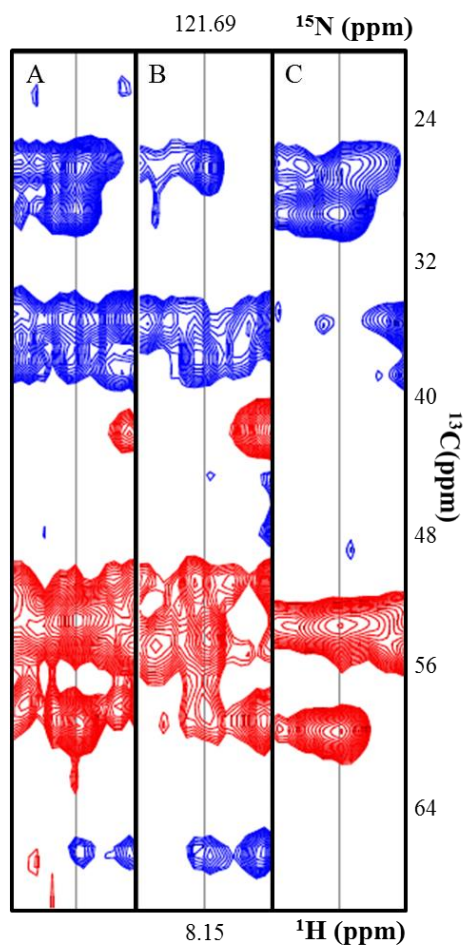
#### **Section 4.2.2 Specific amino acid isotope labeling**

A variety of specifically labeled amino acid samples were prepared in an attempt to assign the resonances unambiguously in the crowded, random coil region of the spectra. Briefly, the technique used to generate specific amino acids: 1.) Opa<sub>60</sub> was expressed in *E. coli* in the presence of natural abundance nitrogen and carbon sources. 2.) Prior to inducing the bacteria to produce Opa<sub>60</sub>, <sup>15</sup>N-labeled amino acids were added to the media. 3.) The *E. coli* incorporates these amino acids, which barring metabolic breakdown and scrambling, will result with specific <sup>15</sup>N labeled amino acids.(25) Four initial samples were generated, each combining two different <sup>15</sup>N-labeled amino acids; arginine and serine (21 and 15 residues in the sequence, respectively), isoleucine and glutamate (16 and 10), valine and lysine (17 and 13), and asparagine and tyrosine (19 and 17). Each of the corresponding <sup>15</sup>N, <sup>1</sup>H-HSQC spectra resulted in a total number of amide resonances greater than expected (Figure 4.5). Analysis of the <sup>13</sup>C chemical shifts acquired from the triple resonance experiments for these respective <sup>15</sup>N, <sup>1</sup>H

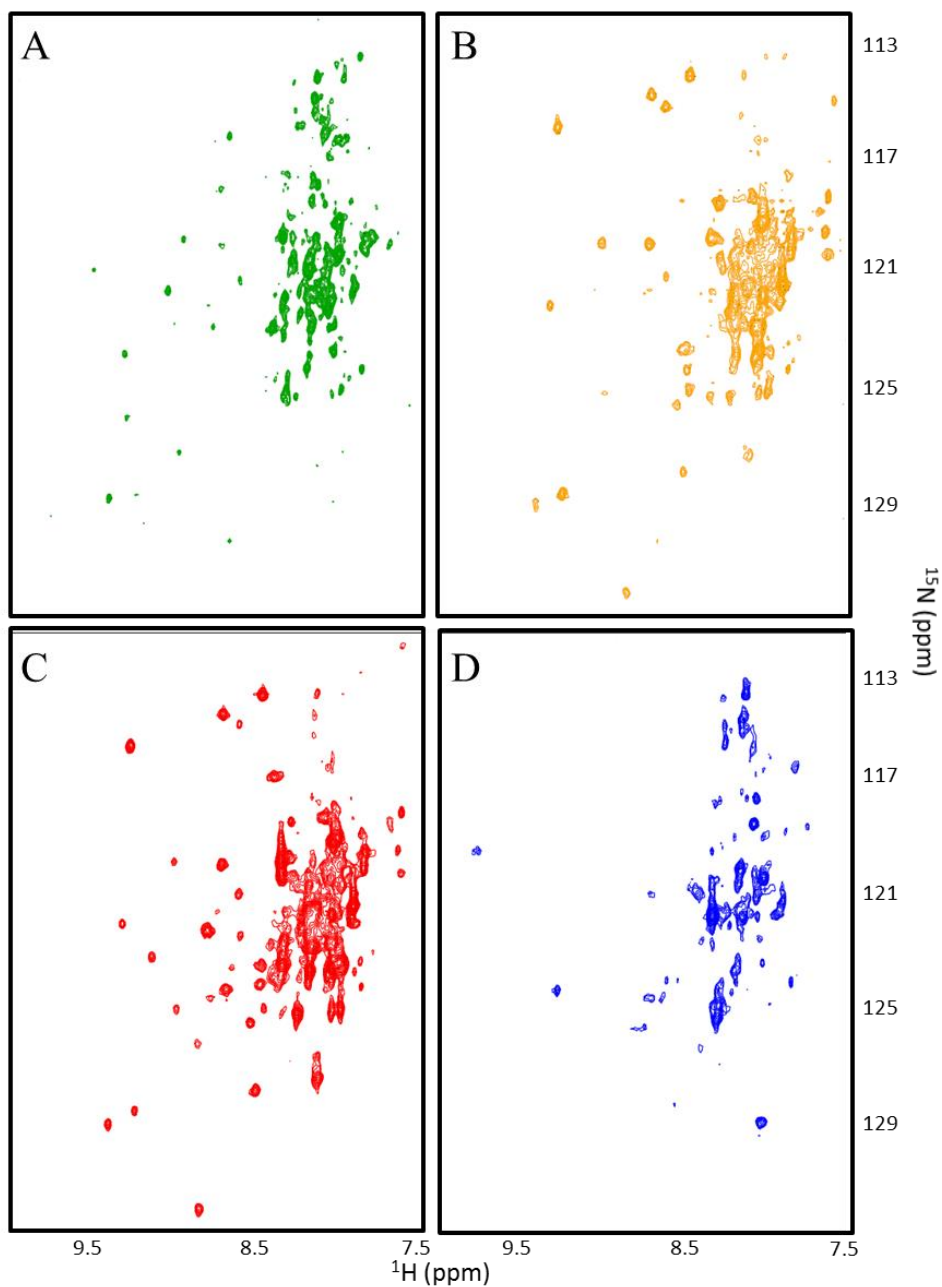


systems led to concluding that varying degrees of metabolic scrambling of the isotopes occurred when the *E. coli* produced Opa<sub>60</sub>.<sup>(26)</sup> Redundant <sup>15</sup>N, <sup>1</sup>H resonances observed in multiple spectra coupled with <sup>13</sup>C chemical shifts were used to identify the amino acid types of many of these resonances. For example, a peak present in both the isoleucine labeling and valine labeling scheme was likely to be Ile, Leu, Val or Ala as both Ile and Val could be easily metabolized to any of those four amino acids, all of which have carbon chemical shifts that can be differentiated from each other, with average chemical shifts of 61.62 ppm ( $\alpha$ -carbon) and 38.61 ppm ( $\beta$ -carbon) for Ile, 55.65 ppm and 42.30 ppm for Leu, 53.17 ppm and 19.07 ppm for Ala, and 62.53 ppm and 32.75 ppm for Val for  $\alpha$ -carbon and  $\beta$ -carbon shifts, respectively.<sup>(27)</sup>

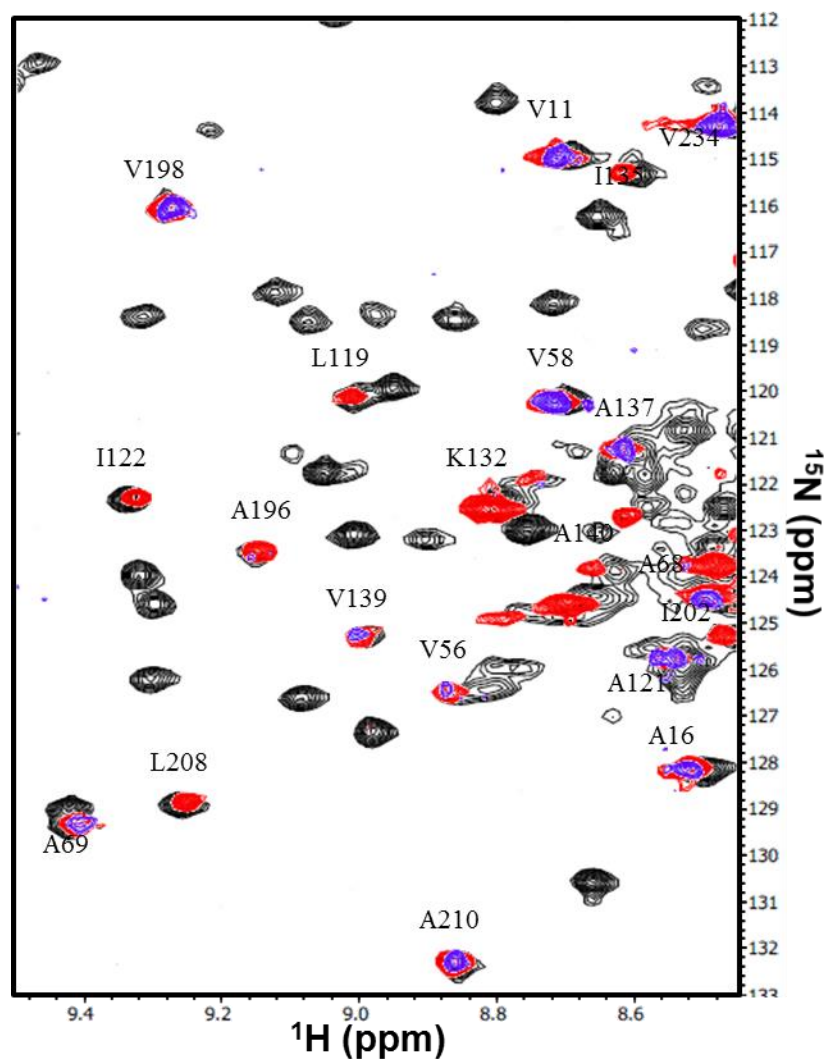
Another isotopic labeling technique used to minimize the effects of metabolic scrambling was the incorporation of NMR inactive, natural abundance <sup>14</sup>N-labeled amino acids into the growth media. A sample of Opa<sub>60</sub> was generated using <sup>15</sup>N-labeled valine with natural abundance <sup>14</sup>N leucine and isoleucine prior to induction. The addition of these unlabeled amino acids favors the more direct metabolic pathway of the *E.coli* to incorporate these amino acids directly into the protein as opposed to metabolizing the <sup>15</sup>N-labeled valine into isoleucine and leucine (Figure 4.6). This technique was useful in assigning nearly every valine present in Opa<sub>60</sub> when combined with the other techniques used in this chapter.



**Figure 4.4: Spectral overlap of full length Opa<sub>60</sub>.** A sample set of  $^{13}\text{C}$  strips in which three spectra are required (A. HNCACB, B. HN(CO)CACB, C. i-HNCACB) to identify resonances from multiple spin systems (as indicated by multiple i-1 peaks observed in B and multiple i peaks in C.  $\alpha$ -carbon shifts are in red and  $\beta$ -carbon shifts appear in blue.



**Figure 4.5: Specific amino acid labeling of Opa<sub>60</sub>.** Four different labeling schemes (A - Asp and Tyr, B – Glu and Ile, C – Val and Lys and D – Ser and Arg) were used to aid in the assignment of Opa<sub>60</sub>.



**Figure 4.6: Supplemented specific amino acid labeling of Opa<sub>60</sub>.** Region of  $^{15}\text{N}$ ,  $^1\text{H}$ -HSQC spectra of Opa<sub>60</sub> with specifically labeled amino acids; uniformly  $^{15}\text{N}$ -labeled sample (black),  $^{14}\text{N}$  ammonium chloride and  $^{15}\text{N}$ -labeled valine and lysine amino acids (red),  $^{14}\text{N}$  ammonium chloride,  $^{15}\text{N}$ -labeled valine and  $^{14}\text{N}$ -labeled leucine and isoleucine amino acids (blue). The addition of the NMR inactive isotopes minimizes the effects amino acid scrambling of the NMR active  $^{15}\text{N}$ -labeled valine.

### Section 4.3 Assignment of the $\beta$ -barrel region

To address the region of the spectra crowded by broadened loop signal, the extracellular loops were cleaved using a protease. Many proteases do not partition well into the lipid bilayer, thus preserving any cleavage sites on the transmembrane region.(10) Opa<sub>60</sub> was treated with three different common proteases at a ratio of 50:1 Opa<sub>60</sub>:protease: chymotrypsin, which catalyzes the hydrolysis of peptide bonds on the C-terminal side of aromatic residues, such as tyrosine, tryptophan and phenylalanine; thermolysin, which cleaves on the C-terminal side of hydrophobic amino amino acids; and trypsin, which cleaves on the C-terminal side of charged amino acids, lysine and arginine. The cleaved samples were visualized via SDS-PAGE after 1, 5, 15, and 60 minutes to determine the extent of proteolysis and the stability of the cleaved  $\beta$ -barrel-detergent complex (Figure 4.7). After an hour, complete digestion of all exposed cleavage sites was observed for both thermolysin and trypsin, whereas chymotrypsin required additional time for full proteolysis. In all three cases, strong bands ranging between 15-22 kDa remained, indicating that the  $\beta$ -barrel of Opa<sub>60</sub> remained stable and intact throughout the proteolysis. Based on the amino acid sequence, the predicted topology of the  $\beta$ -barrel, and the inaccessibility of the  $\beta$ -barrel to proteases, trypsin is predicted to remove 94 residues (Figure 4.8). The remaining micelle-protected  $\beta$ -barrel would have an approximate molecular weight of 15 kDa (~30-35 kDa including the micelle).

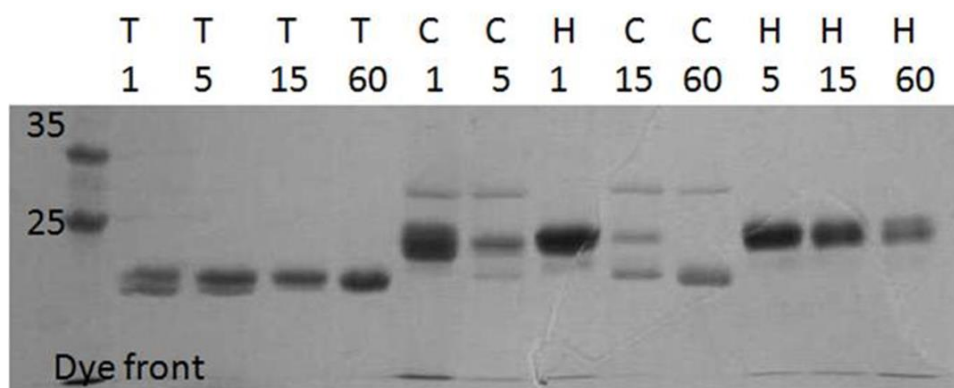
After cleaving a uniformly labeled <sup>15</sup>N, <sup>2</sup>H sample of Opa<sub>60</sub>, a <sup>15</sup>N, <sup>1</sup>H HSQC spectrum was recorded to assess the perturbation of the local environment of the

remaining amino acids (Figure 4.9). 147 unique resonances remain, almost all are identical to the full-length spectrum, specifically in the downfield shifted amide region. The lack of a change in chemical shift implies that the remaining amino acids are structurally unperturbed from the full-length protein. Opa<sub>60</sub> proteolyzed by trypsin can be supplemented with the full-length protein to aid in the assignment. The benefit of this technique can be observed in the <sup>15</sup>N <sup>1</sup>H HSQC spectrum (Figure 4.9), in which many of the larger broadened peaks associated with the extracellular loops are now removed, providing the less intense peaks to be monitored.

Another Opa protein, Opa<sub>50</sub>, was also treated by trypsin by Ryan Lo (University of Virginia Department of Chemistry). Opa<sub>50</sub> is an HSPG-binding protein that has a 96% sequence identity with Opa<sub>60</sub> in the proposed barrel region; but different loops sequences.(28) For many of the downfield shifted peaks, the chemical shifts align very well amongst the two different trypsin-treated Opa proteins (Figure 4.10) indicating that the  $\beta$ -barrel fold for these two different proteins is very similar. Because of the similar overlap, future studies of Opa<sub>50</sub> will benefit from assignment of Opa<sub>60</sub>, and an Opa<sub>60</sub>  $\beta$ -barrel structure can be used as a scaffold to model future Opa  $\beta$ -barrel structures. This overlap is also useful in the assignment of Opa<sub>60</sub>, as resonances that overlap in both protein samples are likely to result from the same amino acid.

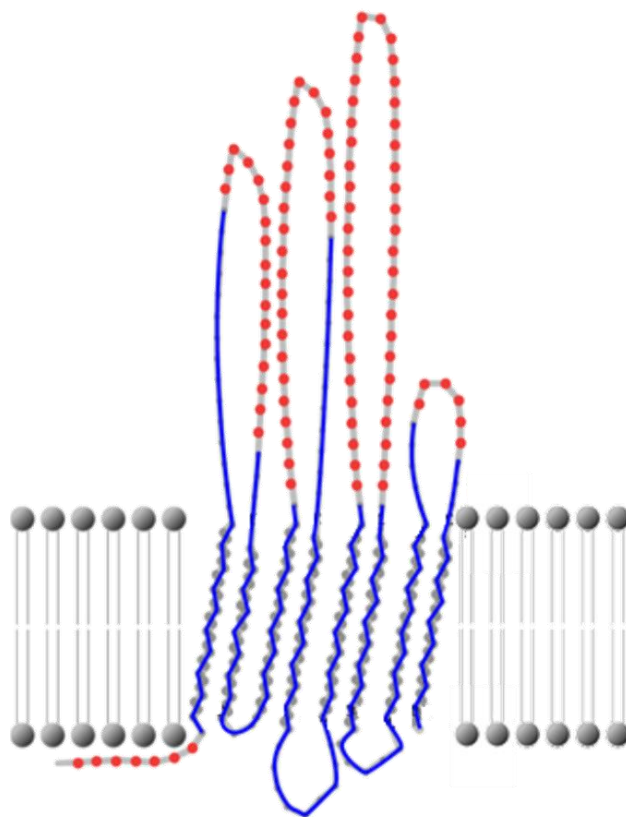
The same suite of triple resonance experiments that were applied to the full-length sample were performed on the trypsin-cleaved Opa<sub>60</sub> sample, resulting in 97% assignment of the transmembrane  $\beta$ -barrel and periplasmic turns of the

Opa<sub>60</sub> (Figure 4.11). Furthermore, resonances for residues in the barrel located near the micelle interface on the extracellular side were either very broad, or not observed in the full-length spectrum. In the cleaved spectrum, these signals drastically improved. The large conformational exchange of a subset of the loop residues is likely to result from chemical exchange near the interface that would previously have rendered those resonances too broad to observe (Figure 4.12).

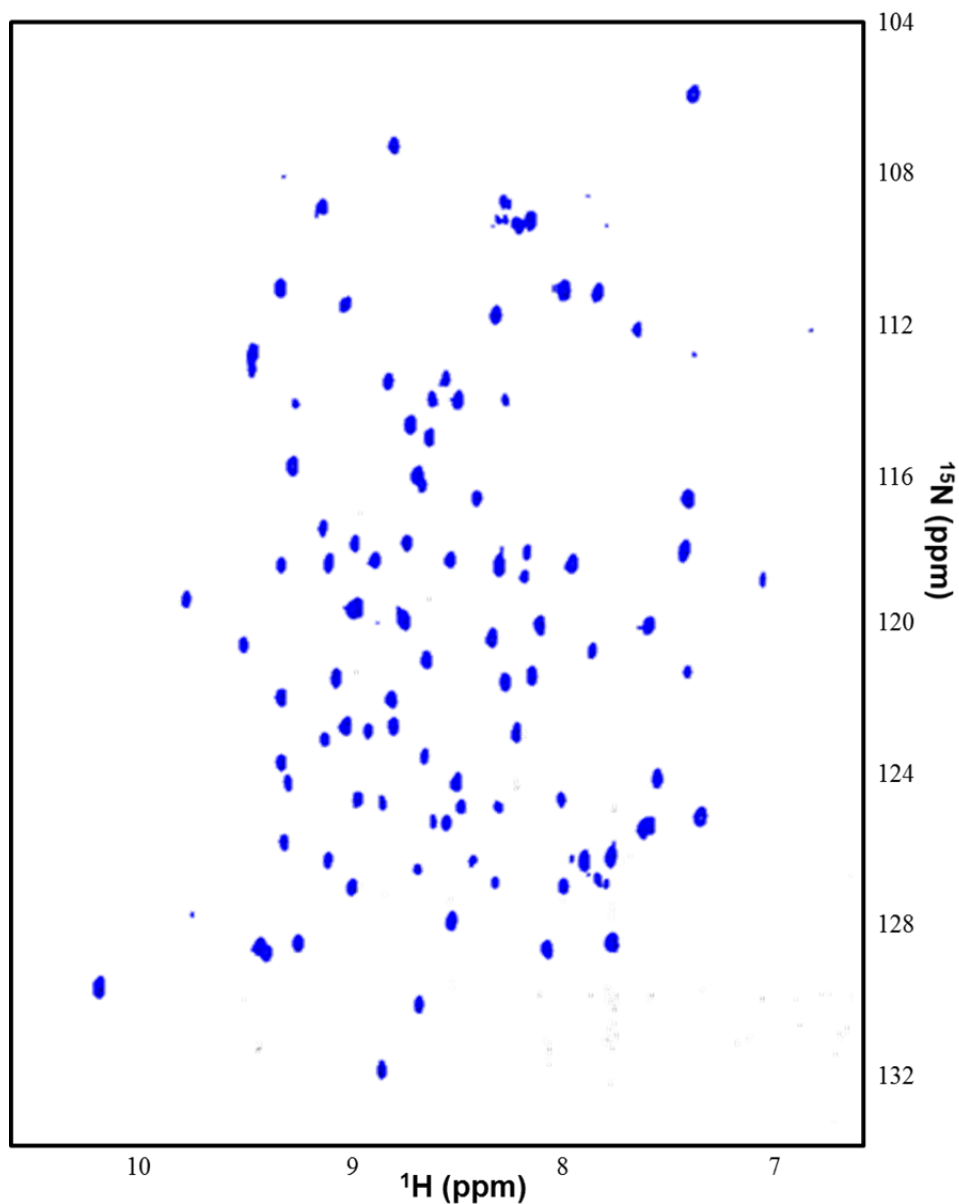


**Figure 4.7: SDS-PAGE of Opa<sub>60</sub> proteolysis.** Protease cleavage of Opa<sub>60</sub> was monitored over the course of an hour for three different proteases (T – trypsin, C – chymotrypsin, H – thermolysin). Time points were taken after 1, 5, 15, and 60 minutes to monitor extent of cleavage.

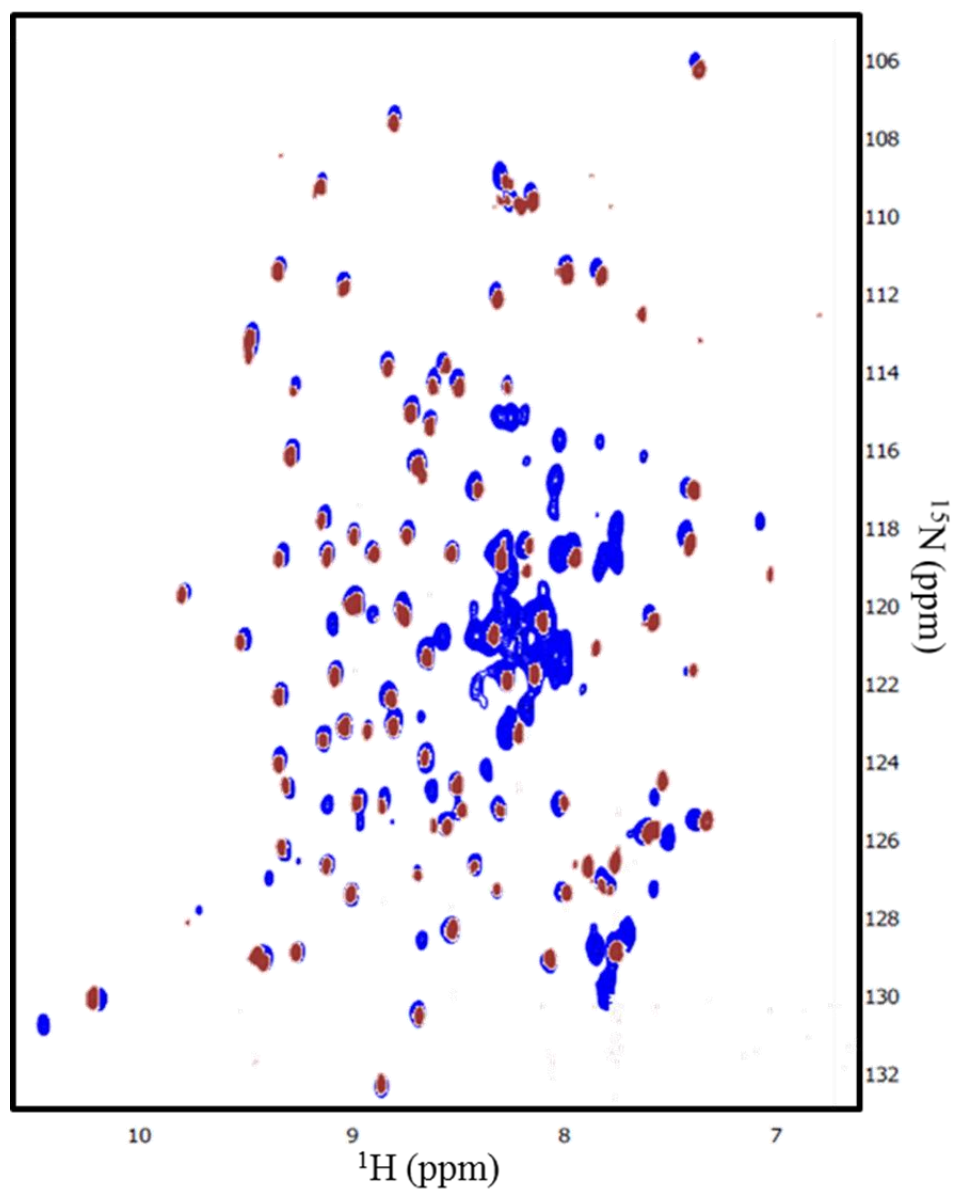




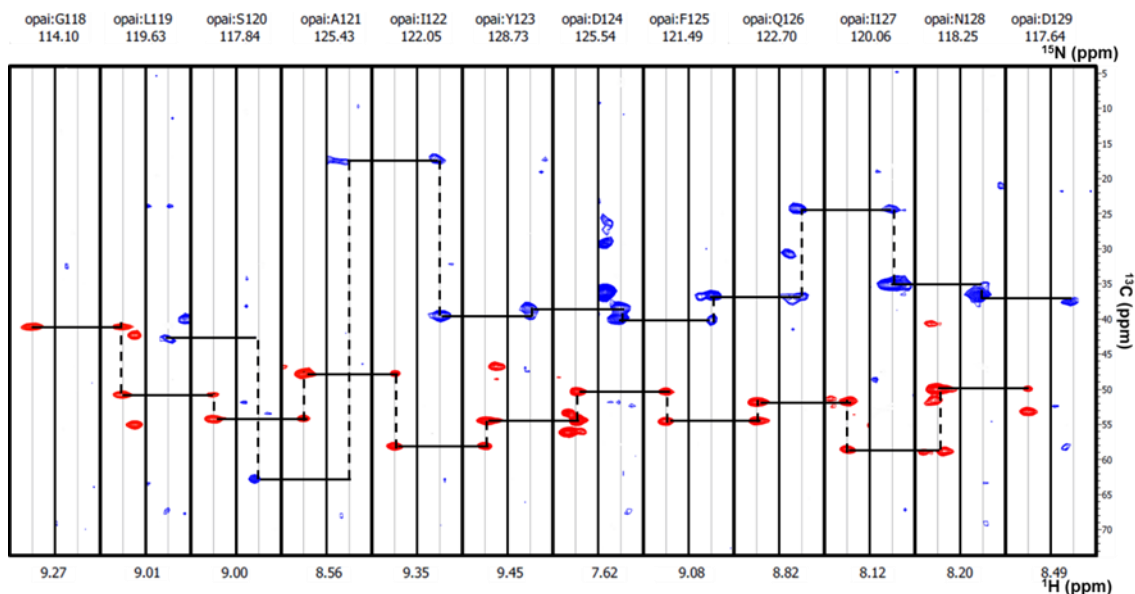
**Figure 4.8: Proposed trypsin-cleaved region of Opa<sub>60</sub>.** Amino acids that are proposed to be cleaved via trypsin digestion are highlighted in red.



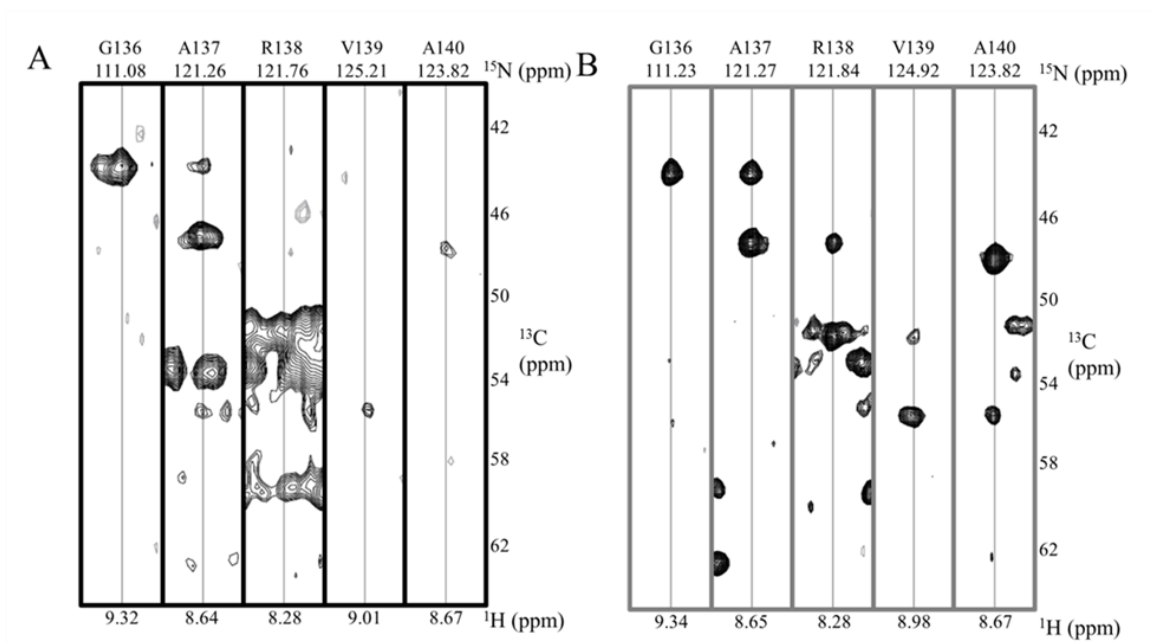
**Figure 4.9:**  $^{15}\text{N}$ ,  $^1\text{H}$  HSQC of trypsin-treated  $\text{Opa}_{60}$ . 147 resonances are observed in this spectrum, all of similar intensity. Many of the peaks observed overlap with peaks on the full length HSQC implying that very little perturbation of the barrel occurs.



**Figure 4.10: Overlay of trypsin-treated Opa<sub>60</sub> with trypsin-treated Opa<sub>50</sub>.** The amino acid sequence of the  $\beta$ -barrel region of both proteins has a very high sequence identity. Given the overlap of many of the chemical shifts of the barrel peaks, both Opa proteins appear to have very similar  $\beta$ -barrel structures.



**Figure 4.11: Select assigned strips for amino acids G118-D129 of trypsin digested  $\text{Opa}_{60}$ .** Strips of TROSY-HNCA (red) and HN(CA)CB (blue) are shown for each spin system. Horizontal lines are shown connecting  $\text{C}\alpha_i$  peaks to  $\text{C}\alpha_{i-1}$  peaks in the HNCA strips and  $\text{C}\beta_i$  peaks to  $\text{C}\beta_{i-1}$  peaks in the HN(CA)CB strips.



**Figure 4.12: Comparison of micellar interface amino acids of full-length (A) and trypsin-cleaved (B) *Opa*<sub>60</sub>.** Resonances of regions near the micellar interface of the full length *Opa*<sub>60</sub> spectra are typically less intense, or too broadened to observe in A (V139-A140) whereas these regions in trypsin-cleaved spectra are observed to exhibit narrower, more intense peaks.

## Section 4.4 Assignment of the extracellular loop region

### Section 4.4.1 Temperature effects on Opa<sub>60</sub> loops

All solution NMR  $\beta$ -barrel membrane protein structures determined to date have contained relatively short extracellular loops (the longest loop is that of OprH, containing 29 amino acids);(10) thus, spectral overlap and line broadening from loop resonances did not hinder the assignment of the majority of the proteins studied. However, Opa<sub>60</sub> has three extracellular loops longer than the longest of OprH, with two of them composed of over 40 amino acids. CD spectra indicate that these loops have no observed secondary structure to them (Figure 4.2). As indicated, many of the resonances observed in the full length spectrum exhibit significant line broadening at 40 °C due to conformational exchange processes.

By adjusting the temperature, in an attempt to adjust both the exchange rates,(29) and relative equilibrium (30) of the conformational states of the loops, the resulting NMR lineshapes of Opa<sub>60</sub> loop resonances can be greatly affected. As the temperature decreases, the overall tumbling of the protein-detergent complex slows, causing the  $\beta$ -barrel peaks (and likely many of the loop residues near the  $\beta$ -barrel) to broaden beyond detection. However, the loop regions most extended from the micelle can fluctuate more rapidly than the  $\beta$ -barrel-micelle complex and, therefore, may remain observable at lower temperatures (Figure 4.13). A  $^{15}\text{N}$ ,  $^1\text{H}$  HSQC spectrum recorded at 10 °C results in 34 distinct resonances. Recording TROSY-based HNCA, HN(CO)CA, HNCO and HN(CA)CO experiments, 27 of those resonances were assigned, corresponding

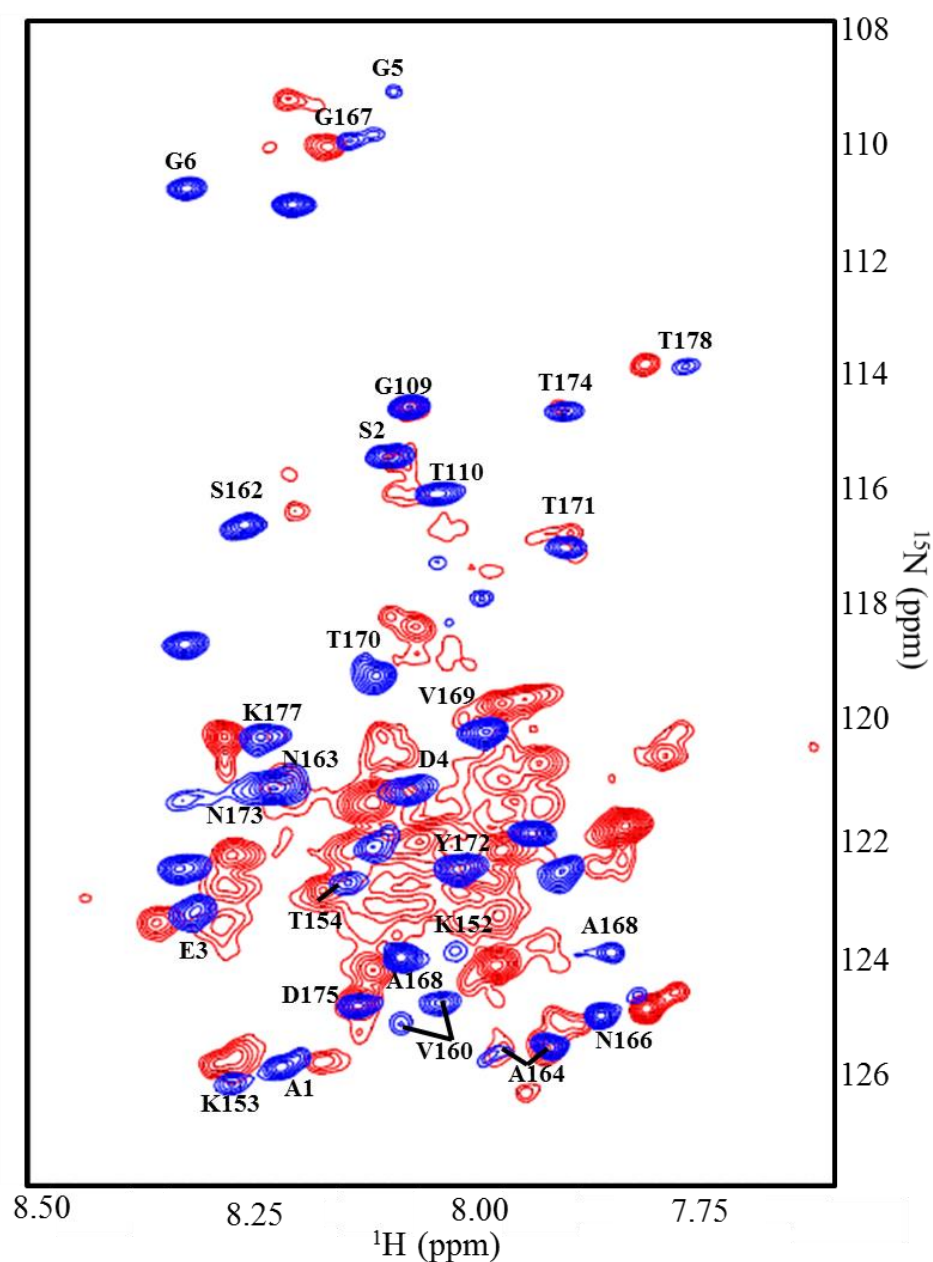
to the N-terminus (residues 1-6); loop 2 (residues 109-110); and loop 3 (residues 152-154, 160, 162-164, 166-175, and 177-178). These assignments correspond with the regions of the loops most extended from the barrel. A series of  $^{15}\text{N}$ ,  $^1\text{H}$ -HSQC spectra were recorded at 10, 20, 30, and 40°C (Figure 4.14) and the peaks were tracked through each spectrum in order to correlate the low temperature assignments to the assignment of the full length Opa<sub>60</sub> at 40°C.

#### **Section 4.4.2 Synthetic peptide analysis**

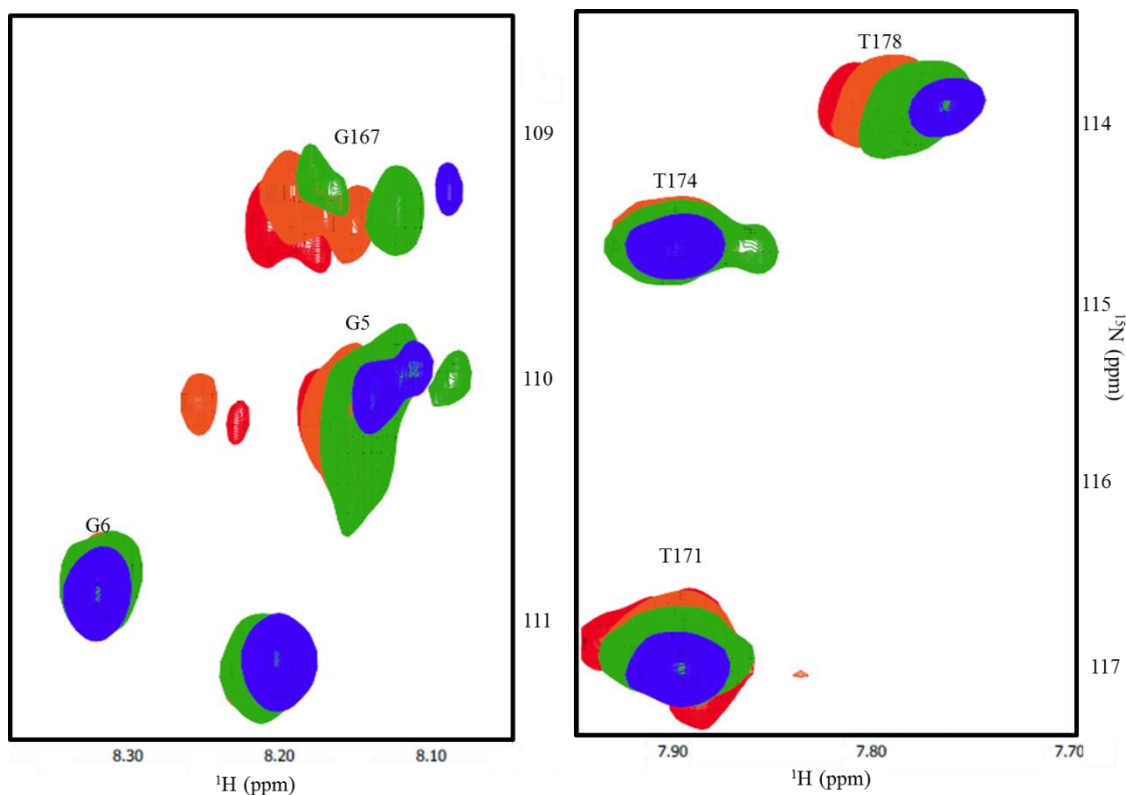
After observing that the most extended regions of the loops lacked secondary structure, unique individual peptides were considered for further structural studies.<sup>(31)</sup> The peptide sequences were chosen for biological significance, solubility, and amino acid composition. The regions of interest are in the two hypervariable regions located on loops two and three. Solubility proved to be more difficult than expected as there are many hydrophobic residues located throughout the loops. Finally, amino acid labeling specificity is difficult for certain residues that cellular machinery can easily convert to other amino acids, specifically Asx and Glx residues.<sup>(25)</sup> A twenty amino acid sequence located in the hypervariable region 2 (HV2), from residues 159-178 (Ac-TVPSNAPNGAVTTYNTDPKT-NH<sub>2</sub>) met all of these criteria, with a GRAVY value of -0.835, and only 4 Asx residues. The peptide was synthesized with selective  $^{15}\text{N}$  labeling in all threonine, valine, serine, glycine, and lysine residues. Near complete spectral overlap was observed in the  $^{15}\text{N}$ ,  $^1\text{H}$  HSQC between the peptide and the full-length spectrum at both 40 °C and 10 °C (Figure 4.15). Only twelve residues were  $^{15}\text{N}$  labeled, however through both heteronuclear and

homonuclear NMR experiments the assignment of the peptide resonances and, the full-length by comparison, could be assigned. The amino acid  $^1\text{H}$  and  $^{15}\text{N}$  spin systems were identified with  $^1\text{H}, ^1\text{H}$ -COSY,  $^1\text{H}, ^1\text{H}$ -TOCSY (Figure 4.16), and  $^{15}\text{N}$ -edited versions of each. Using the spin system identifications from the TOCSY spectrum of the peptide,(27) the amino acid type for HNCA strips was identified and, thus, sequentially assigned (Figure 4.17). The sequential assignment of the peptide could then be transferred to the full length 3D spectra (based on spectral overlap in the  $^{15}\text{N}$ ,  $^1\text{H}$ -HSQC spectra) and verified with the HNCA of the full length protein. Given the amount of spectral overlap, amino acid composition, and chemical shift degeneracy in this loop region, previous attempts to assign the resonances proved too difficult prior to the peptide comparison.

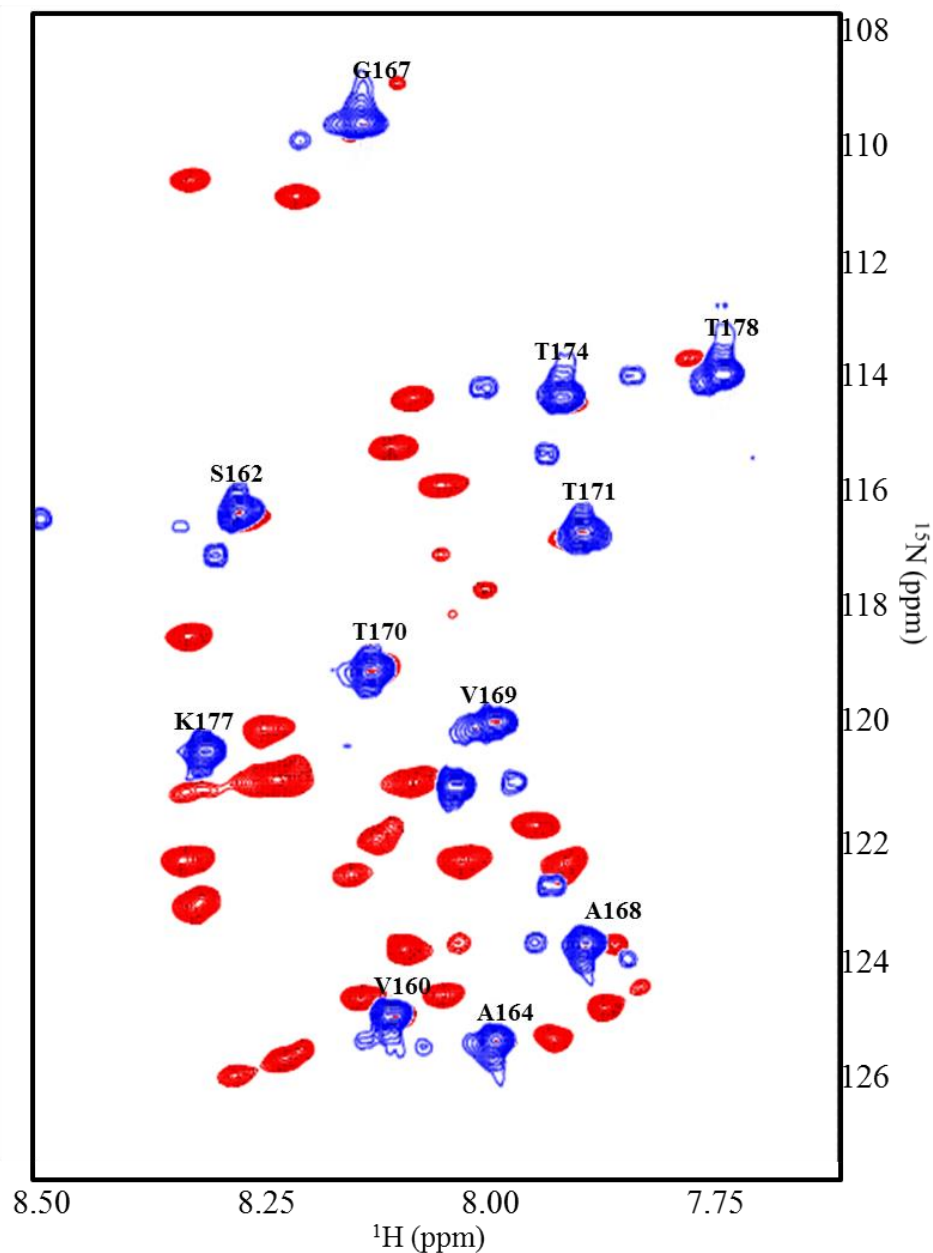




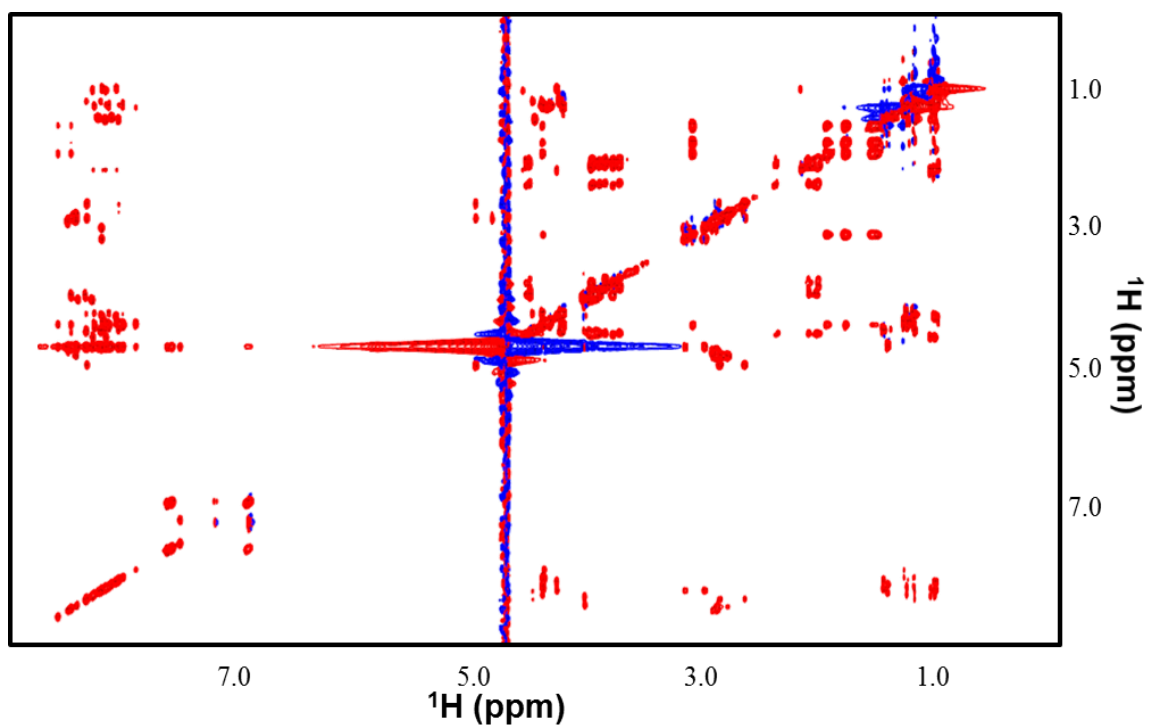
**Figure 4.13: Overlay of Opa<sub>60</sub> <sup>15</sup>N, <sup>1</sup>H-HSQC spectra recorded at 10 and 40 °C.** The cold spectrum (blue) broadens out many of the already broad peaks of the 40 °C spectrum (red) by decreasing the tumbling of the protein. The peaks that remain are from the most mobile regions of the protein.



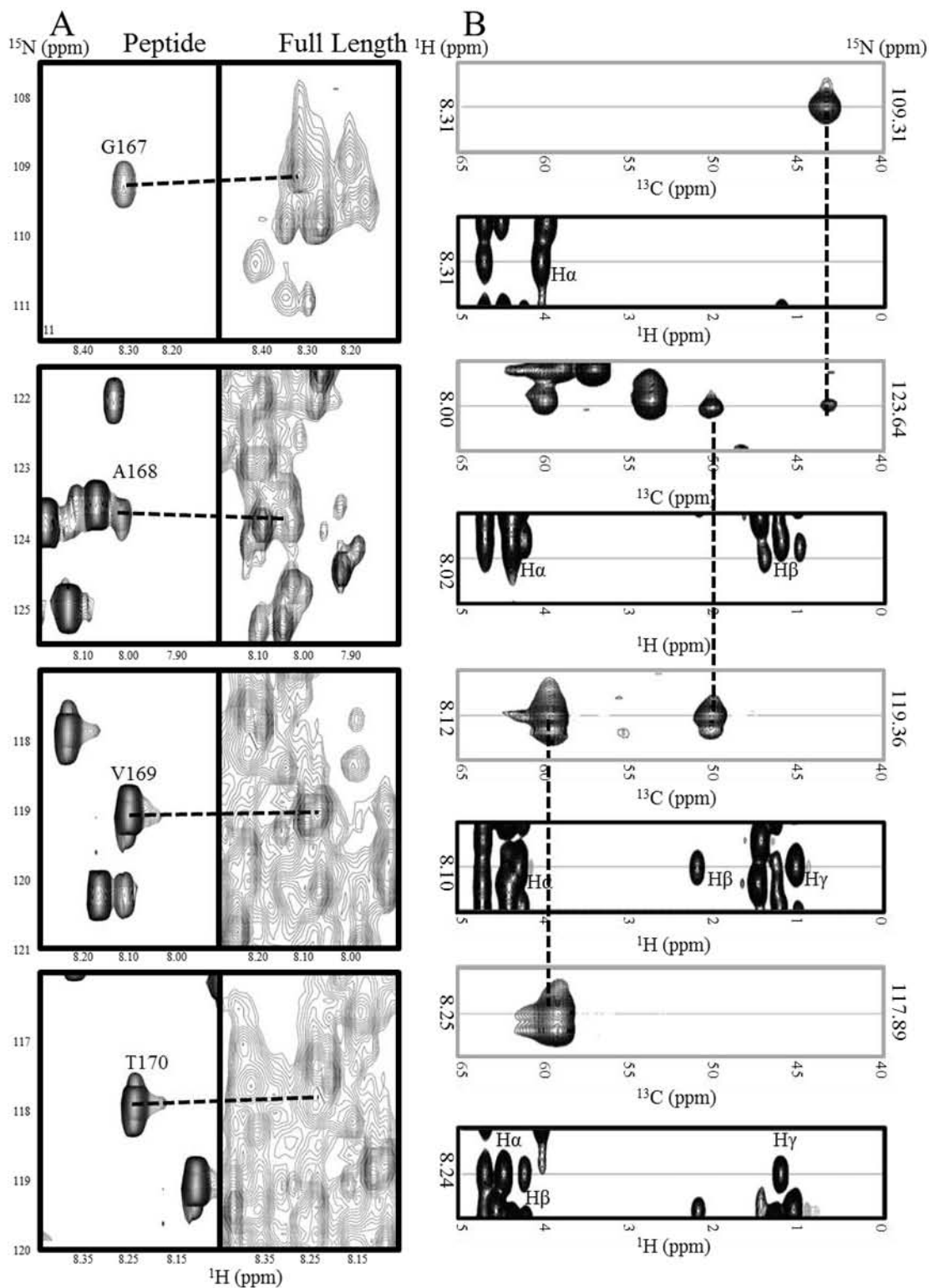
**Figure 4.14: Select regions of  $^{15}\text{N}$ ,  $^1\text{H}$ -HSQC spectra recorded over a series of temperatures.** The resonances that were assigned in the 10 °C spectra were able to be extrapolated over a series of temperatures (red – 40 °C, orange - 30 °C, green – 20 °C, blue – 10 °C).



**Figure 4.15: Overlay of HV2 peptide and Opa<sub>60</sub> <sup>15</sup>N, <sup>1</sup>H HSQC spectra recorded at 10 °C.** The overlap of chemical shifts of the assigned HV2 region of the full length Opa<sub>60</sub> spectrum with the HV2 peptide indicate that both regions are in similar chemical environments, allowing the study of the smaller peptide to be correlated to the full length protein.



**Figure 4.16:**  $^1\text{H}$ ,  $^1\text{H}$  TOSCY spectrum of the 20 amino acid peptide. Sidechain protons were able to be assigned via this spectrum helping in both assignment and structure.



**Figure 4.17: Backbone assignment of loop resonances using a synthetic peptide.** (A)  $^{15}\text{N}$ ,  $^1\text{H}$  HSQC spectra were recorded for both peptide and Opa<sub>60</sub>

(zoomed in regions of specific resonances are shown) and based on overlap (indicated by black dashed line) identical spins in the full length and peptide were identified. (B) The homonuclear 2D  $^{15}\text{N}$ -edited  $^1\text{H}, ^1\text{H}$  TOCSY strips (black strips) identified side chain spin systems and compared to the HNCA strips (gray strips) to obtain the sequential assignment. Connectivity between  $\text{C}_{\alpha, i}$  and  $\text{C}_{\alpha, i-1}$  resonances are shown with a dashed black line.

## Section 4.5 Summary

After refolding Opa<sub>60</sub> into dodecyl-phosphocholine micelles, both CD and NMR confirmed that the protein was homogeneously folded in a  $\beta$ -barrel conformation. Initial studies of the full length protein proved that traditional methods previously used to determine  $\beta$ -barrel structures via solution NMR would prove insufficient for this system. The primary difference between Opa<sub>60</sub> and other solved  $\beta$ -barrel structures is the three long mobile extracellular loops associated with Opa proteins, which result in multiple broadened peaks at conditions necessary to observe  $\beta$ -barrel resonances. Multiple methods were developed to circumvent this, including NMR studies following proteolysis of the loops from the barrel, cold temperature amino acid assignment strategies, and generating synthetic peptides of biological significance for assignment. Overall, 97% of the  $\beta$ -barrel and periplasmic loops were able to be assigned, while 24% of the loops residues were assigned. This total assignment proved to be sufficient to perform structural studies of the protein.

1. Kay, L. E. (2005) NMR studies of protein structure and dynamics, *J Magn Reson* 173, 193-207.
2. Sattler, M., Schleucher, J., and Griesinger, C. (1999) Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulsed field gradients, *Prog Nucl Mag Res Sp* 34, 93-158.
3. Ishima, R., and Torchia, D. A. (2000) Protein dynamics from NMR, *Nat Struct Biol* 7, 740-743.
4. Bruschweiler, R. (2003) New approaches to the dynamic interpretation and prediction of NMR relaxation data from proteins, *Curr Opin Struc Biol* 13, 175-183.
5. 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 Biol* 8, 334-338.
6. Fernandez, C., Adeishvili, K., and Wuthrich, K. (2001) Transverse relaxation-optimized NMR spectroscopy with the outer membrane protein OmpX in dihexanoyl phosphatidylcholine micelles, *P Natl Acad Sci USA* 98, 2358-2363.
7. Hwang, P. M., Choy, W. Y., Lo, E. I., Chen, L., Forman-Kay, J. D., Raetz, C. R. H., Prive, G. G., Bishop, R. E., and Kay, L. E. (2002) Solution structure and dynamics of the outer membrane enzyme PagP by NMR, *P Natl Acad Sci USA* 99, 13560-13565.



8. Liang, B. Y., and Tamm, L. K. (2007) Structure of outer membrane protein G by solution NMR spectroscopy, *P Natl Acad Sci USA* 104, 16140-16145.
9. 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.
10. 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*, *J Biol Chem* 286, 39211-39223.
11. Clore, G. M., and Gronenborn, A. M. (1998) NMR structure determination of proteins and protein complexes larger than 20 kDa, *Curr Opin Chem Biol* 2, 564-570.
12. Fiebig, K. M., Schwalbe, H., Buck, M., Smith, L. J., and Dobson, C. M. (1996) Toward a description of the conformations of denatured states of proteins. Comparison of a random coil model with NMR measurements, *J Phys Chem-Us* 100, 2661-2666.
13. 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.
14. Karlake, C., Piotto, M. E., Pak, Y. K., Weiner, H., and Gorenstein, D. G. (1990) 2d Nmr and Structural Model for a Mitochondrial Signal Peptide Bound to a Micelle, *Biochemistry-Us* 29, 9872-9878.

15. 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, *J Phys Chem B* 111, 12427-12438.
16. Patel, G. J., Behrens-Kneip, S., Holst, O., and Kleinschmidt, J. H. (2009) The Periplasmic Chaperone Skp Facilitates Targeting, Insertion, and Folding of OmpA into Lipid Membranes with a Negative Membrane Surface Potential, *Biochemistry-Us* 48, 10235-10245.
17. Kleinschmidt, J. H., and Tamm, L. K. (1996) Folding intermediates of a beta-barrel membrane protein. Kinetic evidence for a multi-step membrane insertion mechanism, *Biochemistry-Us* 35, 12993-13000.
18. Huysmans, G. H. M., Baldwin, S. A., Brockwell, D. J., and Radford, S. E. (2010) The transition state for folding of an outer membrane protein, *P Natl Acad Sci USA* 107, 4099-4104.
19. Kelly, S. M., Jess, T. J., and Price, N. C. (2005) How to study proteins by circular dichroism, *Bba-Proteins Proteom* 1751, 119-139.
21. Temussi, P. A., and Goodman, M. (1971) Conformational Aspects of Polypeptide Structure .39. Conformational Transition in Oligopeptides - Nmr Spectroscopic Study, *P Natl Acad Sci USA* 68, 1767-&.
22. Grzesiek, S., and Bax, A. (1992) An Efficient Experiment for Sequential Backbone Assignment of Medium-Sized Isotopically Enriched Proteins, *J Magn Reson* 99, 201-207.

23. Grzesiek, S., and Bax, A. (1992) Correlating Backbone Amide and Side-Chain Resonances in Larger Proteins by Multiple Relayed Triple Resonance Nmr, *J Am Chem Soc* 114, 6291-6293.
24. Tossavainen, H., and Permi, P. (2004) Optimized pathway selection in intraresidual triple-resonance experiments, *J Magn Reson* 170, 244-251.
25. Waugh, D. S. (1996) Genetic tools for selective labeling of proteins with alpha-N-15-amino acids, *J Biomol Nmr* 8, 184-192.
26. McIntosh, L. P., and Dahlquist, F. W. (1990) Biosynthetic Incorporation of N-15 and C-13 for Assignment and Interpretation of Nuclear-Magnetic-Resonance Spectra of Proteins, *Q Rev Biophys* 23, 1-38.
27. Biological Magnetic Resonance Data Bank.
28. Hauck, C. R., and Meyer, T. F. (2003) 'Small' talk: Opa proteins as mediators of Neisseria-host-cell communication, *Curr Opin Microbiol* 6, 43-49.
29. Jeener, J., Meier, B. H., Bachmann, P., and Ernst, R. R. (1979) Investigation of Exchange Processes by 2-Dimensional Nmr-Spectroscopy, *J Chem Phys* 71, 4546-4553.
30. Mittermaier, A., and Kay, L. E. (2006) Review - New tools provide new insights in NMR studies of protein dynamics, *Science* 312, 224-228.
31. Szymczyna, B. R., Gan, L., Johnson, J. E., and Williamson, J. R. (2007) Solution NMR studies of the maturation intermediates of a 13 MDa viral capsid, *J Am Chem Soc* 129, 7867-7876.



## Section 5 Structure calculations of Opa<sub>60</sub>

With the assignment of Opa<sub>60</sub>, structural restraints can be obtained. Both distance restraints via nuclear Overhauser effect (NOE) interactions (1) and bond angle restraints based on secondary chemical shifts(2) can be obtained and an ensemble of structures based on these restraints will converge on a high resolution ensemble of structures. The well-structured barrel region of Opa<sub>60</sub> results in a very precise structure, comparable to the resolution of the barrel region of previously studied  $\beta$ -barrel membrane proteins.(3-6) The extracellular loops lack both complete assignment due to conformational dynamics that inhibit high resolution of the NMR spectra and NOE distance and dihedral angle constraints are observed. To overcome a lack of specific NMR-based data for the loop regions, the NMR-derived structures were subject to molecular dynamics (MD) simulations(7) in the presence of a lipid bilayer (in collaboration with Peter Kasson and Per Larrson in the Department of Molecular Physiology and Biological Physics at the University of Virginia). The additional information gained from these studies suggest that the loops remain mostly disordered but form weak, transient interactions with each other, favoring a much smaller sampling space than the NMR-derived structures indicated.

## Section 5.1 NMR derived structural restraints

### 5.1.1 NOE-based restraints of Opa<sub>60</sub>

In order to obtain through space distance restraints as outlined in section 3.1.4, a three dimensional <sup>15</sup>N-edited, <sup>1</sup>H, <sup>1</sup>H NOESY spectrum was collected.(8) To reduce proton signal artifacts (9) of the relatively high concentration of the detergent (detergent monomer concentration over 200 times greater than the protein concentration), Opa<sub>60</sub> was refolded into 98% <sup>2</sup>H-labeled dodecylphosphocholine (FC-12) detergent micelles.(10) Refolding into deuterated micelles also had the added effect of reducing many of the dipole-dipole relaxation pathways that the amide protons on Opa<sub>60</sub> would experience with the micelle, thus optimizing NOE interactions amongst the intra-protein protons.(11) Because both the micelle and the protein were perdeuterated, a longer NOE mixing time of 180 ms was determined to be the most efficient time used to observe NOE interactions.

There were 123 NOE crosspeaks assigned via this experiment. All of which were assigned to  $\beta$ -barrel or periplasmic turn residues. The distinct lack of NOEs observed for the extracellular loops is not surprising. Given the dynamics associated with the assigned regions and the lack of secondary structural elements, significant NOE interactions were not observed during the NOE mixing time. Of the 123 NOE assignments, 62 were assigned to be sequential NOEs. There were 32 intense NOE cross peaks that provided inter-strand NOEs between  $\beta$ -strands (Figure 5.1). Due to the intensity of the peaks compared to

the other inter-strand NOEs, these interactions were determined to be involved in the hydrogen bonding network of  $\beta$ -barrels.(12) These conclusions were used to generate distance and angle restraints for the accompanying O...H-N backbone hydrogen bond interactions of residues between strands. (13)

In addition to the amide-amide NOE assignments, 31 aromatic ring–amide proton NOEs were assigned.(14) The majority of these aromatic residues were located in the aromatic belt on the periplasmic side of the protein (Figure 5.2).(15) These additional restraints were useful in calculating the structure of the periplasmic turns by adding distance restraints greater than one residue apart.

NOEs observed with the local solvent also aided greatly in both restraints and assignment.(16) NOEs were observed for the methyl, methylene, or phosphocholine headgroup protons of the FC-12 micelle as well as exchange peaks with water. Environmental NOE interactions were monitored through sequential  $\beta$ -strand assignments as the strands traverse the micelle; observing the general trend of phosphocoline headgroup protons near the edge of the barrel, methylene protons throughout the transmembrane region and methyl protons towards the center of the barrel, ensuring proper assignment (Figure 5.3).

### **Section 5.1.2 Dihedral angle restraints**

Dihedral angle restraints were obtained for Opa<sub>60</sub> through TALOS+.(17) The chemical shifts of all assigned residues were uploaded to the TALOS+ server (<http://spin.niddk.nih.gov/NMRPipe/talos/>) to be compared to other chemical

shifts of similar peptide triplets with known structures (section 3.1.3).<sup>(18)</sup> There were 128 dihedral angle constraints, 64  $\psi$  and 64  $\phi$  angles, that were reported as reliable dihedral angles based on comparable chemical shifts. As expected, the resonances with amide protons downfield shifted had  $\phi$ ,  $\psi$  dihedral angles around  $135^\circ$ ,  $-135^\circ$ , respectively, consistent with  $\beta$ -strand structure. For the assigned residues in the extracellular loops dihedral angles typically had angle errors greater than  $\pm 40^\circ$ ; therefore, there were minimal dihedral restraints in the loops.

### **Section 5.1.3 Membrane interface restraints**

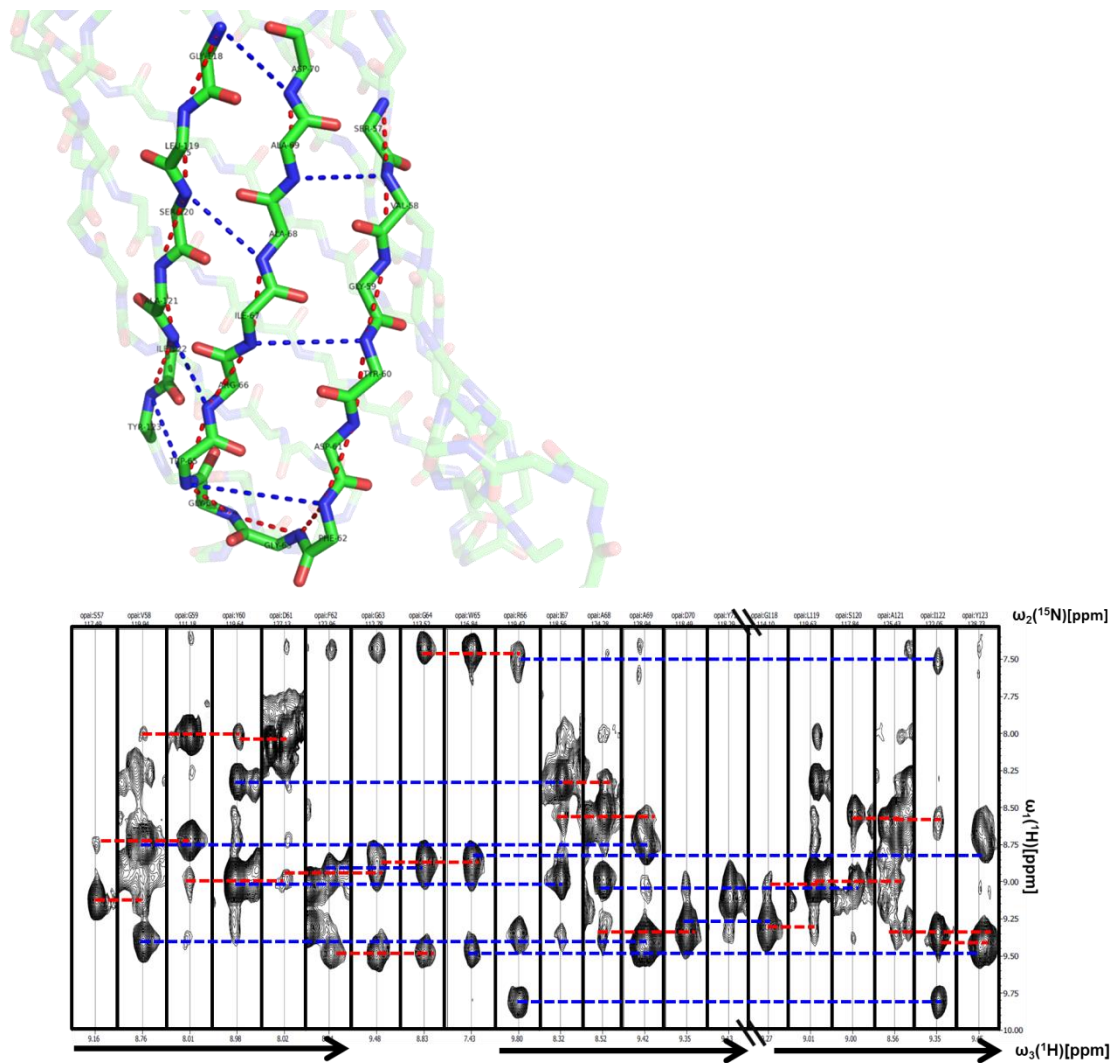
Combining the NOE distance restraints, the hydrogen bonding restraints and the dihedral angle restraints, initial structure calculations were performed using Xplor-NIH;<sup>(19, 20)</sup> a program commonly used to determine biomolecule structures with NMR. After inputting the protein sequence and the distance and angle restraints, the program performs stepwise cooling from a predefined high temperature (4000 K was chosen for Opa<sub>60</sub>). Energetic penalties are assigned to deviations from any of the calculated restraints, as well as predefined bond lengths, bond angles, van der Waals spheres and rigidly defined regions, such as aromatic side chains.

The first round of structure calculations exhibited a unique problem with Opa<sub>60</sub> that was previously not observed for other  $\beta$ -barrel proteins. The extracellular loops of Opa<sub>60</sub> are sampling space that would otherwise be occupied by the micelle or membrane. A lack of both observed NOE distance and dihedral

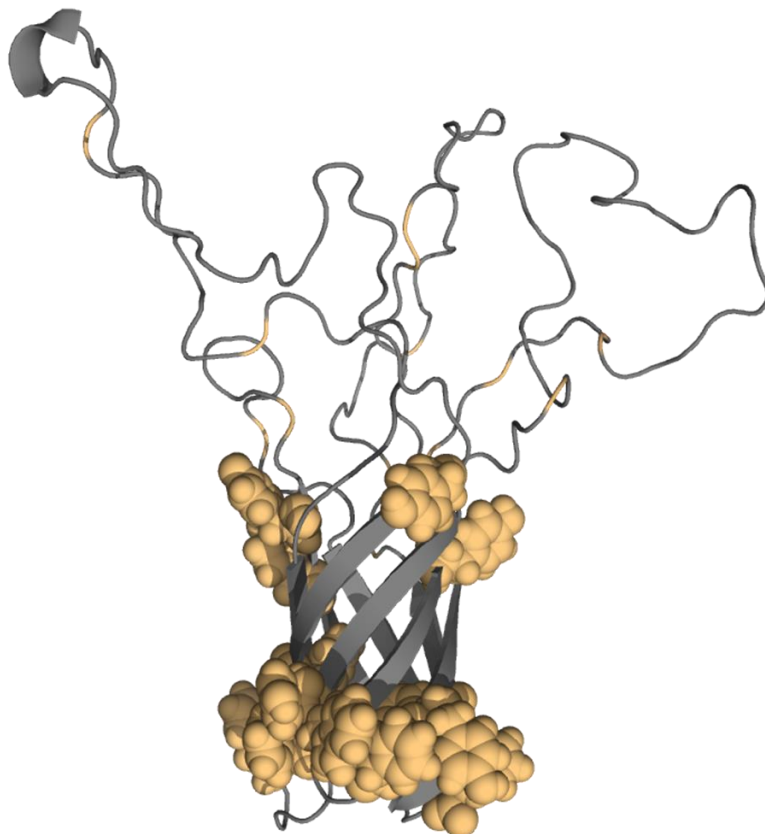


restraints permit these loops to extend into the regions occupied by the micelle which environmental NOEs indicate is unlikely (Figure 5.4). This problem is unique to Opa<sub>60</sub> compared to other  $\beta$ -barrels as a result of the length of the extracellular loops. The other  $\beta$ -barrel structures studied by NMR lacked loops long enough to significantly sample space located in the membrane plane.

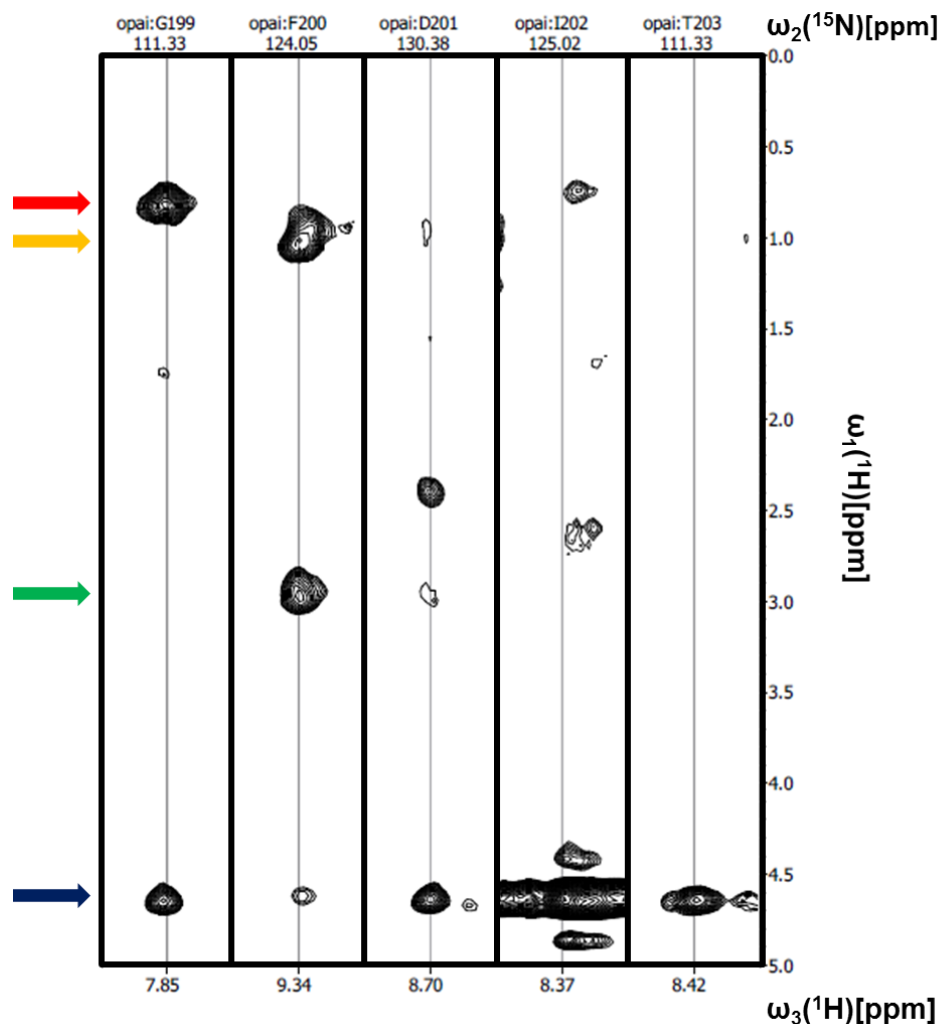
To overcome the spatial sampling violation of the loops, an additional set of planar restraints was added to the structure calculations to prevent the extracellular loops from sampling space below the micellar interface on the extracellular side of the barrel.(21) The first of these planes was determined by the residues that demonstrated NOE crosspeaks with the headgroups of the micelles on the periplasmic side of the protein. The second plane was determined to be 32 Å, apart from the first plane, with an error free penalty allowing residues on either side of the planes to be within 25 Å. This data was determined by both environmental NOEs and the defined hydrophobic thickness of a previously characterized *Neisseria meningitidis* outer membrane protein, PorB. (9)



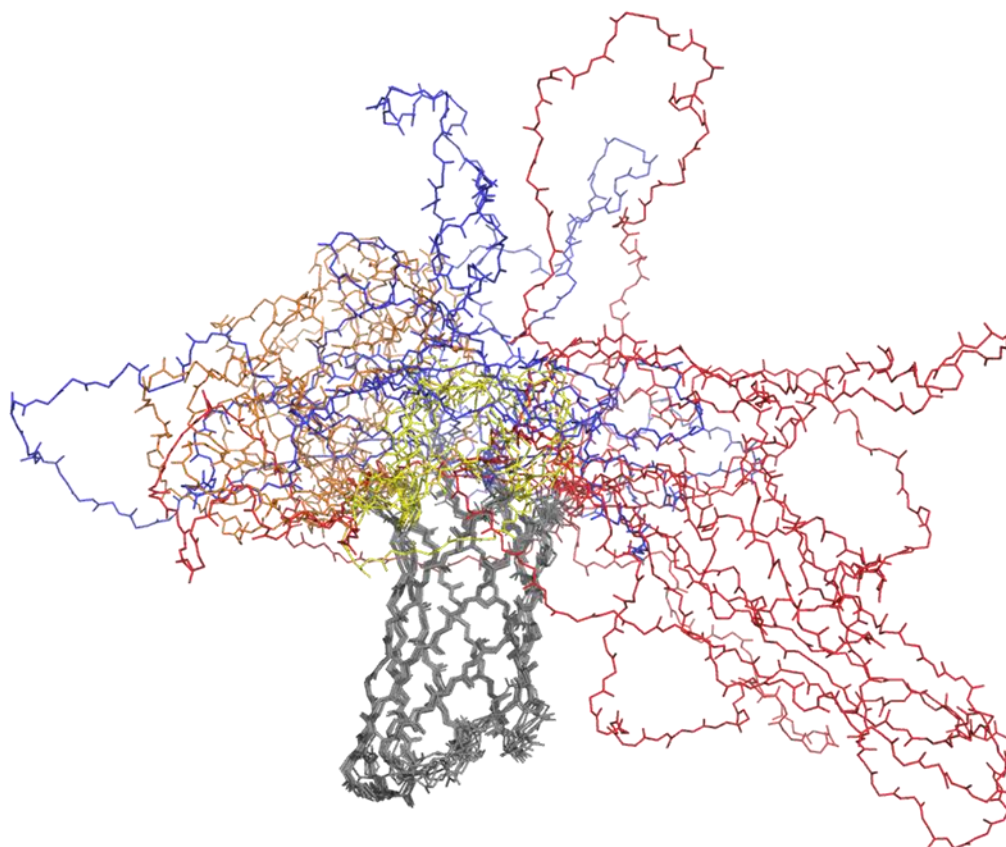
**Figure 5.1:** Select strips from  $^{15}\text{N}$ -edited NOESY spectrum of Opa<sub>60</sub>. Red Dashed lines represent intra-strand NOE interactions and blue strands representing inter-strand NOEs for segments of  $\beta$ -strands 3, 4, and 5.



**Figure 5.2: Aromatic residues of Opa<sub>60</sub> at micellar interface.** Aromatic residues colored tan as van der Walls spheres.



**Figure 5.3: Environment NOE and water exchange peaks.** Select  $^{15}\text{N}$ -NOESY strips for a region of strand 6 extending into periplasmic turn 3 are shown, highlighting the environmental NOEs. Methyl and methylene crosspeaks of the detergent are observed in the first two strips, residues G199 and F200, at approximately 1 ppm (red and orange arrows, respectively). Crosspeaks with the phosphocholine headgroup are present in strands two and three at approximately 3 ppm (green arrow) and residues exposed to aqueous solution observe intense water exchange peaks at approximately 4.7 ppm (blue arrow).



**Figure 5.4: Ensemble of structures generated in the absence of bilayer planar restraints.** The extracellular loops (color coded as loop 1 – orange, loop 2 – blue, loop 3 – red, and loop 4 – yellow) sample space that is physically unrealistic as these extracellular regions would dip into the hydrophobic membrane region.

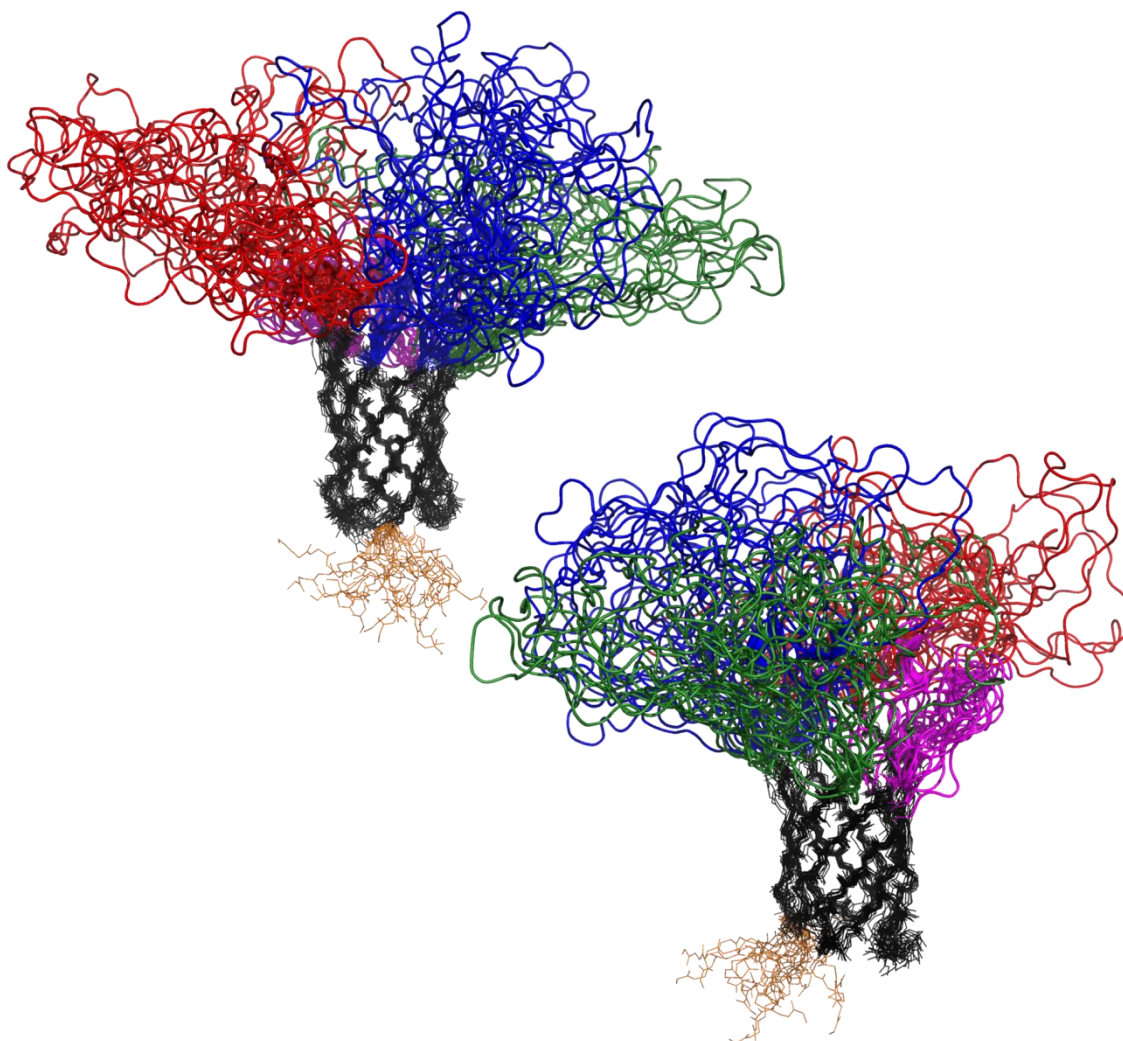
## Section 5.2 NMR-determined structure of Opa<sub>60</sub>

Using, the dihedral, membrane, hydrogen bond, and NOE restraints, four hundred structures were generated using NIH Xplor. (19, 20) The twenty lowest energy structures were collected and analyzed (Figure 5.5) (Table 5.1). The barrel region of these structures converged very well, with a protein backbone rmsd of 0.96 Å. The barrel length displays asymmetry as strands 5 and 6 extend for 11 residues whereas strands 1 and 2 extended 8 residues. Asymmetric  $\beta$ -strand lengths are reported for several  $\beta$ -barrel structures.(22) The axial ratio of the barrel is 1.25, comparable to that of all previously studied eight-stranded  $\beta$ -barrels with the exception of OmpX. (23)

The most consistent violations throughout the ensemble of structures during the calculations were dihedral angles of residues associated in the periplasmic loops. The calculated structure indicates a relatively short turn in periplasmic turn 1, comprised of an F-G-G-W sequence. The aromatic sidechain of W64 shares an NOE with D124 on the neighboring periplasmic turn, indicating that the sidechain is pointing out towards the next periplasmic turn, whereas the aromatic ring of F62 forms NOEs with the amide protons of both glycines in the short turn. The flexible glycines facilitate this short turn to remain stable despite its unusual conformation.(24)

The extracellular loops observed via these structure calculations are very disordered with no structural convergence (Figure 5.5). The dynamics studies and lack of NOE crosspeaks indicate that these regions are both mobile and lack

any observable interactions. However, due to the lack of a complete assignment, conclusions regarding the loop positions and sampled space are unclear from the NMR data alone.



**Figure 5.5: Ensemble of 20 lowest energy structures of Opa<sub>60</sub> derived from NMR restraints.** The barrel residues are displayed in black, extracellular loop 1 – green, loop 2 – blue, loop 3 – red, loop 4 magenta. The backbone rmsd of the barrel was reported to be 0.96 Å. The loops remain largely disordered and sample all available space.



## NMR distance and dihedral angle constraints

### Structure Calculation

Unique HN-HN NOE	123
Sequential	62
Medium Range	11
Long Range	50
Hydrogen Bond Constraints	120
Dihedral Angle Constraints	128

### NMR constraint violations

NOE (Å)	0.002
Dihedral Angle	0.106
NIH-XPLOR energy (kcal mol <sup>-1</sup> )	-830.11 ± 76.74

### Ramachandran map analysis

Favored Region (%)	78.3
Allowed Region (%)	11.9
Outlier Region (%)	9.8

### Ensemble RMSD

#### Mean global backbone rmsd (Å)

β-Sheet residues	0.84 ± 0.12
β-Sheet and turn residues	0.96 ± 0.11
All residues	9.49 ± 1.73

#### Mean global heavy atom rmsd (Å)

β-Sheet residues	1.99 ± 0.20
β-Sheet and turn residues	2.13 ± 0.18
All residues	10.06 ± 1.58

**Table 5.1: Opa<sub>60</sub> NMR structure statistics:** Constraints, violations and ensemble RMSD values from the calculated from Opa<sub>60</sub> NMR ensemble.

### **Section 5.3 Molecular dynamics refinement**

To better understand the extracellular loops of Opa<sub>60</sub>, which lack distance and dihedral restraints, the structures were refined using molecular dynamics (MD) simulation. Other studies have used MD to refine structures, but these studies have been more interested in smaller reorientations of the transmembrane regions of proteins in different membrane mimics and not larger scale extracellular regions.(25-28) These simulations were performed by Per Larsson and Peter Kasson (University of Virginia Membrane Biology and Department of Molecular Physiology and Biological Physics).

#### **Section 5.3.1 Molecular dynamics experiment**

Each of the twenty lowest energy structures solved from the NMR studies were simulated independently under identical conditions. A 512 dimyristoylphosphatidylcholine (DMPC) lipid patch was generated(29) and the protein was embedded into the bilayer, removing approximately 15-20 lipids in the process.(30) Approximately 40,000 TIP3p waters were included with NaCl ions added to the biologically relevant concentration of 150 mM (Figure 5.6). The simulations were performed using Gromacs 4.5 (31) with the Charmm36 forcefield, (32). After an initial energy minimization step, each of the twenty simulations were carried out over 100 ns using a time step of 2 fs. Structural snapshots were recorded every 500 ps, resulting in 200 structures recorded over 100 ns.

The structures were clustered using the k-means algorithm. The 400 structures acquired between 50-70 ns of the experiment were compared to the 400 structures acquired between 70-90 ns to ensure sufficient sampling. The two different sets of data were divided into 50 different clusters based on rmsd of the structures. No substantial drift between the two data sets was observed as the relative occupancy of each of the clusters was comparable amongst the two groups. The 4000 total structures generated from the twenty simulations were clustered into 100 clusters. One structure from each of the twenty most populated clusters that best satisfied the initial NMR restraints was selected and combined to represent the MD refined ensemble. (Figure 5.7).

### **Section 5.3.2 Loop Assessment**

#### *Section 5.3.2.a Loop position*

One of the first structural changes observed during the simulation was the positioning of the loops. In NMR restraint-exclusive structures, (Figure 5.5) the loops extended outwardly parallel the plane of the membrane. As the MD structures indicate (Figure 5.6) these loops contract, occupying more space above the barrel and less extended out. This reorientation occurs quite early in the simulation as assessed by the changes in the radius of gyration ( $R_g$ ) (Figure 5.8). There is a sharp trend early in the experiment, typically within the first 20 ns, of a decreased  $R_g$  of each of the structures. The loops, many which extend far from the barrel in the NMR-derived structures, quickly congregate to the region above the barrel.

Although these loops sample less space than the NMR structures indicated, they do not resolve into to any specific interloop conformation. The simulations indicate that these loops appear to have weak non-specific transient interactions with each other. The biologically significant regions of Opa<sub>60</sub>, the hypervariable loops on average are observed to be in much closer in proximity to each other after the MD refinement (Figure 5.6). Given that both hypervariable regions are necessary to efficiently bind to host CEACAM receptors, the improved spatial association between the two HV regions aids in the likelihood of cooperative interactions between the two regions (Figure 5.9).

#### *Section 5.3.2.b Loop secondary structure*

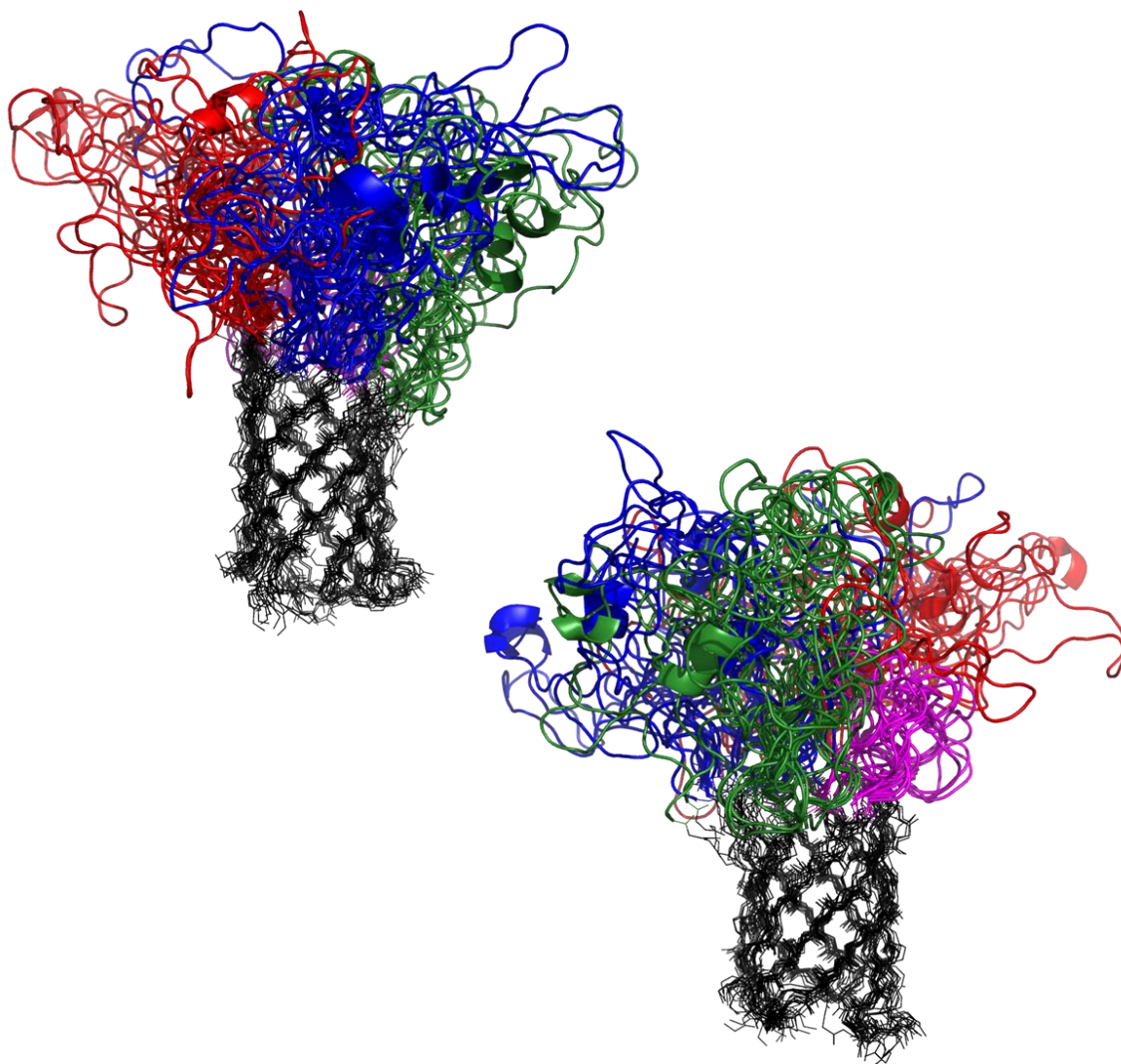
Although the experimental data indicate the loops are random coil, two regions of the loops exhibited transient helix formation over the course of the simulation. Both of these helical regions are located in the two hypervariable (HV) regions of the protein (Figure 5.10).(33) There is a portion from E87 to I97 in HV1 and three small regions in HV2 that exhibit helical character; T154-T158, N166-T174 and T178-Q183. Each of the three regions in HV2 is separated by proline residues at position 161, 164, and 176.

#### **Section 5.3.3 Barrel Assessment**

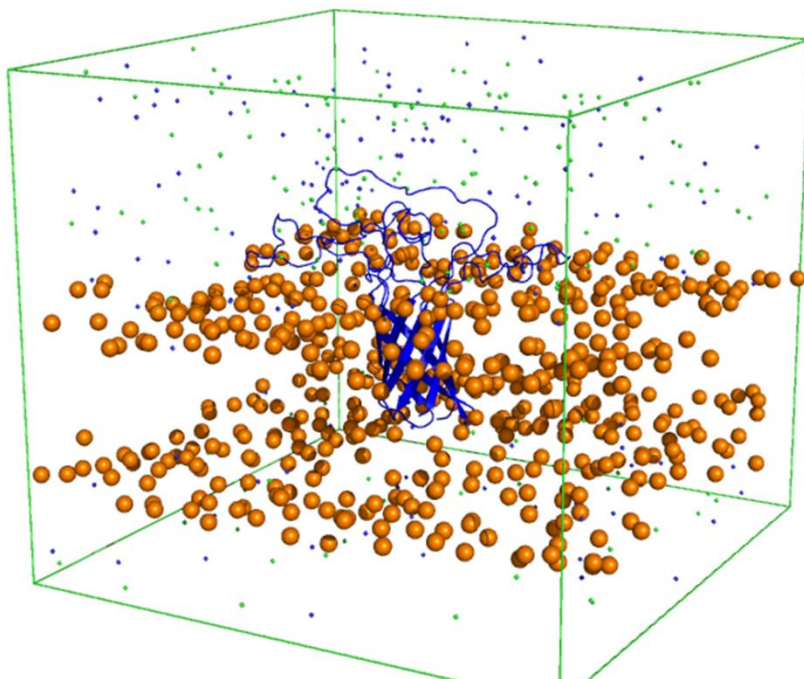
Despite the considerable reorientation that the loops undergo during the simulation, the backbone of the NMR-restraint based barrel structure remains mostly unperturbed throughout the experiment. Since all of the NMR experiments that observed the  $\beta$ -barrel were performed with deuterated Opa<sub>60</sub>, none of the

sidechain protons were used in the structure calculation.(34) As a result, the sidechains of these residues were treated generically, applying default bond angles and lengths, while randomly orienting based almost exclusively on sterics and directionality based on backbone orientation. Electrostatics can strongly favor the interactions of the sidechains, especially the interior of  $\beta$ -barrels where a large collection of charged and ionic sidechains are sequestered. (22) There are five salt bridges that are formed in the barrel (Figure 5.11). This additional tight salt bridge network helps to stabilize the  $\beta$ -barrel (unfolding conditions have yet to be found).

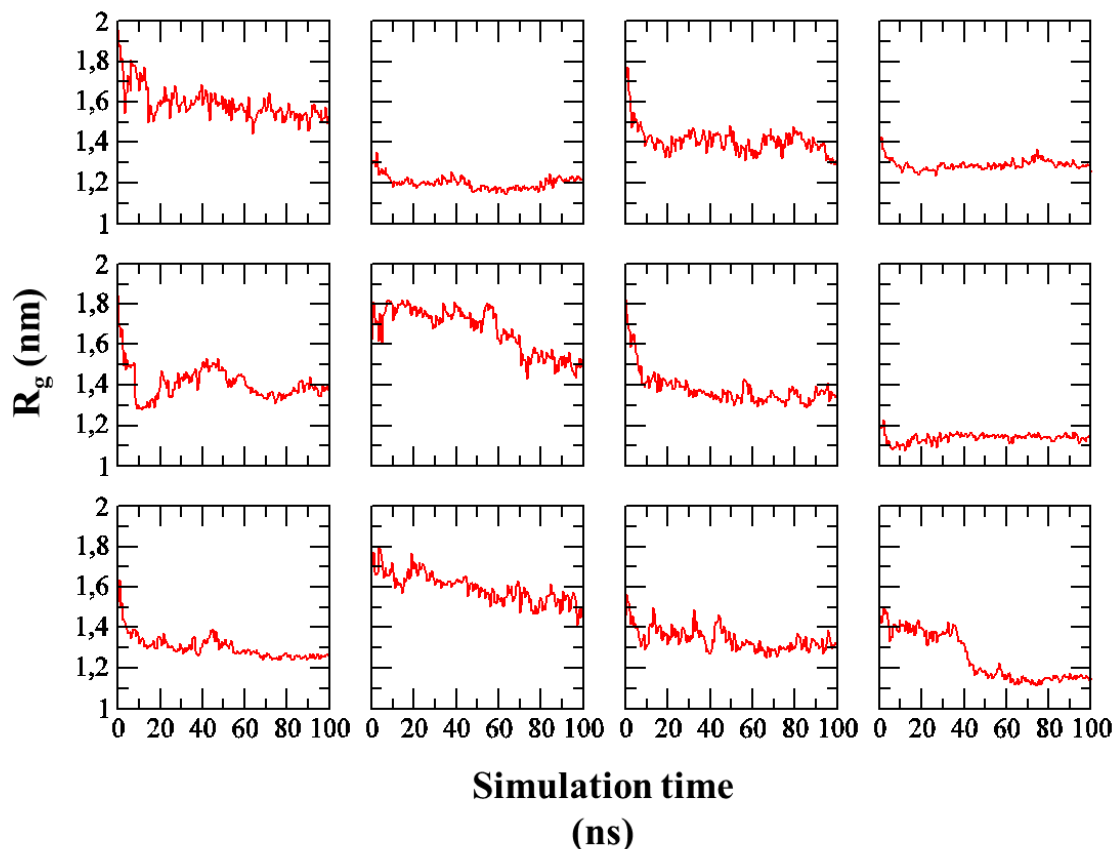
In addition to sidechain reorientation within the barrel, three strands, strands 2, 3, and 4, extend by one residue each in the extracellular dimension with Arg55, Arg72 and Leu 117 exhibiting  $\beta$ -strand character. Arginine residues are shown to form a stable conformation “snorkeling” across the membrane interface. (Section 1.2.1).(35) These residues were not assigned via NMR and were likely in a conformational exchange such that their resonances were not visible in the NMR spectra. These  $\beta$ -barrel characteristics are in agreement with previously studied proteins, with more salt bridges observed in Opa<sub>60</sub> than in the previously characterized eight-stranded  $\beta$ -barrels and are in good support of a valid high resolution structure of Opa<sub>60</sub>.



**Figure 5.6: Ensemble of Opa<sub>60</sub> structures after MD simulation:** The barrel residues are colored black; extracellular loop 1 – green, loop 2 – blue, loop 3 – red, and loop 4 magenta. The structure of the barrel remains relatively unperturbed when compared to Figure 5.5, however the extracellular loops occupy far less volume on average after the molecular dynamics simulation.



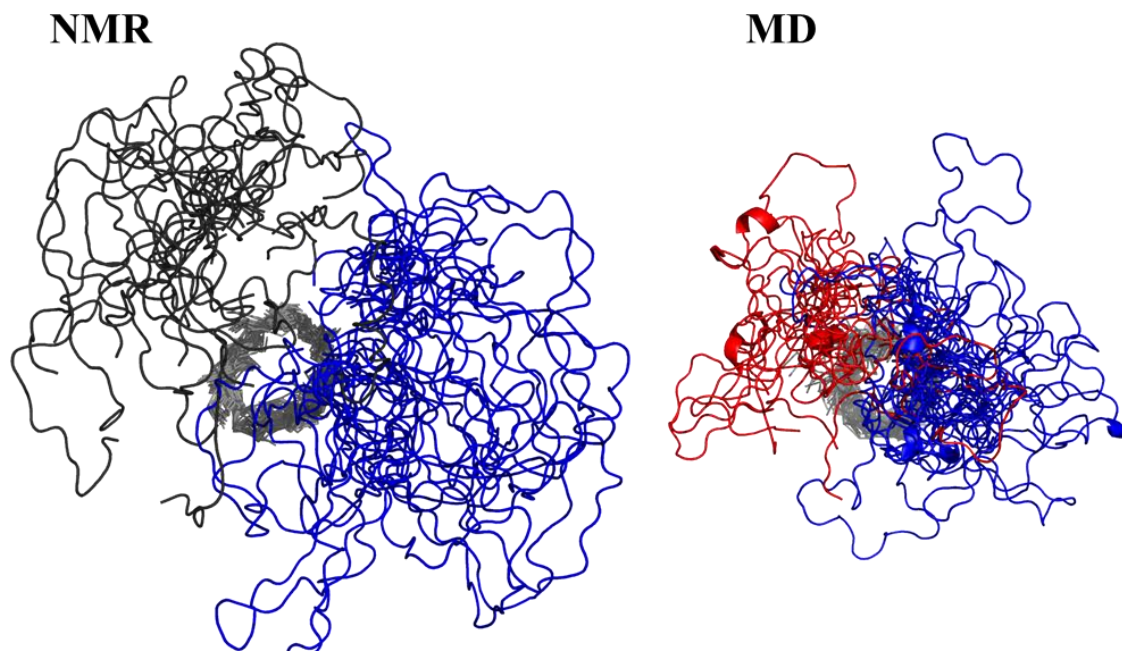
**Figure 5.7: Experimental set up for MD simulations.** Each of the 20 lowest NMR restraint derived structures were placed into a DMPC bilayer (headgroups shown in orange, tails excluded from image for clarity). Each experiment is comprised approximately 180,000 atoms in a box 11.00 nm tall and 12.46 nm wide.



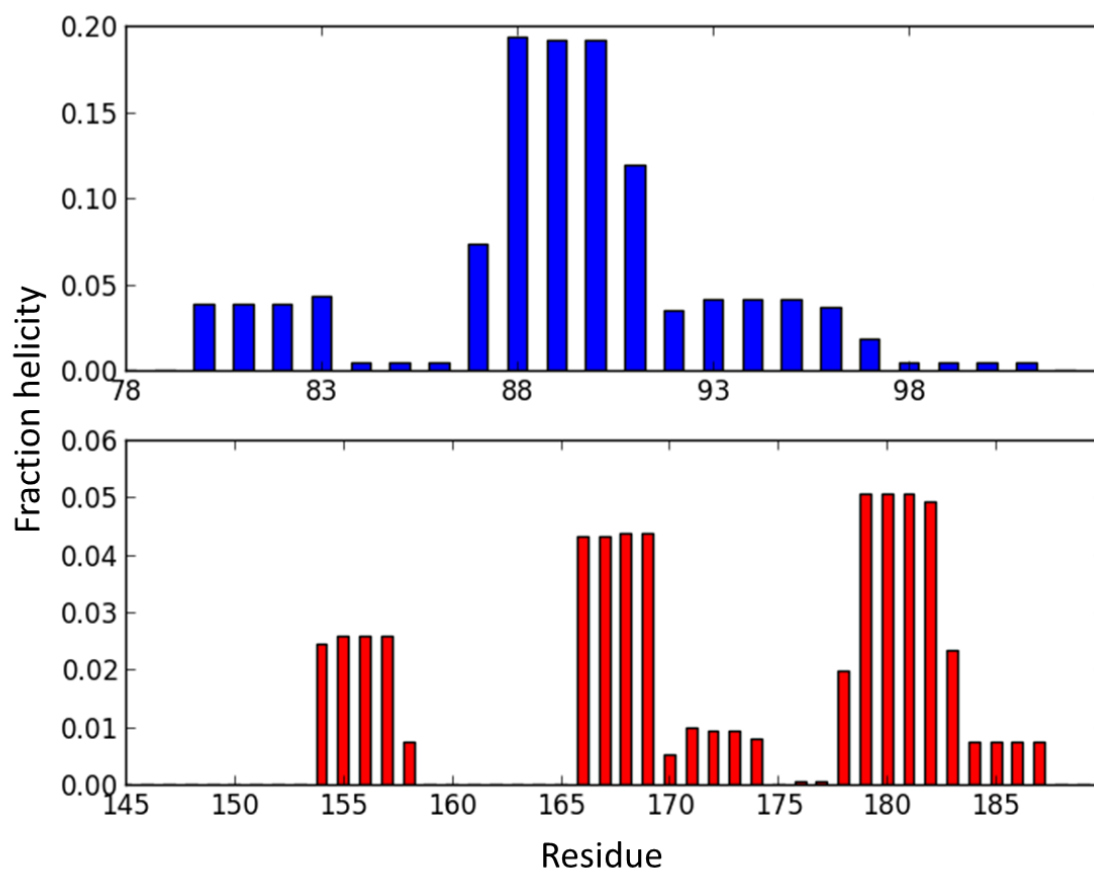
**Figure 5.8 Radius of gyration ( $R_g$ ) of Opa<sub>60</sub> structures during MD simulation.**

The  $R_g$  of 12 different Opa<sub>60</sub> structures throughout the course of the MD simulations. The  $R_g$  is shown to decrease throughout the experiment; the largest component for this change coming from the loops forming a more compact structure above the barrel (Figure 5.6).

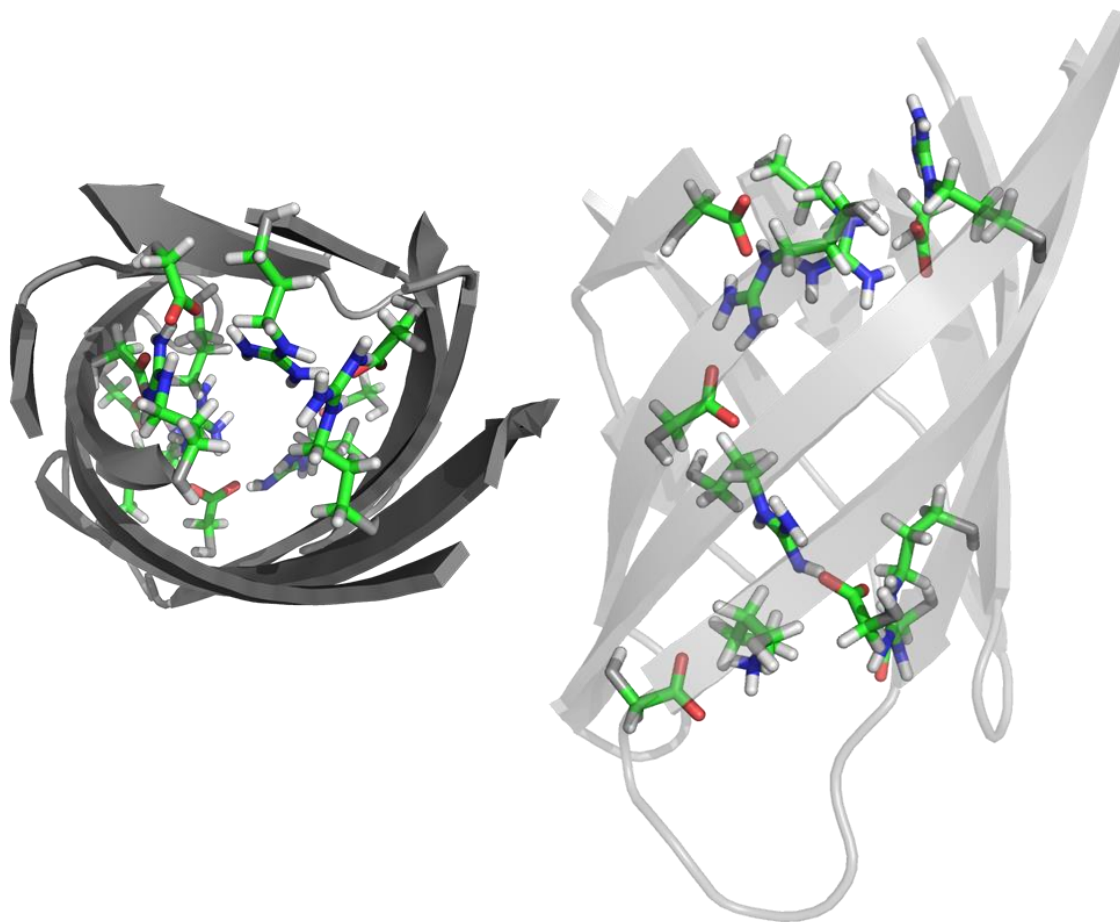




**Figure 5.9: Comparison of hypervariable regions of Opa<sub>60</sub> before and after MD simulation:** HV1 region of loop 2(red) and HV2 region of Loop 2 (blue). The two regions that are integral for receptor binding are considerably closer to each other after MD simulations and sample less volume.



**Figure 5.10: Helical content of HV regions.** Residues in HV1 (blue) and HV2 (red) have helical character during MD simulations. The three helical regions in HV2 are separated by proline residues.



**Figure 5.11: Ionic network in barrel.** Five stable salt bridges are observed within the barrel of Opa<sub>60</sub>. The interactions of these sidechains contribute to the stability of the  $\beta$ -barrel fold.

## 5.4 Functional significance of structure

With the refined structures of Opa<sub>60</sub>, specifically in the loop regions, obtained from these MD simulations, some hypotheses about the binding mechanism of can be inferred. A popular binding kinetics theory for unstructured proteins is the so called “fly casting mechanism.”(2) This mechanism states that in some instances, an unstructured protein can have a greater capture radius of its binding partner than that of a folded state. This initial association of random coiled protein to the binding partner that will result in a secondary structure formation. Previous studies of the phosphorylated kinase inducible activation domain (pKID) of the transcription factor CREB results in an initially unstructured sequence that stabilizes interactions with its binding partner.(1) This interaction is initially stabilized by non-specific hydrophobic contacts before a transient helix of pKID is stabilized in the final bound state.

The hypervariable regions of Opa<sub>60</sub> exhibit several elements similar to this mechanism. These regions spend the vast majority of the time throughout the simulation in random coil, unstructured conformations. Although the fly casting mechanism indicates that the more space sampled by these binding regions, the greater the chance of association with the binding partner, the MD simulations actually show a vastly diminished sampling space compared to the original NMR ensemble. The smaller sampling space is believed to be useful for the two hypervariable domains to remain in close proximity to each other. Because both hypervariable regions are necessary for Opa:CEACAM binding, the fly casting mechanism for Opa<sub>60</sub> requires two steps: loop-loop interactions followed by a

hypervariable 1 and 2 complex interacting with CEACAM. Based on the MD simulations, in the absence of any binding partners, the two hypervariable regions loosely associate with each other, primarily through hydrophobic contacts which are prevalent in both extracellular regions. A sampling from different k-means clustered structures show hydrophobic contacts between V84-I148, I87-V157, I91-I155, and I97-V160.

These hydrophobic residues also coincide with the region in HV1 that shows the most helical content. Interestingly, when these HV1 hydrophobic residues are not making contacts with HV2, they are primarily making contacts with each other, specifically multiple structures with contacts observed between V84-I86, I86-V89, I86-I91, I91-I99 and I97-I99. These hydrophobic contacts may lead to the nascent helices observed through the simulation. When there are hydrophobic contacts observed between the two different loops, helical content is not observed for the structures sampled from these simulations. Whereas helical formation may be necessary to stabilize Opa<sub>60</sub> binding to CEACAM, as observed in pKID,(37) the destabilization of the helices help to stabilize loop-loop interactions. This equilibrium between inter- and intraloop interactions may be one of the main causes of the exchange broadening observed in the random coil region of the NMR spectra.

Given the hydrophobic surface of the binding face of CEACAM1,(7) these hydrophobic residues may also be integral in CEACAM binding interactions. In the simulations without CEACAM, helical content in the hypervariable regions of Opa<sub>60</sub> never exceeded 20% for any residue whereas in the final ensemble of

structures, hydrophobic interactions between the hypervariable regions were observed in approximately 50% of the structures, with no specific interaction observed in more than 15% of the structures. The hydrophobic residues spend the majority of their time either forming intraloop interactions with neighboring hydrophobic sidechains or exposed to the solution. Favorable interactions with the hydrophobic CEACAM face are likely to shift the equilibrium of the intra- or interloop interactions into a more stable intraloop interaction that could stabilize the transient helices observed in the simulations, thus potentially generating a stable helical binding domain, as observed in pKID binding studies.

The high content of hydrophobic residues in the hypervariable regions of Opa<sub>60</sub> is uncommon for water-exposed regions of proteins that lack prevalent secondary structure. The interactions of these residues appear to drive the loose association hypervariable loops 1 and 2. By first relegating the two loops spatially near each other, a possible cooperative interaction between the two loops that may be necessary to bind to CEACAM is more likely to occur. The hydrophobic face of CEACAM can then provide a platform to stabilize the transient helices observed in the loops necessary to bind to CEACAM.

1. Sugase, K., Dyson, H. J., and Wright, P. E. (2007) Mechanism of coupled folding and binding of an intrinsically disordered protein, *Nature* 447, 1021-U1011.
2. Shoemaker, B. A., Portman, J. J., and Wolynes, P. G. (2000) Speeding molecular recognition by using the folding funnel: The fly-casting mechanism, *P Natl Acad Sci USA* 97, 8868-+.
3. 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*, *J Biol Chem* 286, 39211-39223.
4. Fernandez, C., Hilty, C., Wider, G., Guntert, P., and Wuthrich, K. (2004) NMR structure of the integral membrane protein OmpX, *J Mol Biol* 336, 1211-1221.
5. Hwang, P. M., Choy, W. Y., Lo, E. I., Chen, L., Forman-Kay, J. D., Raetz, C. R. H., Prive, G. G., Bishop, R. E., and Kay, L. E. (2002) Solution structure and dynamics of the outer membrane enzyme PagP by NMR, *P Natl Acad Sci USA* 99, 13560-13565.
6. 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 Biol* 8, 334-338.
7. Billker, O., Popp, A., Gray-Owen, S. D., and Meyer, T. F. (2000) The structural basis of CEACAM-receptor targeting by neisserial Opa proteins, *Trends Microbiol* 8, 258-260.

8. Williamson, M. P. (2009) Applications of the NOE in Molecular Biology, *Annu Rep Nmr Spectro* 65, 77-109.
9. Tanabe, M., Nimigean, C. M., and Iverson, T. M. (2010) Structural basis for solute transport, nucleotide regulation, and immunological recognition of *Neisseria meningitidis* PorB, *P Natl Acad Sci USA* 107, 6811-6816.
10. Brown, L. R. (1979) Use of Fully Deuterated Micelles for Conformational Studies of Membrane-Proteins by High-Resolution H-1 Nuclear Magnetic-Resonance, *Biochim Biophys Acta* 557, 135-148.
11. Ikura, T., Go, N., and Inagaki, F. (1991) Refined Structure of Melittin Bound to Perdeuterated Dodecylphosphocholine Micelles as Studied by 2d-Nmr and Distance Geometry Calculation, *Proteins* 9, 81-89.
12. Liang, B. Y., and Tamm, L. K. (2007) Structure of outer membrane protein G by solution NMR spectroscopy, *P Natl Acad Sci USA* 104, 16140-16145.
13. Sessions, R. B., Gibbs, N., and Dempsey, C. E. (1998) Hydrogen bonding in helical polypeptides from molecular dynamics simulations and amide hydrogen exchange analysis: Alamethicin and melittin in methanol, *Biophys J* 74, 138-152.
14. Wagner, G., and Wuthrich, K. (1982) Sequential Resonance Assignments in Protein H-1 Nuclear Magnetic-Resonance Spectra - Basic Pancreatic Trypsin-Inhibitor, *J Mol Biol* 155, 347-366.



15. Ippolito, J. A., Alexander, R. S., and Christianson, D. W. (1990) Hydrogen-Bond Stereochemistry in Protein-Structure and Function, *J Mol Biol* 215, 457-471.
16. Fernandez, C., Hilty, C., Wider, G., and Wuthrich, K. (2002) Lipid-protein interactions in DHPC micelles containing the integral membrane protein OmpX investigated by NMR spectroscopy, *P Natl Acad Sci USA* 99, 13533-13537.
17. Shen, Y., Delaglio, F., Cornilescu, G., and Bax, A. (2009) TALOS plus : a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts, *J Biomol Nmr* 44, 213-223.
18. Cornilescu, G., Delaglio, F., and Bax, A. (1999) Protein backbone angle restraints from searching a database for chemical shift and sequence homology, *J Biomol Nmr* 13, 289-302.
19. Schwieters, C. D., Kuszewski, J. J., Tjandra, N., and Clore, G. M. (2003) The Xplor-NIH NMR molecular structure determination package, *J Magn Reson* 160, 65-73.
20. Schwieters, C. D., Kuszewski, J. J., and Clore, G. M. (2006) Using Xplor-NIH for NMR molecular structure determination, *Prog Nucl Mag Res Sp* 48, 47-62.
21. Jo, S., Kim, T., and Im, W. (2007) Automated Builder and Database of Protein/Membrane Complexes for Molecular Dynamics Simulations, *Plos One* 2.

22. Wimley, W. C. (2003) The versatile beta-barrel membrane protein, *Curr Opin Struc Biol* 13, 404-411.
23. Fernandez, C., Hilty, C., Bonjour, S., Adeishvili, K., Pervushin, K., and Wuthrich, K. (2001) Solution NMR studies of the integral membrane proteins OmpX and OmpA from Escherichia coli, *Febs Lett* 504, 173-178.
24. Vertessy, B. G. (1997) Flexible glycine rich motif of Escherichia coli deoxyuridine triphosphate nucleotidohydrolase is important for functional but not for structural integrity of the enzyme, *Proteins* 28, 568-579.
25. Liang, J., Naveed, H., Jimenez-Morales, D., Adamian, L., and Lin, M. S. (2012) Computational studies of membrane proteins: Models and predictions for biological understanding, *Bba-Biomembranes* 1818, 927-941.
26. Choutko, A., Glattli, A., Fernandez, C., Hilty, C., Wuthrich, K., and van Gunsteren, W. F. (2011) Membrane protein dynamics in different environments: simulation study of the outer membrane protein X in a lipid bilayer and in a micelle, *Eur Biophys J Biophy* 40, 39-58.
27. Lumb, C. N., He, J., Xue, Y., Stansfeld, P. J., Stahelin, R. V., Kutateladze, T. G., and Sansom, M. S. P. (2011) Biophysical and Computational Studies of Membrane Penetration by the GRP1 Pleckstrin Homology Domain, *Structure* 19, 1338-1346.
28. Stansfeld, P. J., and Sansom, M. S. P. (2011) Molecular Simulation Approaches to Membrane Proteins, *Structure* 19, 1562-1572.

29. Jo, S., Lim, J. B., Klauda, J. B., and Im, W. (2009) CHARMM-GUI Membrane Builder for Mixed Bilayers and Its Application to Yeast Membranes, *Biophys J* 97, 50-58.
30. Wolf, M. G., Hoefling, M., Aponte-Santamaria, C., Grubmuller, H., and Groenhof, G. (2010) g\_membed: Efficient Insertion of a Membrane Protein into an Equilibrated Lipid Bilayer with Minimal Perturbation, *J Comput Chem* 31, 2169-2174.
31. Hess, B., Kutzner, C., van der Spoel, D., and Lindahl, E. (2008) GROMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation, *J Chem Theory Comput* 4, 435-447.
32. Piggot, T. J., Pineiro, A., and Khalid, S. (2012) Molecular Dynamics Simulations of Phosphatidylcholine Membranes: A Comparative Force Field Study, *J Chem Theory Comput* 8, 4593-4609.
33. Hauck, C. R., and Meyer, T. F. (2003) 'Small' talk: Opa proteins as mediators of Neisseria-host-cell communication, *Curr Opin Microbiol* 6, 43-49.
34. Kay, L. E. (2005) NMR studies of protein structure and dynamics, *J Magn Reson* 173, 193-207.
35. Chamberlain, A. K., Lee, Y., Kim, S., and Bowie, J. U. (2004) Snorkeling preferences foster an amino acid composition bias in transmembrane helices, *J Mol Biol* 339, 471-479.

## Chapter 6 Conclusions and Future Work

### Section 6.1 Overview

Opa<sub>60</sub> is a member of the outer membrane protein Opa family of proteins found in the pathogenic bacteria *Neisseria meningitis* and *N. gonorrhoeae*. The primary function of this family of protein is to interact with receptor binding target. These interactions induce a signaling cascade that induces the host cell to engulf the bacterium, even host cells that do not usually undergo phagocytosis. The regions of Opa that interacts specifically with the receptor are called hypervariable loops 1 and 2 (HV1 and HV2), in which both are necessary to induce bacterial uptake.

To understand this interaction, NMR structural studies were performed. Because the loops of Opa<sub>60</sub> are considerably longer and more flexible than the loops of previously studied  $\beta$ -barrel membrane proteins, new techniques were developed in membrane protein studies. To isolate the less intense NMR resonances associated with the  $\beta$ -barrel, Opa<sub>60</sub> was treated with a protease to remove the extracellular loops. In order to isolate the extracellular loops for NMR studies, cold temperature studies were used to observe the dynamic regions of the protein. Finally small peptides were synthesized to study the regions in the loops furthest away from the barrel. These techniques were used to obtain 97% assignment of the barrel region and 24% coverage of the extracellular loops.

Incorporating distance restraints obtained from nuclear Overhauser (NOE) interactions and dihedral angle constraints based on secondary chemical shifts, an ensemble of structures were calculated using simulated annealing methods.

The structure of the barrel regions in the ensemble agreed very well, with a backbone rmsd of 0.96 Å. The loops, lacking assignments and restraints, were very disordered occupying a large volume. Further structural refinement was used on the ensemble of 20 structures by performing a 100 ns long molecular dynamic (MD) simulation of the structures in a lipid bilayer. Results from these simulations indicate that the loops still remain relatively unstructured, with some latent helical content observed in both hypervariable loops. The extracellular loops sampled a far smaller space during the MD simulations, primarily the space above barrel instead of spreading out across the membrane. The loops appear to have low energy, non-specific interactions with each other, never observing a long term stable interaction between the loops.

## **Section 6.2 Applications Towards Other Opa Proteins**

Opa<sub>60</sub> is one of approximately 25 Opa proteins characterized in terms of receptor engagement. Structural and dynamic insight to multiple Opa proteins can result in a greater understanding to the sequence identity and mechanisms present in binding to the host receptors. The trypsin cleaved  $\beta$ -barrel has shown minimal structural changes compared to the full length Opa<sub>60</sub> barrel based on chemical shift perturbation. The cleaved Opa<sub>60</sub> barrel also overlaps very well with a cleaved barrel of Opa<sub>50</sub>, another Opa protein with a completely different binding partner. If this trend is continued through other Opa protein barrel structures, as the sequential similarity of Opa proteins suggests, then the determined barrel structure of Opa<sub>60</sub> can be treated as a scaffold for structural studies of other Opa proteins. Once confirmed, one could model the unique hypervariable loops of

different Opa proteins onto this barrel scaffold and perform molecular dynamics simulations to identify unique Opa loop properties. Latent secondary structure may be involved in the binding mechanism, specifically in the proline rich regions of CEACAM binding HV2 loops. The interaction between Opa and heparin-sulfate proteoglycan molecules is believed to be driven by electrostatics, which can efficiently be modeled *in silico*. A relatively high throughput of unique loops on a predetermined barrel scaffold could result in a wealth of loop interaction data.

### **Section 6.3 CEACAM Binding Studies**

With a much greater knowledge of the role that structure and dynamics plays on the hypervariable regions of Opa<sub>60</sub>, understanding of the interactions between these regions with their binding partner, CEACAM, can be accomplished. Typically binding mechanisms can be elucidated by titrating in the binding partner and monitoring the chemical shift perturbations of different protons. These perturbations can map out specific regions of the three dimensional structure that are involved in binding, leading to proposed mechanisms. Despite the crowded region in NMR spectra where the loop resonances have been assigned, both cold temperature studies and synthetic peptides can generate simplified spectra to study these interactions. Understanding how Opa proteins can bind to very specific targets in the human host and facilitate phagocytosis will provide insights to *Neisseria* pathogenesis and potential mechanisms for targeted delivery.

## Appendix: Materials and Methods

### *Expression and purification of Opa<sub>60</sub>*

The gene for Opa<sub>60</sub> was sub-cloned into pET28b from pEX vector constructs (provided by Martine Bos, Utrecht University) and transformed into BL21 (DE3) cells. Opa<sub>60</sub> was sub-cloned such that both N- and C-terminal fusion tags were included (MGSSHHHHHHSSGLVPRGSHM and KLAAALEHHHHHH, respectively). Cells were grown in D<sub>2</sub>O (99.8%) minimal medium containing 4 g/liter <sup>13</sup>C(99%)-glucose and 1 g/liter <sup>15</sup>N(99%)-ammonium chloride (Cambridge Isotopes Lab) at 310°C to an OD<sub>600</sub> of ~0.8 then induced with 1 mM isopropyl-β-thio-D-galactoside for 8 hours. The cells were resuspended and lysed in 50 mM Tris-HCl and 150 mM NaCl (lysis buffer). The lysate was then centrifuged at 12,000 x g for 30 min to remove cell debris. The pellet was resuspended in extraction buffer (lysis buffer with 8 M urea) and solubilized overnight. The resuspension was centrifuged at 12,000 x g for 30 min and the soluble fraction was loaded onto a Co<sup>2+</sup> immobilized metal affinity chromatography (IMAC) column equilibrated with 10 column volumes (CV) of extraction buffer. The column was then washed with 15 CV of wash buffer (20 mM sodium phosphate, pH 7.8, 150 mM NaCl, 20 mM imidazole, 8 M urea) followed by an elution with 5 CV of elution buffer (20 mM sodium phosphate, pH 7.0, 150 mM NaCl, 680 mM imidazole).

***Refolding of Opa<sub>60</sub>***

The elution was concentrated to 200  $\mu$ M and subsequently diluted 20-fold with 20 mM Tris-HCl, pH 8.0, 500 mM NaCl with 4.5 mM n-dodecyl-phosphocholine (FC-12, Anatrace). The sample was incubated at room temperature for five days. Protein folding was monitored based on the shift of apparent molecular weight on SDS-PAGE, until the final sample lacked the higher apparent molecular weight band of the unfolded species. The solution was concentrated and dialyzed against 4 L of 20 mM sodium phosphate, pH 6.2, and 150 mM NaCl; three times for an hour each. Final NMR samples were concentrated to 400-800  $\mu$ M and contained 110-150 mM FC-12 as measured by comparing NMR peak intensities with standard FC-12 concentrations.

***Trypsin proteolysis***

Trypsin from porcine pancreas (Sigma-Aldrich) was added to the proteins in their respective final buffers at a trypsin:sample ratio between 1:50 – 1:100. After incubating overnight at room temperature, trypsin-treated samples were assessed with SDS-PAGE. Gels were stained with Coomassie brilliant blue. With confirmation of complete proteolysis from the SDS-PAGE gel, trypsin was removed by flowing the solution over 0.5 mL of *p*-aminobenzamidine-agarose resin (Sigma-Aldrich). The flow-through was then dialyzed against 4 L of 20 mM sodium phosphate, pH 6.2, and 150 mM NaCl; three times for an hour each. Final NMR samples were concentrated to 400-800  $\mu$ M and contained 110-150 mM FC-12 as measured by comparing NMR peak intensities with standard FC-12 concentrations.



### ***Preparation of peptide***

A 20 amino acid peptide corresponding to a region in the third extracellular loop of Opa<sub>60</sub> (Ac-TVPSNAPNGAVTTYNTDPKT-NH<sub>2</sub>) was synthesized by Anaspec with <sup>15</sup>N amide nitrogen incorporation for all threonine, valine, alanine, and lysine residues. The lyophilized peptide was resuspended in 20 mM sodium phosphate, pH 6.2, and 150 mM NaCl at a concentration of 1.0 mM.

### ***NMR spectroscopy***

NMR spectra were collected on Bruker AVANCE spectrometers operating at proton frequencies of 600 MHz and 800 MHz equipped with Bruker 5 mm TXI cryoprobes and recorded at 10, 20, 30 and 40°C for Opa<sub>60</sub> and the synthetic peptide. Spectra were processed with Topspin and assigned using CARRA. Relaxation data was obtained using NMRPipe. In order to assign the backbone, TROSY versions of HNCA, HN(CO)CA, HNCO, HN(CA)CO, HN(CA)CB, and HN(COCA)CB pulse sequences were recorded for both the full-length and trypsin-treated samples. <sup>15</sup>N, <sup>1</sup>H TROSY-HSQC spectra were recorded over a series of temperatures from 40°C to 10°C to observe chemical shifts changes of assigned resonances for both Opa<sub>60</sub> and the synthetic peptide. 2D <sup>1</sup>H, <sup>1</sup>H TOCSY and COSY spectra (both full <sup>1</sup>H and <sup>15</sup>N-edited H in the direct axis) were recorded for the peptide to assign the side chains.

### ***NMR Structure calculations***

The TALOS+ program was used to obtain backbone dihedral angle restraints. Assigned NOE peak heights were measured and binned into strong, medium or weak interactions. These were assigned upper limits of 3.5, 5.0 and 6.5 Å. Where

applicable, hydrogen bonds were used with upper limits of 2.5 and 3.5 Å for HN...O and N...O, respectively. Additionally, planar restraints from the NIH-Xplor library were used to represent the lipid bilayer to limit the spatial sampling of the extracellular loops. The structure calculations were performed using NIH-XPLOR v2.31. Starting at 3000 K, 5000 steps of high-temperature annealing was used to fold the initial extended structure. Twenty of the lowest overall violation energies of the 300 calculated structures were selected for further MD simulations.

### ***MD simulations***

All simulations were performed using Gromacs 4.5 and the Charmm36 forcefield for protein and lipid interactions. Temperature equilibration used the velocity-rescaling thermostat using a temperature of 300 K with a time-constant of 0.1 ps, and pressure was controlled semi-isotropically using the Parrinello-Rahman barostat at 1 bar. All covalent bonds were constrained using LINCS, and long-range electrostatics were computed every step using Particle Mesh Ewald (PME).

An initial 512 dimyristoylphosphatidylcholine (DMPC) lipid patch was obtained from the CHARMM-GUI membrane builder tool. Including approximately 40000 TIP3p waters, the total system size before protein insertion was approximately 180000 atoms. Ions were added to neutralize the system at a concentration of 150 mM NaCl. The membrane patch was equilibrated as described above, the starting area per lipid before protein insertion was 0.60 nm<sup>2</sup>, close to the experimentally determined value (0.606±0.5 nm<sup>2</sup>). The system dimensions were 12.46 nm (sides) and 11 nm (height).

The 20 lowest energy NMR were inserted in the equilibrated membrane using the Gromacs tool `g_membed`, removing approximately 15-20 lipids in the process. Each system was then energy minimized for 500 steps using steepest descents to relieve any bad contacts possibly induced in the insertion process. Production runs were then carried out for 100 ns using a time step of 2 fs. Snapshots were saved every 500 ps yielding 200 structures for each simulation. Experimental distance restraints were used on the barrel with a force constant of 1000 kJ/mol. To perform the c-alpha density analysis, each trajectory was first fitted to the same reference structure using only the atoms in the barrel in the process (ie atoms in the loops were excluded from the fit). Then the density was calculated on a 3D-grid with the MDAnalysis toolkit using a grid-spacing of 0.1 nm. The resulting density grid was visualized using the volume visualization capabilities of PyMol.

Clustering was performed using the Gromacs tool `g_cluster`, which was extended to include the k-means algorithm. Snapshots from 50-70 ns of simulation time were extracted and used to obtain 50 cluster centers. Then, using RMSD as a distance measure, snapshots from 70-90 ns were assigned to the corresponding closest cluster center and cluster occupancies could thus be calculated and compared.