The Role of Proline Isomerization as a Route for Opa-CEACAM Specificity

Ethan Cole Sesco
Morganton, NC

B.S. Wake Forest University, 2020

A Thesis presented to the Graduate Faculty of the University of Virginia in Candidacy for the Degree of Master of Science

Department of Chemistry

University of Virginia
May 2022

Contents

Abstract

Due to the prevalence of antibiotic use in recent decades, there has been a rise in antibiotic resistant bacterial infections. Bacterial gonorrhea is one such infection and the responsible bacterium, *Neisseria gonorrhoeae*, has reached superbug status with strains that are resistant to all available antibiotics. *Neisseria* triggers uptake by human cells by binding to carcinoembryonic antigen-like cell adhesion molecules (CEACAMs) on the surface of host cells via opacity associated proteins (Opa) located in the outer membrane of *Neisseria*. Opa proteins are eight-stranded β-barrel proteins that have four extracellular loops. Extracellular loops two and three contain hypervariable regions. These hypervariable regions are responsible for the binding specificity between Opa and the N-terminal Ig-like V-domain conserved across all CEACAMs. Opa_{60}, an Opa protein expressed in N. *gonorrhoeae*, binds CEACAM1, 3, 5, and 6. While it is established that Opa proteins bind the N-terminal domain in CEACAM1, the molecular determinants of binding specificity remain unknown. This study will investigate the role of proline isomerization in Opa_{60} using mutagenesis studies, solution NMR and in vitro binding assays.

Acknowledgements

First and foremost, I would like to thank my advisor Dr. Linda Columbus. Your commitment to your lab and students is incredible. Thank you for supporting, pushing and believing in me during my graduate career. I could not have asked for a better graduate advisor and mentor while at UVA.

Secondly, I would like to thank all members of the Columbus Lab for their support and friendship. Thank you to previous members who I did not meet, but who built the lab and paved the way for the research I was fortunate to work on.

And former/current lab members who I did meet, Dr. Meagan Belcher Dufrisne, Dr. Chris Baryiames, Dr. Tracy Caldwell, Connor McDermott and Matt Necelis, thank you for the training, laughs and always being encouraging no matter how tough some days could be. I consider each of you friends and look forward to keeping in touch.

Thirdly, I would like to thank Dr. Meagan Belcher Dufrisne and Dr. Jeffery Ellena for their help experimentally. I would like to thank Meagan for allowing me to use her PhoE-Opa_{60} constructs and training me. Your support and training was incredibly valuable during my time in the Columbus Lab. I would like to thank Jeff for his assignment of the 2D and 3D Opa_{60} spectra, as well as his support and training me from using the NMR spectrometer to running NMR experiments. I could not have finished this project without Meagan and Jeff.

Finally, I would like to thank my friends and family for believing and supporting me, not only in my academic but my personal pursuits as well. I could not have made it this far without you.

Section 1: Protein Structure and Function

1.1 Relation between protein structure and function

Proteins are vital macromolecules that play an integral role in biological systems and are instrumental in a variety of functions. Enzymes, catalytic proteins, catalyze reactions essential to processes like cellular respiration, scaffolding proteins serve as platforms for cellular signaling, and adhesion proteins allow cells to facilitate communication between each other. Protein function is dictated by protein structure. For example, in the well-studied alpha helical membrane protein, bacteriorhodopsin, the residue lysine 216 is vital for forming a Schiff's base with the chromophore in the complex which allows for the absorption of light and the signal to be transduced in the cell[1,2]. Understanding protein structure increases fundamental knowledge of biological processes that can guide applications like drug design.

1.2 Building Proteins

1.2.a The amino acids

The basic unit of protein structure is the amino acid. There are twenty standard amino acids that are commonly observed in biological systems. Each of these proteogenic amino acids are considered α-amino acids, which consist of an α carbon bonded to a primary amine ($-NH_2$), carboxylic acid (-COOH), an α proton, and side chain (R group). Since each functional group bonded to the α carbon is chemically unique, the amino acids are chiral and reflect plane-polarized light. Generally, all amino acids obtained from proteins contain the L-configuration[3,4].

The twenty amino acids are categorized into three groups based on the chemical composition of their R group (**Figure 1**). The four aliphatic amino acids are alanine, valine, leucine, and isoleucine. The R group of these amino acids are carbons linked in an open or

branched chain. The aromatic amino acids are histidine, tyrosine, phenylalanine, tryptophan. These amino acids contain a closed, aromatic ring system. Histidine contains an imidazole side chain, phenylalanine contains a benzene ring, tyrosine contains a phenol ring, and tryptophan contains an indole ring. Ten of the proteogenic amino acids contain a functionalized aliphatic side chain. These amino acids are serine, threonine, glutamate, aspartate, glutamine, asparagine, cysteine, methionine, lysine, and arginine. These side chains vary in structure but are generally hydrophilic and polar, except for methionine and cysteine. The remaining amino acids, proline and glycine are the outliers to these three general categories of protein structure. Since the backbone nitrogen of proline is in the closed ring when bonded to other amino acids, proline does not contain an amino group but instead has a backbone imino group. Glycine contains two hydrogen atoms connected to the α carbon and therefore, it has no distinct R group and is achiral[4].

1.2.b Protein structure

The hierarchy of protein structure begins with the amino acid sequence of the protein (**Figure 2**). The amino acid sequence, also known as the primary structure, does not account for the three-dimensional orientation of the amino acids relative to one another. Primary amino acid structure is represented as a linear chain of amino acids from the amino group (N-terminus) of the first amino acid to the carboxylic acid (C-terminus) of the last amino acid[3]. Peptides and proteins are polymers of amino acids bonded together via amide bonds. The amino group (NH_2) of amino acid (i) is bonded to the carbonyl carbon of amino acid (i-1). This amide bond, (C=O) – NH is often referred to as a peptide bond. Proteins are generally considered to have more than 40 amino acids while peptides typically have less than 40 amino acids[3,4].

Secondary structure of proteins begins to consider the 3D orientation of the amino acids in relation to one another. Due to partial double bond character of the amide bond between the C=O and N, peptide bonds have a planar orientation, because the partial double bond maximizes pi-orbital overlap. The remaining bonds of the main chain, or protein backbone, contain two dihedral angles that vary in degrees of freedom about the N-C_α (Φ angle) bond and C-C=O bond (Ψ angle) depending on the amino acid. Of the twenty standard amino acids, proline is the most sterically constrained amino acid, while glycine is the least sterically constrained.

There are two common secondary structural motifs observed in protein structure. These two motifs are the α-helix and the β-sheet (**Figure 2**). The helix first observed by Linus Pauling and co-workers in 1951 ideally contains amino acids every 3.4 Å, Φ = -57, and Ψ = -47. An α-helix ideally contains a hydrogen bond every 3.4 Å between the amide nitrogen of amino acid i and the carbonyl of amino acid, i-3/4[3,5]. The β-sheet, also proposed by Pauling and coworkers in 1951, ideally has $\Phi = \Psi = \pm 180^\circ$ and can vary in the amount of amino acids in each involved strand[3,6]. A structural difference between the α-helix and the β-sheet is how many strands of a polypeptide chain are contained in the motif. In α-helices hydrogen bonding is within the same strand in a polypeptide chain. While hydrogen bonding in β-sheets is between separate strands in a polypeptide chain (**Figure 2**). The composition of various strands allows for two different orientations of β-sheets to be observed. β-sheets can be presented as parallel or antiparallel. Parallel β-sheets arise when both polypeptide strands are arranged in the same orientation, i.e. both are oriented from N-terminus to C-terminus or vice versa. Antiparallel β-sheets are the opposite case. Antiparallel β-sheets arise when the separate polypeptide chains are oriented in opposite directions, i.e. one chain is arranged from N- to C-terminus while the other chain is arranged C- to N-terminus[3,6].

The next two levels of protein structure are tertiary and quaternary structure. Tertiary structure considers the 3D shape of the protein and is the ensemble of all secondary structures in a protein, including regions of intrinsic disorder, regions that have no secondary structure within a protein. Quaternary is the final level of protein structure and combines all the protein subunits or protomers. Some proteins do not have multiple subunits and only have tertiary structure. An example of a protein with only tertiary structure is myoglobin, which is composed of one polypeptide chain that is folded into a three-dimensional shape[7]. An example of a protein with quaternary structure is hemoglobin. Hemoglobin is a heterotetrametric protein composed of four subunits, two alpha and two β subunits[8,9]. Proteins that have quaternary structure are classified on their number of subunits, monomer = 1 subunit, dimer = 2 subunits, trimer = 3 subunits, tetramer = 4 subunits, as well as if these subunits are similar or different, homo or hetero.

Tertiary and quaternary structure are stabilized by covalent interactions like disulfide bonds and/or through noncovalent interactions. Covalent bonds are stronger than noncovalent interactions, but the structure of biomolecules is dependent on the combined influence of many noncovalent interactions. Noncovalent interactions include electrostatic interactions and van Der Waals forces: hydrogen bonding, dipole-dipole interactions and London dispersion forces, which arise from permanent dipole interactions. Noncovalent interactions vary in strength, electrostatic interactions ≈ 18 kcal/mol, hydrogen bonding ≈ 4 kcal/mol, dipole-dipole interactions ≈ 2 kcal/mol and London dispersion forces ≈ 0.06 kcal/mol[3].

Protein folding is important aspect of protein structure. Folded proteins are in a low energy, low entropy state but arrive to this state from a high energy, high entropy state. Driving forces for protein folding include the hydrophobic effect, the tendency for nonpolar surfaces to interact with one another, and interactions like hydrogen bonding[3]. Hydrogen bonding has a

variable role in protein folding. The formation of an H-bond in a vacuum is approximately -6 kcal/mol, but is only -0.9 kcal/mol in polyalanine (an α-helix composed mainly of alanine) in an aqueous environment[10]. The energy for the hydrogen bond formation becomes even higher, +2.5 kcal/mol, when the polyalanine helix is introduced into a protein interior. Therefore, when a protein is folding the desolvation of the peptide bond must be paired with introducing the peptide bond into the protein interior[10].

Membrane protein folding is governed by the same interactions as soluble proteins, but the combination of them is more complex because of the lipid environment surrounding membrane proteins. The hydrophobic effect for soluble proteins concerns hydrophobic interactions increasing the entropy of the surrounding water molecules. While in membrane proteins, the hydrophobic effect is mostly focused on the protein surfaces increasing the entropy of the surrounding lipids[11,12].

It is important to note that not all proteins have secondary, tertiary or quaternary structure. These proteins are called intrinsically disordered proteins (IDP's) and are typically long, unstructured chains of amino acids. Proteins can also have structured and disordered domains[13,14]. An example of a protein with disordered regions are the opacity associated protein (Opa) family found in the outer membrane of *Neisseria*. These proteins contain a structured beta barrel motif as well as loop regions that are unstructured, lacking observable secondary structure[15–17].

1.3 Membrane protein structure

Located in the lipid bilayer, membrane proteins serve an essential role in overall cellular function. These roles include signaling, adhering to neighboring cells, defense against the

environment and entry to host cells for pathogenic bacteria like *Neisseria*. Due to their importance in function, approximately 20% of the human genome encodes for membrane proteins[18].

Membrane proteins are generally categorized as peripheral or integral membrane proteins. Peripheral membrane proteins are associated with only one side, or leaflet, of the lipid bilayer, while integral membrane proteins span the entire lipid bilayer. Structural motifs observed in integral membrane proteins are the α helix and β barrel (**Figure 3**).

Alpha helical transmembrane proteins are found in both eukaryotes and prokaryotes. The first crystal structure of an a-helical membrane protein was determined by Michel and co-workers in 1985 of the photoreaction center in *Rhodopseudomonas viridis*[19]. Integral α-helical membrane proteins can have a single transmembrane domain or multiple pass membrane helices. These transmembrane helices are oriented so that they make hydrophobic contacts to shield the polar protein backbone from the hydrophobic environment of the lipid bilayer[3,20–22]

Another common transmembrane protein motif observed is the β barrel. β barrel proteins have been largely observed in gram-negative bacteria, prokaryotes with an outer membrane. The structure of β barrel proteins was later after alpha helical membrane proteins. The first β barrel protein structures were determined with X-ray crystallography by Schiltz & coworkers, as well as Rosenbusch & coworkers in 1991 and 1992, respectively[22–24].

1.3.a Challenges determining structure of membrane proteins

Given their importance to cellular function, membrane proteins provide valuable targets to understand structurally. However, membrane proteins present challenges to structural biologists. Due to their location in the lipid bilayer, membrane proteins are not stable in an

aqueous environment. Therefore, membrane proteins need to be extracted from the membrane with detergents to surround the hydrophobic contacts of the protein. Various detergents like dodecylphosphocholine (FC-12) or dodecyl maltoside (DDM) are used to extract a membrane protein from the lipid environment. Often a screening of different detergents is necessary to determine which detergent extracts the most active, soluble, and stable protein[25].

Expression and purification of membrane proteins can also be challenging. β barrel membrane proteins can be expressed into inclusion bodies (soluble protein aggregates in the cellular cytoplasm of bacteria) without their signal sequence to traffic the β barrel to the lipid bilayer. This allows the protein to be purified without detergent extraction and be refolded subsequently in detergent or synthetic lipids *in vitro*[25–28]. However, alpha helical membrane proteins are more difficult to purify from inclusion bodies and are typically preferred to be overexpressed to the lipid bilayer and extracted by detergents[25,29,30].

1.4 How protein structure is determined

Various methods are used to determine protein structure (**Figure 4**). Historically, x-ray crystallography has been the "gold standard" for protein structure determination. The first protein structure was determined by x-ray crystallography and many of the structures deposited in the Protein Data Bank (PDB) are x-ray crystal structures, since 2012 at least 8,000 x-ray crystal structures have been entered into the PDB annually[31]. Despite being available as a structural technique for decades, cyro-electron microscopy (Cryo-EM) was recently developed into a powerful structural technique especially for the study of large protein complexes. The recent rapid development of Cryo-EM is the result of advances in microscopes, detectors and data processing[32]. Nuclear magnetic resonance is also a valuable technique for structural biology.

While protein NMR spectroscopy is typically limited to smaller proteins methods do exist to study larger proteins. NMR is a valuable tool for understanding protein dynamics because NMR can observe proteins in different conformations since residue specific and atomic level information can be gathered[33–35].

1.5 Protein structure and function determination in this thesis

This thesis focuses on furthering previous structural work of the outer membrane protein, Opa_{60} found in *Neisseria gonorrhoeae*, by solution NMR. The solution NMR structure of Opa_{60} was determined in 2014 and revealed Opa_{60} to be an eight-stranded β-barrel protein with four unstructured loop regions (**Figure 5**). β-barrel proteins are more convenient to study with NMR than alpha helical membrane proteins[16,36]. The peaks observed in an alpha helical membrane NMR spectrum are typically clustered together leading to overlap making peak assignment with NMR difficult and NOE derived distances from backbone assignments, typically only define secondary structure in α-helical membrane proteins[25]. This project focuses on mutating proline residues to glycine residues to understand the functional implication of mutating residues located in the loop region of Opa_{60}. The loop regions are necessary for the function of Opa_{60} binding to its target protein in humans, carcinoembryonic antigen-like cell adhesion molecules (CEACAMs).

Section 1 Figures

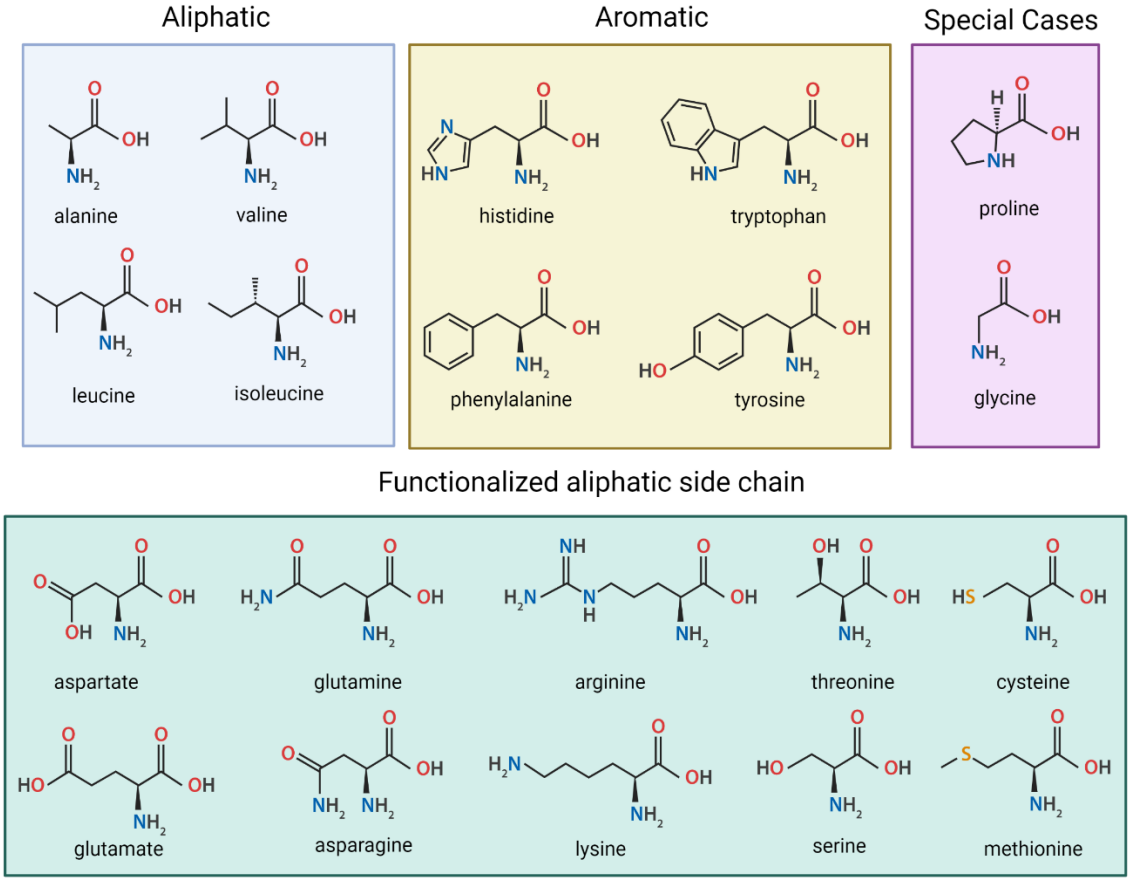


Figure 1. Amino acid chart showing the 20 standard proteogenic amino acids. The amino acids are broadly grouped based on the chemical composition of their R group. The aliphatic amino acids are shown in blue, the hydrophobic amino acids are shown in yellow, and the functionalized aliphatic side chain amino acids are shown in green. Proline and glycine are shonw in purple since they are exceptions to these classifications. Proline is considred an imino acid and glycine has no distinctive R group[3,4]. Created with BioRender.com.

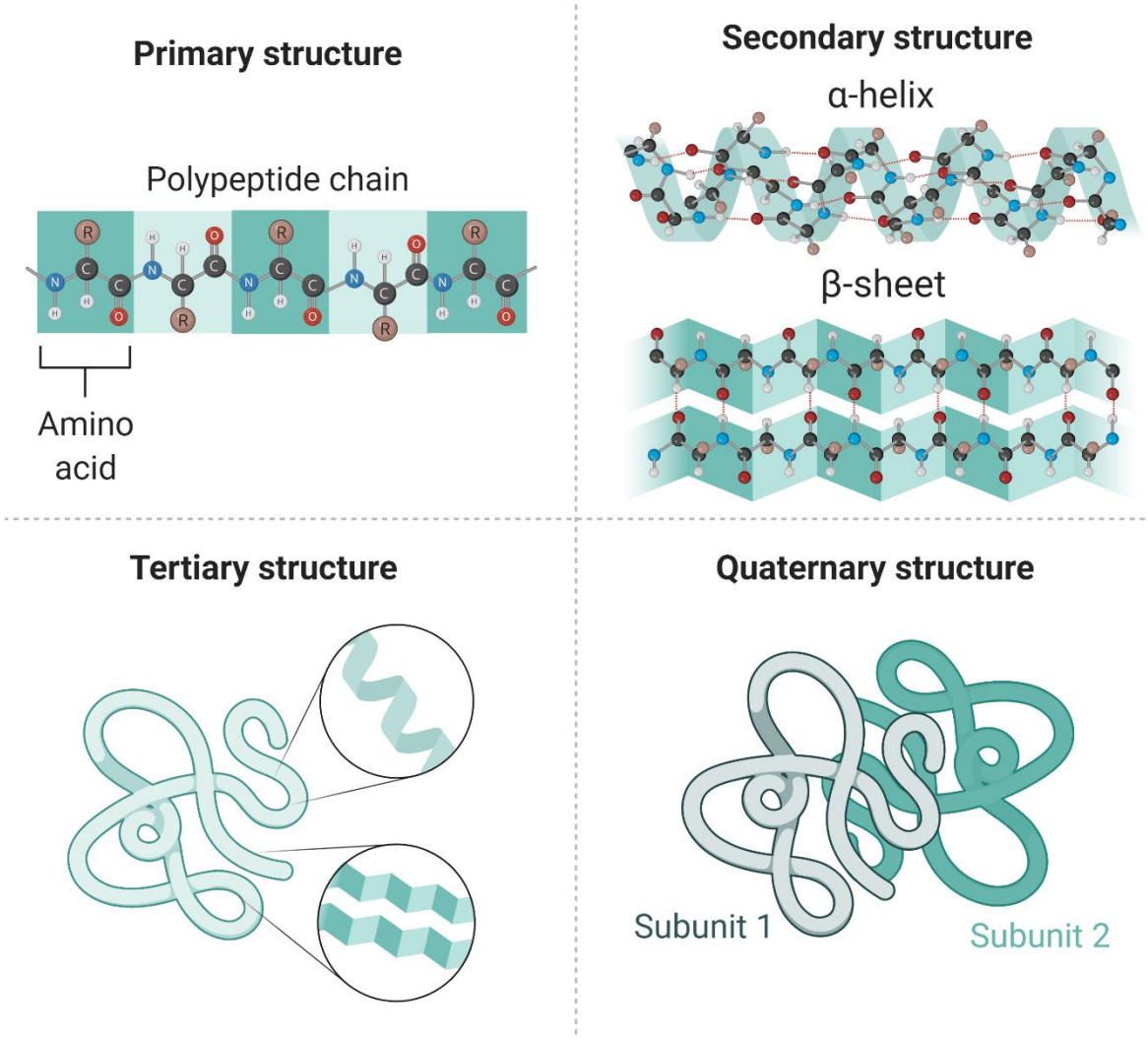


Figure 2. Hierarchy of protein structure. A polypeptide chain linked from the amine group (N-terminus) of the first amino acid to the carboxylic acid (C-terminus) of the last amino is shown under primary structure. An example of an α helix and antiparallel β sheet is shown for secondary structure. The tertiary and quaternary panels show the three-dimensional shape of a folded protein and subunits[3]. Created with BioRender.com.

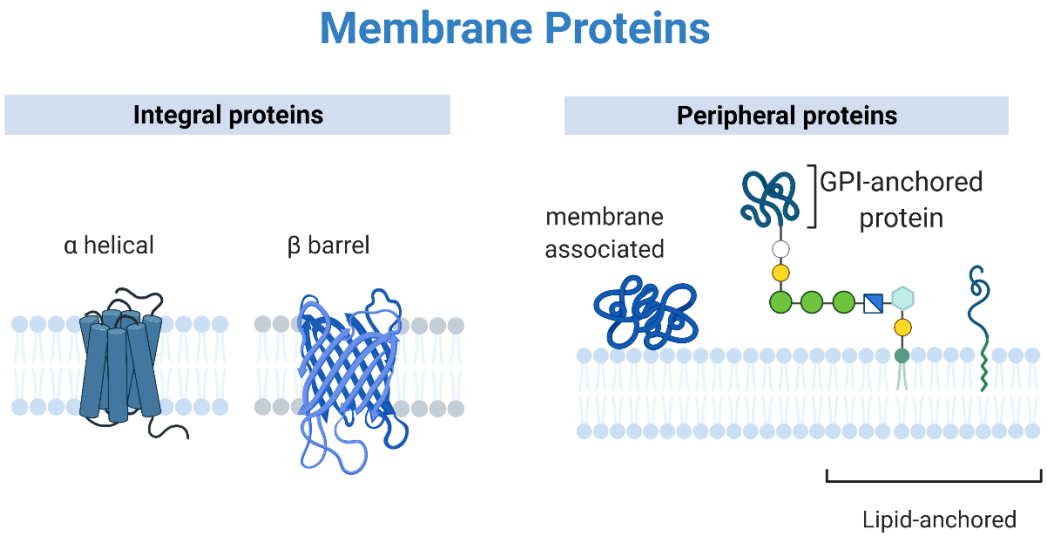


Figure 3. Types of membrane proteins. Integral membrane proteins have either alpha helical or beta barrel transmembrane domains. Peripheral membrane proteins can vary in their attachment to lipid membrane by being noncovalently associated or anchored into one leaflet of the lipid bilayer[3,37]. Created with BioRender.com.

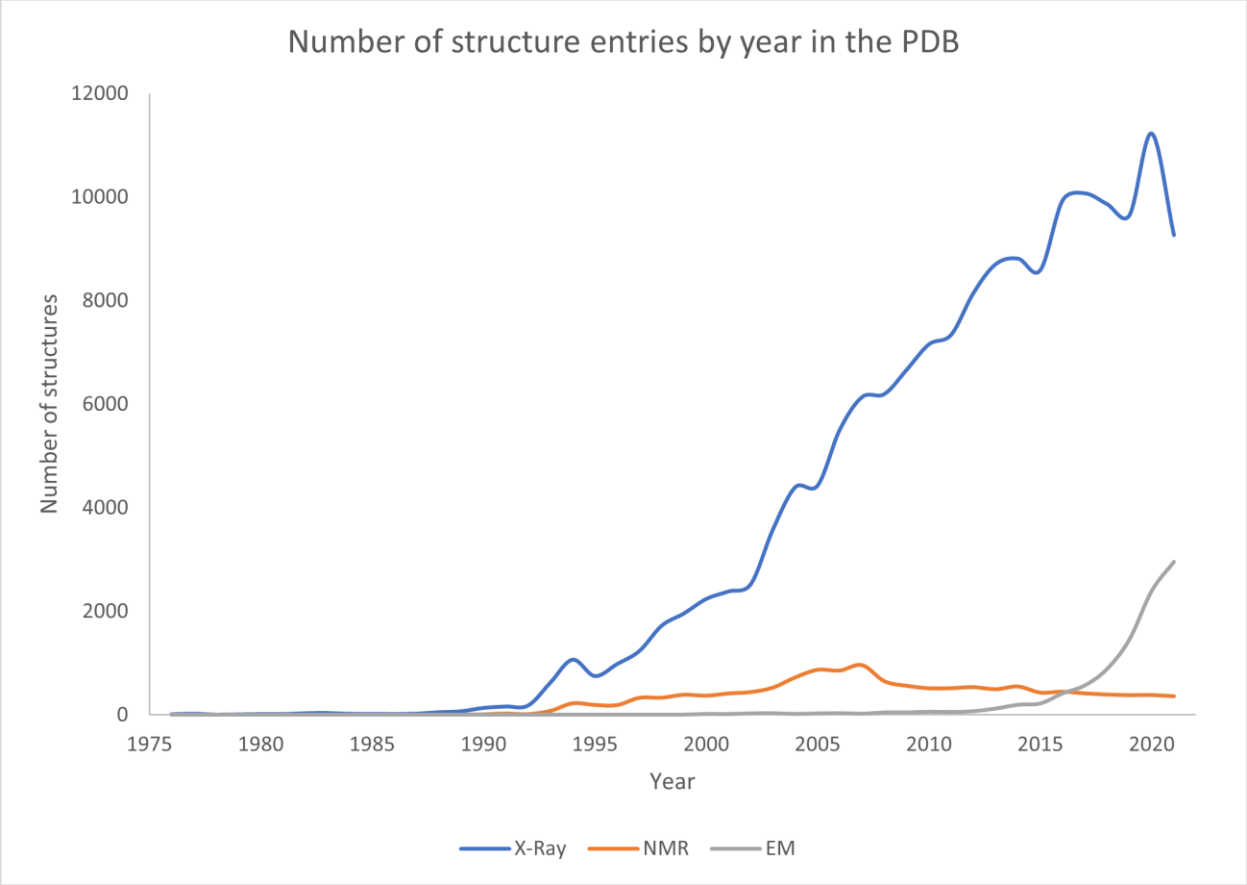


Figure 4. Number of strucures deposited in the Protein Data Bank (PDB) by year from 1970 to 2021. X-ray crystallograhy is shown in blue, NMR is shown in orange, and Cyro-EM is shown in gray. Statistics downloaded from the PDB[31].

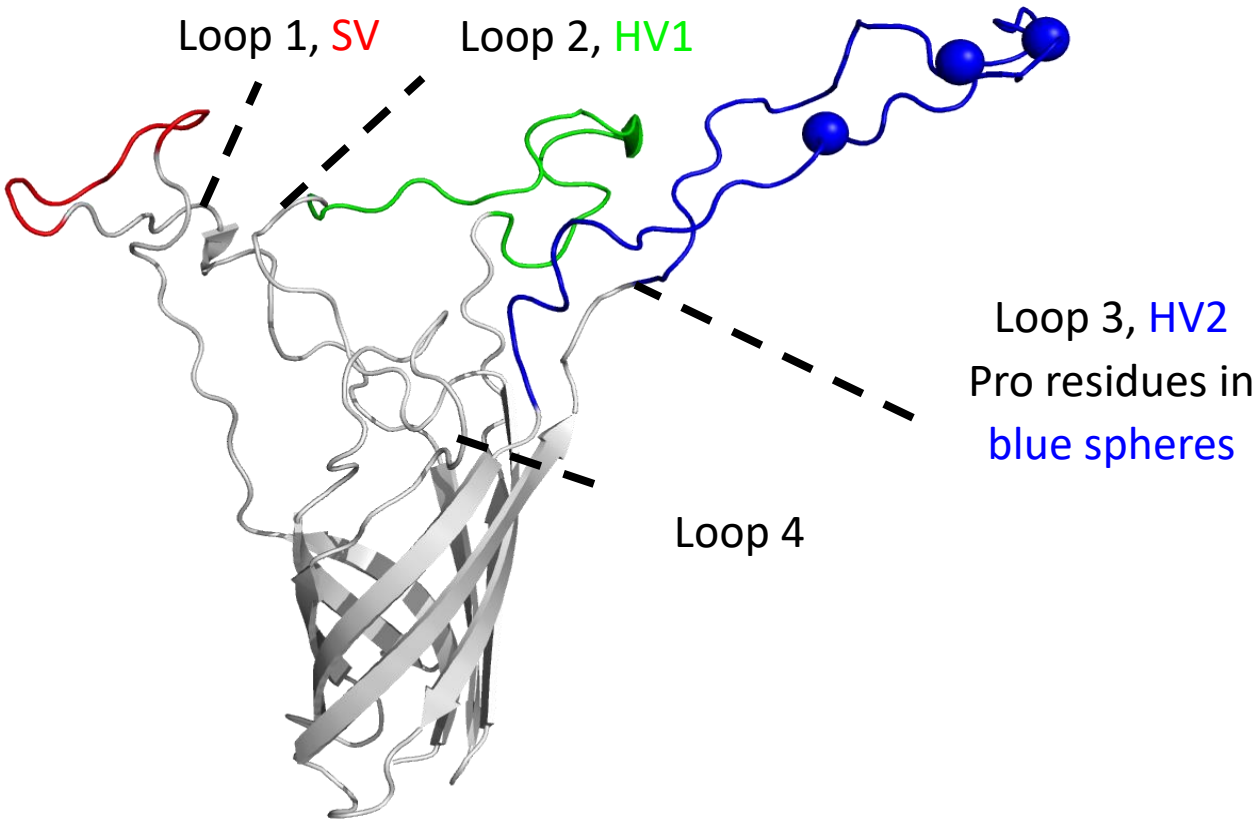


Figure 5. Opa proteins consist of an eight stranded β-barrel, which anchors the protein to the outer membrane of *Neisseria* and four, highly dynamic extracellular loops that mediate Opa-receptor interactions[36]. Within these four loops, there are three regions of sequence variability: a semi-variable (SV, red) region in loop 1, one hypervariable region (HV1, green) in loop 2, and a second hypervariable region (HV2, blue) in loop 3. The fourth loop is conserved across Opa variants. The β-barrel and loop 4 are shown in gray[36]. Mutated proline residues represented as blue spheres. PDB 2MAF and figure created with PyMol[38].

References

(1) Bayley H; Huang K S; Radhakrishnan R; Ross A H; Takagaki Y; Khorana H G. Site of Attachment of Retinal in Bacteriorhodopsin. *Proceedings of the National Academy of Sciences* **1981**, *78* (4), 2225–2229. https://doi.org/10.1073/pnas.78.4.2225.

(2) Gautier, A.; Mott, H. R.; Bostock, M. J.; Kirkpatrick, J. P.; Nietlispach, D. Structure Determination of the Seven-Helix Transmembrane Receptor Sensory Rhodopsin II by Solution NMR Spectroscopy. *Nature Structural & Molecular Biology* **2010**, *17* (6), 768–774. https://doi.org/10.1038/nsmb.1807.

(3) Voet, D., Voet, Judith G. ,. Pratt, Charlotte W. ,. *Fundamentals of Biochemistry Life at the Molecular Level*; 2016.

(4) Barrett, G. C.; Elmore, D. T. *Amino Acids and Peptides*; Cambridge University Press: Cambridge, 1998. https://doi.org/10.1017/CBO9781139163828.

(5) Pauling Linus; Corey Robert B.; Branson H. R. The Structure of Proteins: Two Hydrogen-Bonded Helical Configurations of the Polypeptide Chain. *Proceedings of the National Academy of Sciences* **1951**, *37* (4), 205–211. https://doi.org/10.1073/pnas.37.4.205.

(6) PAULING, L.; COREY, R. B. The Pleated Sheet, a New Layer Configuration of Polypeptide Chains. *Proc Natl Acad Sci U S A* **1951**, *37* (5), 251–256. https://doi.org/10.1073/pnas.37.5.251.

(7) KENDREW, J. C.; DICKERSON, R. E.; STRANDBERG, B. E.; HART, R. G.; DAVIES, D. R.; PHILLIPS, D. C.; SHORE, V. C. Structure of Myoglobin: A Three-Dimensional Fourier Synthesis at 2 Å. Resolution. *Nature* **1960**, *185* (4711), 422–427. https://doi.org/10.1038/185422a0.

(8) PERUTZ, M. F.; ROSSMANN, M. G.; CULLIS, A. F.; MUIRHEAD, H.; WILL, G.; NORTH, A. C. T. Structure of Hæmoglobin: A Three-Dimensional Fourier Synthesis at 5.5-Å. Resolution, Obtained by X-Ray Analysis. *Nature* **1960**, *185* (4711), 416–422. https://doi.org/10.1038/185416a0.

(9) Schechter, A. N. Hemoglobin Research and the Origins of Molecular Medicine. *Blood* **2008**, *112* (10), 3927–3938. https://doi.org/10.1182/blood-2008-04-078188.

(10) Baldwin, R. L. In Search of the Energetic Role of Peptide Hydrogen Bonds. *Journal of Biological Chemistry* **2003**, *278* (20), 17581–17588. https://doi.org/10.1074/jbc.X200009200.

(11) Helms, V. Attraction within the Membrane. *EMBO reports* **2002**, *3* (12), 1133–1138. https://doi.org/10.1093/embo-reports/kvf245.

(12) Corin, K.; Bowie, J. U. How Physical Forces Drive the Process of Helical Membrane Protein Folding. *EMBO reports* **2022**, *23* (3), e53025. https://doi.org/10.15252/embr.202153025.

(13) Wright, P. E.; Dyson, H. J. Intrinsically Disordered Proteins in Cellular Signalling and Regulation. *Nat Rev Mol Cell Biol* **2015**, *16* (1), 18–29. https://doi.org/10.1038/nrm3920.

(14) van der Lee, R.; Buljan, M.; Lang, B.; Weatheritt, R. J.; Daughdrill, G. W.; Dunker, A. K.; Fuxreiter, M.; Gough, J.; Gsponer, J.; Jones, D. T.; Kim, P. M.; Kriwacki, R. W.; Oldfield, C. J.; Pappu, R. V.; Tompa, P.; Uversky, V. N.; Wright, P. E.; Babu, M. M.

Classification of Intrinsically Disordered Regions and Proteins. *Chem Rev* **2014**, *114* (13), 6589–6631. https://doi.org/10.1021/cr400525m.

(15) Bos Martine P.; Kao David; Hogan Daniel M.; Grant Christopher C. R.; Belland Robert J. Carcinoembryonic Antigen Family Receptor Recognition by Gonococcal Opa Proteins Requires Distinct Combinations of Hypervariable Opa Protein Domains. *Infection and Immunity* **2002**, *70* (4), 1715–1723. https://doi.org/10.1128/IAI.70.4.1715-1723.2002.

(16) Fox, D. A.; Columbus, L. Solution NMR Resonance Assignment Strategies for β-Barrel Membrane Proteins. *Protein Sci* **2013**, *22* (8), 1133–1140. https://doi.org/10.1002/pro.2291.

(17) Dewald, A. H.; Hodges, J. C.; Columbus, L. Physical Determinants of β-Barrel Membrane Protein Folding in Lipid Vesicles. *Biophysical Journal* **2011**, *100* (9), 2131–2140. https://doi.org/10.1016/j.bpj.2011.03.025.

(18) Piccoli, S.; Suku, E.; Garonzi, M.; Giorgetti, A. Genome-Wide Membrane Protein Structure Prediction. *Curr Genomics* **2013**, *14* (5), 324–329. https://doi.org/10.2174/13892029113149990009.

(19) Deisenhofer, J.; Epp, O.; Miki, K.; Huber, R.; Michel, H. Structure of the Protein Subunits in the Photosynthetic Reaction Centre of Rhodopseudomonas Viridis at 3Å Resolution. *Nature* **1985**, *318* (6047), 618–624.

(20) Vinothkumar, K. R.; Henderson, R. Structures of Membrane Proteins. *Q Rev Biophys* **2010**, *43* (1), 65–158. https://doi.org/10.1017/S0033583510000041.

(21) Li, F.; Egea, P. F.; Vecchio, A. J.; Asial, I.; Gupta, M.; Paulino, J.; Bajaj, R.; Dickinson, M. S.; Ferguson-Miller, S.; Monk, B. C.; Stroud, R. M. Highlighting Membrane Protein Structure and Function: A Celebration of the Protein Data Bank. *Journal of Biological Chemistry* **2021**, *296*. https://doi.org/10.1016/j.jbc.2021.100557.

(22) Luckey, M. Chapter 1 Introduction to the Structural Biology of Membrane Proteins. In *Computational Biophysics of Membrane Proteins*; The Royal Society of Chemistry, 2017; pp 1–18. https://doi.org/10.1039/9781782626695-00001.

(23) Weiss, M. S.; Abele, U.; Weckesser, J.; Welte, W.; Schiltz, E.; Schulz, G. E. Molecular Architecture and Electrostatic Properties of a Bacterial Porin. *Science* **1991**, *254* (5038), 1627–1630.

(24) Cowan, S. W.; Schirmer, T.; Rummel, G.; Steiert, M.; Ghosh, R.; Pauptit, R. A.; Jansonius, J. N.; Rosenbusch, J. P. Crystal Structures Explain Functional Properties of Two E. Coli Porins. *Nature* **1992**, *358* (6389), 727–733. https://doi.org/10.1038/358727a0.

(25) Carpenter, E. P.; Beis, K.; Cameron, A. D.; Iwata, S. Overcoming the Challenges of Membrane Protein Crystallography. *Curr Opin Struct Biol* **2008**, *18* (5), 581–586. https://doi.org/10.1016/j.sbi.2008.07.001.

(26) Wagner, G.; Wüthrich, K. Sequential Resonance Assignments in Protein 1H Nuclear Magnetic Resonance Spectra: Basic Pancreatic Trypsin Inhibitor. *Journal of Molecular Biology* **1982**, *155* (3), 347–366. https://doi.org/10.1016/0022-2836(82)90009-2.

(27) Bannwarth, M.; Schulz, G. E. The Expression of Outer Membrane Proteins for Crystallization. *Biochimica et Biophysica Acta (BBA) - Biomembranes* **2003**, *1610* (1), 37–45. https://doi.org/10.1016/S0005-2736(02)00711-3.

(28) Rinas, U.; Garcia-Fruitós, E.; Corchero, J. L.; Vázquez, E.; Seras-Franzoso, J.; Villaverde, A. Bacterial Inclusion Bodies: Discovering Their Better Half. *Trends in Biochemical Sciences* **2017**, *42* (9), 726–737. https://doi.org/10.1016/j.tibs.2017.01.005.

(29) Wagner, S.; Bader, M. L.; Drew, D.; de Gier, J.-W. Rationalizing Membrane Protein Overexpression. *Trends in Biotechnology* **2006**, *24* (8), 364–371. https://doi.org/10.1016/j.tibtech.2006.06.008.

(30) Kiefer, H. In Vitro Folding of Alpha-Helical Membrane Proteins. *Biochimica et Biophysica Acta (BBA) - Biomembranes* **2003**, *1610* (1), 57–62. https://doi.org/10.1016/S0005-2736(02)00717-4.

(31) Protein Data Bank Statistics https://www.rcsb.org/stats.

(32) Nogales, E.; Scheres, S. H. W. Cryo-EM: A Unique Tool for the Visualization of Macromolecular Complexity. *Molecular Cell* **2015**, *58* (4), 677–689. https://doi.org/10.1016/j.molcel.2015.02.019.

(33) Kanelis, V.; Forman-Kay, J. D.; Kay, L. E. Multidimensional NMR Methods for Protein Structure Determination. *IUBMB Life* **2001**, *52* (6), 291–302. https://doi.org/10.1080/152165401317291147.

(34) Ruschak, A. M.; Kay, L. E. Methyl Groups as Probes of Supra-Molecular Structure, Dynamics and Function. *Journal of Biomolecular NMR* **2009**, *46* (1), 75. https://doi.org/10.1007/s10858-009-9376-1.

(35) Barrett, P. J.; Chen, J.; Cho, M.-K.; Kim, J.-H.; Lu, Z.; Mathew, S.; Peng, D.; Song, Y.; Van Horn, W. D.; Zhuang, T.; Sönnichsen, F. D.; Sanders, C. R. The Quiet Renaissance of Protein Nuclear Magnetic Resonance. *Biochemistry* **2013**, *52* (8), 1303–1320. https://doi.org/10.1021/bi4000436.

(36) Fox, D. A.; Larsson, P.; Lo, R. H.; Kroncke, B. M.; Kasson, P. M.; Columbus, L. Structure of the Neisserial Outer Membrane Protein Opa_{60}: Loop Flexibility Essential to Receptor Recognition and Bacterial Engulfment. *J Am Chem Soc* **2014**, *136* (28), 9938–9946. https://doi.org/10.1021/ja503093y.

(37) Goodman, S. R. *Medical Cell Biology*; Essays in World History; J.B. Lippincott, 1994.

(38) *The PyMOL Molecular Graphics System*; Schrödinger, LLC.

Section 2: Protein Nuclear Magnetic Resonance Spectroscopy

2.1. Essentials of Nuclear Magnetic Resonance Spectroscopy

Nuclear magnetic resonance (NMR) spectroscopy has developed from theory to a robust structural technique. While NMR has been commonly used to determine the structure of small molecules, biomolecular NMR has developed into a valuable tool to determine the structure and dynamics of proteins, nucleic acids and carbohydrates.

To determine the structure of small molecules or proteins, NMR exploits the quantum mechanical property of spin, a source of angular momentum not produced by rotation but is intrinsic to the nuclei of interest. When the nuclei is placed in a strong, external magnetic field, the spins align along the direction of the field. The nuclear spins are perturbed by a weak oscillating field near the resonant, i.e., equivalent, frequency of the spins. This causes the spins to change orientation and precess, move about, the direction of the external magnetic field. This precession about the magnetic field allows for a signal to be detected, which is proportional to the strength of the external field[1–3]. NMR is named so because it subjects **nuclear** spins in an external **magnetic** field to weak oscillating fields at the spin's **resonant** frequency.

The NMR phenomenon results from the Zeeman splitting between energy levels of nuclei with detectable spin. The spin angular momentum of the nucleus presents a magnetic moment on the nucleus, the magnetic moment μ, is given by $\mu = I\hbar\gamma$. Where I = nuclear spin quantum number, γ = the gyromagnetic ratio of the nucleus, ħ = Planck's constatnt[4]. The energy levels of the Zeeman splitting are energetically equivalent without the presence of a magnetic field. However, in the presence of an external magnetic field (B_0) the magnetic moment of the nucleus experiences a torque about the applied B_0 field. Since spin is a quantum mechanical property, the torque experienced is quantized. As a result, the classic mechanical model of a spin undergoing

precession about its axis does not occur, but the model helps to describe the phenomenon, (**Figure 1**)[4]. The nuclear angular momentum, J and the magnetic moment of the nucleus are related by the vector product:

$$\frac{dJ}{dt} = \mu \times B_0 \ (\mathbf{1})$$

$$J = I\hbar \ (\mathbf{2})$$

$$\frac{d\mu}{dt} = \gamma\mu \times B_0, using\ \mu = I\hbar\gamma = \gamma J \ (\mathbf{3})$$

$$\frac{dL}{dt} = r \times mg \ (\mathbf{4})$$

The spin angular momentum experienced by the nucleus is analogous to the angular momentum experienced by a body in a gravitational force[4], equation 4. If J is equated to L, B_0 is equated to g and r x m as analogous to μγ. The energy of this interaction is proportional to μ and B_0, (equation 5)[4].

$$E = -\gamma\hbar\, m_I B_0 \ (\mathbf{5})$$

The external field, B_0, is often illustrated along the Z-axis, therefore spins oriented along the B_0 field are often called "spin-up" and spins opposing the field are "spin-down." Most nuclei commonly used in NMR are I = ½ and only 2I +1 spin states are allowed. Therefore, most nuclei used in NMR only have two states I = ± ½. Spin-up nuclei are at a lower energy state, α, and spin down nuclei are in a higher energy state, β (**Figure 2**)[1,2]. The population of the two energy states between spins I = ± ½ is dictated by the Boltzmann equation $\frac{N_\beta}{N_\alpha} = e^{-\Delta E/k_B T}$.

The energy difference for the two spin states α and β for a I = ± ½ nucleus is given by equation 6 and 7.

$$since\ m_I = 1\ for\ spins\ I = \frac{1}{2},\ E = -\gamma\hbar B_0 \ (\mathbf{6})$$

$$\Delta E = \hbar v \ \therefore v = -\frac{\gamma B_0}{2\pi} \ (in\ Hz)\ or\ \omega = -\gamma B_0 \ (\mathbf{7})$$

The value, ω, in equation 7 is known as the Larmor frequency which is proportional to the gyromagnetic ratio, γ and the external magnetic field, B_0[1,2]. The Larmor frequency is the frequency at which a nucleus precesses about the magnetic field and is unique to each nuclei[1,2].

2.1.a Chemical Shift

While individual nuclei such as ^1H have the same Larmor frequency (equation 7), the signal detected via an NMR experiment can be different between two ^1H nuclei due to differences in the local (through bond and through space) environment of the nuclei. For example, a ^1H in a -CH_3 group and a ^1H in an aromatic ring system will give rise to different NMR signals. The differences observed in the Larmor frequency result in a chemical shift, related by $\omega^0_j = -\gamma B^0(1+\delta_j^{iso})$, where ω^0_j is chemically shifted Larmor frequency, γ is the gyromagnetic ratio of the nucleus, B^0 is the strength of the external magnetic field and δ_j^{iso} is the isotropic chemical shift[2]. The chemical shift of a nucleus arises from differences in the local environment due to the induced current on electrons from the external B_0 field. The induced current in the electron cloud generates an induced magnetic field. The induced field is typically only 10^{-4} of the external field, but is enough to shift the precession frequencies of the spins[2]. Chemical shift typically occurs when the nuclei is either shielded or de-shielded from the magnetic field which is due to the electrons moving about the nucleus to oppose the applied field. The opposition to the applied field changes the effective field experienced by the nucleus which shifts the Larmor frequency of the nucleus[2]. Various functional groups have distinct regions of chemical shifts, for example, protons in aromatic functional groups are typically observed from 6.8 ppm to 8.0 ppm. These characteristic shifts can allow for a “fingerprint” to

observe in an NMR spectrum[1,3,5]. Chemical shifts are also useful for identifying amino acids in proteins. Specific atoms in amino acids have characteristic chemical shift ranges[2,5].

2.1.b Relaxation (T_1 and T_2)

In spectroscopy, relaxation is the term used for a perturbed system to re-establish equilibrium. Without the presence of a magnetic field, all spin orientations are energetically degenerate. However, when the magnetic field is applied the spins orient along the direction of the B_0 field and causes the spins to relax to thermal equilibrium along the B_0 field with the spins being distributed along the direction of the magnetic field. This type of relaxation is known as T_1 or spin lattice relaxation[2]. T_1 relaxation occurs on the order of milliseconds to seconds for nuclei and is dependent on the type of nucleus[2].

Additionally when a RF pulse is applied, the spins in a sample precess about the plane perpendicular to the applied field at their Larmor frequency[2]. During the experiment, the spins in a sample experience slightly different fields resulting in the spins to precess at different frequencies relative to one another. The out-of-phase spins eventually lose net transverse magnetization (magnetization in the xy plane) due to the loss of synchrony between spins via homogenous decay[2]. This is known as T_2 relaxation or spin-spin relaxation. T_2 relaxation is directly proportional to the linewidth observed in a NMR spectrum and can effect spectral quality of the system of interest[6].

For small molecules T_1 and T_2 are of the same order of magnitude – spins perform millions of precession cycles before they begin to lose phase. However, for larger molecules like proteins, T_2 can become very short due to molecular motion of the protein. This change in relaxation can lead to line broadening and change the quality of the spectrum recorded[2].

2.2 Types of NMR Experiments

2.2.a One Dimensional NMR experiments

The basic NMR experiment is a one-dimensional (1D) experiment where one nucleus of interest is observed which includes ^{1}H or ^{13}C. A 1D NMR experiment is broken up into three components: preparation, detection, and Fourier transformation[1,7]. During the preparation stage, the nucleus of interest is excited with a 90 deg radio frequency (RF) pulse along the -x-axis. This excitation causes magnetization of the nuclei to be placed along the +y axis and the spins undergo precession perpendicular to the external magnetic field along, which is along the z-axis, according to their Larmor frequency. During the detection stage, the spins begin to relax due to T1 and T2 and the free induction decay (FID) is recorded. At the end of the experiment, the FID undergoes Fourier transformation to report on the Larmor frequencies, i.e., chemical shifts, and the spectrum is processed and analyzed. The Fourier transformed spectrum is a plot of frequency (chemical shift) and intensity[8].

One-dimensional NMR is useful for the structure determination of small organic molecules, but less so for larger biomolecules. Proteins are polymers of different amino acids linked by repeated peptide bonds. Due to the number of amino acids, 1D NMR spectra of proteins are not interpretable because of chemical shift overlap. For example, amide protons have chemical shifts between of 6-8 ppm[5]. This region for proteins becomes crowded and becomes difficult to distinguish between amide bonds of different amino acids. Proteins also relax faster than small organic molecules. The slow tumbling of proteins leads to line broadening and a lack of distinguishable signal due to T_2 relaxation[2]. Because of these challenges, two- and three-dimensional NMR experiments are more useful for protein structure determination.

2.2.b Two-Dimensional NMR Experiments

While NMR techniques and instrumentation have greatly improved since the first protein NMR spectrum was recorded in 1957, 1D NMR experiments are not sufficient to gain atomic resolution information about protein structure and dynamics[9]. Two-dimensional (2D) NMR experiments are instrumental to determine protein structure and dynamics. The development of two-dimensional (2D) NMR techniques was made possible by Richard Ernst and co-workers. Ernst was awarded the 1991 Nobel Prize in Chemistry "for his development of the methodology of high resolution nuclear magnetic resonance (NMR) spectroscopy."[10]

2.2.b.1 Two-Dimensional Homonuclear NMR

The first notable 2D NMR experiments were homonuclear experiments (^{1}H-^{1}H) due to the higher gyromagnetic ratio, and, thus, increased sensitivity, of ^{1}H nuclei. The first notable experiment was the COSY (Correlated **S**pectroscopy) pulse sequence, first described by Jean Jeener in 1971 and published by Ernst and co-workers five years later[1,11–13]. Like 1D experiments, 2D NMR experiments follow a general framework. The first period is preparation, where the spins are excited and precess about the external field. The following step is the evolution period where spins evolve for a time period, t_1. After the evolution period, there is a mixing period which consists of a pulse or pulses. Finally, the signal is recorded during the detection period as a function of the time variable t_2, (**Figure 3**)[1].

Protein structure determined via 2D experiments was first demonstrated by Wuthrich and co-workers in 1982 of the protein bovine pancreas trypsin inhibitor, BPTI[14]. Their assignment of the protein was made possible using COSY experiments and NOESY (Nuclear Overhauser Effect Spectroscopy) experiments. Unlike COSY experiments, NOESY experiments are used for

through-space or dipolar interactions by taking advantage of the nuclear Overhauser effect (NOE), a phenomenon that describes the dipolar relaxation of a two-spin system after excitation (**Figure 3**)[15,16]. NOE interactions are only detected within ≈4 Å between ^{1}H-^{1}H nuclei[14]. The combination of through-bond interactions in a COSY experiment and the through space interactions in a NOESY experiment allowed for the first protein to be sequentially assigned by NMR, and eventually lead to the first 3D structure a 1985[17,18]. This general framework of combing through bond and through space NMR experiments to solve protein structure is still used by researchers today.

2.2.b.2 Two-Dimensional Heteronuclear NMR

Protein structural determination by NMR was accelerated by the development of 2D heteronuclear NMR experiments. Unlike homonuclear experiments, heteronuclear experiments correlate the chemical shifts of different nuclei. In biological systems these nuclei are commonly ^{1}H, ^{13}C and ^{15}N.

The development of heteronuclear experiments was greatly aided by the development of the INEPT experiment. INEPT, Insensitive Nucleus Enhancement by Polarization Transfer, allows for the indirect detection of less sensitive nuclei. Indirect detection involves modulating the transverse magnetization on the more sensitive nucleus, ^{1}H for biological systems[8,19]. The polarization transfer between ^{1}H nuclei and less sensitive nuclei including ^{13}C or ^{15}N occurs due to through bond coupling between the two nuclei and the application of two simultaneous 180 degree pulses in the corresponding frequencies of the ^{1}H and insensitive nucleus. After the application of these two pulses there is a mixing time that refocuses the chemical shift but not the coupling of the two nuclei[8]. Following the mixing time, 90 degree pulses, one 90 degree pulse at

each nucleus are applied to invert the magnetization of the two nuclei. The FID is recorded after these 90 degree pulses[8,19].

While developed later than other 2D NMR experiments in 1980, the HSQC, Heteronuclear Single Quantum Transfer, experiment has become the preferred 2D experiment for initial protein structure assessment[20]. HSQC pulse sequences use two INEPT transfer steps[21]. The first transfer polarizes the less sensitive ^{13}C or ^{15}N nucleus and the spins are allowed to precess for a time period t. Afterwards, the second INEPT transfer is used to transfer polarization back to the sensitive nucleus. These set of INEPT transfers allow for the two nuclei to report on their respective chemical shifts and be correlated since the coupling is preserved (**Figure 4**)[21]. HSQC experiments can be combined with a NOESY experiment to associate the through bond information from a HSQC experiment and the through space information from a NOESY experiment to assign protein structure[22,23]. The combining of HSQC with other pulse sequences is the foundation for many multi-dimensional NMR experiments.

2.2.c Three-Dimensional NMR Experiments

To aid in the assignment of biomolecules, three-dimensional (3D) or triple resonance NMR experiments were developed by Ad Bax and co-workers[24–28]. Triple resonance experiments are a suite of NMR experiments that correlate ^{1}H, ^{13}C, and ^{15}N nuclei through scalar couplings (through bond interactions) of residues in a protein. Examples of triple resonance experiments are the HNCA and HNCOCA. The HNCA experiment correlates the amide nitrogen of one residue (i) with the alpha carbon of the previous amino acid (i-1). Magnetization transfer begins on the amide proton, is transferred to the amide nitrogen of amino acid i and finally transferred to the alpha carbon of residue i and i-1. HNCA experiments are combined with other experiments

including the HNCOCA experiment which correlates the amide nitrogen of residue i with the carbonyl carbon and alpha carbon of residue i-1 (**Figure 5**)[24,25,27]. Three-dimensional NMR experiments are combined with one another to conduct a "backbone walk" to "walk" down the backbone of the protein of interest to sequentially assign the residues[29]. Three-dimensional NMR experiments have become the preferred method to obtain resonance assignments for proteins.

With assignments, multi-dimensional NOESY spectra can provide through space restraints. In folded proteins, one proton can be surrounded by as many as 15 other protons, which allows for many possible NOE correlations[29]. These correlations are crucial for NMR structural determination and many multi-dimensional NOE experiments have been developed to aid in acquiring NOE restraints[29].

2.2.d Studying Larger Systems in NMR

While the above techniques can be used generally in all NMR experiments, when studying proteins additional steps are taken to study the system of interest. Firstly, isotopic labeling must be used for heteronuclear experiments to correlate ^{1}H with ^{15}N and ^{13}C nuclei. While ^{1}H is the only isotope for hydrogen, ^{13}C and ^{15}N are not naturally abundant as the more common isotopes ^{12}C and ^{14}N. Isotopic enrichment is achieved by over-expressing the protein recombinantly in the cell line of choice, typically *E. coli*, in minimal media along with ^{13}C-glucose, ^{15}N ammonium chloride and/or specifically ^{13}C and ^{15}N labeled amino acids. The organism incorporates the stable isotopes in the proteins synthesized during growth in the enriched medium[29].

In addition, larger biological (greater than 30 kDa) systems usually require the use of pulse sequences optimized to study large systems[30]. Transverse relaxation optimized

spectroscopy (TROSY) experiments, developed by Wutrich and co-workers, are designed to provide better signal for larger systems by taking advantage of the dipole-dipole relaxation and chemical shift anisotropy in these systems[29,31].

2.3 Conclusion

In this thesis, one, two and three- dimensional NMR experiments were used in combination with mutagenesis studies to extend previous structure and function work of the outer membrane protein, Opa_{60}.

Section 2 Figures

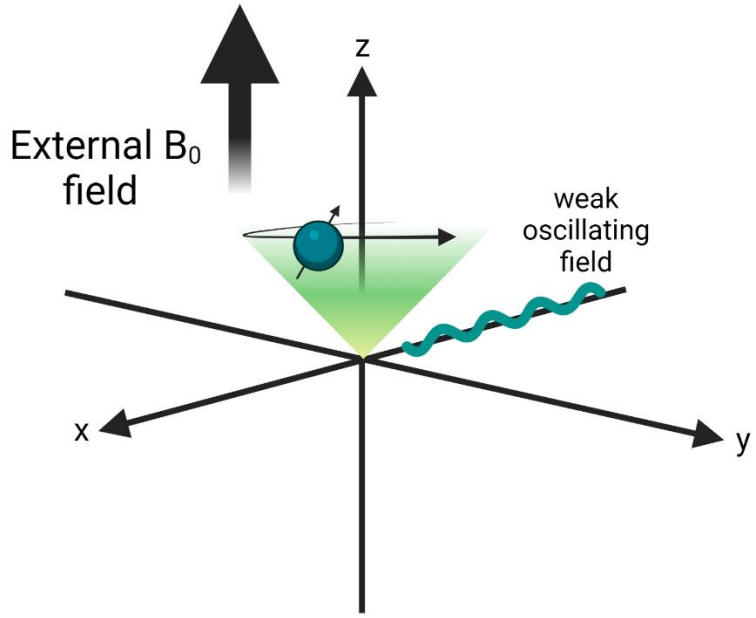


Figure 1. Classic vector model of spin precession about the direction of an external magnetic field. The external magnetic field (B_0) is along the z-axis and the spin is precessing in the xy plane[1,4]. Created with BioRender.com

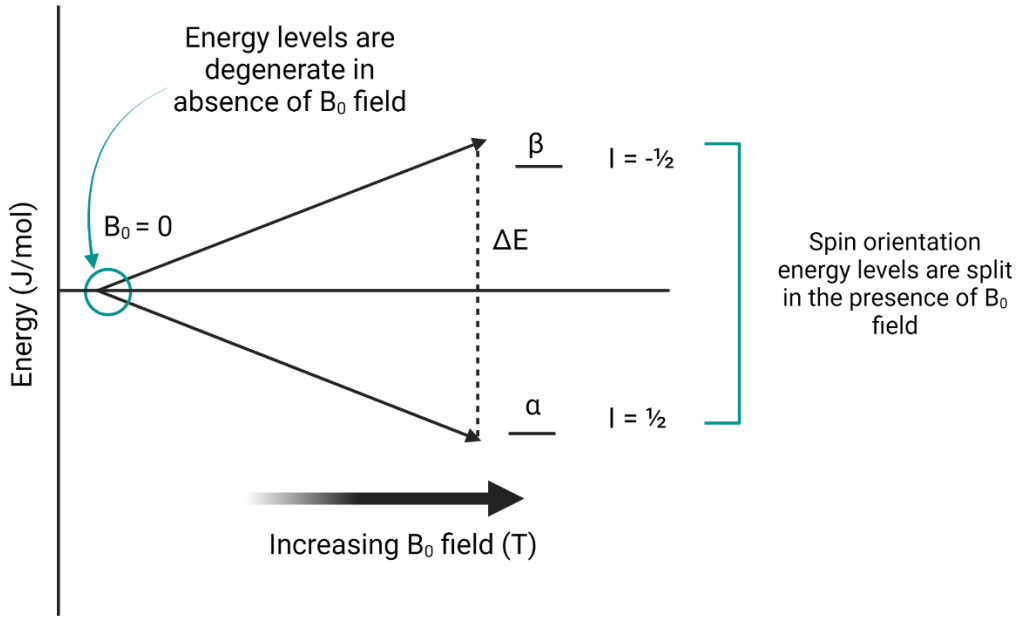


Figure 2. Energy level diagram showing the ΔE of an I = ½ nucleus. Spin orientations are energetically equivalent, i.e. degenerate, without the precsence of an external magnetic field (B_0). The ΔE is proportional to the strength of the B_0 field and the gyromagnetic ratio of the nucleus (γ)[1,2,4]. Created with BioRender.com.

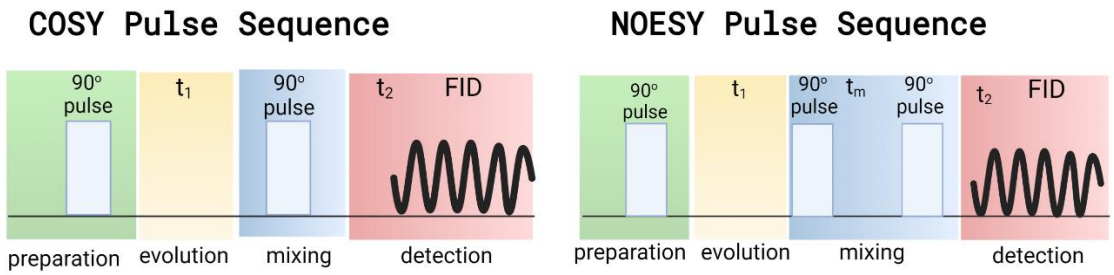


Figure 3. Conventional COSY and NOESY pulse sequences[1,4,14-21]. The general framework of preparation, evolution, mixing and detection for a two-dimensional experiment is shown in green, yellow, blue and red, respectively. Created with BioRender.com.

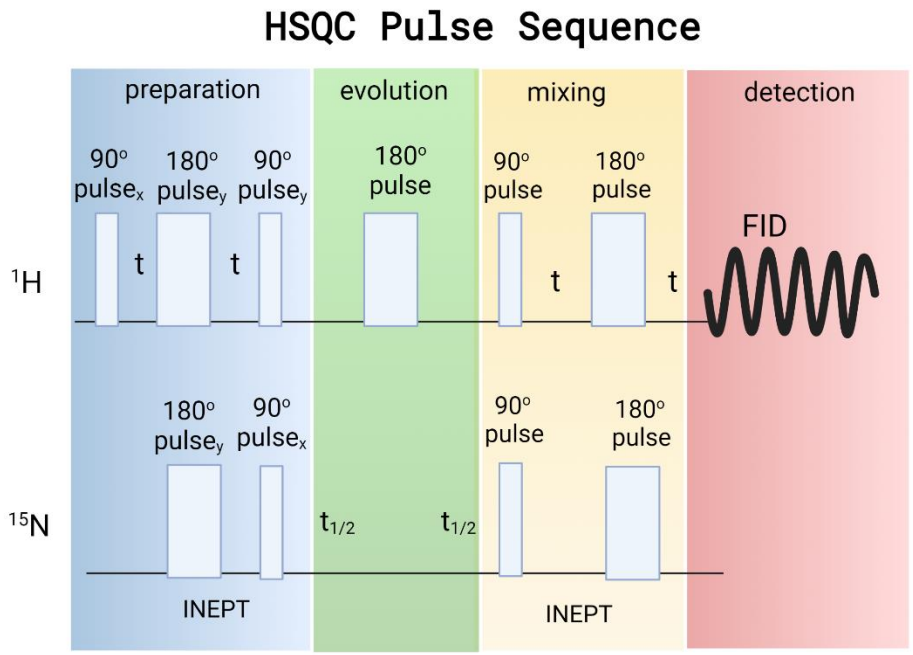


Figure 4. Conventional HSQC pulse sequence[20, 23,24]. The general framework of preparation, evolution, mixing and detection for a two-dimensional experiment is shown in green, yellow, blue and red, respectively. INEPT transfer steps are labeled in the preparation and mixing steps. Created with BioRender.com.

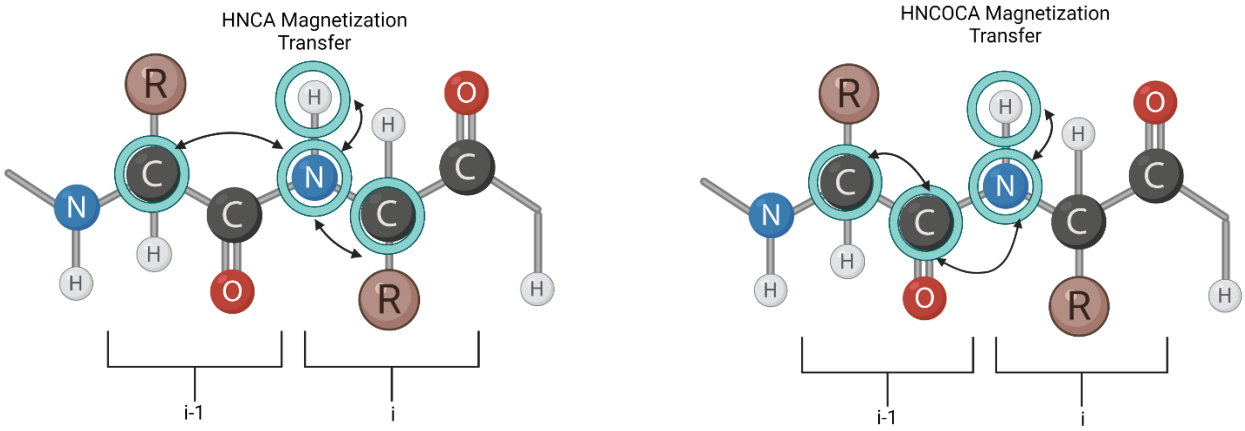


Figure 5. HNCA and HNCOCA magnetization transfer pathways. Involved nuclei are indicated in blue circles and magnetization transfer is shown in double headed arrows. Magnetization begins on the amide proton in both experiments[4, 27-31]. Created with BioRender.com.

References

(1) Keeler, J. *Understanding NMR Spectroscopy*; Wiley, 2011.

(2) Levitt, M. H. *Spin Dynamics: Basics of Nuclear Magnetic Resonance*; Spin Dynamics: Basics of Nuclear Magnetic Resonance; Wiley, 2001.

(3) Hoult, D. I.; Bhakar, B. NMR Signal Reception: Virtual Photons and Coherent Spontaneous Emission. *Concepts in Magnetic Resonance* **1997**, *9* (5), 277–297. https://doi.org/10.1002/(SICI)1099-0534(1997)9:5<277::AID-CMR1>3.0.CO;2-W.

(4) Evans, J. N. S. *Biomolecular NMR Spectroscopy*; Oxford University Press, 1995.

(5) Wishart, D. S.; Bigam, C. G.; Holm, A.; Hodges, R. S.; Sykes, B. D. 1H, 13C and 15N Random Coil NMR Chemical Shifts of the Common Amino Acids. I. Investigations of Nearest-Neighbor Effects. *Journal of Biomolecular NMR* **1995**, *5* (1), 67–81. https://doi.org/10.1007/BF00227471.

(6) Charlier, C.; Cousin, S. F.; Ferrage, F. Protein Dynamics from Nuclear Magnetic Relaxation. *Chem. Soc. Rev.* **2016**, *45* (9), 2410–2422. https://doi.org/10.1039/C5CS00832H.

(7) Benn, R.; Günther, H. Modern Pulse Methods in High-Resolution NMR Spectroscopy. *Angewandte Chemie International Edition in English* **1983**, *22* (5), 350–380. https://doi.org/10.1002/anie.198303501.

(8) Benn, R.; Gunther, H. Modern Pulse Methods in High Resolution NMR Spectroscopy.

(9) Saunders, M.; Wishnia, A.; Kirkwood, J. G. THE NUCLEAR MAGNETIC RESONANCE SPECTRUM OF RIBONUCLEASE1. *J. Am. Chem. Soc.* **1957**, *79* (12), 3289–3290. https://doi.org/10.1021/ja01569a083.

(10) The Nobel Prize in Chemistry 1991 https://www.nobelprize.org/prizes/chemistry/1991/summary/.

(11) Jeener, J. Ampere International Summer School, Basko Pojle, Yugoslavia, 1971.

(12) Aue, W. P.; Bartholdi, E.; Ernst, R. R. Two-dimensional Spectroscopy. Application to Nuclear Magnetic Resonance. *J. Chem. Phys.* **1976**, *64* (5), 2229–2246. https://doi.org/10.1063/1.432450.

(13) Ferretti, J. A.; Ernst, R. R. Interference Effects in NMR Correlation Spectroscopy of Coupled Spin Systems. *J. Chem. Phys.* **1976**, *65* (10), 4283–4293. https://doi.org/10.1063/1.432837.

(14) Wagner, G.; Wüthrich, K. Sequential Resonance Assignments in Protein 1H Nuclear Magnetic Resonance Spectra: Basic Pancreatic Trypsin Inhibitor. *Journal of Molecular Biology* **1982**, *155* (3), 347–366. https://doi.org/10.1016/0022-2836(82)90009-2.

(15) Overhauser, A. W. Polarization of Nuclei in Metals. *Phys. Rev.* **1953**, *92* (2), 411–415. https://doi.org/10.1103/PhysRev.92.411.

(16) Kumar, A.; Ernst, R. R.; Wüthrich, K. 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* **1980**, *95* (1), 1–6. https://doi.org/10.1016/0006-291X(80)90695-6.

(17) Williamson, M. P.; Havel, T. F.; Wüthrich, K. Solution Conformation of Proteinase Inhibitor IIA from Bull Seminal Plasma by 1H Nuclear Magnetic Resonance and

Distance Geometry. *Journal of Molecular Biology* **1985**, *182* (2), 295–315. https://doi.org/10.1016/0022-2836(85)90347-X.

(18) Wüthrich, K. Protein Structure Determination in Solution by NMR Spectroscopy. *Journal of Biological Chemistry* **1990**, *265* (36), 22059–22062. https://doi.org/10.1016/S0021-9258(18)45665-7.

(19) Morris, G. A.; Freeman, R. Enhancement of Nuclear Magnetic Resonance Signals by Polarization Transfer. *J. Am. Chem. Soc.* **1979**, *101* (3), 760–762. https://doi.org/10.1021/ja00497a058.

(20) Bodenhausen, G.; Ruben, D. J. Natural Abundance Nitrogen-15 NMR by Enhanced Heteronuclear Spectroscopy. *Chemical Physics Letters* **1980**, *69* (1), 185–189. https://doi.org/10.1016/0009-2614(80)80041-8.

(21) Mandal, P. K.; Majumdar, A. A Comprehensive Discussion of HSQC and HMQC Pulse Sequences. *Concepts in Magnetic Resonance Part A* **2004**, *20A* (1), 1–23. https://doi.org/10.1002/cmr.a.10095.

(22) Marion, D.; Driscoll, P. C.; Kay, L. E.; Wingfield, P. T.; Bax, A.; Gronenborn, A. M.; Clore, G. M. 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* **1989**, *28* (15), 6150–6156. https://doi.org/10.1021/bi00441a004.

(23) Zuiderweg, E. R. P.; Fesik, S. W. Heteronuclear Three-Dimensional NMR Spectroscopy of the Inflammatory Protein C5a. *Biochemistry* **1989**, *28* (6), 2387–2391. https://doi.org/10.1021/bi00432a008.

(24) Bax, A.; Ikura, M. An Efficient 3D NMR Technique for Correlating the Proton And15N Backbone Amide Resonances with the α-Carbon of the Preceding Residue in Uniformly15N/13C Enriched Proteins. *Journal of Biomolecular NMR* **1991**, *1* (1), 99–104. https://doi.org/10.1007/BF01874573.

(25) Grzesiek, S.; Bax, A. Improved 3D Triple-Resonance NMR Techniques Applied to a 31 KDa Protein. *Journal of Magnetic Resonance (1969)* **1992**, *96* (2), 432–440. https://doi.org/10.1016/0022-2364(92)90099-S.

(26) Grzesiek, S.; Bax, A. An Efficient Experiment for Sequential Backbone Assignment of Medium-Sized Isotopically Enriched Proteins. *Journal of Magnetic Resonance (1969)* **1992**, *99* (1), 201–207. https://doi.org/10.1016/0022-2364(92)90169-8.

(27) Ikura, M.; Kay, L. E.; Bax, A. A Novel Approach for Sequential Assignment of Proton, Carbon-13, and Nitrogen-15 Spectra of Larger Proteins: Heteronuclear Triple-Resonance Three-Dimensional NMR Spectroscopy. Application to Calmodulin. *Biochemistry* **1990**, *29* (19), 4659–4667. https://doi.org/10.1021/bi00471a022.

(28) Kay, L. E.; Ikura, M.; Tschudin, R.; Bax, A. Three-Dimensional Triple-Resonance NMR Spectroscopy of Isotopically Enriched Proteins. *Journal of Magnetic Resonance (1969)* **1990**, *89* (3), 496–514. https://doi.org/10.1016/0022-2364(90)90333-5.

(29) Kanelis, V.; Forman-Kay, J. D.; Kay, L. E. Multidimensional NMR Methods for Protein Structure Determination. *IUBMB Life* **2001**, *52* (6), 291–302. https://doi.org/10.1080/152165401317291147.

(30) Pervushin, K. Impact of Transverse Relaxation Optimized Spectroscopy (TROSY) on NMR as a Technique in Structural Biology. *Quarterly Reviews of Biophysics* **2000**, *33* (2), 161–197. https://doi.org/10.1017/S0033583500003619.

(31) Pervushin Konstantin; Riek Roland; Wider Gerhard; Wüthrich Kurt. Attenuated T2 Relaxation by Mutual Cancellation of Dipole–Dipole Coupling and Chemical Shift Anisotropy Indicates an Avenue to NMR Structures of Very Large Biological Macromolecules in Solution. *Proceedings of the National Academy of Sciences* **1997**, *94* (23), 12366–12371. https://doi.org/10.1073/pnas.94.23.12366.

Section 3: Previous Work and Project Overview

3.1 Previous Work on Opa-CEACAM Specificity

Due to the increased use of antibiotics, drug resistance has been observed among various pathogenic strains of bacteria including strains of *Neisseria*. *Neisseria gonorrhoeae,* the primary pathogen responsible for gonorrhea infections, obtained "super bug" status in 2012 with some strains of *N. gonorrhoeae* showing resistance to all therapeutics introduced to treat gonorrhea[1,2]. In 2016, it was estimated that 87 million new gonococcal infections occurred and if left untreated cervical gonorrhea can impose significant reproductive health complications[3,4].

Pathogenic strains of *Neisseria*, including *N. gonorrhoeae*, bind to carcino-embyronic antigen-like cell adhesion molecules (CEACAM's) to mediate their infection into human cells. CEACAM's are a part of the immunoglobulin superfamily responsible for vital cellular functions such as proliferation, cell adhesion and differentiation[5]. CEACAM's contain a highly conserved, extracellular N-terminal immunoglobulin variable like domain (N-IgV) that is responsible for various interactions[6]. Twelve CEACAM variants have been identified in humans[7,8]. CEACAM N-IgV's have a glycosylated and non-glycosylated face and pathogens like *Neisseria* interact with the non-glycosylated face, specifically of CEACAM's 1, -3, -5 and -6[6,9–11]. *Neisseria* interact with human CEACAM's via opacity associated proteins (Opa), (**Figure 1)**. Opa proteins are classified by the human targets they bind. Opa_{HSPG} bind heparan sulfate proteoglycans (HSPG) whereas the more abundant class Opa_{CEA} bind human CEACAM's[12–14]. Importantly, not all Opa_{CEA} proteins bind all CEACAM's. For example, Opa_{60} binds CEACAM1, -3, -5, and -6, whereas Opa_D only binds CEACAM1 and -3[15].

Previous structural work showed Opa_{60}, an Opa protein expressed in *N. gonorrhoeae,* to be an eight-stranded beta-barrel protein that contains 4 extracellular loops, (**Figure 2)**[16,17]. The

structure of the beta barrel and loop 4 of Opa_{60} are conserved between other Opa proteins However, loops 1, 2 and 3 contain regions of sequence variability which are the semi-variable region (SV) located in loop 1, the hypervariable region 1 (HV1) located in loop 2, and the hypervariable region 2 (HV2) located in loop 3 (**Figure 3**). The HV1 and HV2 regions are primarily responsible for binding to CEACAM's and promoting bacterial colonization of human cells[13,15,16].

For the NMR structure of Opa_{60}, almost all of the beta barrel and loop 4 backbone resonances were assigned to amino acids in the sequence of Opa_{60}. However, the other loop regions of Opa_{60} were more difficult to assign. The loops of Opa_{60} are flexible and highly dynamic with motions measured on the nanosecond timescale t = 2-5 ns[16]. In addition microsecond backbone dynamics increased broadening with only 27% of the loop backbone resonances being assigned, compared to 97% of the beta barrel backbone resonances (**Figure 4**)[16,17]. Most of the loop backbone resonances were in the HV2 region of loop 3. Since, there is limited structural information of the loops in Opa_{60} much is not known about the specifics for Opa-CEACAM binding specificity.

3.2 Possible Route for Conformational Selection

Since there is no underlying sequence motif that indicates how Opa proteins can bind specifically to human CEACAM's, we aimed to investigate the prevalence of particular residues in the HV2 sequence, despite their locations not being conserved. Specifically, HV2 has a set of proline residues in most HV2 sequences. Previous work from Andreotti and coworkers have shown that proline isomerization can be a route for protein conformational selection (**Figures 5a and 5b**)[18,19]. They observed extra resonances in the spectrum of the Src homology 2 domain

(SH2) of interleukin kinase 2 (Itk2). They concluded that isomerization of the Asn286-Pro287 bond was occurring since mutation to glycine residues removed the extra cross peaks[19]. The isomerization of the Asn286-Pro287 bond in the CD loop of Itk-SH2 lead to structural changes extending from the isomerized bond[19]. Populations of Asn286-Pro287 cis/trans isomers were measured by NMR peak integration to be in almost equal populations, 40% cis and 60% trans, in a 1H-^{15}N HSQC of Itk-SH2[18–20]. For small peptides the population of cis proline bonds is 10-40%[21]. SH2 domains canonically bind phosphorylated tyrosine residues and SH2-Itk preferentially bound phosphotyrosine when the Asn286-Pro287 bond was in the trans position[18,20]. The preferential binding led the researchers to conclude that proline isomerization can act as a molecular switch to recognize ligands or targets[18–20].

Of the loop assignments in the structure of Opa_{60}, 21 out of 42 residues in the HV2 region of loop 3 were assigned[16]. In the assigned HV2 region, there are three proline residues, P161, P165 and P176, of which two are near two alanine residues, Ala164 and Ala168. In the previously assigned ^{1}H-^{15}N HSQC wild-type (WT) spectrum of Opa_{60}, two peaks of Ala164 and Ala168 were observed (**Figures 6a and 6b**)[16,17]. This could indicate the presence of neighboring proline residues sampling different conformations. The conformational selection could be a mechanism that promotes binding of Opa proteins to human CEACAM receptors.

This project seeks to understand the physical and biological role of proline isomerization as a route for conformational selection of the HV2 region in Opa_{60} to promote binding to human CEACAM receptors. Protein NMR was used to assess isomerization of proline to glycine (P to G) mutants of WT Opa_{60}. To obtain concentrations required for NMR, P161G, P165G and P176G mutations were individually introduced into the WT Opa_{60} background. These mutations are in constructs that express to inclusion bodies allowing for sufficient protein concentrations

necessary for biomolecular NMR studies. We hypothesize that replacing the proline residues, P161, P165 and P1976 with glycine, we will only observe one ^{1}H-^{15}N resonance for either Ala164 or Ala168 as the prolines were mutated and will not be sampling different isomers. Glycine was chosen since its ^{15}N chemical shift (109.1 ppm) is notably lower than the other amino acids. This would facilitate efficient assignment since the cross peak would be observed outside the chemical shift region typically observed for amide nitrogen's (120ppm) in a ^{1}H-^{15}N HSQC[22].

To understand the biological role of proline isomerization, membrane expressing P to G mutants were generated from WT membrane expressing Opa_{60} constructs. The goal of the membrane expressing mutants was to determine if proline isomerization plays a role in binding to CEACAM receptors. The selectivity for human CEACAM receptors was examined by *E. coli* whole cell pull-down assays. Both WT Opa_{60} and P to G mutants were expressed in *E. coli*, incubated with GST-hCEACAM1 N-terminal domain fusion protein (GST-NCCM1), and probed for binding by western blotting. Our hypothesis is the P to G mutants will have a marked decline in the amount of GST-CEACAM1 bound compared to the wild-type.

Section 3 Figures

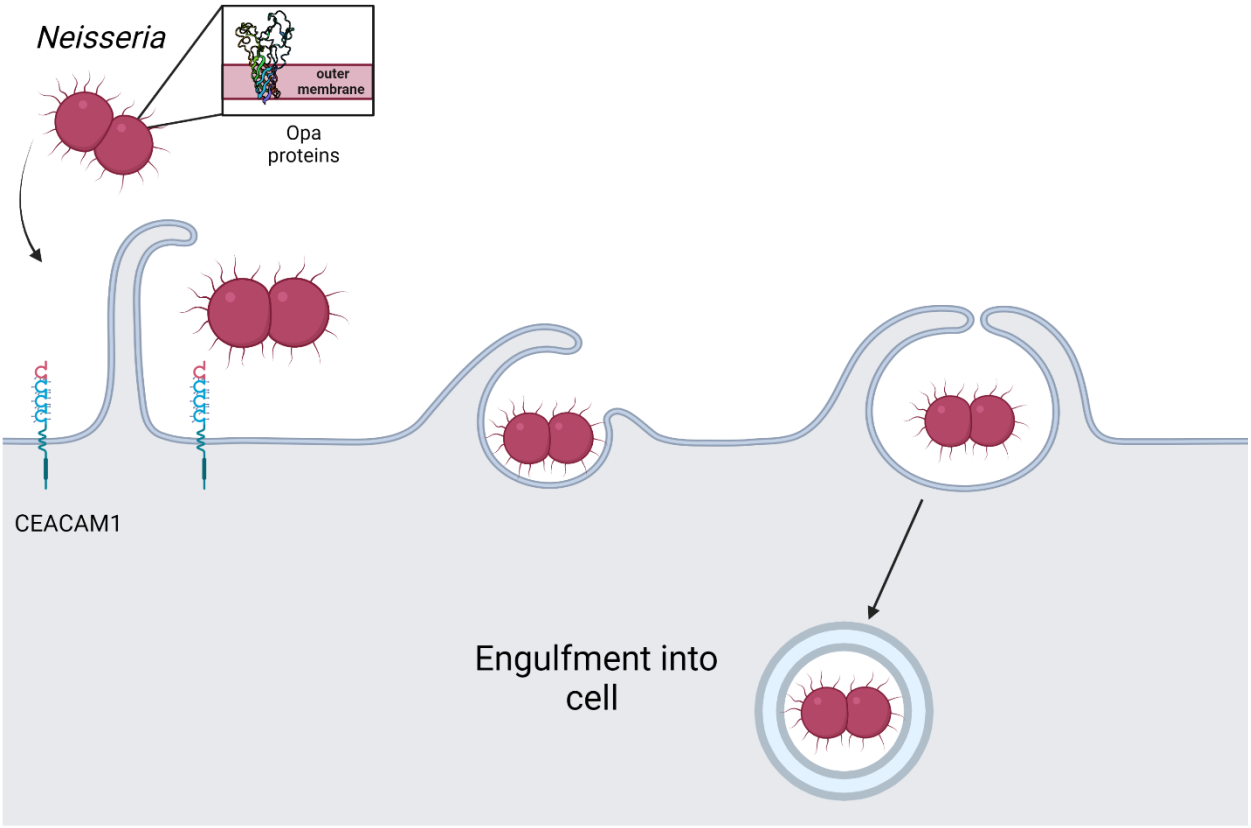


Figure 1. *Neisseria* initiates engulfment into the host cell via the binding of opacity-associated (Opa) proteins to human carcinoembryonic antigen-related cell adhesion molecules (CEACAMs)[4,11]. Created with BioRender.com.

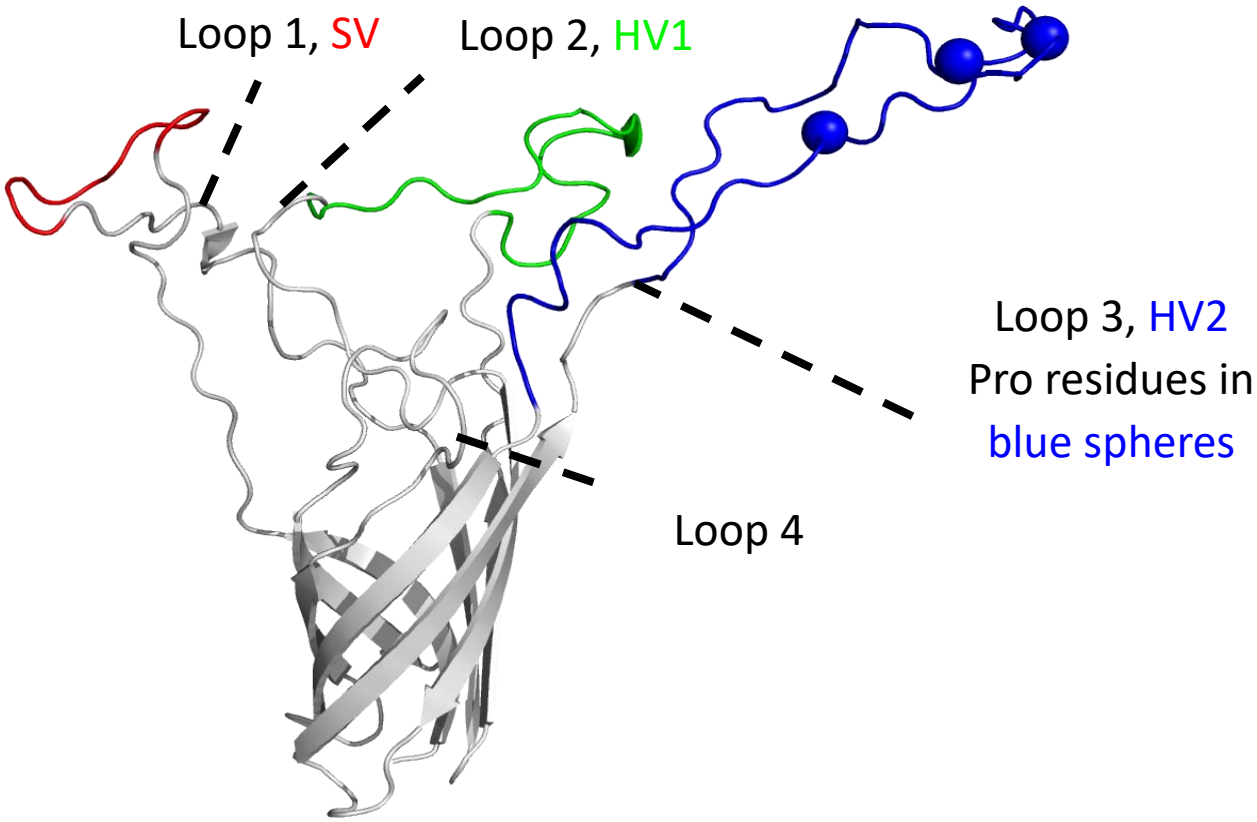


Figure 2. Opa proteins consist of an eight stranded β-barrel and four, highly dynamic extracellular loops that mediate Opa-receptor interactions. Within these four loops, there are three regions of sequence variability: a semi-variable (SV, red) region in loop 1, hypervariable region 1 (HV1, green) in loop 2, and a second hypervariable region (HV2, blue) in loop 3. The fourth loop is conserved across Opa variants. The β-barrel and loop 4 are shown in gray[16,17]. Mutated proline sites represented as blue spheres. PDB 2MAF and figure created with PyMol[23].

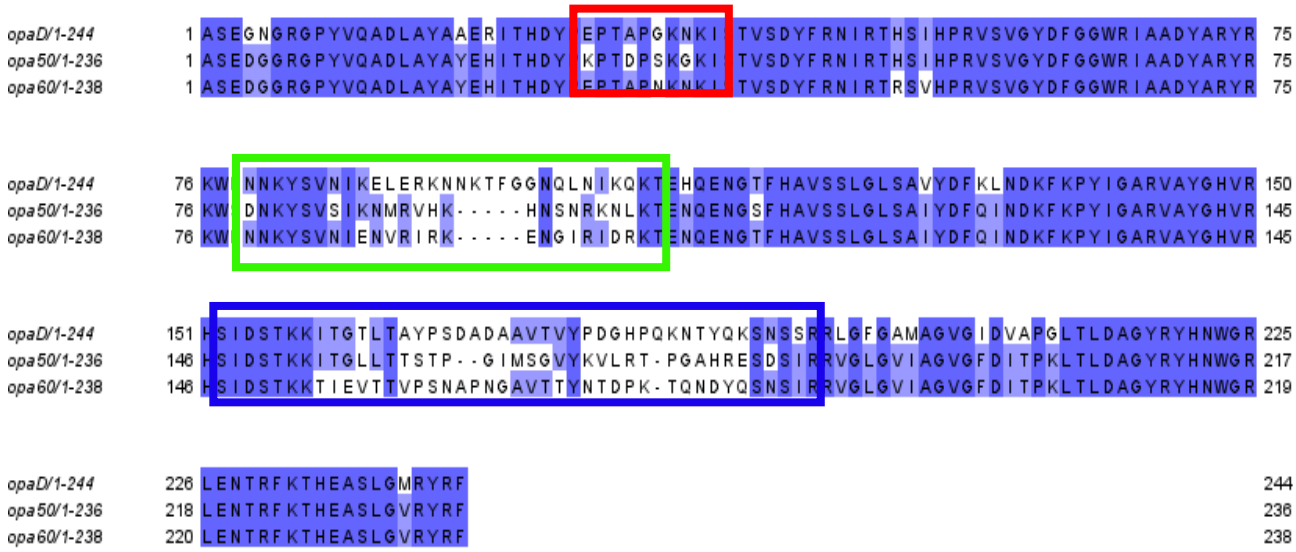


Figure 3. Sequence alignment of three Opa variants (OpaD, Opa_{60}, Opa_{50}) demonstrates the sequence variability between SV (red), HV1 (green), and HV2 (blue) regions. Overall, 26 different SV sequences have been identified as well as 96 HV1 sequences and 127 HV2 sequences across the 338 sequenced and distinct opa alleles[15,16]. Created with Jalview.[24]

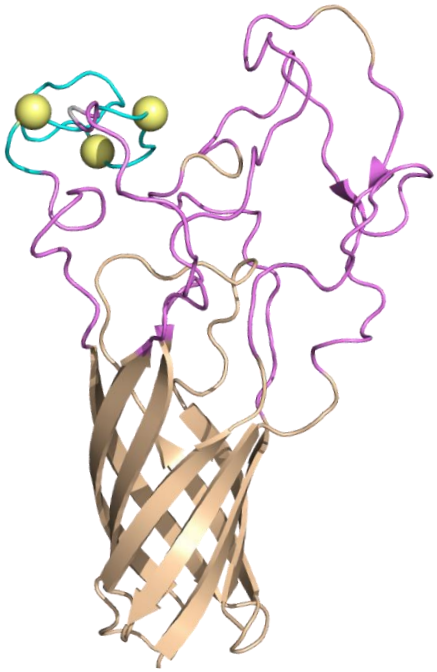
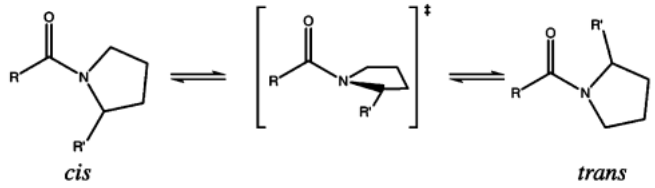


Figure 4. NMR structure of Opa_{60} showing the assigned (beige and cyan) and unassigned residues (violet)[16,17]. Assigned region in HV2 is shown in cyan with mutated prolines shown as pale-yellow spheres. PDB 2MAF and figure created with PyMol[23].

5a.



5b.

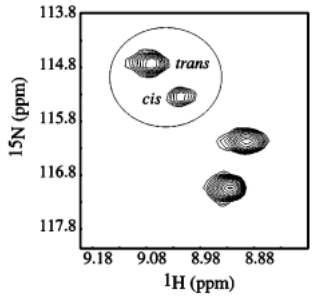


Figure 5a. Cis- and trans-isomers of proline. Proline isomerization has been demonstrated to be a means of conformational selection for some proteins to recognize their substrates[18,19]. **5b**. ^{1}H-^{15}N HSQC spectrum showing proline isomerization[18–20]. Reprinted (adapted) with permission from: Andreotti, A. H. Native State Proline Isomerization: An Intrinsic Molecular Switch. *Biochemistry* **2003**, *42* (32), 9515–9524. https://doi.org/10.1021/bi0350710.. Copyright 2003 American Chemical Society.

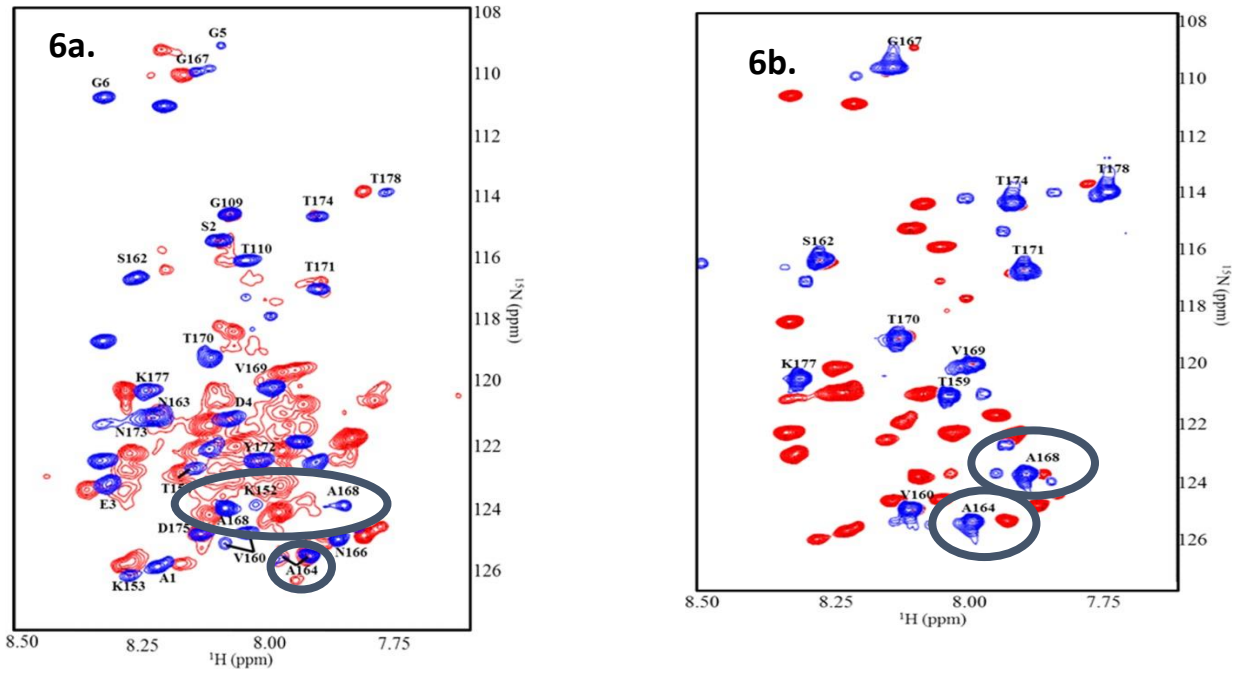


Figure 6. (a) ^{1}H-^{15}N HSQC of Opa_{60} at 40 °C (red) and 10 °C blue.[4] **(b)** ^{1}H-^{15}N HSQC of Opa_{60} at 10 °C (red) and synthesized peptide T*-V*-P-S*-N-A*-P-N-G*-A*-V*-T*-T*-Y-N-T*-D-P-K*-T* (shown in blue; * denotes ^{15}N isotope labeling).Two cross peaks are assigned for A164 and A168, both residues are near proline residues. Isomerization of neighboring prolines could be a contributing factor for Opa-CEACAM specificity[17]. Spectra was recorded at 10 ℃ to allow for a slower exchange of loops in Opa_{60}. Reprinted (adapted) with permission from: Fox, D. A.; Columbus, L. Solution NMR Resonance Assignment Strategies for β-Barrel Membrane Proteins. *Protein Sci* **2013**, *22* (8), 1133–1140. https://doi.org/10.1002/pro.229.

References

(1) Unemo Magnus; Shafer William M. Antimicrobial Resistance in Neisseria Gonorrhoeae in the 21st Century: Past, Evolution, and Future. *Clinical Microbiology Reviews* **2014**, *27* (3), 587–613. https://doi.org/10.1128/CMR.00010-14.

(2) Suay-García, B.; Pérez-Gracia, M. T. Future Prospects for Neisseria Gonorrhoeae Treatment. *Antibiotics (Basel)* **2018**, *7* (2), 49. https://doi.org/10.3390/antibiotics7020049.

(3) Kirkcaldy, R. D.; Weston, E.; Segurado, A. C.; Hughes, G. Epidemiology of Gonorrhoea: A Global Perspective. *Sex Health* **2019**, *16* (5), 401–411. https://doi.org/10.1071/SH19061.

(4) Unemo, M.; Seifert, H. S.; Hook, E. W.; Hawkes, S.; Ndowa, F.; Dillon, J.-A. R. Gonorrhoea. *Nature Reviews Disease Primers* **2019**, *5* (1), 79. https://doi.org/10.1038/s41572-019-0128-6.

(5) Martin, J. N.; Ball, L. M.; Solomon, T. L.; Dewald, A. H.; Criss, A. K.; Columbus, L. Neisserial Opa Protein–CEACAM Interactions: Competition for Receptors as a Means of Bacterial Invasion and Pathogenesis. *Biochemistry* **2016**, *55* (31), 4286–4294. https://doi.org/10.1021/acs.biochem.6b00124.

(6) Fedarovich, A.; Tomberg, J.; Nicholas, R. A.; Davies, C. 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* **2006**, *62* (9), 971–979.

(7) Zebhauser, R.; Kammerer, R.; Eisenried, A.; McLellan, A.; Moore, T.; Zimmermann, W. Identification of a Novel Group of Evolutionarily Conserved Members within the Rapidly Diverging Murine Cea Family. *Genomics* **2005**, *86* (5), 566–580.

(8) Kuespert, K.; Pils, S.; Hauck, C. R. CEACAMs: Their Role in Physiology and Pathophysiology. *Current opinion in cell biology* **2006**, *18* (5), 565–571.

(9) McCaw, S. E.; Liao, E. H.; Gray-Owen, S. D. Engulfment of Neisseria Gonorrhoeae: Revealing Distinct Processes of Bacterial Entry by Individual Carcinoembryonic Antigen-Related Cellular Adhesion Molecule Family Receptors. *Infection and immunity* **2004**, *72* (5), 2742–2752.

(10) Belcher Dufrisne, M.; Swope, N.; Kieber, M.; Yang, J.-Y.; Han, J.; Li, J.; Moremen, K. W.; Prestegard, J. H.; Columbus, L. Human CEACAM1 N-Domain Dimerization Is Independent from Glycan Modifications. *Structure* **2022**. https://doi.org/10.1016/j.str.2022.02.003.

(11) Gray-Owen, S. D.; Blumberg, R. S. CEACAM1: Contact-Dependent Control of Immunity. *Nature Reviews Immunology* **2006**, *6* (6), 433–446. https://doi.org/10.1038/nri1864.

(12) Dehio, C.; Gray-Owen, S. D.; Meyer, T. F. The Role of Neisserial Opa Proteins in Interactions with Host Cells. *Trends in microbiology* **1998**, *6* (12), 489–495.

(13) Sadarangani, M.; Pollard, A. J.; Gray-Owen, S. D. Opa Proteins and CEACAMs: Pathways of Immune Engagement for Pathogenic Neisseria. *FEMS microbiology reviews* **2011**, *35* (3), 498–514.

(14) Billker, O.; Popp, A.; Gray-Owen, S. D.; Meyer, T. F. The Structural Basis of CEACAM-Receptor Targeting by Neisserial Opa Proteins. *Trends in microbiology* **2000**, *8* (6), 258–260.

(15) Bos Martine P.; Kao David; Hogan Daniel M.; Grant Christopher C. R.; Belland Robert J. Carcinoembryonic Antigen Family Receptor Recognition by Gonococcal Opa Proteins Requires Distinct Combinations of Hypervariable Opa Protein Domains. *Infection and Immunity* **2002**, *70* (4), 1715–1723. https://doi.org/10.1128/IAI.70.4.1715-1723.2002.

(16) Fox, D. A.; Larsson, P.; Lo, R. H.; Kroncke, B. M.; Kasson, P. M.; Columbus, L. Structure of the Neisserial Outer Membrane Protein Opa_{60}: Loop Flexibility Essential to Receptor Recognition and Bacterial Engulfment. *J Am Chem Soc* **2014**, *136* (28), 9938–9946. https://doi.org/10.1021/ja503093y.

(17) Fox, D. A.; Columbus, L. Solution NMR Resonance Assignment Strategies for β-Barrel Membrane Proteins. *Protein Sci* **2013**, *22* (8), 1133–1140. https://doi.org/10.1002/pro.2291.

(18) Mallis, R. J.; Brazin, K. N.; Fulton, D. B.; Andreotti, A. H. Structural Characterization of a Proline-Driven Conformational Switch within the Itk SH2 Domain. *Nature Structural Biology* **2002**, *9* (12), 900–905. https://doi.org/10.1038/nsb864.

(19) Andreotti, A. H. Native State Proline Isomerization: An Intrinsic Molecular Switch. *Biochemistry* **2003**, *42* (32), 9515–9524. https://doi.org/10.1021/bi0350710.

(20) Brazin Kristine N.; Mallis Robert J.; Fulton D. Bruce; Andreotti Amy H. Regulation of the Tyrosine Kinase Itk by the Peptidyl-Prolyl Isomerase Cyclophilin A. *Proceedings of the National Academy of Sciences* **2002**, *99* (4), 1899–1904. https://doi.org/10.1073/pnas.042529199.

(21) Cheng, H.; Bovey, F. Cis-trans Equilibrium and Kinetic Studies of Acetyl-L-proline and Glycyl-L-proline. *Biopolymers: Original Research on Biomolecules* **1977**, *16* (7), 1465–1472.

(22) Statistics Calculated for All Chemical Shifts from Atoms in the 20 Common Amino Acids' https://bmrb.io/ref_info/csstats.php.

(23) *The PyMOL Molecular Graphics System*; Schrödinger, LLC.

(24) Waterhouse, A. M.; Procter, J. B.; Martin, D. M. A.; Clamp, M.; Barton, G. J. Jalview Version 2—a Multiple Sequence Alignment Editor and Analysis Workbench. *Bioinformatics* **2009**, *25* (9), 1189–1191. https://doi.org/10.1093/bioinformatics/btp033.

Section 4: Results and Future Work

4.1 Results

Uniformly labeled ^{15}N, ^{13}C Opa_{60} wild-type (WT) and proline to glycine (P to G) mutants were grown in minimal media expressed and purified from inclusion bodies to perform NMR studies. The molecular weight of WT Opa_{60} is 29.37 kDa and is observed at this molecular weight before refolding. The P to G mutants are observed at the same molecular weight as the unfolded wild-type indicating no changes in observed molecular weight (**Figure 1)**. The P to G mutants also expressed to similar levels as the wild-type. The amount of protein for the WT and P to G mutants ranged from 25-30 mg per liter of growth media as measured using the theoretical extinction coefficient at A_{280}.

Opa_{60} WT and glycine mutants were refolded in dodecylphosphocholine (FC-12) micelles for five days at room temperature. Refolding of the wild-type and mutant proteins was assessed by SDS-PAGE gel electrophoresis since the folded protein migrates at a lower observed molecular weight than the unfolded protein (**Figure 2).** The migration indicates the P to G mutants are refolding when compared to previous studies and the wild-type Opa_{60} protein[1,2].

After purification and refolding, the Opa_{60} samples were dialyzed to remove imidazole and concentrated to approximately 500 µM. The FC-12 monomer and micelle concentration were measured by 1D ^{1}H NMR. The monomer concentration ranged from 120-160 mM and the micelle concentration ranged from 2-3 mM (**Figure 3**).

^{1}H, ^{15}N TROSY- HSQC spectra of P182G (orange), P186G (blue) and P197G (green) overlayed on the WT Opa_{60} spectrum (red) (**Figures 4-6**, respectively). The HSQC spectra were recorded at 10 ˚C to reduce line broadening in the loop regions of Opa_{60}. The backbone resonance assignment of the ^{1}H,^{15}N HSQC spectra was completed using 3D HNCA and

HNCOCA experiments. Cross peaks in the HNCA and HNCOCA were used to perform a backbone "walk" to assign the resonances in the ^{1}H, ^{15}N HSQC spectra for the wild-type Opa_{60} protein (**Figure 8**). Most of the cross peaks in the glycine mutant spectra overlay consistently with the wild-type HSQC. However, there are some peaks that are shifted, likely due to the added flexibility of the protein backbone because of the glycine mutation. Some amino acids near the P to G mutation are not observed. For example, the V160 cross peaks are not observed in the P161G spectrum and the Ala164 cross peaks are not observed. Cross peaks in the P182G spectrum may show a shift of cross peaks near Ala164 (**Figure 7)**. Further 3D experiments are needed to assign the P182G spectrum to determine the identity of the shifted peak or determine if one of the peak populations disappeared due to the lack of a proline residue.

E. coli whole cell pull down assays were performed on PhoE-Opa_{60} membrane expressing constructs of the WT Opa_{60} and P to G mutants. Each culture of *E. coli* was inoculated with GST-hCEACAM1 N-terminal domain fusion protein (GST-NCCM1), incubated, centrifuged, and probed for analysis by western blotting (**Figures 9 and 10**). More GST-NCCM1 was bound to WT Opa_{60} compared to the empty vector in the pull down assays (**Figure 9**). Of the membrane expressing constructs, only the WT PhoE-Opa_{60} construct was expected to not perturb GST-NCCM1 binding. However, all P to G PhoE-Opa_{60} mutant constructs bound GST-NCCM1 (**Figure 10**). Further experiments are needed to understand the role of proline isomerization on Opa-CEACAM binding. Specifically, to understand if binding is perturbed in the P to G Opa_{60} mutants.

4.2 Discussion and Future Directions

To understand the role of proline isomerization in Opa_{60}, proline to glycine (P to G) mutants were created to observe isomerization and the biological impact of proline isomerization on Opa-CEACAM binding. The mutants expressed and refolded similarly to the wild-type Opa_{60} protein. The ^{1}H, ^{15}N HSQC spectra of the P to G mutants overlayed consistently with the wild-type Opa_{60} spectrum, with exceptions that are likely the result of added flexibility to the protein backbone due to the glycine mutation**.**

Of the mutants, the P182G mutant showed possible signs of isomerization. One of the Ala164 cross peaks is not observed in the overlayed spectrum with the wild-type. However, further experiments are needed to confirm if the residue is isomerizing, or the cross peak was shifted because of the glycine mutation. Further 3D NMR experiments are also needed to confirm the identity of the shifted resonances in the 3D spectra for the P to G mutants.

E. coli whole cell pull down assays were performed on WT PhoE-Opa_{60} and PhoE-Opa_{60} P to G mutants. All constructs including the P to G mutants bound GST-NCCM1.

Possible future directions of this work would be to perform relaxation experiments of the P to G mutants by protein NMR. All of the mutants experienced the disappearance of at least one cross peak when compared to the wild-type Opa_{60} protein. This can indicate a change in backbone dynamics due to the glycine mutation. In addition, residues next to the proline residues could be mutated to tryptophan residues. Previous work by Wuthrich and coworkers, showed that the rate of proline isomerization was slowed when phenylalanine, tryptophan, tyrosine residues were bonded next to a proline in peptides[3]. The possible benefit of this work over the current work is the preservation of the proline residues. The proline residues would still be present with the addition of the tryptophan residues and the alanine residues could still be used as

reporters of proline isomerization. The tryptophan residue could also be efficiently assigned since the indole nitrogen and hydrogen of tryptophan have distinctive chemical shifts, 129 and 10 ppm respectively[4].

Section 4 Figures

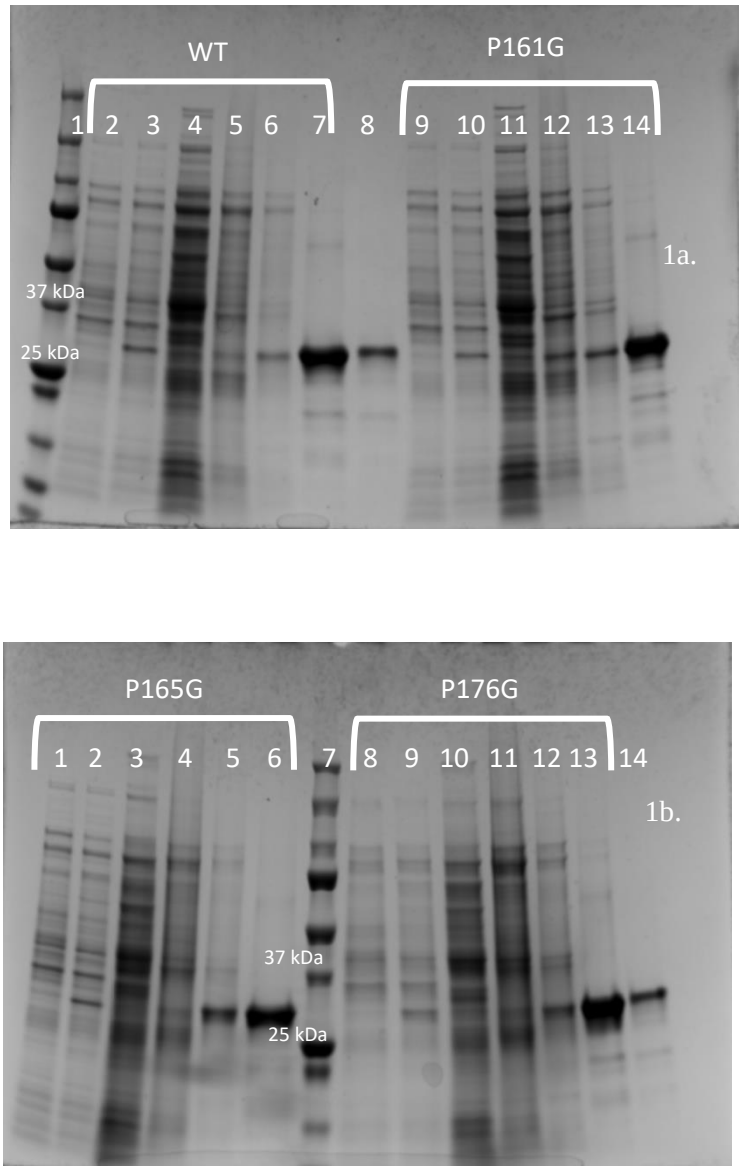


Figure 1. SDS-PAGE analysis of Opa_{60}, WT and proline to glycine mutants, P161G, P165G and P176G. **a.)** Gel of WT and P161G. 1- ladder, 2 – before induction with IPTG, 3- after induction with IPTG, 4 - supernatant after first centrifugation at 12000g for 30 minutes at 16 ˚C, 5- flow through after loading on Co^{2+} immobilized metal affinity column, 6 – wash fraction 7- elution fraction, 8 – unfolded Opa_{60} control. Lanes 9-14 are the steps as lanes 2-7. **b.)** Gel of P165G and P176G. 1 – before induction with IPTG, 2- after induction with IPTG, 3 - supernatant after first centrifugation at 12000g for 30 minutes at 16 ˚C, 4- flow through after loading on Co^{2+} immobilized metal affinity column, 5 – wash fraction 6- elution fraction, 7 – ladder and lane 14 is the unfolded Opa_{60} control. Lanes 8-13 after the same as lanes 1-6.

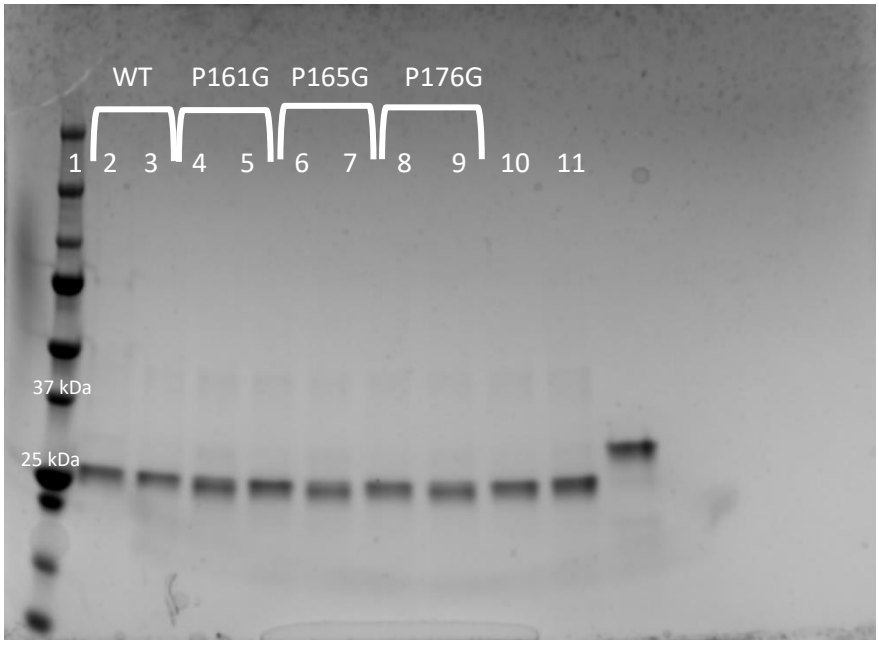
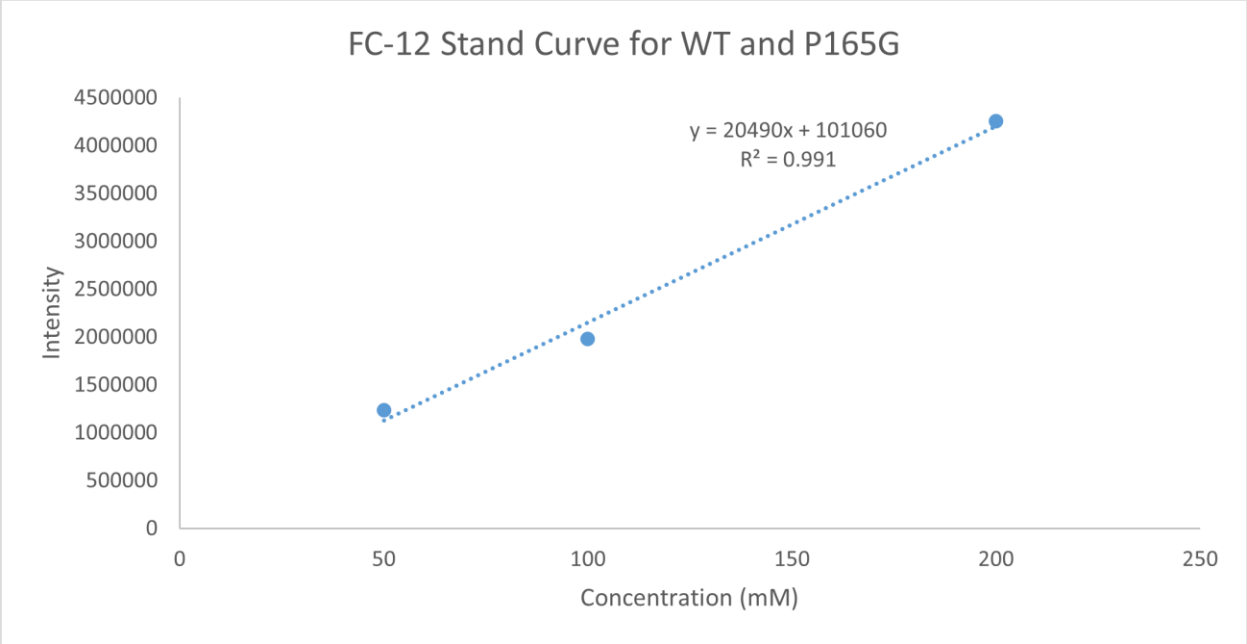


Figure 2. SDS-PAGE analysis of refolded WT Opa_{60} and proline to glycine mutants. Refolded Opa_{60} samples in FC-12 micelles migrate farther than unfolded samples due to incomplete denaturation by SDS. 1 –ladder, 2 – WT Opa_{60} **day 1**, 3 – WT Opa_{60} **day 5**, 4 – P161G Opa_{60} **day 1**, 5 – P161G Opa_{60} **day 5**, 6 – P165G Opa_{60} **day 1**, 7 – P165G Opa_{60} **day 5**, 8 – P176G Opa_{60} **day 1**, 9 – P176G Opa_{60} **day 5**, 10 – refolded Opa_{60} control and 11 – unfolded Opa_{60} control.



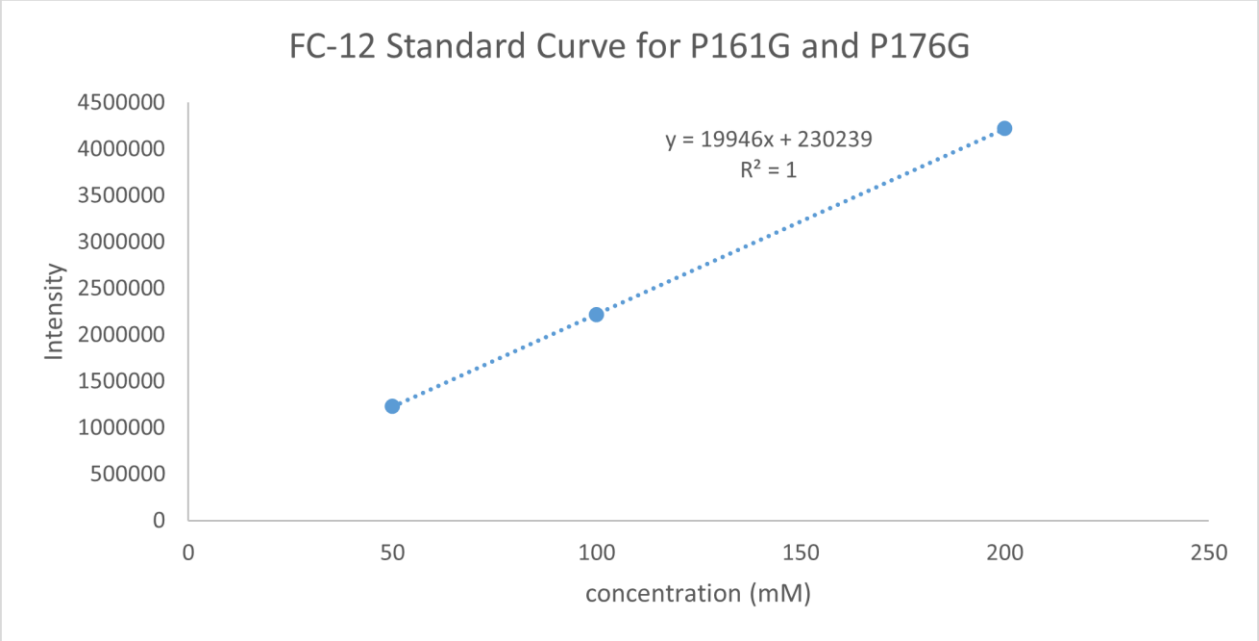


Figure 3. Linear plots of 1D ^{1}H intensity vs. FC-12 concentration. The regression equation was used to calculate the FC-12 monomer concentration in the WT and P to G samples.

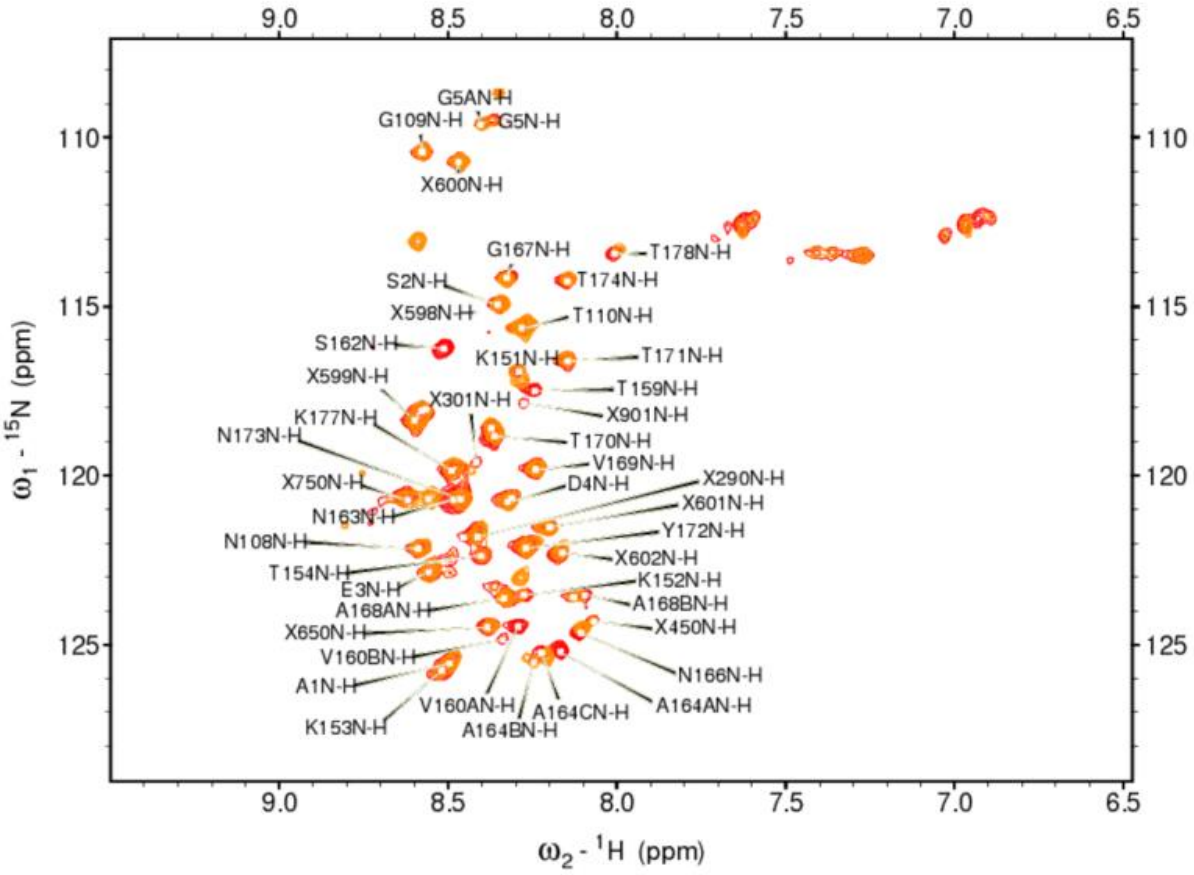


Figure 4. ^{1}H-^{15}N TROSY HSQC of WT Opa_{60} (red) vs. P161G Opa_{60} (orange) at 10 ˚C. TROSY HSQC spectra were recorded on a Bruker Avance 800 MHz spectrometer and processed in Bruker Topsin 3.2. and analyzed in NMRFAM-SPARKY[5].

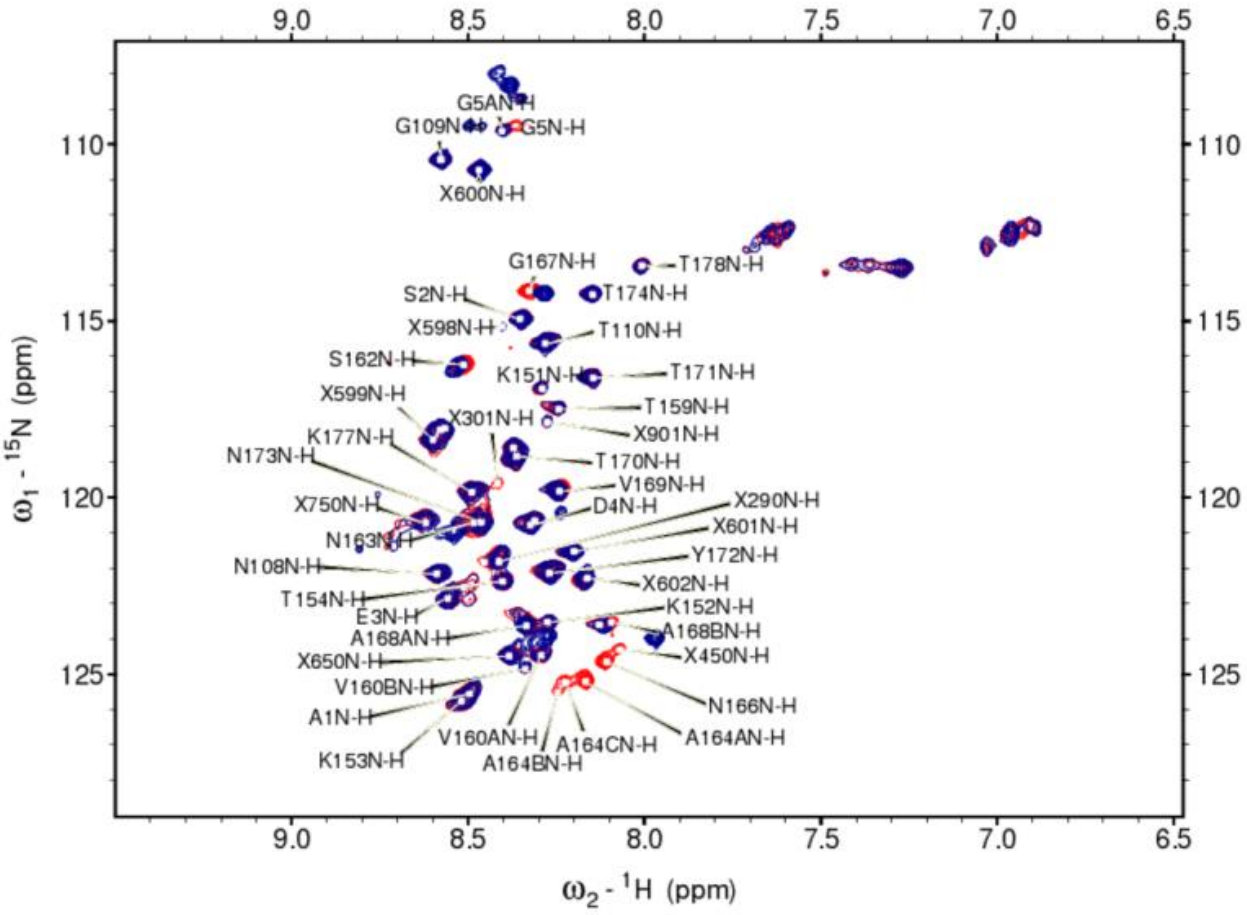


Figure 5. ^{1}H-^{15}N TROSY HSQC of WT Opa$_{60}$ (red) vs. P165G Opa$_{60}$ (blue) at 10 ˚C. TROSY HSQC spectra were recorded on a Bruker Avance 800 MHz spectrometer and processed in Bruker Topsin 3.2. and were analyzed in NMRFAM-SPARKY[5].

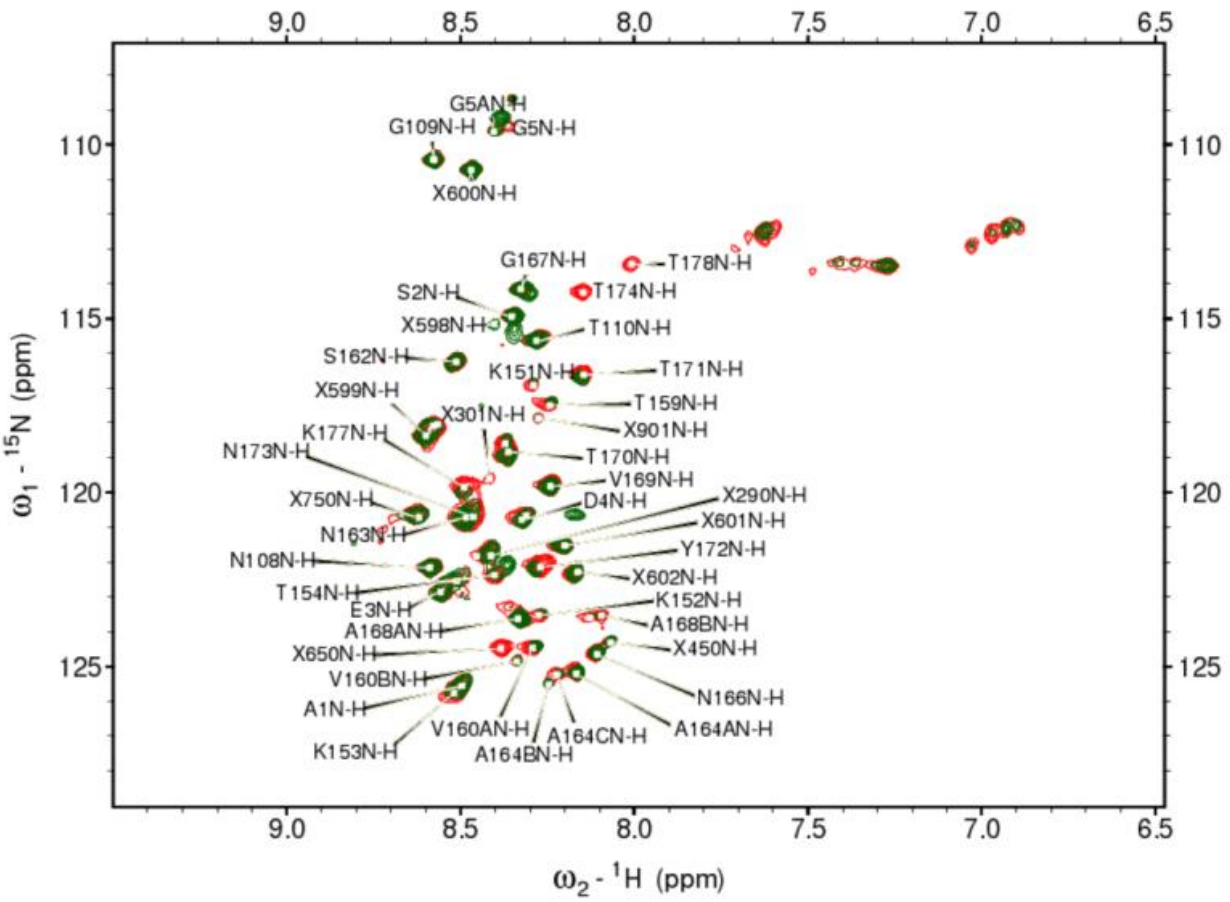


Figure 6. ^{1}H-^{15}N TROSY HSQC of WT Opa$_{60}$ (red) vs. P176G Opa$_{60}$ (green) at 10 ˚C. TROSY HSQC spectra were recorded on a Bruker Avance 800 MHz spectrometer and processed in Bruker Topsin 3.2. and analyzed in NMRFAM-SPARKY[5].

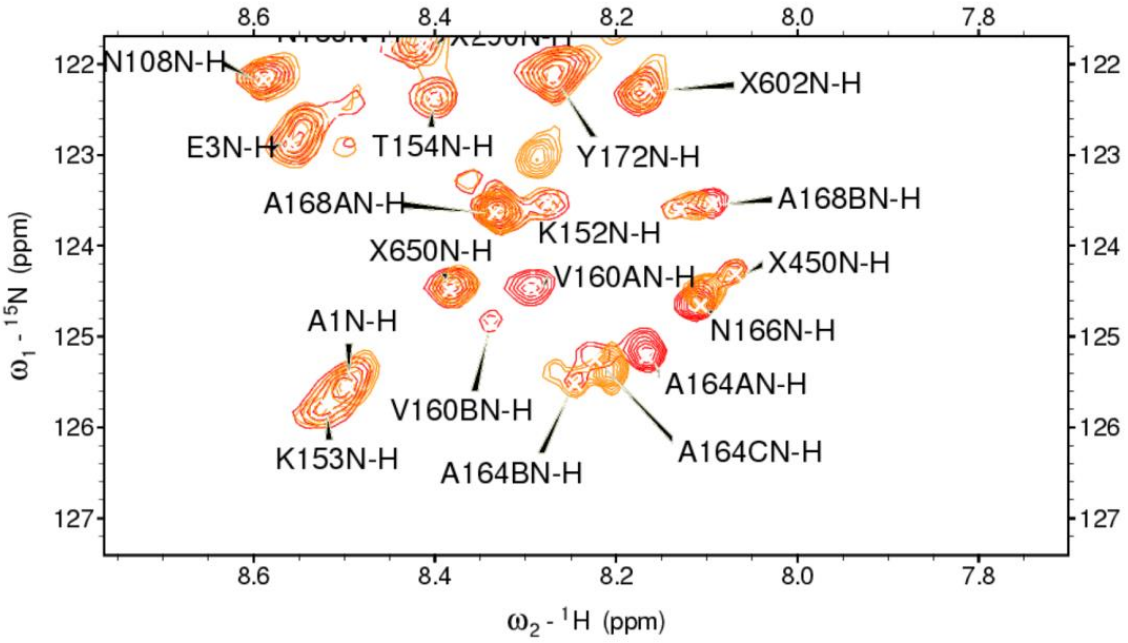


Figure 7. ^{1}H-^{15}N TROSY HSQC of WT Opa$_{60}$ (red) vs. P161G Opa$_{60}$ (orange) at 10 ˚C, zoomed view of Ala164 cross peaks. TROSY HSQC spectra were recorded on a Bruker Avance 800 MHz spectrometer and processed in Bruker Topsin 3.2 and analyzed in NMRFAM-SPARKY[5].

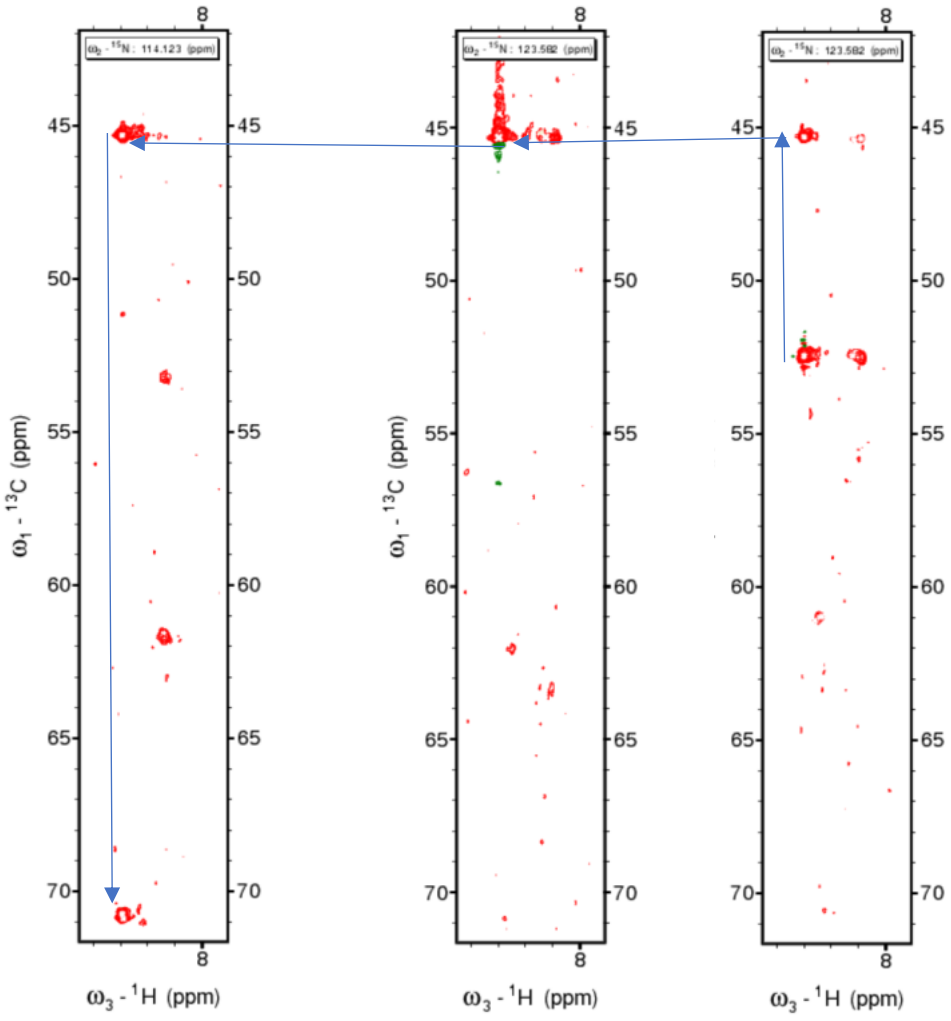


Figure 8. HNCA and HNCOCA strips of WT Opa_{60} of Ala168 and G167. Spectra was recorded on a Bruker Avance 800 MHz spectrometer and processed in Bruker Topsin 3.2, reconstructed using NMRPipe and SMILE, and analyzed in NMRFAM-SPARKY[5-7].

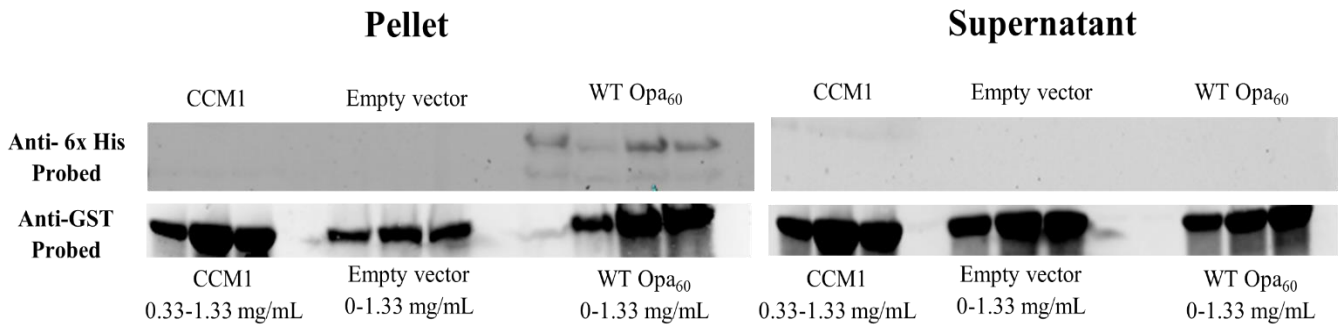


Figure 9. Western blot probing PhoE-WT Opa_{60} and empty Opa_{60} vector binding GST-NCCM1. Concentrations of GST-NCCM1 used were 0.33, 0.66 and 1.33 mg/mL. Purified standards of GST-NCCM1 were used to compare relative binding of Opa_{60} to GST-NCCM1. Opa_{60} constructs were probed with a primary anti-6x His tag mouse monoclonal antibody and detected by a secondary anti-mouse goat secondary monoclonal antibody. The GST-NCCM1 fusion constructs were probed with a primary anti-GST tag mouse monoclonal antibody and detected by a secondary anti-mouse goat secondary monoclonal antibody.

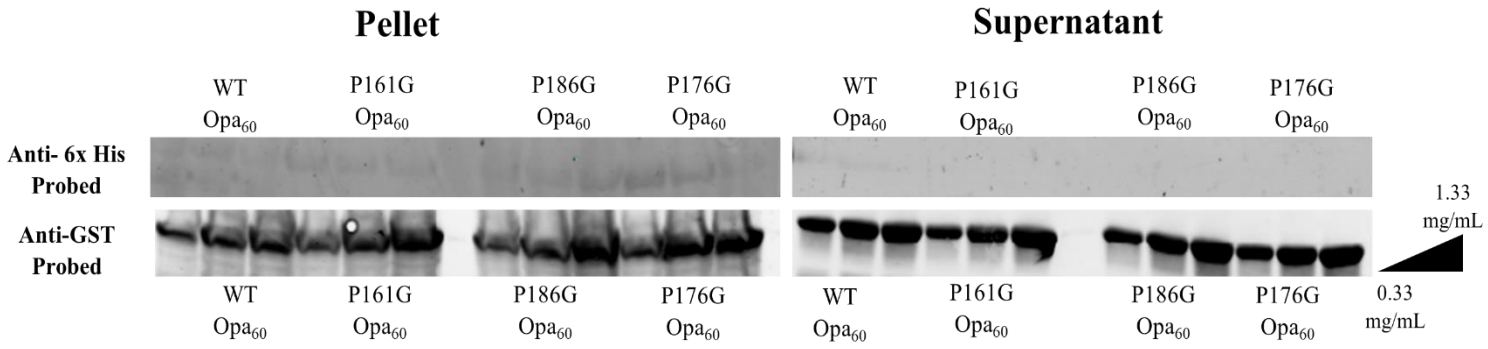


Figure 10. Western blot probing PhoE-WT Opa_{60} and PhoE-P to G Opa_{60} mutants (P161G, P165G and P197G) binding to GST-NCCM1. Concentrations of GST-NCCM1 used were 0.33, 0.66 and 1.33 mg/mL. Opa_{60} constructs were probed with a primary anti-6x His tag mouse monoclonal antibody and detected by a secondary anti-mouse goat secondary monoclonal antibody. The GST-NCCM1 fusion constructs were probed with a primary anti-GST tag mouse monoclonal antibody and detected by a secondary anti-mouse goat secondary monoclonal antibody.

References

(1) Fox, D. A.; Columbus, L. Solution NMR Resonance Assignment Strategies for β-Barrel Membrane Proteins. *Protein Sci* **2013**, *22* (8), 1133–1140. https://doi.org/10.1002/pro.2291.

(2) Dewald, A. H.; Hodges, J. C.; Columbus, L. Physical Determinants of β-Barrel Membrane Protein Folding in Lipid Vesicles. *Biophysical Journal* **2011**, *100* (9), 2131–2140. https://doi.org/10.1016/j.bpj.2011.03.025.

(3) Grathwohl, C.; Wüthrich, K. Nmr Studies of the Rates of Proline Cis–Trans Isomerization in Oligopeptides. *Biopolymers* **1981**, *20* (12), 2623–2633. https://doi.org/10.1002/bip.1981.360201209.

(4) Statistics Calculated for All Chemical Shifts from Atoms in the 20 Common Amino Acids' https://bmrb.io/ref_info/csstats.php.

(5) Goddard, T.; Kneller, D. Sparky 3; University California, San Francisco; 2004. *Google Scholar There is no corresponding record for this reference* **2021**.

(6) Delaglio, F.; Grzesiek, S.; Vuister, G. W.; Zhu, G.; Pfeifer, J.; Bax, A. NMRPipe: A Multidimensional Spectral Processing System Based on UNIX Pipes. *Journal of Biomolecular NMR* **1995**, *6* (3), 277–293. https://doi.org/10.1007/BF00197809.

(7) Ying, J.; Delaglio, F.; Torchia, D. A.; Bax, A. Sparse Multidimensional Iterative Lineshape-Enhanced (SMILE) Reconstruction of Both Non-Uniformly Sampled and Conventional NMR Data. *J Biomol NMR* **2017**, *68* (2), 101–118. https://doi.org/10.1007/s10858-016-0072-7.

Section 5: Materials and Methods

5.1 Site Directed Mutagenesis of P161G, P165G and P176G

Constructs of P161G, P165G and P176G Opa_{60} in pET28b and pET22b vectors were created using an Agilent Quick Change®, Site Directed Mutagenesis kit using custom primers designed on the sequence of interest and ordered from Eurofins. Primers for P161G were GAGGTTACTACCGTC**GGC**AGCAATGCTCCTAAC (forward) and GTTAGGAGCATTGCTG**CCG**ACGGTAGTAACCTCA (reverse). Primers for P165G were GTCCCCAGCAATGCT**GGT**AACGGAGCAGTTACA (forward) and TGTAACTGCTCCGTT**ACC**AGCATTGCTGGGGAC (reverse). Primers for P176G were ACTTATAATACTGAT**GGA**AAGACGCAAAACGAT (forward) and ATCGTTTTGCGTCTT**TCC**ATCAGTATTATAAGT (reverse). Bases in bold were the mutated codons. Afterwards, the primers, wild type (WT) Opa_{60} template DNA (pET28b or pET22b), 10x buffer, dNTP's, water, Quik® solution and DNA polymerase were combined and amplified using PCR. Once PCR was complete, the samples underwent a DpnI digest for 15 minutes at 37 ˚C to remove methylated template DNA. The DpnI treated DNA samples were transformed into Agilent XL-Gold cells via heat shock. The XL Gold® cells were plated on LB-kanamycin (pET28b) or LB-ampicillin (pET22b) media and incubated overnight at 37 ˚C. Colonies were selected from the plate and grown overnight in 5 mL culture of liquid LB-ampicillin media at 37 ˚C, 225 rpm. The overnight cultures were mini-prepped using a Qiagen QIAprep® Spin Miniprep Kit. Mutations were confirmed by Sanger Sequencing (Genewiz).

5.2 WT and Proline to Glycine mutant Opa_{60} growth, expression, and purification for NMR studies

Opa_{60} WT or mutants in a pET28b vector were transformed into *E.coli* BL21(DE3) cells to be purified from inclusion bodies. The opa60 construct was previously cloned into a pET28b vector and contains a thrombin cleavable N-terminal 6x-His tag (MGSHHHHHHSSGLVPRGSHM)[1]. A 5 mL starter culture of LB-kanamycin media was inoculated with one colony from the BL21 transformation and was grown overnight for approximately 16 hours at 37 ˚C, 225 rpm. The overnight cultured cells were used to inoculate 1L of minimal medium containing H_2O, 2 g/liter of ^{13}C (99%)-glucose and 1 g/liter of ^{15}N (99%)- ammonium chloride both isotopes were purchased from Cambridge Isotope Labs, kanamycin (50 mg/mL), solution Q (5M HCl, 25 mM $FeCl_2 \cdot 4H_2O$, 1.25 mM $CaCl_2 \cdot 2H_2O$, 1 mM H_3BO_3, 75 mM $CoCl_2 \cdot 6H_2O$, 23 mM $CuCl_2 \cdot 2H_2O$, 2.5 mM $ZnCl_2$, 2.5 mM $Na_2MoO_4 \cdot 2H_2O$ and 0.2 mM $MnCl_2 \cdot 4H_2O$), and 20 mM $MgSO_4$. Cells were allowed to grow to an OD_{600} between 0.8-1.0 at 37 ˚C. Afterwards, the protein production was induced with 1 mM isopropyl-beta-thio-D-galactoside (IPTG) for 8-10 hours at 37 ˚C. The cells were harvested by centrifugation at 4000g, 4 ˚C for 20 minutes and resuspended in lysis buffer containing 50 mM Tris pH=8.0 and 150 mM NaCl. Cells were lysed using a NanoDEBEE (Bee International). The cell lysate was then centrifuged at 12000g, 12 ˚C for 30 minutes. The supernatant was removed and the cell pellet was resuspended in lysis buffer. The resuspended pellet was then centrifuged again at 12000g, 12 ˚C for 30 minutes. Afterwards, the supernatant was removed, the pellet was resuspended in extraction buffer (lysis buffer including 8M urea) and solubilized overnight at room temperature. The overnight solubilization was centrifuged at 12000g for 30 minutes at 12 ˚C. The supernatant was loaded on to a Co^{2+} immobilized metal affinity chromatography column

equilibrated with approximately 10 column volumes (CV) of extraction buffer. The column was washed with approximately 6 CV of wash buffer (20 mM sodium phosphate buffer pH = 7.8, 150 mM NaCl, 20 mM imidazole, 8M urea) and eluted with 4 CV of elution buffer (20 mM sodium phosphate buffer pH = 7.0 150 mM NaCl, 680 mM imidazole, 8M urea). Elution samples of Opa_{60} were concentrated to approximately 5 mg/mL and stored in 10 mg aliquots at -20 ˚C.

5.3 Refolding and Preparation for NMR

One purified 10mg aliquot of Opa_{60} was diluted into 40 mL of refolding buffer (20 mM sodium phosphate buffer pH = 8.0 and 150 mM NaCl) with 4.4 mM of dodecylphosphocholine (FC-12). Opa samples were added dropwise into the refolding buffer with FC-12 and incubated at room temperature for 5 days. Folding of Opa was assessed using SDS PAGE because the refolded Opa migrates faster than the unfolded Opa on SDS-PAGE gel [1,2].

The refolding reaction was concentrated to approximately 500 µM and dialyzed successively against three separate 4 liters of dialysis buffer (20 mM phosphate buffer pH = 6.2 and 150 mM NaCl) for 1 hour each.

5.4 Measurement of FC-12 concentration

The concentration of FC-12 in solution with Opa_{60} was assessed by 1D ^{1}H NMR on a Bruker Avance 600 MHz spectrometer. Standards of FC-12 at 50, 100 and 200 mM were prepared in dialysis buffer with 10% v/v D_2O. The standards were used to create a linear plot of 1D ^{1}H peak intensities vs. concentration. Spectra were visualized in MNova® (Mestre Labs). A linear regression equation was extrapolated to calculate the FC-12 monomer concentration and micelle concentration of FC-12 was estimated using the aggregation number of 54[3].

5.5 NMR studies

Two-dimensional and three-dimensional NMR spectra for Opa_{60} were recorded on a Bruker Avance 800 MHz spectrometer. Spectra were collected at 10 ˚C to visualize the loop regions of interest. Two dimensional spectra were processed on Topspin 3.2 and spectra were analyzed using NMRFAM-SPARKY[4]. Three dimensional spectra were collected utilizing non-uniform sampling and reconstructed with the SMILE extension in NMRPipe[5,6]. The spectra was then analyzed and assigned utilizing NMRFAM-SPARKY[4]. The assignment and reconstruction was completed by Dr. Jeffery Ellena.

5.6 GST-hCEACAM1 N-terminal domain fusion protein growth, expression, and purification

Previously cloned GST-hCEACAM1 N-terminal domain fusion protein (GST-NCCM1) in a pGEX-2T vector was transformed into *E. coli* BL21 (DE3) cells. The construct contained a C-terminal glutathione S-transferase (GST) tag and a tobacco etch virus (TEV) cleavage site between the GST tag and hCEACAM1 N-terminal domain protein. One colony from the transformation was selected to be grown overnight in liquid LB-ampicillin media at 37 ˚C, 225 rpm. Overnight cultures were added to 1L of LB-ampcillin and grown at 37 ˚C to an OD_{600} = 0.6. Cells were cooled at 25 ˚C for 30 minutes then induced with 0.5 mM IPTG. Induced cultures were grown for 18 hours at 25 ˚C. Cells were harvested by centrifugation at 4,000g for 20 minutes at 4 ˚C. The protein production was resuspended in lysis buffer (20mM Tris pH = 8.0, 150 mM NaCl, 2mM EDTA, and 10% glycerol) and lysed using a NanoDEBEE (Bee International). The cell lysate was then centrifuged at 18,000g for 1 hour at 4 ˚C. The supernatant was retained and precipitated with ammonium sulfate to 55% saturation with stirring for 1 hour

at 4 ˚C. The precipitated mixture was centrifuged at 12,000g for 30 minutes at 4 ˚C. Afterwards, the pellet was retained, re-resuspended in lysis buffer, and homogenized. The resuspension was centrifuged at 12,000g for 30 mins at 4 ˚C. Supernatant was retained and loaded onto a glutathione affinity column previously equilibrated with equilibration buffer (20mM Tris pH = 7.3, 150 mM NaCl, 2mM DTT, and 10% glycerol). The column was washed with 75 mL of equilibration buffer (20mM Tris pH = 7.3, 150 mM NaCl, 2mM DTT, and 10% glycerol) and eluted with 50 mL of elution buffer (20mM Tris pH = 7.3, 150 mM NaCl, 2mM DTT, 10 mM glutathione and 10% glycerol). Elution samples were concentrated to 3 mg/mL and stored in 3 mg aliquots at -80 ˚C.

5.7 E. coli Whole Cell Pull-down Assays

Previously, Dr. Meagan Belcher Dufrisne cloned Opa_{60} into a pET22b vector. The opa60 construct contained a PhoE signal sequence (MKKSTLALVVMGIVASASVQA) with an N-terminal 6x-His tag. The WT and P to G mutant Opa_{60} constructs were transformed into *E. coli* BL21 (DE3) cells. One colony was selected to be grown overnight in liquid LB-ampicillin media at 37 ˚C, 225 rpm. Overnight cultures were added to 250 mL of LB-ampcillin and grown at 37 ˚C to an OD_{600} = 0.8. Cells were cooled at 22 ˚C for 30 minutes then induced with 0.5 mM IPTG. Induced cultures were grown for 18 hours at 22 ˚C. The OD_{600} of the culture was re-measured and cells were diluted with fresh LB media to reach a concentration of 3×10^8 CFU/mL. Afterwards, 1 mL of the diluted culture was inoculated with 0.33 mg/mL, 0.66 mg/mL and 1.33 mg/mL of GST-NCCM1 fusion protein and incubated together for 1 hour at 37 ˚C, 225 rpm. The incubated cultures were centrifuged at 4000g for 20 minutes at room temperature. The supernatant was removed from the pellet. The pellet was resuspended in 1x PBS with 20 µg/mL

of DNase and 0.20 mg/mL lysozyme. Then, 4X SDS loading buffer was added to samples of the pellet and supernatant.

Pellet and supernatant samples were run on 4-20% Tris-Glycine SDS-PAGE BioRad gels, transferred to mini BioRad nitrocellulose membrane, with a BioRad Turbo Blotter for 10 minutes at 1.3 A, 25V. Afterwards, the western blot membranes were blocked overnight at 4 ˚C in a 5% milk solution in 1x PBST (1X PBS with 0.1% Tween). Opa_{60} constructs were probed with a primary anti-6x His tag mouse monoclonal antibody (Invitrogen) and detected by a secondary anti-mouse goat secondary monoclonal antibody (IRDye 800 Li-COR). The GST-NCCM1 fusion constructs were probed with a primary anti-GST tag mouse monoclonal antibody (ABcolonal) and detected by a secondary anti-mouse goat secondary monoclonal antibody with (IRDye 800 Li-COR). Western blots were imaged with a Li-COR Odyssey® Imager using the accompanying software.

References

(1) Dewald, A. H.; Hodges, J. C.; Columbus, L. Physical Determinants of β-Barrel Membrane Protein Folding in Lipid Vesicles. *Biophysical Journal* **2011**, *100* (9), 2131–2140. https://doi.org/10.1016/j.bpj.2011.03.025.

(2) Fox, D. A.; Columbus, L. Solution NMR Resonance Assignment Strategies for β-Barrel Membrane Proteins. *Protein Sci* **2013**, *22* (8), 1133–1140. https://doi.org/10.1002/pro.2291.

(3) Anatrace -F308S - Fos-Choline-12, Sol-Grade n-Dodecylphosphocholine https://www.anatrace.com/Products/Lipids/FOS-CHOLINE/F308S.

(4) Goddard, T.; Kneller, D. Sparky 3; University California, San Francisco; 2004. *Google Scholar There is no corresponding record for this reference* **2021**.

(5) Delaglio, F.; Grzesiek, S.; Vuister, G. W.; Zhu, G.; Pfeifer, J.; Bax, A. NMRPipe: A Multidimensional Spectral Processing System Based on UNIX Pipes. *Journal of Biomolecular NMR* **1995**, *6* (3), 277–293. https://doi.org/10.1007/BF00197809.

(6) Ying, J.; Delaglio, F.; Torchia, D. A.; Bax, A. Sparse Multidimensional Iterative Lineshape-Enhanced (SMILE) Reconstruction of Both Non-Uniformly Sampled and Conventional NMR Data. *J Biomol NMR* **2017**, *68* (2), 101–118. https://doi.org/10.1007/s10858-016-0072-7.