

**Neisserial Opa protein dynamics and interactions with human  
CEACAM1**

Marissa Kathryn Kieber

Wilmington, NC

B.S. The University of North Carolina at Chapel Hill, 2012

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

Department of Chemistry

University of Virginia

May, 2018

© Copyright by  
Marissa Kathryn Kieber  
All rights reserved  
May 2018

## Abstract

*Neisseria gonorrhoeae* and *N. meningitides* are human pathogens which infect millions each year and are becoming increasingly antibiotic resistant. These bacteria utilize opacity-associated (Opa) proteins to mediate bacterial uptake into non-phagocytic and phagocytic cells. Opa proteins primarily engage carcinoembryonic antigen-related cellular adhesion molecules (CEACAMs), hijacking host cellular mechanisms to induce bacterial engulfment. The Opa family of proteins are outer membrane  $\beta$ -barrels with eight transmembrane strands and four extracellular loops. The loops of Opa contain regions of high sequence variability that engage specific CEACAM receptors (termed hypervariable (HV) regions 1 and 2). Multiple sequence alignment of the HV regions does not reveal a conserved CEACAM-binding motif, and the molecular determinants of the Opa-CEACAM interaction are unknown.

Results from CW-EPR and previous NMR relaxation experiments indicate Opa loops are dynamic on the nanosecond timescale, which may enable a sampling of transient structures and loop interactions that promote receptor engagement. A combination of Double Electron Electron Resonance (DEER) and molecular dynamics (MD) were utilized to determine the unbound conformational ensemble of the dynamic Opa loops, where spin label pairs were selected as optimal MD restraints. Experimentally-restrained simulations of Opa show an association of the extracellular loops.

The N-terminal domain of CEACAM proteins (nCEACAM) dimerize *in vivo* and *in vitro*, which may impact Opa binding. Multiple crystal structures of the nCEACAM1 homodimer have been determined, so DEER distances of the dimer were acquired to refine the homodimer structure in solution. CEACAM proteins are also highly glycosylated *in*

*in vivo*, yet the role of these glycans on CEACAM oligomerization and interactions with Opa is unknown. Recombinant N-linked glycosylated nCEACAM1 proteins were generated using human cell cultures, and the addition of N-linked glycans did not alter the homodimerization propensity of nCEACAM1 proteins. N-glycosylated nCEACAM1 proteins are currently being used in Opa binding experiments to identify the effect of glycosylation on Opa interactions. Understanding the molecular mechanism of Opa-CEACAM binding will elucidate the manner by which pathogenic bacteria induce phagocytosis into human cells.

## Acknowledgements

Thank you to Dr. Linda Columbus for always offering words of encouragement and support and for providing me the opportunity to grow as a scientist.

Thank you to the past and present members of the Columbus lab for their training, support, and friendship. Many thanks in particular to Jennifer Martin, Ashton Brock, Jason Kuhn, Steven Keller, Nicole Swope, and Tracy Caldwell for making being in lab enjoyable, even on the most trying days.

Thank you to our collaborators, Jennifer Hays and Peter Kasson, for all of their hard work and helpful discussions.

Thank you to my parents for teaching me that Chemistry was cool as soon as I could walk.

Thank you to Dan for always making me laugh, even while I was writing my dissertation.

## Table of Contents

Copyright page.....	ii
Abstract.....	iii
Acknowledgements.....	v
Table of contents .....	vi
List of figures.....	x
List of tables.....	xiv
<b>Chapter 1: Introduction.....</b>	<b>1</b>
<b>1.1 Protein structure.....</b>	<b>1</b>
<b>1.2 Protein dynamics.....</b>	<b>2</b>
1.2.1 Binding events in dynamic protein systems.....	7
<b>1.3 Cell membranes and membrane proteins.....</b>	<b>10</b>
1.3.1 Liposomes and detergent micelles to study membrane proteins...11	
1.3.2 Outer membrane and outer membrane proteins.....	15
<b>1.4 <i>Neisseria meningitidis</i> and <i>N. gonorrhoeae</i>.....</b>	<b>16</b>
1.4.1 Opacity associated proteins.....	19
1.4.2 Carcino-embryonic antigen-like cellular adhesion molecules....	25
<b>1.5 Dissertation overview.....</b>	<b>30</b>
<b>1.6 References.....</b>	<b>31</b>

<b>Chapter 2: Electron Paramagnetic Resonance Spectroscopy</b> .....	41
<b>2.1 Introduction to EPR</b> .....	41
<b>2.2 Free electron in an applied magnetic field</b> .....	42
<b>2.3 A-tensor and hyperfine interactions</b> .....	46
<b>2.4 Relaxation</b> .....	47
<b>2.5 Applications of EPR</b> .....	48
2.5.1 Continuous wave EPR and protein dynamics.....	52
2.5.2 Double electron electron resonance to measure distances.....	55
<b>2.6 Utilizing EPR to study Opa-CEACAM interactions</b> .....	62
<b>2.7 References</b> .....	63
<b>Chapter 3: Opa loop dynamics and conformational ensemble</b> .....	67
<b>3.1 Introduction</b> .....	67
<b>3.2 Structure of Opa</b> .....	67
3.2.1 Frequent, short lived contacts between Opa loops.....	70
<b>3.3 Folding and site directed spin labeling of Opa proteins</b> .....	76
<b>3.4 Mapping Opa loop dynamics with CW-EPR</b> .....	78
<b>3.5 Combining DEER and MD to determine the Opa loop conformational ensemble</b> .....	81
3.5.1 Identifying highly informative DEER pairs.....	84

3.5.2 Measuring informative DEER pairs.....	87
<b>3.6 Opa loop conformational ensemble.....</b>	<b>90</b>
<b>3.7 Conclusions.....</b>	<b>93</b>
<b>3.8 References.....</b>	<b>95</b>
<b>Chapter 4: Homodimerization of nCEACAM1 proteins.....</b>	<b>98</b>
<b>4.1 CEACAMs are adhesion molecules.....</b>	<b>98</b>
<b>4.2 Developing methods to disrupt CEACAM homodimerization.....</b>	<b>105</b>
4.2.1 nCEACAM1 with no glycosylation or attached affinity tag.....	107
4.2.2 MBP-nCEACAM1.....	109
4.2.3 BAP-nCEACAM1.....	111
<b>4.3 N-linked glycosylation may disrupt nCEACAM1 homodimerization....</b>	<b>114</b>
4.3.1 N-linked glycosylation.....	116
4.3.2 Mammalian recombinant expression to generate glycosylated nCEACAM1.....	119
<b>4.4 Characterizing nCEACAM1 homodimerization.....</b>	<b>120</b>
4.4.1 SEC-MALS to identify oligomeric states of nCEACAM1 proteins.....	125
4.4.2 DEER distances of nCEACAM1 homodimer../.....	127
<b>4.5 Future research directions.....</b>	<b>132</b>

4.5.1 SAXS to determine the shape of the nCEACAM1 homodimer.....	132
4.5.2 X-ray crystallography of nCEACAM1 GlcNAc.....	133
<b>4.6 Conclusions.....</b>	<b>136</b>
<b>4.7 References.....</b>	<b>140</b>
<b>Chapter 5: Prospects for future research into Opa-CEACAM interactions.....</b>	<b>146</b>
<b>5.1 Introduction to Opa-CEACAM interactions.....</b>	<b>146</b>
<b>5.2 Effect of CEACAM binding on Opa loop dynamics.....</b>	<b>149</b>
5.2.1 Opa loop dynamics upon CEACAM binding with CW- EPR.....	152
5.2.2 Modeling the Opa-CEACAM complex with DEER.....	158
5.2.3 Opa loop conformational ensemble upon binding CEACAMs using DEER.....	162
<b>5.3 Impact of LOS on Opa loop dynamics and CEACAM binding.....</b>	<b>163</b>
<b>5.4 Effect of CEACAM glycosylation on Opa binding.....</b>	<b>165</b>
<b>5.5 Opa residues involved in CEACAM interactions.....</b>	<b>166</b>
<b>5.6 Mechanism of Opa-CEACAM binding among different Opa variants..</b>	<b>168</b>
<b>5.7 Concluding remarks.....</b>	<b>168</b>
<b>5.8 References.....</b>	<b>171</b>
<b>Materials and Methods.....</b>	<b>192</b>

## List of Figures

### Chapter 1

1.1 Dynamic protein motions over a range of timescales.....	5
1.2 Biophysical methods utilized to study protein structure and dynamics.....	6
1.3 Types of dynamic protein binding interactions.....	9
1.4 Detergent micelle and liposome membrane mimics.....	14
1.5 Micrograph of <i>Neisseria gonorrhoeae</i> being phagocytosed into human cells.....	18
1.6 NMR structure of Opa <sub>60</sub> in detergent micelles (2MLH).....	22
1.7 Multiple sequence alignment of the semivariable and hypervariable regions of selected Opa proteins.....	23
1.8 Hydrophobic residues in the extracellular loops of Opa <sub>60</sub> .....	24
1.9 Schematic of the CEACAM family of proteins.....	28
1.10 N-terminal domain of CEACAM1 (2GK2).....	29

### Chapter 2

2.1 Energy level scheme of a nitroxide spin system in a magnetic field.....	44
2.2 Magnetic frame of a nitroxide.....	45
2.3 Schematic of site directed spin labeling proteins.....	50
2.4 EPR experiments and their applications.....	51
2.5 EPR lineshapes reflect the protein environment of the spin label.....	54
2.6 Four-pulse DEER experiment setup.....	59
2.7 Schematic of a two-pulse sequence to generate a spin echo.....	60

2.8 DEER signal as a function of the mean distance and width around the mean.....	61
---	----

### **Chapter 3**

3.1 MD-refined NMR structures of Opa <sub>60</sub> proteins.....	69
3.2 Opa HV regions have a high probability to interact over short periods of time.....	72
3.3 Hydrophobic residues in Opa loops are hypothesized to promote loop interactions and a receptor-competent conformational ensemble.....	73
3.4 NMR and MD predict that Opa extracellular loops are dynamic on the ns-timescale .....	74
3.5 HV1-HV2 contact maps for structural snapshots of Opa loops.....	75
3.6 Opa loop dynamics via CW-EPR.....	80
3.7 Integrating DEER and MD to generate a conformational ensemble of Opa loops....	83
3.8 Spectroscopist-selected DEER pairs.....	86
3.9 First round of DEER measurements of mRMR pairs.....	88
3.10 Second round of DEER measurements of mRMR pairs.....	89
3.11 Refinement of Opa <sub>60</sub> loops from the first round of mRMR DEER pairs.....	91
3.12 Refinement of Opa <sub>60</sub> loops from the second-round mRMR pairs.....	92

### **Chapter 4**

4.1 Residues involved in nCEACAM1 binding interactions.....	102
4.2 Arrangements of CEACAM1 ectodomains.....	103

4.3 Potential nCEACAM1 homodimer crystal structures.....	104
4.4 Schematic of nCEACAM1 constructs used to study nCEACAM1 homodimerization .....	106
4.5 DEER distributions of non-glycosylated nCEACAM1.....	108
4.6 DEER distributions of MBP-nCEACAM1.....	110
4.7 Generating BAP-nCEACAM1 proteins that are immobilized on avidin resin using in- cell biotinylation with BirA.....	112
4.8 Immobilized BAP-nCEACAM1 DEER distributions.....	113
4.9 nCEACAM1 contains three N-linked glycosylation sites.....	115
4.10 Lipid-linked oligosaccharide (LLO) glycan core assembled in N-linked glycosylation.....	118
4.11 MALDI-TOF of nCEACAM1 GlcNAc, nCEACAM1 high mannose, and nCEACAM1 with complex glycosylation.....	124
4.12 SEC-MALS elution profiles of various nCEACAM1 constructs.....	126
4.13 DEER distributions of nCEACAM1 in its glycosylated and non-glycosylated forms .....	131
4.14 Initial nCEACAM1 GlcNAc crystal screens.....	135
4.15 Overlay of nCEACAM1, nCEACAM5, and nCEACAM6 homodimer structures .....	139

## Chapter 5

5.1 Sequence alignment of human CEACAM N-terminal domains.....	148
5.2 CW-EPR lineshapes of single labeled Opa proteins in detergent with excess nCEACAM1 GlcNAc proteins.....	151
5.3 CW-EPR lineshapes of Opa 152R1 micelles upon the addition of varying concentrations of nCEACAM1 GlcNAc.....	152
5.4 DEER distributions of MBP-nCEACAM1 Q103R1 + Opa 177R1.....	154
5.5 DEER distributions of BAP-nCEACAM1 L20R1 proteins immobilized onto avidin resin + Opa R1 micelles.....	155
5.6 DEER distributions of nCEACAM1 GlcNAc proteins + Opa R1 micelles or liposomes.....	156
5.7 DEER distributions of double spin labeled Opa proteins + unlabeled nCEACAM1 GlcNAc proteins.....	160
5.8 DEER distributions of Opa 31-166R1 micelles with nCEACAM1 GlcNAc or BSA .....	161
5.9 DEER distributions of Opa 80-158R1 liposomes with and without LPS.....	164

## List of Tables

4.1 Average distance between non-glycosylated nCEACAM1 residues in DEER vs. homodimer crystal structures.....	123
5.1 Selected mean DEER distances between single labeled Opa and CEACAM proteins .....	157



## Chapter 1: Introduction

### 1.1 Protein structure

Proteins are highly diverse biomolecules which are involved in most biological processes that constitute life. Proteins have a remarkable variety of functions that range from maintaining the structural integrity of the cell to intracellular signaling and enzyme catalysis. At their base, proteins are composed of various combinations of 20 different natural amino acids, connected through peptide bonds (the peptide backbone) to form a polypeptide. This amino acid sequence, or the primary structure, is dictated by a DNA sequence, which is ultimately translated into a polypeptide. Protein secondary structure is defined as the conformation of the peptide backbone, which can be classified by specific folding patterns, namely helices,  $\beta$ -strands, and turns. The tertiary structure is the folding pattern of secondary structural elements and is thus the three-dimensional arrangement of a protein. Finally, the quaternary structure includes the arrangement of multiple protein subunits.

Studying the overall folded structure of a protein is central to the understanding of a protein's biological function, which may in turn have medicinal or biotechnological significance. Research on protein fold was revolutionized in 1958 with the determination of the first protein structure, myoglobin, by Kendrew and Perutz using X-ray crystallography [1]. This groundbreaking work won the pair a Nobel Prize in 1962. Several protein structures were determined shortly thereafter, with the first enzyme structure (lysozyme) being solved in 1965 [2]. Nuclear magnetic resonance (NMR) spectroscopy was developed as an alternative method to X-ray crystallography for protein structure

determination in the 1970's [3]. While there was great progress in soluble protein structure determination, it took until 1985 for the atomic detail of the first integral membrane protein to be solved using X-ray crystallography [4]. This lag in the determination of membrane protein structures was largely due to challenges which arise from providing a membrane-like environment for these proteins, some of which will be discussed in this dissertation.

## 1.2 Protein dynamics

The first protein structure determined in the 1950's led to the hypothesis that a specific protein sequence dictates a specific protein fold, which in turn is critical to the function of a protein. The static capture of a single fold led to an interpretation of protein function as a pre-existing protein shape that interacts only with a complementary ligand, as keys are designed to fit specific locks (thus, the lock-and-key model) [5]. As our knowledge of proteins increased over time, it became clear that proteins are not static entities, and that dynamic motions within a protein are often critical to a protein's function and cannot be captured with a static structure. Proteins with highly dynamic backbone fluctuations are implicated in a number of diseases, such as Parkinson's and various cancers [6]. It is estimated that over 25% of proteins in complex organisms undergo large scale dynamic motion [6], and such unstructured regions or large conformational changes cannot easily be predicted by a static structure alone, as many proteins are instead an ensemble of different protein conformations.

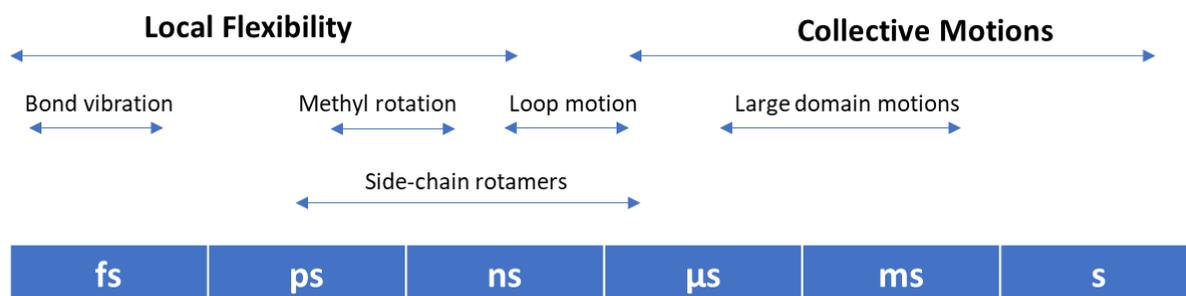
Dynamic movements within a protein arise from motions along  $C\alpha$ -N bonds ( $\phi$ ),  $C\alpha$ -C bonds ( $\psi$ ), and bonds within an amino acid side chain (whereas the C-N peptide bond itself is rigid). These dynamic fluctuations range from small scale motions of an individual side chain to large scale conformational rearrangements of a protein's 3D structure, and can be

characterized by their timescale of motion, as well as their amplitude and directionality (Figure 1.1). These motions yield an ensemble of conformational states, whose multidimensional energy landscape is defined by the probability between states (thermodynamics), as well as energy barriers between states (kinetics) [7]. Ligand binding, sequence mutations, and altered external conditions can all affect the protein's energy landscape and shift the equilibrium between conformational states.

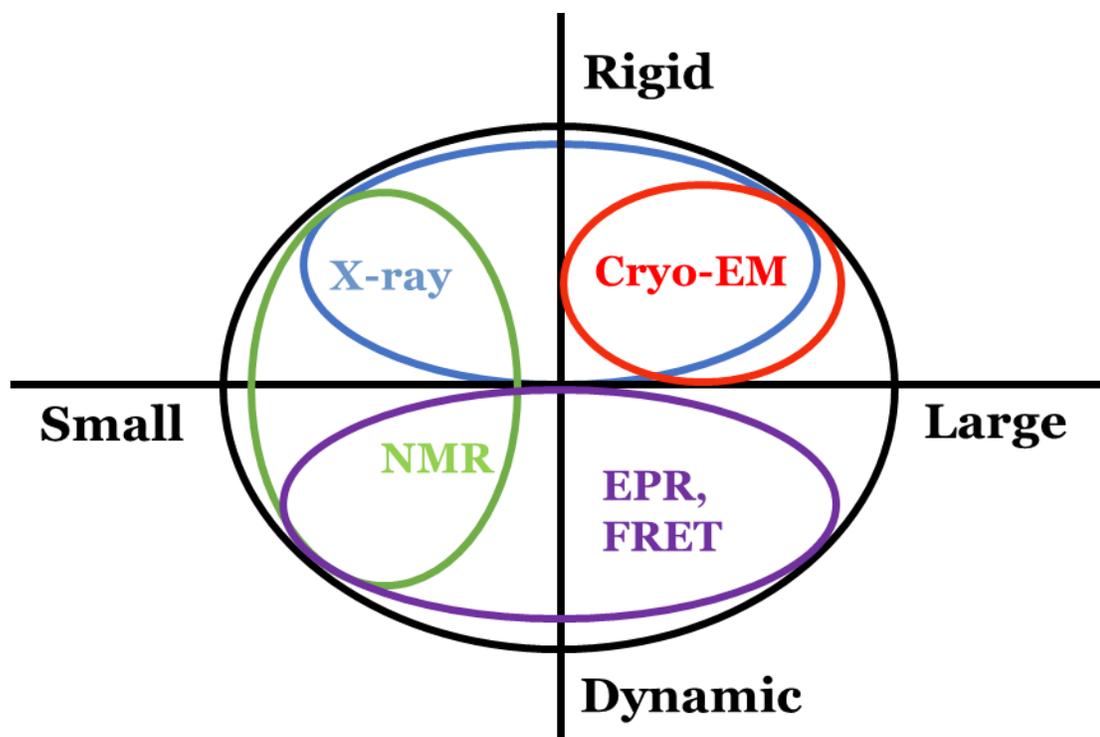
Though comparatively rare, dynamic protein motions on a  $\mu\text{s}$  or slower timescale occur on the order of enzyme catalysis, protein-protein interactions, and signal transduction timescales. These motions yield distinctive, long-lifetime protein states that can often be captured experimentally using techniques such as X-ray crystallography, Nuclear Magnetic Resonance spectroscopy (NMR), Cryo-Electron Microscopy (Cryo-EM), and Small Angle X-ray Scattering (SAXS) (Figure 1.2). X-ray crystallography provides high-resolution information on a 3D protein structure, but requires a homogenous crystal which is often only possible with the addition of reagents which 'trap' the protein in a specific conformation. SAXS and Cryo-EM are lower resolution alternatives to X-ray crystallography which do not require this homogenous crystal, making it possible to determine a structural ensemble. NMR is advantageous in that it provides both atomic resolution and timescales of protein motion in solution.

Fast motions in the ps-ns timescale include local, small amplitude fluctuations in conformation. Protein loop motion occurs on the ns timescale, whereas local atomic fluctuations, such as side chain rotations, typically occur on the ps-ns timescale. While both X-ray crystallography and NMR have methods to gather information on these fast timescale motions, molecular dynamics (MD) simulations and electron paramagnetic

resonance (EPR) spectroscopy are ideal techniques to measure protein fluctuations in this ps-ns time range. The advantages and applications of EPR to measure such local and loop motions, as well as different conformational states, will be discussed extensively in later chapters.



**Figure 1.1: Dynamic protein motions over a range of timescales** *Adapted from [8] with permission.* Protein dynamic motions occur across a range of timescales, which can in turn be examined via a variety of experimental techniques.



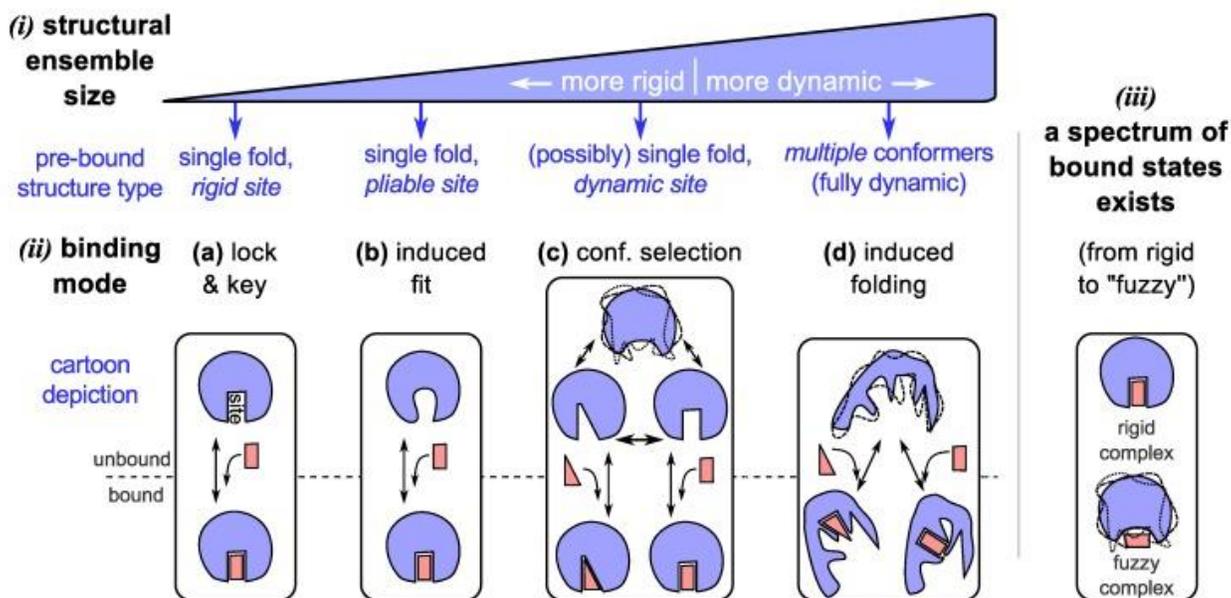
**Figure 1.2: Biophysical methods utilized to study protein structure and dynamics.**

*Adapted from [9] with permission.* X-ray crystallography, Cryo-EM, NMR, EPR, and FRET can probe protein systems of various sizes and mobilities. X-ray crystallography is an ideal method to determine a high-resolution structure of a static protein, whereas Cryo-EM is better suited to larger protein systems. NMR is an ideal technique to study both rigid and dynamic proteins, but is limited to smaller systems of approximately  $< 50$  kDa. Conversely, EPR and FRET can probe dynamic proteins without restrictions on the size or complexity of the system. EPR will be the primary technique utilized in this dissertation.

### 1.2.1 Binding events in dynamic protein systems

Protein-ligand interactions were first described as a lock-and-key system, where a specific substrate engages a static protein binding pocket [5]. However, as our knowledge of the dynamic nature of proteins has increased, it is apparent that these motions often play a significant role in protein-ligand interactions. Structural flexibility allows protein binding pockets to adapt upon engaging specific receptors (Figure 1.3). The earliest deviation from the lock-and-key idea was a model described as “induced-fit,” where the protein binding pocket is pliable and shifts to complement a ligand upon binding [10]. Alternatively, a protein may sample different conformational states in its unbound form, where a ligand binds selectively to one of these pre-existing conformations, termed “conformational selection.” Often this allows proteins to sample a range of conformations which each engage specific receptors. Ubiquitin, for example, binds a multitude of receptors because of large-scale backbone “pincer-like” motions which undergo conformational selection [11]. An extreme example of protein dynamics lies in a class of proteins called intrinsically disordered proteins (IDPs), which have little to no defined conformation, but often collapse into a dominant structure upon engaging a binding partner or undergo conformational selection. While these dynamic binding models largely describe a system where a single dominant conformation is maintained upon binding, dynamic disorder can also be retained in a protein binding event. This is described as a “fuzzy complex,” where a number of binding partners interact with more than one conformation or are disordered in their bound state [12]. This allows transient interactions and alternate conformations to form between binding partners.

It is often critical for protein complexes to maintain structural malleability to enable a response to the protein's environment and signals. A protein channel, for example, often requires large scale alternating opened and closed conformations upon engaging substrate in the extracellular environment to transmit information or contents into or out of the cell. Likewise, antibodies are a prime example of structural dynamics, as antibodies can bind a myriad of ligands, which allows the recognition of a myriad of antigens by the immune system. Intrinsically disordered regions within these antibodies (called hypervariable regions) [13, 14] allow binding promiscuity via conformational selection of these disordered regions, with distinct antibody conformations binding specific antigen partners [15, 16]. Protein dynamics also potentially lends itself to higher evolvability – structural dynamics provide a protein with conformations that are alternative to the functional fold, which represent the potential for evolution to select a new function from receptors that may bind an alternative conformation [17, 18].



**Figure 1.3: Types of dynamic protein binding interactions, re-printed from [19] with permission.** Many types of protein binding interactions exist, where the degree of receptor flexibility varies from rigid (lock & key), to a fully dynamic binding site of multiple conformers (induced folding). An induced fit model describes a pliable receptor binding pocket of a single fold, and conformational selection describes a dynamic binding site, which typically reduces to a single fold upon binding. A fuzzy complex exists with a spectrum of bound conformational states or a disordered bound state.

### 1.3 Cell membranes and membrane proteins

Proteins can be classified into two overarching categories – soluble and membrane bound. Both types of proteins contain a mixture of hydrophobic and hydrophilic amino acid residues; soluble proteins primarily occlude hydrophobic residues from their polar environment by burying them within the core of the protein, while membrane proteins arrange hydrophobic residues within a lipid bilayer.

The cell or lipid membrane itself provides a physical barrier separating a cell from its external environment and divides eukaryotic cells into various compartments. Membranes are composed of an amphipathic lipid bilayer, where nonpolar lipid tails associate leaving polar, often charged, lipid head groups exposed to the external aqueous environment. A simplified model of the membrane contains a mixture of lipids, sterols, and proteins. This provides a highly complex, heterogeneous amphipathic environment that is largely impermeable to external molecules. As such, signaling molecules and nutrients are transmitted across this membrane barrier via membrane proteins.

Membrane proteins make up roughly 50% of the mass of a membrane and have a diverse range of roles including transporting molecules, catalysis, and signal transduction. There are two primary classifications of membrane proteins: (1) Integral membrane proteins, which completely traverse the lipid bilayer, and often contain multiple transmembrane segments connected by external and internal domains of various sizes, and (2) membrane-associated proteins, which are not integrated into the membrane bilayer, are instead anchored to the membrane via fatty acid chains covalently attached to the protein or by noncovalent interactions between lipids and exposed surfaces of the protein. The specific lipid composition of the membrane itself may also affect the function of membrane

proteins by affecting membrane protein dynamics, stabilizing various conformations, and modulating protein clustering and localization.

Approximately 30% of the coding genome encodes membrane proteins, which often have medical implications arising from protein mutations [20, 21]. It is estimated that approximately 50% of drugs on the market target membrane proteins, making them essential areas of study [21, 22]. Researching membrane proteins is immensely challenging, however, due to difficulties regarding their reconstitution into membrane mimics. Such membrane mimics make studying the structure and function of membrane proteins challenging or impossible with many techniques, causing a significant lag in membrane protein structure determinations. Of the structures deposited in the protein data bank (PDB), less than 1% are of membrane proteins [23].

### 1.3.1 Liposomes and detergent micelles to study membrane proteins

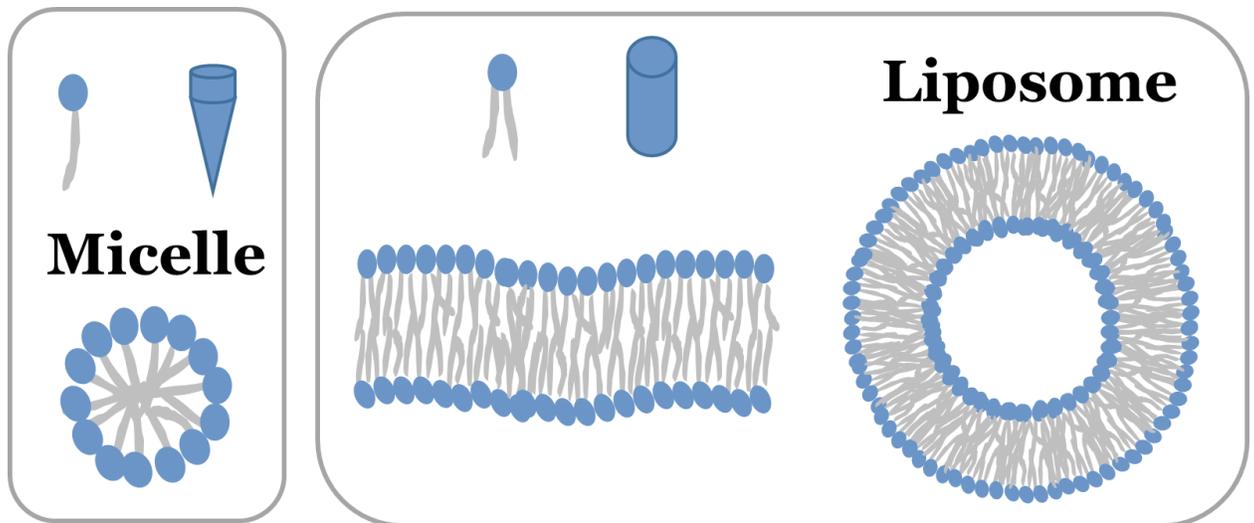
Isolating membrane proteins for structural and functional studies is challenging for several reasons. Membrane proteins are typically expressed at low levels in biological membranes, making isolating sufficient amounts of these proteins from their native environment unfeasible. Overexpression of recombinant membrane proteins in a cell system, using *E. coli* for example, provides an alternative solution, although this often leads to high levels of membrane protein aggregation in the cytoplasm, making generating high yields of functional protein difficult. Additionally, many mammalian membrane proteins require post-translational modifications that are critical to the function of a protein but cannot be incorporated using bacterial expression systems. The highly heterogeneous and complex biological membrane is also impossible to incorporate when using techniques that require a homogeneous or small system, such as NMR. Perhaps the greatest challenge to

studying membrane proteins lies in the fact that membrane proteins are not soluble and functional in an environment devoid of amphipathic molecules to satisfy the hydrophobic nature of a protein's transmembrane regions. As such, synthetic lipid and detergent systems provide a workable solution for *in vitro* studies of membrane proteins, though not without their own challenges as non-native membrane mimics (Figure 1.4).

Detergent micelles are often utilized to isolate and purify membrane proteins. Detergent monomers are amphipathic molecules with a polar head group and a single non-polar hydrocarbon tail. Above a certain concentration (called the critical micelle concentration or CMC, unique to each type of detergent), detergent monomers spontaneously aggregate to form micelles. Detergent micelles are often stable for long periods of time, making them commonly used membrane mimics. A micelle shape and composition can also be modified to solubilize various membrane proteins. While a protein may be successfully solubilized in a specific detergent, it is often difficult to determine if the protein is folded as it would be *in vivo*, or if it is exhibiting a non-functional, stable fold. Ultimately, however, detergent micelles are hugely beneficial in solubilizing membrane proteins, as they allow a relatively simple and small system to study membrane proteins.

Liposomes provide an alternative membrane mimic to detergent micelles. Liposomes are composed of spherical lipid bilayer, whose properties, such as charge and hydrophobic thickness, can be tailored using various lipids. Liposomes can range in size from approximately 50 nm to larger than 1  $\mu\text{m}$  and can be classified by vesicle size, forming either multilamellar vesicles (MLVs), large unilamellar vesicles (LUVs), or small unilamellar vesicles (SUVs). Because liposomes are so highly customizable to mimic a

native bilayer, they are a widely used system to study membrane proteins. Liposomes also have promising potential as therapeutic delivery vessels, as soluble cargo can be encapsulated in their aqueous core and hydrophobic drugs can be integrated into the bilayer. However, targeted delivery and a short shelf life remains a challenge in such liposomal systems [24].



**Figure 1.4: Detergent micelle and liposome membrane mimics.** Micelles are formed via an aggregation of detergent monomers, which contain a polar head group and a single non-polar tail. Conversely, lipids contain two hydrocarbon tails, causing their overall shape to be roughly cylindrical. This shape results in a packing that forms a lipid bilayer and liposomes.

### 1.3.2 Outer membrane and outer membrane proteins

Chloroplast, mitochondria, and gram-negative bacteria contain two lipid membranes separated by a periplasmic space filled with a peptidoglycan layer [25]. The outer membrane protects the bacteria from a harsh external environment, however, the bacteria must sense the extracellular space and transmit nutrients across this additional membrane barrier. Outer membrane  $\beta$ -barrel proteins provide this function, along with  $\alpha$ -helical proteins which reside in the inner membrane. 2-3% of the coding genome in gram-negative bacteria encodes outer membrane proteins [26, 27].  $\beta$ -barrels serve many roles in gram negative bacteria, including porins, transporters, enzymes, receptors, channels, and membrane anchors [28]. Despite this diversity in function, outer membrane proteins have a common  $\beta$ -barrel structure, with a varying number of  $\beta$ -strands (typically between 8 and 24 strands).  $\beta$ -barrels are very stable due to an extensive backbone hydrogen bonding network between anti-parallel  $\beta$ -strands [29]. Larger  $\beta$ -barrel proteins are often further stabilized with a plug like structure which folds into the interior of the barrel, creating stabilizing interactions with interior facing sidechains, as the radius of a large barrel can span 30 Å or more [30]. Smaller eight-stranded  $\beta$ -barrels can instead form salt bridges across the interior of the barrel for further stabilization [31].

The lipid moieties within the inner and outer membranes also varies dramatically. The inner membrane is composed exclusively of phospholipids, with 70-80% being phosphatidylethanolamine and the rest phosphatidylglycerol and cardiolipin. These phospholipids are equally distributed among both the inner and the outer leaflets of the inner membrane. Conversely, the outer membrane contains a highly asymmetric lipid composition between its two leaflets, with the inner leaflet comprising a similar

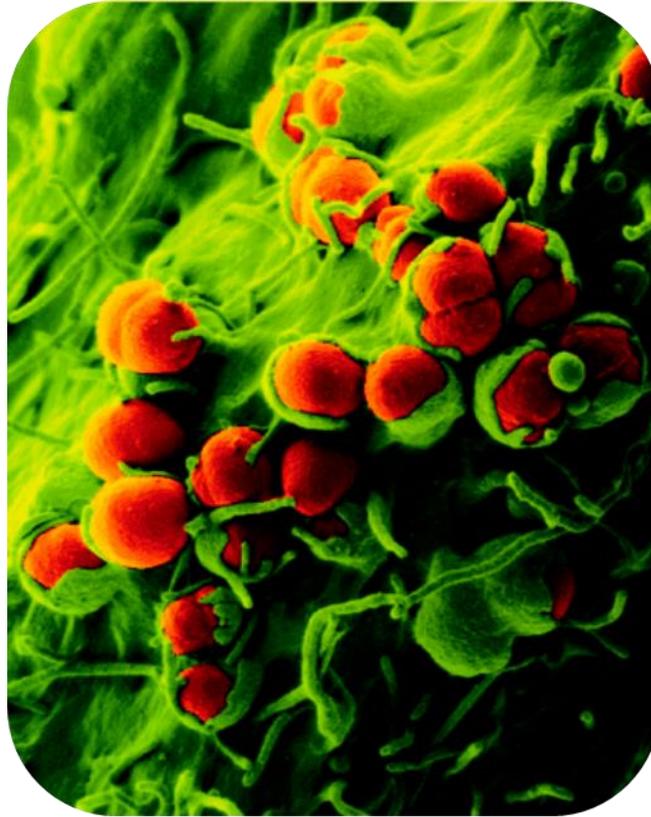
composition to the inner membrane, and the outer leaflet containing primarily lipopolysaccharides (LPS) or lipooligosaccharides (LOS).

LPS and LOS provide a crucial external barrier for gram-negative bacteria, preventing the passive diffusion of hydrophobic solutes, such as antibiotics, into the cell (LPS-devoid bacteria have been shown to be hyper-susceptible to antibiotics [32, 33]). LPS is composed of lipid A, a core oligosaccharide, and an o-antigen polysaccharide at various lengths. This polysaccharide composition and length varies and is often specific to the bacteria species itself. LOS contains lipid A, but lacks the O-antigen unit and the polysaccharide chain is limited to 10 saccharide units. Lipid A and the core oligosaccharide of both LPS and LOS are negatively charged, which, along with the polysaccharide chain, may interact directly with outer membrane proteins. For example, OmpC and OmpF require LPS for trimerization [34-36] while OmpT required LPS for its proteolytic activity [37].

#### 1.4 *Neisseria meningitidis* and *N. gonorrhoeae*

This work specifically focuses on the gram-negative bacteria *Neisseria gonorrhoeae* (Gc) and *N. meningitides* (Nm), which colonize human mucosal tissue, causing the diseases gonorrhea and meningococcal meningitis respectively [38, 39]. Gc is of particular interest due to its increasing antibiotic resistance, prompting the CDC to classify the bacteria as an urgent threat to public health in 2013. Of the hundreds of thousands of new Gc infections in the United States each year, roughly 30% will be resistant to antibiotics, and a particular strain has emerged that has been classified a “super bug” due to its resistance to all available antibiotics [40]. This rise of antibiotic resistance conveys a need for alternative Gc and Nm treatments, and with it, the necessity for an increased understanding of the mechanism of Gc and Nm infections and colonization on the molecular level.

*Neisseria* interact with and colonize host tissues via a number of adhesions on their cell surface, namely pili and the proteins Opa and Opc [41]. Opa proteins (or opacity-associated proteins) and pili dictate tissue and host specificity, while Opa and Opc proteins aid in evasion of the host immune system via phase variations of these pathogen surface structures [42, 43]. Expressing multiple adhesions simultaneously promotes attachment and allows the colonization of different niches. Neisserial entry into host cells occurs via several adhesion-receptor interactions, including Opa-CEACAM binding which will be the focus of this work [41] (Figure 1.5). This Opa-CEACAM binding event triggers the phagocytosis of the bacteria into the host cell. Even low levels of internalization into a healthy host is remarkably beneficial for bacterial survival by providing an additional method to evade the host immune response. Internalization of the bacteria into host cells also provides the pathogen with an ideal nutrient source. Gonorrhoea and meningococcal meningitis onset occurs following bacterial colonization, where the bacteria enter systemic circulation and multiply within the host, likely involving many of the same adhesions which enable colonization [38, 39]. However, the role of Opa proteins in this disease progression is less studied than their role in Neisserial phagocytosis.



**Figure 1.5: Micrograph of *Neisseria gonorrhoeae* being phagocytosed into human cells.** Cover image of Billker, et al., reprinted with permission [44]. Falsely colored scanning electron micrograph of Gc bacteria (red) being phagocytosed into HeLa cells (green) that have been transfected to express CEACAM1 receptors, which can bind Opa proteins.

#### 1.4.1 Opacity associated proteins

The family of outer membrane proteins called Opa, or opacity associated proteins, were named as such due to an opaque phenotype of cells expressing Opa [45-47]. Opa proteins are found exclusively in the human pathogens Gc and Nm. These proteins mediate bacterial uptake into non-phagocytic and phagocytic human cells upon engaging host receptors - heparan sulfate proteoglycans (HSPGs) or carcino-embryonic antigen-like cellular adhesion molecules (CEACAMs) [41, 48, 49]. The expression of Opa proteins to the outer membrane of *E. coli* is sufficient to promote bacterial uptake into HeLa cells (transfected to express the respective host receptors), suggesting that Opa proteins are necessary to trigger the phagocytosis of *Neisseria* into host cells [50].

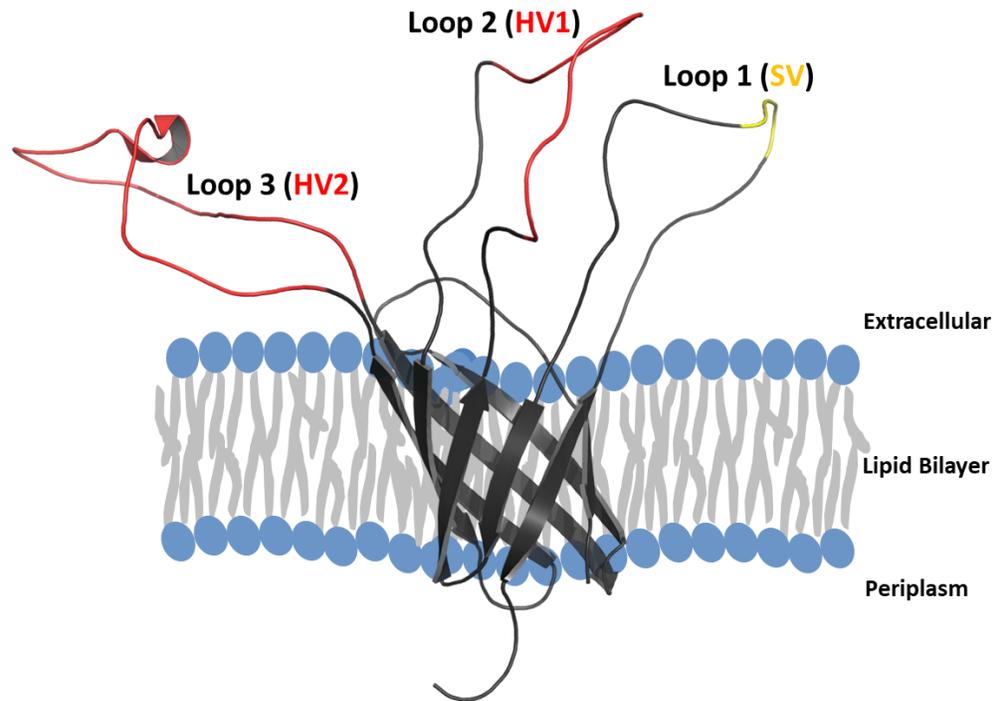
Opa proteins are encoded by a set of unlinked genes, where Gc contains up to eleven copies of *opa* genes and Nm contains three to four [46]. Point mutations and a high rate of recombination events within these genes enable a tremendous variety of Opa proteins (hundreds of different Opa sequences have been identified to date) [49, 51]. *Neisseria* are highly competent bacteria, enabling the cells to easily absorb and incorporate exogenous DNA, allowing *opa* genes to recombine with DNA from other organisms, as well as other varieties of *Neisseria* with different sequences of *opa* [49]. While this causes a tremendous range of possible *opa* alleles, specific Opa variants are often prevalent in *Neisseria*, likely driven by immunological selection and functional selection (with only specific Opa variants being able to bind host receptors), which would enhance bacterial colonization and survival [52].

The Opa family of proteins are remarkably varied in sequence, which may aid in diverse receptor engagement as well as immune evasion. This high sequence diversity is

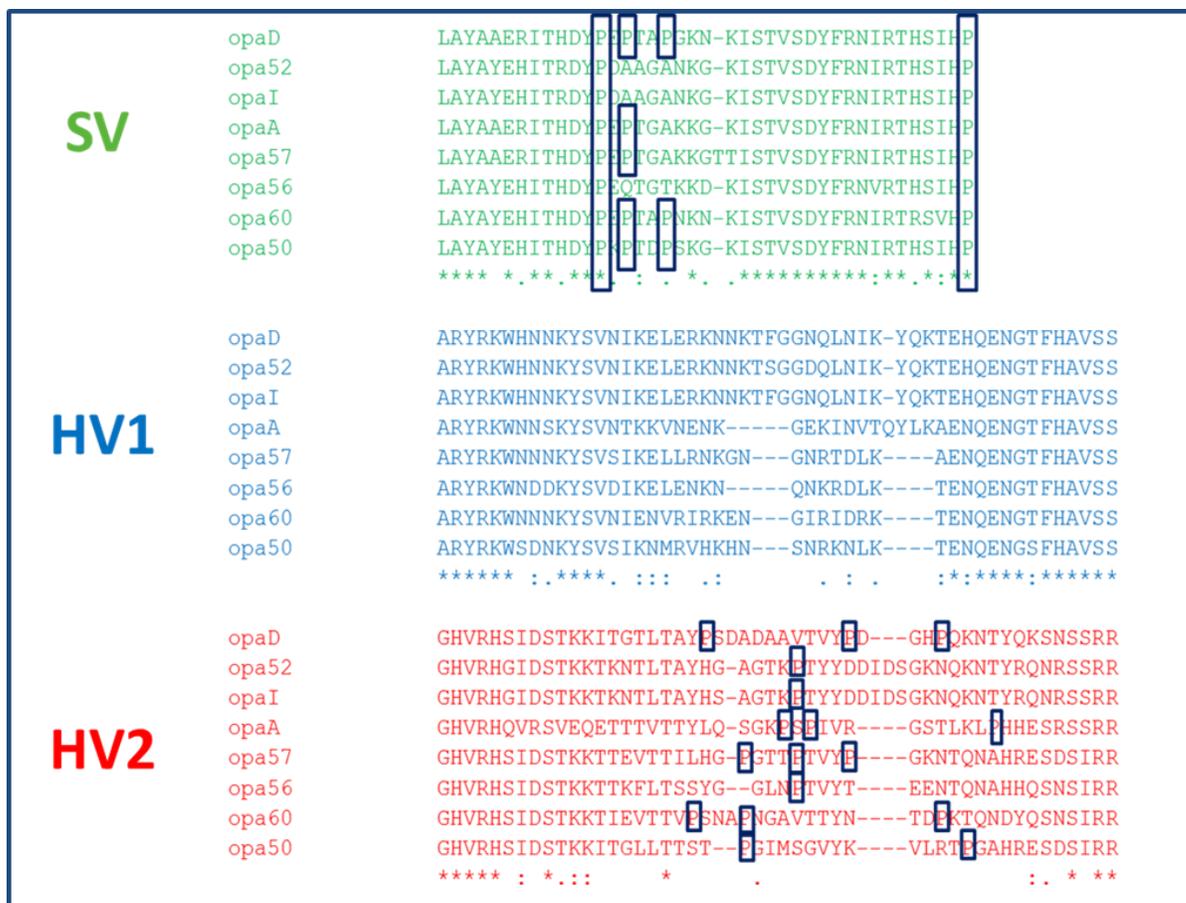
primarily evident in the extracellular loops of Opa. Opa proteins are eight stranded  $\beta$ -barrels with four extracellular loops [53] (Figure 1.6). Two of the loops contain regions termed hyper-variable (HV), due to their high sequence diversity among Opa variants [49]. Another loop has a region that is semi-variable (SV) among Opa proteins, while the fourth loop and barrel are largely conserved [54]. Hundreds of variable sequences have been identified, yielding thousands of possible Opa protein sequences. Regions HV1 and HV2 (located near the top of loops two and three) are believed to confer receptor specificity [55]. Switching these regions among Opa proteins eliminates receptor binding [55], and as such, both a defined sequence and a specific loop conformational ensemble must be required for receptor recognition. Both aspects of Opa receptor specificity and engagement are unknown, as multiple sequence alignments of Opa variable regions do not reveal a consensus binding motif (Figure 1.7). However, although conserved residues are not observed, the properties of many loop residues is maintained among different Opa proteins (though the specific locations and identities of these residues always vary). The most striking aspect of these loops is their large number of hydrophobic residues, which is highly unusual for the soluble domain of a membrane protein (Figure 1.8). As such, we hypothesize that these hydrophobic residues come together between different loops and loop strands, generating an ensemble of Opa loops, and thus, its receptor binding motif. Likewise, these loops always contain one to three proline residues within the HV regions, which may also promote a local ordering or secondary structure which is required for binding.

Opa expression is regulated through translational phase variable expression. The N-terminal region of *opa* genes contains tandem repeats of the leader peptide [CTCTT]<sub>n</sub>,

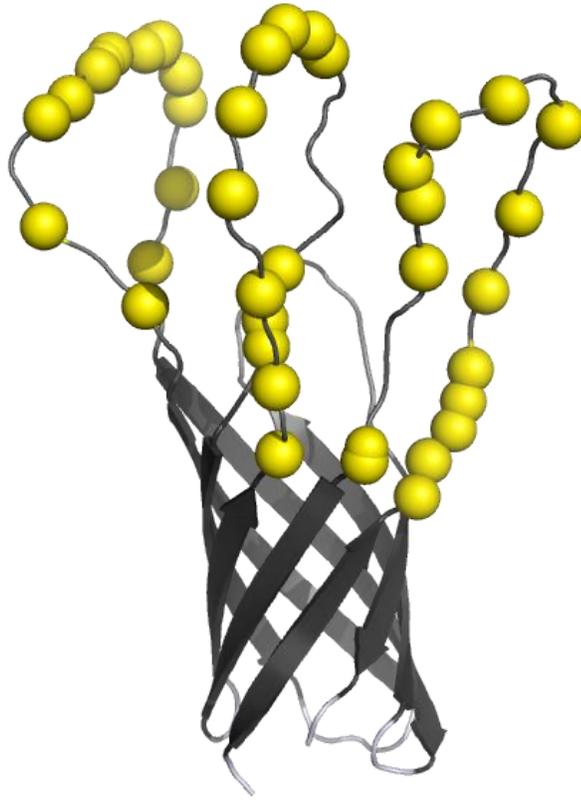
where the number of pentameric repeats within the *opa* gene determines whether the gene is in frame [38]. During translation, slipped-strand mispairings change the number of repeats, which can shift the sequence out of the reading frame, turning off the expression of Opa proteins. As such, a single *Neisseria* bacterium can express one, many, or no Opa variants at any time. This phase variation also causes Opa antigenic variation, where altering the surface expression of Opa proteins (which are also widely variable in sequence) aids in immune evasion. However, this also poses the question of how such variation in Opa expression and sequence yields species and tissue specific receptor engagement for *Neisseria* colonization of host cells. To begin to answer this complex question, we must first understand the molecular mechanism by which Opa proteins interact with host receptors.



**Figure 1.6: NMR structure of Opa<sub>60</sub> in detergent micelles (2MLH).** [53] Opa proteins are eight stranded  $\beta$ -barrels with four extracellular loops found in the outer membranes of Gc and Nm. Loop 1 contains a region that is semivariable in sequence among the Opa family of proteins (yellow) and loops 2 and 3 contain hypervariable regions (red) which interact with respective host receptors. The barrel and loop 4 are largely conserved among Opa proteins (black).



**Figure 1.7: Multiple sequence alignment of the semivariable and hypervariable regions of selected Opa proteins.** Sequence alignments of proposed receptor-engaging regions on Opa loops (HV1 and HV2) reveals minimal sequence conservation among eight Opa proteins (stars indicate conserved residues; a period or colon represents residues with weakly or very similar properties). As such, no receptor binding motifs have been identified in these regions. These regions contain an unusually high number of hydrophobic residues and all have several prolines (boxed), although the specific locations of these residues vary within the loops. These residues likely impact Opa loop conformations and thus receptor binding.



**Figure 1.8: Hydrophobic residues in the extracellular loops of Opa<sub>60</sub>.** Opa loops contain a high number of extracellular hydrophobic residues (shown as yellow spheres) that are diverse in sequence identity and location but are always present.

#### 1.4.2 Carcino-embryonic antigen-like cellular adhesion molecules

This work will focus on Opa interactions with CEACAMs, which are a set of mammalian proteins associated with numerous intercellular binding interactions affecting both normal and pathogenic processes. Homophilic and heterophilic CEACAM interactions mediate cellular adhesion, where CEACAMs play roles in angiogenesis [56], cell proliferation [57], cell motility [58], invasion [59], infection and inflammation [60]. Twelve CEACAMs have been identified in humans to date, all of which contain one variable V-like Ig domain (termed the N-domain) followed by a varying number (none or up to six) of constant C2-like Ig domains (A or B domains) and a membrane anchor (either a single helical transmembrane domain or a glycosylphosphatidylinositol (GPI) anchor) (Figure 1.9) [61, 62]. The extracellular N and A/B domains of CEACAMs are involved in the homophilic and heterophilic adhesion between CEACAMs within the same cell and between neighboring cells [63]. These domains also serve as targets for human and mammalian pathogens, including Neisserial Opa proteins, which bind specifically to the N-domain of various CEACAMs (CEACAM1, 3, 5, or 6) [64]. The C2-like domains of CEACAM1, CEACAM5, and CEACAM6 also mediate association to other receptors, primarily those involved in phosphatidylinositide 3-kinase (PI3K) induced bacterial uptake [65]. These domains of CEACAMs are also highly glycosylated, with more than half of the receptor's weight being from carbohydrates. The role of this glycosylation in CEACAM extracellular domains will be discussed further in Chapter 4.

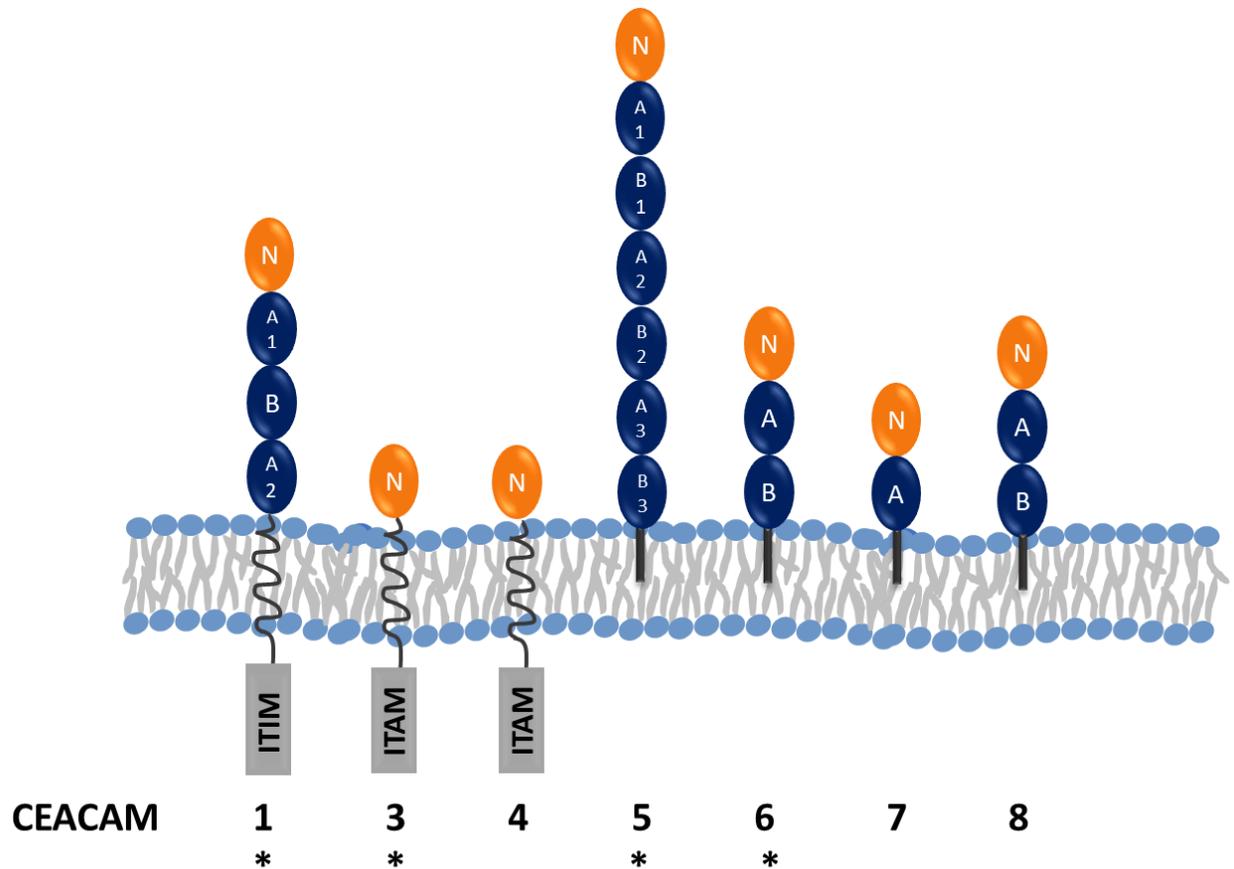
CEACAMs have a range of tissue localizations, where one or more types of CEACAM are present on most human cell types. CEACAM1 is the most widely expressed CEACAM protein, and is expressed in epithelial and endothelial cells, as well as lymphoid and

myeloid cells [66-68]. Both the downregulation and overexpression of CEACAM1 have been found in a variety of cancer cells, including colon [69, 70], prostate [71], breast [72], and bladder cancer [73] among others [74]. As such, CEACAM1 may play a role during tumor progression, although its precise role in cancer cells is unclear. In normal human tissues, CEACAM5 is localized primarily to epithelial cells of the colon [66]. CEACAM5 has also been identified in the carcinomas of lung, pancreatic, gallbladder, and bladder, with significant overexpression in colorectal and gastric cancers [66, 75, 76]. CEACAM5 can be a marker to identify colorectal tumors relative to normal tissue [77]. CEACAM6 is also widely distributed in many epithelial cells [78], and is over expressed in colorectal, breast, pancreatic, ovarian, gastric and lung cancer cells, making it another potential tumor biomarker [79].

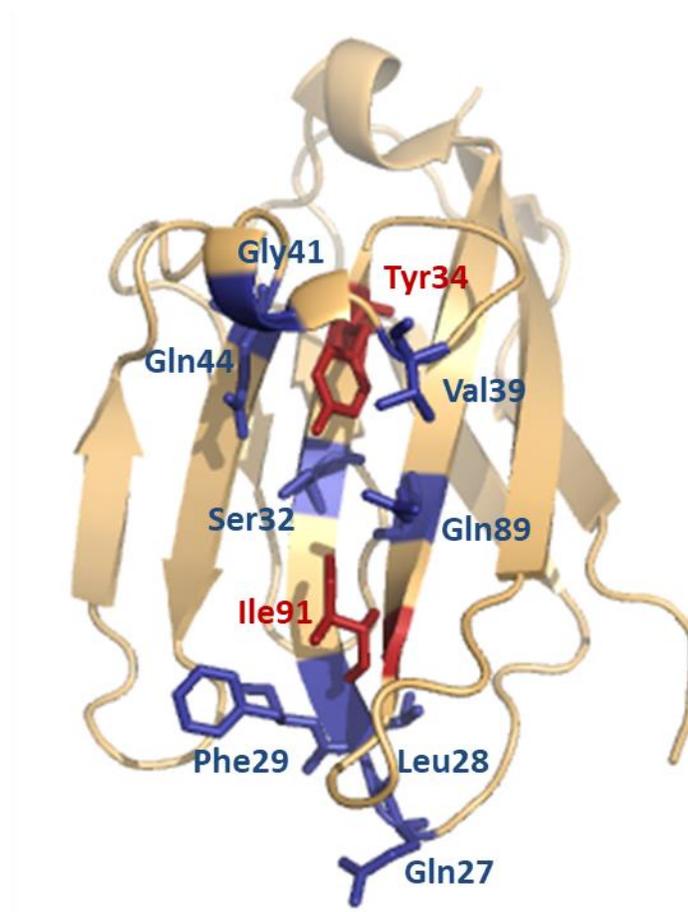
This work will focus specifically on CEACAM1, which has been found to be a target of numerous Opa proteins. CEACAM1 undergoes alternative splicing of its mRNA, resulting in 12 different human isoforms with different numbers of immunoglobulin-like domains, as well as two possible cytoplasmic lengths, termed the long (L) and short (S) tails (resulting from either the inclusion or exclusion of a specific exon) [80]. The functional role of a variation in the number of extracellular domains remains uncertain, but the chain length is known to have marked effects on the cell's response to CEACAM1 binding. CEACAM1-L isoforms also contain two immunoreceptor tyrosine-based inhibitor motifs (ITIMs), which transmit inhibitory signals in the cell. Conversely, CEACAM1-S isoforms lack this ITIM sequence and instead contain regions that can bind calmodulin, tropomyosin, and actin, indicating that these short chain isoforms are involved in interactions with the cytoskeleton. This cytoplasmic tail length is also highly regulated, as

dysregulation of L and S tail expression is found in colorectal, breast, and lung cancers [81, 82]. The N-domain of these different CEACAM1s (NCEACAM1, which bind Opa proteins) is conserved among CEACAM1s.

*Neisseria*, along with other pathogens which bind CEACAM proteins or their carbohydrates (*Moraxella catarrhalis* [83], *Haemophilus influenzae* [84], and *Escherichia coli* [85]), hijack these native roles of CEACAMs to promote internalization into host cells (or other pathogenic processes). Opa proteins bind specifically to one, many, or no variants of CEACAM [45]. Residues which bind Opa have been identified on NCEACAM1 via alanine-scanning mutagenesis and invasion assays [46, 86] (Figure 1.10). Ultimately, however, much is still unknown regarding the Opa-CEACAM binding interaction and how it relates to *Neisseria* phagocytosis into host human cells.



**Figure 1.9: Schematic of the CEACAM family of proteins.** CEACAMs contain an N-terminal IgV domain (orange), followed by a variable number of IgC domains (blue) of either A or B types. CEACAMs are linked to the membrane by either a single transmembrane helical domain or a GPI anchor (black). The cytoplasmic region of the transmembrane domains can have either an ITIM (immunoreceptor tyrosine-based inhibitory motif) or an ITAM (immunoreceptor tyrosine-based inhibitory activation) region. All domains of CEACAMs are highly glycosylated. Starred CEACAMs have been shown to interact with Opa proteins *in vivo* [45].



**Figure 1.10: N-terminal domain of CEACAM1 (2GK2).** [86] Residues which are necessary to engage all CEACAM-binding Opa proteins are shown in red, and those which are essential to bind only specific Opa proteins are shown in blue.

## 1.5 Dissertation overview

This dissertation will focus primarily on a specific Opa protein, Opa<sub>60</sub>, whose specific binding partners have been determined previously, primarily by invasion assays. The Columbus lab has previously determined the structure of Opa<sub>60</sub> and measured its binding affinity towards several variants of CEACAM *in vitro*. The work presented in this thesis provides insight into Opa<sub>60</sub> loop dynamics as well as Opa<sub>60</sub> interactions with CEACAM1. The primary method utilized in these studies of Opa<sub>60</sub> is Electron Paramagnetic Resonance (EPR), where the theory and applications of EPR will be discussed in Chapter 2. Chapter 3 will outline Opa<sub>60</sub> loop dynamics and the development of a hybrid EPR and MD technique to determine the conformational ensemble of the dynamic Opa loops. Chapter 4 will focus on methods to generate a monomeric, glycosylated N-terminal domain of CEACAM1, as well as measuring its homophilic dimerization and its binding to Opa<sub>60</sub> proteins. Finally, Chapter 5 will outline future directions with regards to Opa and CEACAM, the results of which will ultimately further our understanding of receptor-mediated pathogen uptake into phagocytic and non-phagocytic human cells.

1. Kendrew, J.C., et al., *A Three-Dimensional Model of the Myoglobin Molecule Obtained by X-Ray Analysis*. Nature, 1958. **181**: p. 662.
2. Blake, C.C.F., et al., *Structure of Hen Egg-White Lysozyme: A Three-dimensional Fourier Synthesis at 2 Å Resolution*. Nature, 1965. **206**: p. 757.
3. Braun, W., et al., *COMBINED USE OF PROTON-PROTON OVERHAUSER ENHANCEMENTS AND A DISTANCE GEOMETRY ALGORITHM FOR DETERMINATION OF POLYPEPTIDE CONFORMATIONS - APPLICATION TO MICELLE-BOUND GLUCAGON*. Biochimica Et Biophysica Acta, 1981. **667**(2): p. 377-396.
4. Deisenhofer, J., et al., *Structure of the protein subunits in the photosynthetic reaction centre of Rhodospseudomonas viridis at 3Å resolution*. Nature, 1985. **318**(6047): p. 618-24.
5. Fischer, E., *Einfluss der Configuration auf die Wirkung der Enzyme*. Berichte der deutschen chemischen Gesellschaft, 1894. **27**(3): p. 2985-2993.
6. Oldfield, C.J., et al., *Comparing and combining predictors of mostly disordered proteins*. Biochemistry, 2005. **44**(6): p. 1989-2000.
7. Austin, R.H., et al., *Dynamics of ligand binding to myoglobin*. Biochemistry, 1975. **14**(24): p. 5355-5373.
8. Henzler-Wildman, K. and D. Kern, *Dynamic personalities of proteins*. Nature, 2007. **450**(7172): p. 964-972.
9. Claxton, D.P., et al., *Chapter Twelve - Navigating Membrane Protein Structure, Dynamics, and Energy Landscapes Using Spin Labeling and EPR Spectroscopy*, in *Methods in Enzymology*, P.Z. Qin and K. Warncke, Editors. 2015, Academic Press. p. 349-387.

10. Koshland, D.E., *Application of a Theory of Enzyme Specificity to Protein Synthesis*. Proceedings of the National Academy of Sciences of the United States of America, 1958. **44**(2): p. 98-104.
11. Lindorff-Larsen, K., et al., *Simultaneous determination of protein structure and dynamics*. Nature, 2005. **433**(7022): p. 128-32.
12. Tompa, P. and M. Fuxreiter, *Fuzzy complexes: polymorphism and structural disorder in protein-protein interactions*. Trends Biochem Sci, 2008. **33**(1): p. 2-8.
13. Zimmermann, J., et al., *Antibody evolution constrains conformational heterogeneity by tailoring protein dynamics*. Proceedings of the National Academy of Sciences of the United States of America, 2006. **103**(37): p. 13722-13727.
14. Thorpe, I.F. and C.L. Brooks, *Molecular evolution of affinity and flexibility in the immune system*. Proceedings of the National Academy of Sciences of the United States of America, 2007. **104**(21): p. 8821-8826.
15. Keskin, O., *Binding induced conformational changes of proteins correlate with their intrinsic fluctuations: a case study of antibodies*. BMC Structural Biology, 2007. **7**: p. 31-31.
16. James, L.C., P. Roversi, and D.S. Tawfik, *Antibody multispecificity mediated by conformational diversity*. Science, 2003. **299**(5611): p. 1362-7.
17. O'Brien, P.J. and D. Herschlag, *Catalytic promiscuity and the evolution of new enzymatic activities*. Chem Biol, 1999. **6**(4): p. R91-r105.
18. James, L.C. and D.S. Tawfik, *Conformational diversity and protein evolution--a 60-year-old hypothesis revisited*. Trends Biochem Sci, 2003. **28**(7): p. 361-8.
19. Mannige, R.V., *Dynamic New World: Refining Our View of Protein Structure, Function and Evolution*. Proteomes, 2014. **2**(1): p. 128-153.

20. Wallin, E. and G. von Heijne, *Genome-wide analysis of integral membrane proteins from eubacterial, archaean, and eukaryotic organisms*. Protein Science, 1998. **7**(4): p. 1029-1038.
21. Fagerberg, L., et al., *Prediction of the human membrane proteome*. Proteomics, 2010. **10**(6): p. 1141-1149.
22. Overington, J.P., B. Al-Lazikani, and A.L. Hopkins, *Opinion - How many drug targets are there?* Nature Reviews Drug Discovery, 2006. **5**(12): p. 993-996.
23. Patching, S.G., *NMR structures of polytopic integral membrane proteins*. Molecular Membrane Biology, 2011. **28**(6): p. 370-397.
24. Dutta, R.C., *Drug carriers in pharmaceutical design: Promises and progress*. Current Pharmaceutical Design, 2007. **13**(7): p. 761-769.
25. Kellenberger, E. and A. Ryter, *Cell Wall and Cytoplasmic Membrane of Escherichia coli*. The Journal of Biophysical and Biochemical Cytology, 1958. **4**(3): p. 323-326.
26. Zhai, Y.F. and M.H. Saier, *The beta-barrel finder (BBF) program, allowing identification of outer membrane beta-barrel proteins encoded within prokaryotic genomes*. Protein Science, 2002. **11**(9): p. 2196-2207.
27. Wimley, W.C., *Toward genomic identification of beta-barrel membrane proteins: Composition and architecture of known structures*. Protein Science, 2002. **11**(2): p. 301-312.
28. Wimley, W.C., *The versatile beta-barrel membrane protein*. Current Opinion in Structural Biology, 2003. **13**(4): p. 404-411.
29. Chou, K.C. and L. Carlacci, *ENERGETIC APPROACH TO THE FOLDING OF ALPHA-BETA BARRELS*. Proteins-Structure Function and Genetics, 1991. **9**(4): p. 280-295.

30. van den Berg, B., *The FadL family: unusual transporters for unusual substrates*. Current Opinion in Structural Biology, 2005. **15**(4): p. 401-407.
31. Choutko, A., et al., *Membrane protein dynamics in different environments: simulation study of the outer membrane protein X in a lipid bilayer and in a micelle*. European Biophysics Journal with Biophysics Letters, 2011. **40**(1): p. 39-58.
32. Tamaki, S., T. Sato, and M. Matsushashi, *Role of Lipopolysaccharides in Antibiotic Resistance and Bacteriophage Adsorption of Escherichia coli K-12*. Journal of Bacteriology, 1971. **105**(3): p. 968-975.
33. Sanderson, K.E., et al., *Permeability of lipopolysaccharide-deficient (rough) mutants of Salmonella typhimurium to antibiotics, lysozyme, and other agents*. Can J Microbiol, 1974. **20**(8): p. 1135-45.
34. Ried, G., I. Hindennach, and U. Henning, *Role of lipopolysaccharide in assembly of Escherichia coli outer membrane proteins OmpA, OmpC, and OmpF*. Journal of Bacteriology, 1990. **172**(10): p. 6048-6053.
35. Sen, K. and H. Nikaido, *In vitro trimerization of OmpF porin secreted by spheroplasts of Escherichia coli*. Proceedings of the National Academy of Sciences of the United States of America, 1990. **87**(2): p. 743-747.
36. Sen, K. and H. Nikaido, *Lipopolysaccharide structure required for in vitro trimerization of Escherichia coli OmpF porin*. Journal of Bacteriology, 1991. **173**(2): p. 926-928.
37. Kramer, R.A., et al., *Lipopolysaccharide regions involved in the activation of Escherichia coli outer membrane protease OmpT*. Eur J Biochem, 2002. **269**(6): p. 1746-52.
38. Merz, A.J. and M. So, *Interactions of pathogenic Neisseriae with epithelial cell membranes*. Annual Review of Cell and Developmental Biology, 2000. **16**: p. 423-+.

39. Diaz Romero, J. and I.M. Outshoorn, *Current status of meningococcal group B vaccine candidates: capsular or noncapsular?* Clinical Microbiology Reviews, 1994. **7**(4): p. 559-575.
40. Ohnishi, M., et al., *THE NEW SUPERBUG NEISSERIA GONORRHOEAE MAKES GONORRHOEA UNTREATABLE?-FIRST HIGH-LEVEL CEFTRIAXONE RESISTANCE WORLDWIDE AND PUBLIC HEALTH IMPORTANCE.* Sexually Transmitted Infections, 2011. **87**: p. A76-A76.
41. Hauck, C.R. and T.F. Meyer, *'Small' talk: Opa proteins as mediators of Neisseria-host-cell communication.* Current Opinion in Microbiology, 2003. **6**(1): p. 43-49.
42. Segal, E., et al., *Antigenic variation of gonococcal pilus involves assembly of separated silent gene segments.* Proceedings of the National Academy of Sciences of the United States of America, 1986. **83**(7): p. 2177-2181.
43. Stern, A., et al., *Opacity genes in Neisseria gonorrhoeae: control of phase and antigenic variation.* Cell, 1986. **47**(1): p. 61-71.
44. Billker, O., et al., *Distinct mechanisms of internalization of Neisseria gonorrhoeae by members of the CEACAM receptor family involving Rac1- and Cdc42-dependent and -independent pathways.* Embo Journal, 2002. **21**(4): p. 560-571.
45. Dehio, C., S.D. Gray-Owen, and T.F. Meyer, *The role of neisserial Opa proteins in interactions with host cells.* Trends in Microbiology, 1998. **6**(12): p. 489-495.
46. Virji, M., et al., *Critical determinants of host receptor targeting by Neisseria meningitidis and Neisseria gonorrhoeae: identification of Opa adhesiotopes on the N-domain of CD66 molecules.* Molecular Microbiology, 1999. **34**(3): p. 538-551.

47. Ball, L.M. and A.K. Criss, *Constitutively Opa-Expressing and Opa-Deficient Neisseria gonorrhoeae Strains Differentially Stimulate and Survive Exposure to Human Neutrophils*. Journal of Bacteriology, 2013. **195**(13): p. 2982-2990.
48. Criss, A.K. and H.S. Seifert, *A bacterial siren song: intimate interactions between Neisseria and neutrophils*. Nature Reviews Microbiology, 2012. **10**(3): p. 178-190.
49. Sadarangani, M., A.J. Pollard, and S.D. Gray-Owen, *Opa proteins and CEACAMs: pathways of immune engagement for pathogenic Neisseria*. Fems Microbiology Reviews, 2011. **35**(3): p. 498-514.
50. Bos, M.P., F. Grunert, and R.J. Belland, *Differential recognition of members of the carcinoembryonic antigen family by Opa variants of Neisseria gonorrhoeae*. Infection and Immunity, 1997. **65**(6): p. 2353-2361.
51. Bilek, N., C.A. Ison, and B.G. Spratt, *Relative Contributions of Recombination and Mutation to the Diversification of the opa Gene Repertoire of Neisseria gonorrhoeae*. Journal of Bacteriology, 2009. **191**(6): p. 1878-1890.
52. Callaghan, M.J., K.A. Jolley, and M.C.J. Maiden, *Opacity-Associated Adhesin Repertoire in Hyperinvasive Neisseria meningitidis*. Infection and Immunity, 2006. **74**(9): p. 5085-5094.
53. Fox, D.A., et al., *Structure of the Neisserial Outer Membrane Protein Opa(60): Loop Flexibility Essential to Receptor Recognition and Bacterial Engulfment*. Journal of the American Chemical Society, 2014. **136**(28): p. 9938-9946.
54. Malorny, B., et al., *Sequence diversity, predicted two-dimensional protein structure, and epitope mapping of neisserial Opa proteins*. Journal of Bacteriology, 1998. **180**(5): p. 1323-1330.

55. Bos, M.P., et al., *Carcinoembryonic antigen family receptor recognition by gonococcal Opa proteins requires distinct combinations of hypervariable Opa protein domains*. *Infection and Immunity*, 2002. **70**(4): p. 1715-1723.
56. Horst, A.K., et al., *Carcinoembryonic antigen-related cell adhesion molecule 1 modulates vascular remodeling in vitro and in vivo*. *The Journal of Clinical Investigation*, 2006. **116**(6): p. 1596-1605.
57. Scheffrahn, I., et al., *Control of density-dependent, cell state-specific signal transduction by the cell adhesion molecule CEACAM1, and its influence on cell cycle regulation*. *Experimental Cell Research*, 2005. **307**(2): p. 427-435.
58. Klaile, E., et al., *CEACAM1 functionally interacts with filamin A and exerts a dual role in the regulation of cell migration*. *Journal of Cell Science*, 2005. **118**(23): p. 5513-5524.
59. Ebrahimnejad, A., et al., *CEACAM1 Enhances Invasion and Migration of Melanocytic and Melanoma Cells*. *The American Journal of Pathology*, 2004. **165**(5): p. 1781-1787.
60. Gray-Owen, S.D. and R.S. Blumberg, *CEACAM1: contact-dependent control of immunity*. *Nature Reviews Immunology*, 2006. **6**: p. 433.
61. Beauchemin, N., et al., *Redefined nomenclature for members of the carcinoembryonic antigen family*. *Exp Cell Res*, 1999. **252**(2): p. 243-9.
62. Hefta, S.A., et al., *Carcinoembryonic antigen is anchored to membranes by covalent attachment to a glycosylphosphatidylinositol moiety: identification of the ethanolamine linkage site*. *Proc Natl Acad Sci U S A*, 1988. **85**(13): p. 4648-52.
63. Obrink, B., *CEA adhesion molecules: multifunctional proteins with signal-regulatory properties*. *Curr Opin Cell Biol*, 1997. **9**(5): p. 616-26.
64. McCaw, S.E., E.H. Liao, and S.D. Gray-Owen, *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): p. 2742-2752.
65. Voges, M., et al., *Extracellular IgC2 constant domains of CEACAMs mediate PI3K sensitivity during uptake of pathogens*. PLoS One, 2012. **7**(6): p. e39908.
66. Hammarstrom, S., *The carcinoembryonic antigen (CEA) family: structures, suggested functions and expression in normal and malignant tissues*. Semin Cancer Biol, 1999. **9**(2): p. 67-81.
67. Frangsmyr, L., et al., *Cell- and region-specific expression of biliary glycoprotein and its messenger RNA in normal human colonic mucosa*. Cancer Res, 1995. **55**(14): p. 2963-7.
68. Prall, F., et al., *CD66a (BGP), an adhesion molecule of the carcinoembryonic antigen family, is expressed in epithelium, endothelium, and myeloid cells in a wide range of normal human tissues*. J Histochem Cytochem, 1996. **44**(1): p. 35-41.
69. Neumaier, M., et al., *Biliary glycoprotein, a potential human cell adhesion molecule, is down-regulated in colorectal carcinomas*. Proc Natl Acad Sci U S A, 1993. **90**(22): p. 10744-8.
70. Rosenberg, M., et al., *The expression of mouse biliary glycoprotein, a carcinoembryonic antigen-related gene, is down-regulated in malignant mouse tissues*. Cancer Res, 1993. **53**(20): p. 4938-45.
71. Busch, C., et al., *Down-regulation of CEACAM1 in human prostate cancer: correlation with loss of cell polarity, increased proliferation rate, and Gleason grade 3 to 4 transition*. Hum Pathol, 2002. **33**(3): p. 290-8.
72. Huang, J., et al., *Expression of biliary glycoprotein (CD66a) in normal and malignant breast epithelial cells*. Anticancer Res, 1998. **18**(5a): p. 3203-12.

73. Oliveira-Ferrer, L., et al., *Dual role of carcinoembryonic antigen-related cell adhesion molecule 1 in angiogenesis and invasion of human urinary bladder cancer*. *Cancer Res*, 2004. **64**(24): p. 8932-8.
74. Lu, R., et al., *Tumor Angiogenesis Mediated by Myeloid Cells Is Negatively Regulated by CEACAM1* (vol 72, pg 2239, 2012). *Cancer Research*, 2016. **76**(4): p. 984-985.
75. Jothy, S., S.Y. Yuan, and K. Shiota, *Transcription of carcinoembryonic antigen in normal colon and colon carcinoma. In situ hybridization study and implication for a new in vivo functional model*. *Am J Pathol*, 1993. **143**(1): p. 250-7.
76. Kodera, Y., et al., *Expression of carcinoembryonic antigen (CEA) and nonspecific crossreacting antigen (NCA) in gastrointestinal cancer; the correlation with degree of differentiation*. *Br J Cancer*, 1993. **68**(1): p. 130-6.
77. Tiernan, J.P., et al., *Carcinoembryonic antigen is the preferred biomarker for in vivo colorectal cancer targeting*. *Br J Cancer*, 2013. **108**(3): p. 662-7.
78. Scholzel, S., et al., *Carcinoembryonic antigen family members CEACAM6 and CEACAM7 are differentially expressed in normal tissues and oppositely deregulated in hyperplastic colorectal polyps and early adenomas*. *Am J Pathol*, 2000. **156**(2): p. 595-605.
79. Kim, K.S., et al., *Overexpression and clinical significance of carcinoembryonic antigen-related cell adhesion molecule 6 in colorectal cancer*. *Clinica Chimica Acta*, 2013. **415**: p. 12-19.
80. Kammerer, R. and W. Zimmermann, *Coevolution of activating and inhibitory receptors within mammalian carcinoembryonic antigen families*. *BMC Biol*, 2010. **8**: p. 12.
81. Turbide, C., et al., *Optimal ratios of biliary glycoprotein isoforms required for inhibition of colonic tumor cell growth*. *Cancer Res*, 1997. **57**(13): p. 2781-8.

82. Wang, L., et al., *C-CAM1, a candidate tumor suppressor gene, is abnormally expressed in primary lung cancers*. Clin Cancer Res, 2000. **6**(8): p. 2988-93.
83. Hill, D.J. and M. Virji, *A novel cell-binding mechanism of Moraxella catarrhalis ubiquitous surface protein UspA: specific targeting of the N-domain of carcinoembryonic antigen-related cell adhesion molecules by UspA1*. Mol Microbiol, 2003. **48**(1): p. 117-29.
84. Hill, D.J., et al., *The variable P5 proteins of typeable and non-typeable Haemophilus influenzae target human CEACAM1*. Molecular Microbiology, 2001. **39**(4): p. 850-862.
85. Sauter, S.L., et al., *Identification of the specific oligosaccharide sites recognized by type 1 fimbriae from Escherichia coli on nonspecific cross-reacting antigen, a CD66 cluster granulocyte glycoprotein*. J Biol Chem, 1993. **268**(21): p. 15510-6.
86. Fedarovich, A., et al., *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**: p. 971-979.

## Chapter 2: Electron Paramagnetic Resonance Spectroscopy

### 2.1 Introduction to EPR

The interaction of the magnetic moment of electrons with oscillating external electromagnetic radiation results in Electron Paramagnetic Resonance (EPR). Electromagnetic radiation is regarded as coupled electric ( $\mathbf{E}_1$ ) and magnetic ( $\mathbf{B}_1$ ) fields perpendicular to the direction of propagation. Electromagnetic radiation is composed of photons with no mass or net charge and can be thought of as wave packets with electromagnetic fields and spin angular momentum. Electric and magnetic components of their fields oscillate in a narrow frequency range,  $\nu$ , where the quantity  $h\nu$  (where  $h$  is the Planck constant) describes the energy of any given photon. For EPR, these fields oscillate at a frequency  $\nu$ , which ranges from  $10^9 - 10^{11} \text{ s}^{-1}$  (1-100 GHz). The electric field component of radiation interacts with molecules in most spectroscopic experiments (other than magnetic resonance); and for absorption to occur (1) the energy ( $h\nu$ ) of the quantum of radiation must equal the separation between energy levels of the molecule, and (2) the oscillating electric field component  $\mathbf{E}_1$  must interact with an oscillating electric dipole moment. Likewise, molecules with a magnetic dipole may interact with the magnetic component  $\mathbf{B}_1$  of electromagnetic radiation, which is the basis of magnetic resonance spectroscopy. In magnetic resonance experiments, an external magnetic field ( $\mathbf{B}$ ) is applied in addition to the intrinsic  $\mathbf{B}_1$ . This results in a splitting of energy levels where the electron spin is oriented either parallel or anti-parallel to the external magnetic field (Figure 2.1).

Electrons have an intrinsic magnetic dipole that originates from its spin [87], however in most systems electrons occur in pairs, causing the net moment to be zero. Thus, only

systems containing at least one unpaired electron possess the necessary net spin magnetic moment to interact with an electromagnetic field and be studied using EPR. EPR is applicable to such paramagnetic species, which have the required net electron angular momentum to be observed by EPR spectroscopy.

## 2.2 Free electron in an applied magnetic field

An electron can be thought of as a charged sphere with an intrinsic spin angular momentum ( $S$ ) and a permanent magnetic dipole moment ( $\mu$ ), which are related by:

$$\mu = -g_e\beta S \quad (2.1)$$

where  $g_e$  is the “g-factor” (which defines the electron’s response to an applied magnetic field) and  $\beta$  is the Bohr magneton:

$$\beta = \frac{e\hbar}{2mc} \quad (2.2)$$

where  $m$  and  $e$  are the mass and charge of an electron respectively,  $c$  is the velocity of light, and  $\hbar$  is Planck’s constant divided by  $2\pi$ . The intrinsic magnetic dipole of an electron on its own may be oriented in any direction; however, upon the application of an external magnetic field  $\mathbf{B}$ , the dipole experiences a torque in the direction perpendicular to  $\mathbf{B}$ . This torque experienced by the magnetic dipole moment causes the dipole to precess at a frequency proportional to  $\mathbf{B}$ . The electron’s precessional frequency induced by  $\mathbf{B}$  is the Larmor frequency,  $\omega$ . The magnetic dipole of an unpaired electron in  $\mathbf{B}$  rotates around the axis of the applied field at frequency  $\omega$  and can be oriented either along the axis of the field or anti-parallel to  $\mathbf{B}$ . EPR spectra are obtained by applying a constant wavelength while scanning the external magnetic field from low to high. When an electron is subjected to an

external magnetic field, the energy levels of degenerate spin states split. The split energy levels are characterized by the spin quantum numbers  $m_s = +1/2$  and  $m_s = -1/2$  and are dependent on the strength of  $\mathbf{B}$ , where the energies for the two spin states are:

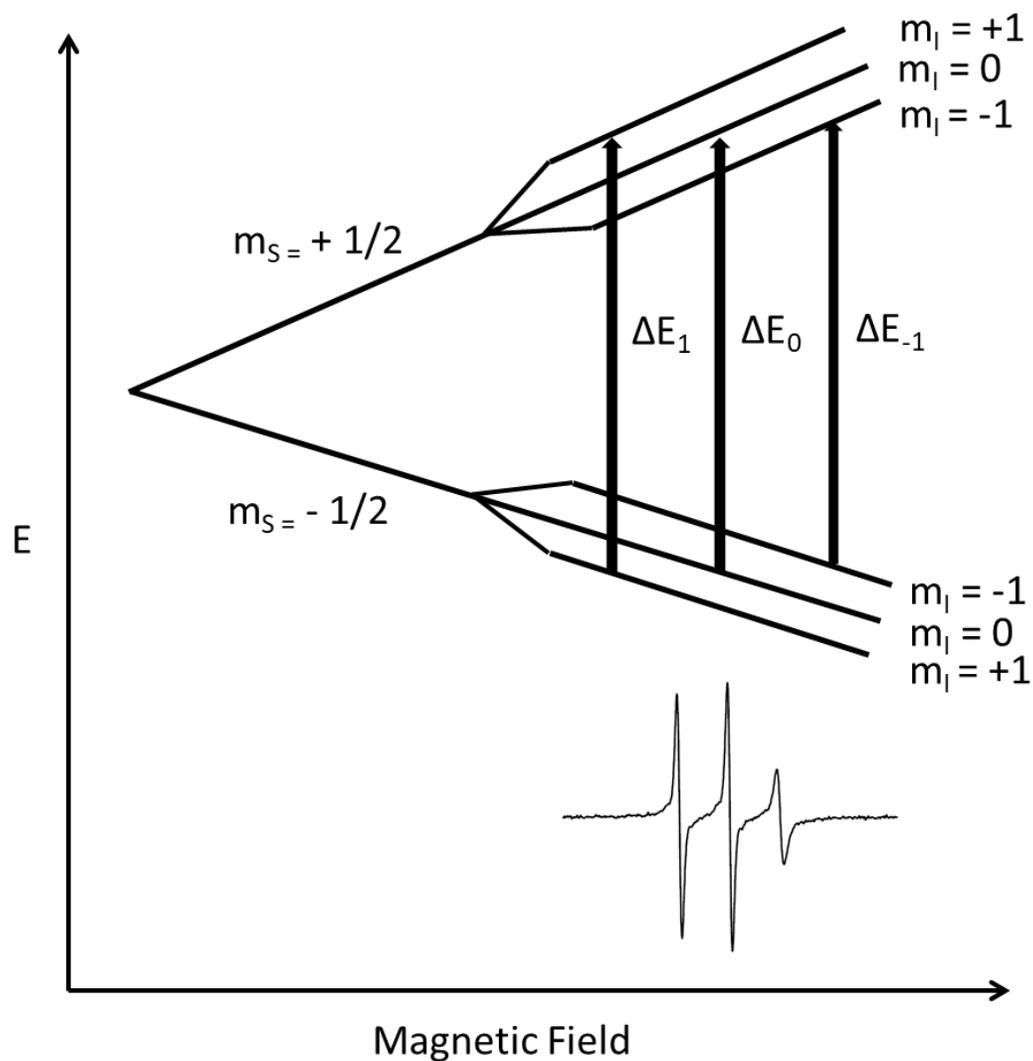
$$E = \pm \frac{1}{2} g_e |\beta| \mathbf{B} \quad (2.3)$$

Transitions between the two energy levels are induced with an applied electromagnetic field  $\mathbf{B}$  at a specific frequency  $\nu$  where the photon energy  $h\nu$  matches the energy level separation (Figure 2.1):

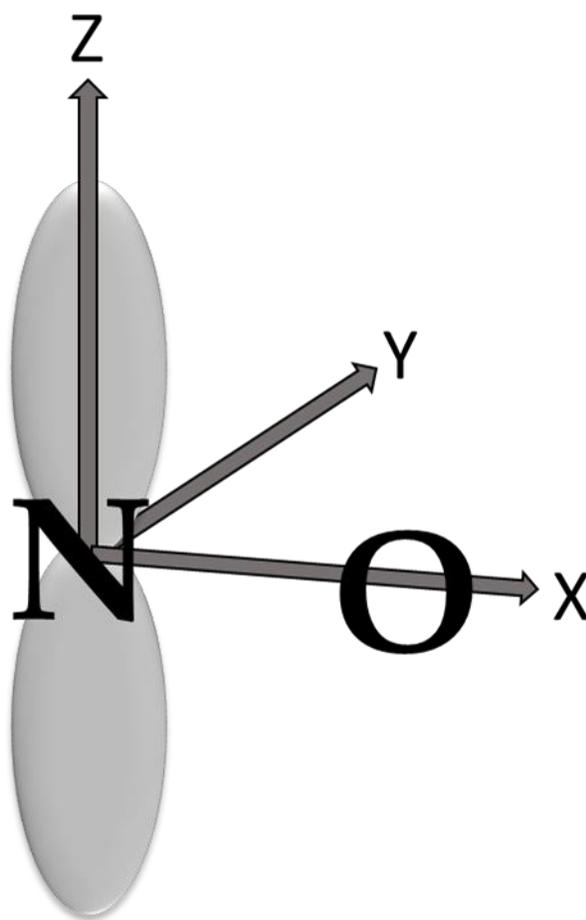
$$\Delta E = h\nu = g_e |\beta| \mathbf{B} \quad (2.4)$$

For electrons, and hence in EPR, the energy of the magnetic dipole requires frequencies in the microwave region (at  $\omega$ ) for resonance to occur.

The g-factor (equal to 2.0023 for a free electron) is a measure of the local field experienced by an electron. The combined effect of an external magnetic field and spin-orbit coupling (the interaction between the spin and orbital magnetic moments) introduces a small amount of orbital angular momentum, which alters the effective magnetic moment of the electron, causing the g value to deviate from the free spin value. The magnitude of the deviation depends on the orientation of the external field with respect to the molecular axis of the system (Figure 2.2). In solution, the orientation dependence of g is averaged by fast molecular motion.



**Figure 2.1: Energy level scheme of a nitroxide spin system in a magnetic field.** In the presence of an external magnetic field  $\mathbf{B}$ , a free electron splits into two energy levels,  $m_s = -1/2$  and  $m_s = +1/2$ . Hyperfine interactions (Section 2.4) with the neighboring  $^{14}\text{N}$  nucleus further splits the energy levels, resulting in three possible transitions between spin states. Therefore, the resulting EPR spectrum (inset) consists of three peaks.



**Figure 2.2: Magnetic frame of a nitroxide.** The z-axis is parallel to the p-orbital of the N atom, the x-axis runs along the N-O bond, and the y axis is perpendicular to the x- and z- axes.

### 2.3 A-tensor and hyperfine interactions

The field experienced by each spin species is due to both the externally applied magnetic field  $\mathbf{B}$  and any local fields  $\mathbf{B}_{local}$  to produce the total effective field  $\mathbf{B}_{eff}$ :

$$\mathbf{B}_{eff} = \mathbf{B} + \mathbf{B}_{local} \quad (2.5)$$

Primary sources that contribute to  $\mathbf{B}_{local}$  are nearby electronic or nuclear dipoles, which generate local magnetic fields that act on the local  $\mathbf{B}_{eff}$  of the electron. The electron spin magnetic dipole can interact with neighboring nuclear dipole moments, which results in a splitting of energy levels called the *nuclear hyperfine interaction* or *hyperfine splitting*. The neighboring nuclei have a spin quantum number  $I$  and split into  $2I + 1$  energy levels. For the commonly used nitroxide spin label, the neighboring nuclei is a  $^{14}\text{N}$  atom, which has an  $I = 1$  and a multiplicity of 3 ( $M_I = +1, 0, \text{ and } -1$ ). The nuclear hyperfine interaction further splits the energy levels into three in this case, so three possible EPR transitions are observed (Figure 2.1).

The hyperfine interaction is typically anisotropic, with the exception being samples where the anisotropy is averaged via fast molecular motion (when the spin label correlation time  $\tau_c$  is less than approximately 100 ps). This hyperfine anisotropy occurs because the majority of the electron density is distributed in the p-orbital of the N atom that is aligned parallel to the z-axis of a nitroxide, so the dipolar coupling to the nuclei in that direction dominates (Figure 2.2). Isotropic EPR spectra are the result of a rapid, random reorientation of solute molecules. In such isotropic systems, changing the sample orientation relative to  $\mathbf{B}$  does not alter the spectrum, as all orientation specific magnetic fields (g and A tensors) are averaged. Systems with anisotropic magnetic properties, however, result in spectra that

have a dependence on the orientation relative to **B**. The three inherent perpendicular principle axes (x-, y-, and z-axes) together with the results measured along these directions describe the anisotropic properties of a sample, and thus the EPR line positions and splittings.

#### 2.4 Relaxation

Relaxation processes return an excited state to equilibrium, where more electrons transition from the high energy, antiparallel state ( $n_A$ ) to the parallel, lower energy state ( $n_B$ ) with the population between the two energy levels being described by the Boltzmann distribution:

$$\frac{n_A}{n_B} = e^{-E/kT} \quad (2.6)$$

where  $E = h\nu$ ,  $k$  is the Boltzmann constant, and  $T$  is the temperature in Kelvin. Irradiation of the spins at their resonant frequency yields non-equilibrium populations (absorption), where the magnitude of the EPR signal is proportional to the magnetization along the z-axis ( $M_z$ ). To maintain equilibrium with a slight population excess in the lower level, electrons in the higher energy state give up an  $h\nu$  quantum of energy to return to the lower level via spin relaxation processes. The processes by which spins return to their equilibrium populations along the +z axis is called spin-lattice or  $T_1$  relaxation, where energy is dissipated within the surroundings (lattice).

A second form of relaxation, called transverse relaxation or  $T_2$ , is the time necessary for dephasing to occur along the x-y axis as individual spins oscillate at different Larmor frequencies due to interactions with other spin systems (such as dipolar and exchange interactions). This changes the phases of individual electrons, and thus the net

magnetization which causes a reduction in signal intensity. Together these relaxation processes contribute to the EPR resonance linewidth ( $\Delta B$ ) where

$$\Delta B \propto \frac{1}{T_{1e}} + \frac{1}{T_{2e}} \quad (2.7)$$

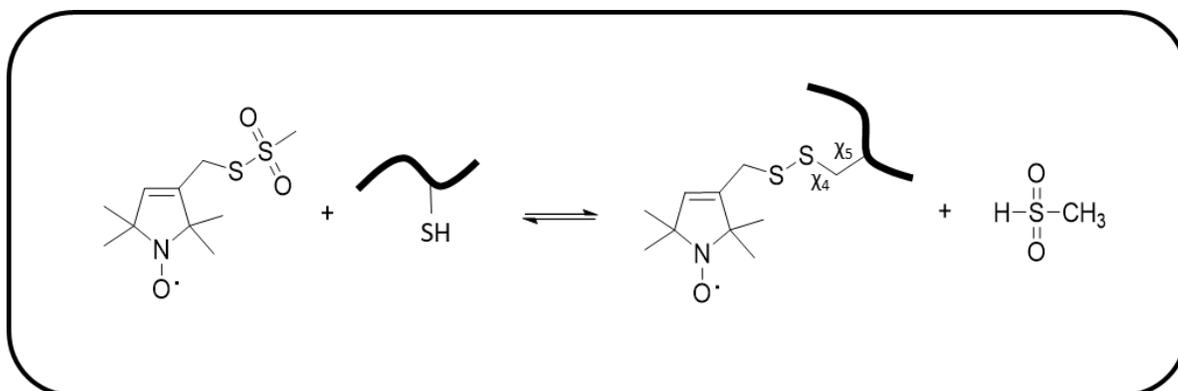
$T_{2e}$  is typically smaller than  $T_{1e}$ , thus the Lorentzian linewidth is generally determined by  $T_{2e}$ .

## 2.5 Applications of EPR

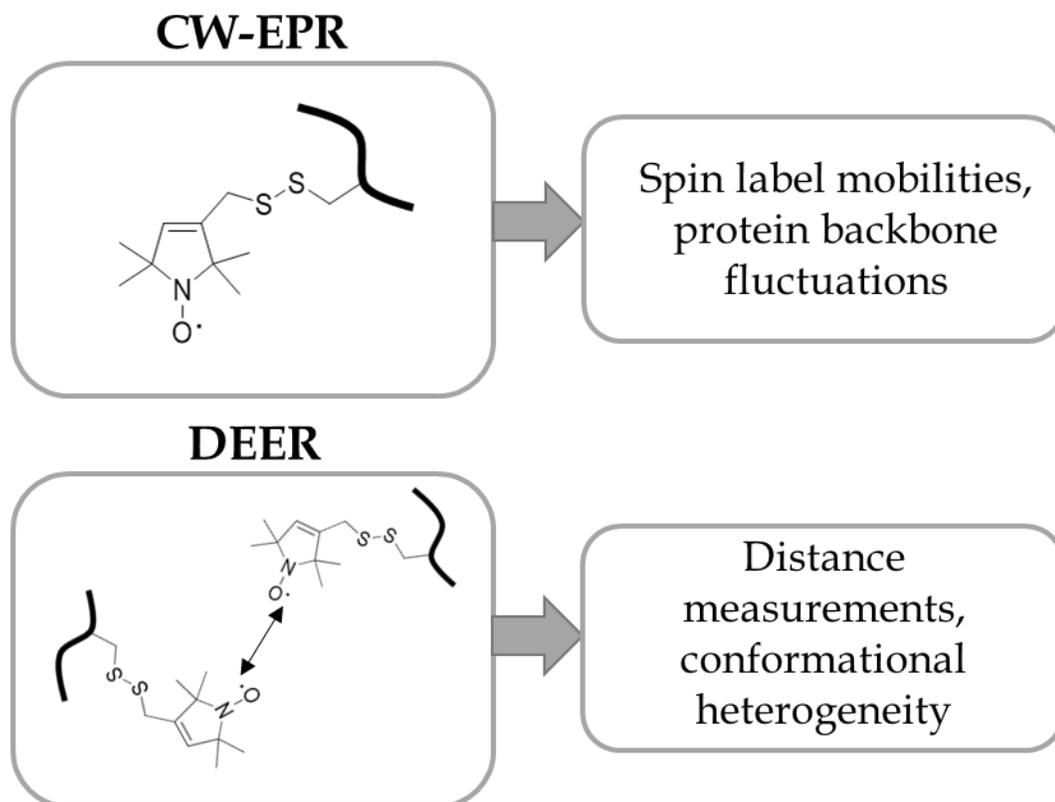
EPR is a highly advantageous technique, however most biological systems, such as proteins without bound metals, do not have the required paramagnetic species required to examine using EPR. Site-directed spin labeling (SDSL) provides a route to study such systems, where a sulfhydryl-reactive paramagnetic spin label reacts with a mutagenesis-introduced cysteine residue within the protein (Figure 2.3) [88, 89]. Any native, non-critical cysteine residues can be substituted with serine or alanine residues to allow site specific cysteine scanning. Thus, the lack of a native paramagnetic center in proteins provides an advantage, as researchers can selectively probe regions of interest (via SDSL) within the protein using EPR. Methanethiosulfonate (MTSL) spin labels are widely used for SDSL. When covalently attached to the protein (R1), the spin label is comparable in size to a tryptophan or phenylalanine side chain, and as such has a minimal impact on the native protein structure and biological activity compared other labeling methods [90, 91] (Figure 2.3).

Two types of EPR experiments will be explored in this dissertation: Continuous-Wave EPR (CW-EPR) and Double Electron Electron Resonance (DEER). These

techniques are utilized to probe dynamics and distance measurements respectively (Figure 2.4). Applying these EPR techniques to protein systems allows the observation of domain movements or motion of secondary structural elements from conformational transitions [92]. EPR methods are low-throughput and provide only moderate structural resolution, due to the selective introduction of a limited number of spin labels. However, EPR provides a marked advantage when studying membrane proteins due to an absence of size restrictions, enabling membrane proteins to be studied in a variety of membrane mimics.



**Figure 2.3: Schematic of site-directed spin labeling proteins.** A cysteine residue is introduced via mutagenesis at a region of interest within a protein. Thiol-reactive spin labels (typically methanethiosulfonate spin label, or MTSSL/R1) bind the cysteine residues, allowing EPR to be applied to proteins. Internal motions within the nitroxide largely originate from torsional oscillations about the dihedral angles  $\chi_4$  and  $\chi_5$ .



**Figure 2.4: EPR experiments and their applications.** *Adapted from [9].* Two EPR methods will be utilized in this dissertation. CW-EPR of a single nitroxide spin label is used to measure protein backbone fluctuations and DEER is used to generate distance distributions between two spin labels.

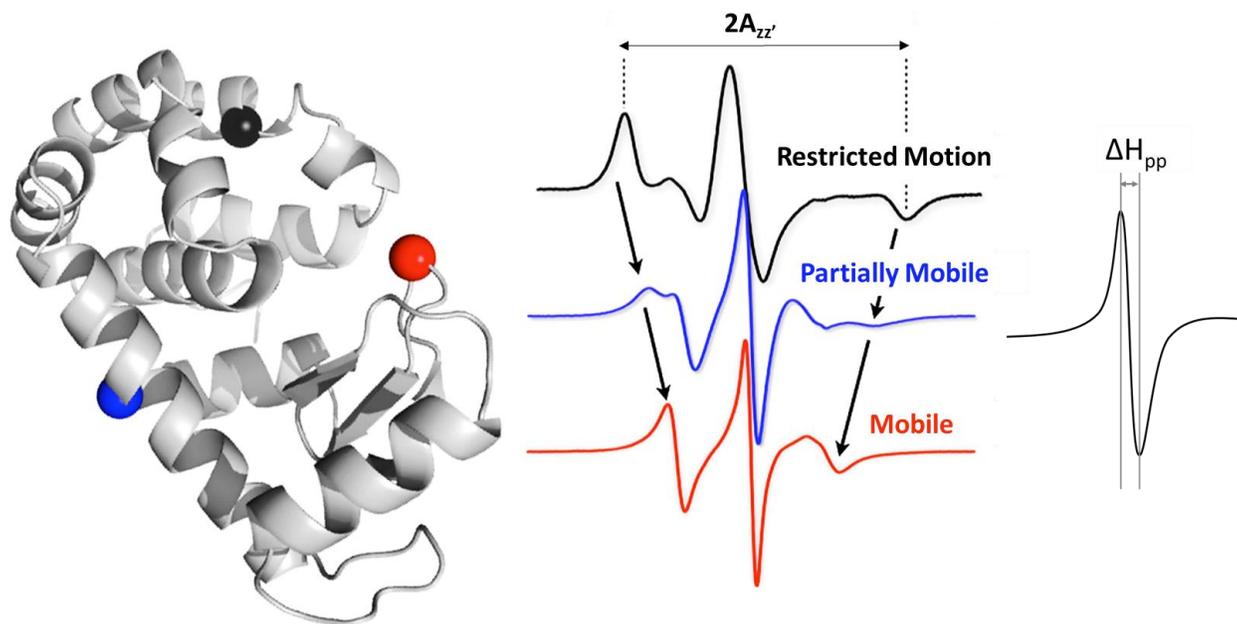
### 2.5.1 Continuous wave EPR and protein dynamics

Continuous wave (CW) EPR is a type of EPR spectroscopy where the external magnetic field  $\mathbf{B}$  is varied with a constant applied microwave frequency (due to limits in microwave electronics). The microwave diode detector of CW-EPR is sensitive to a broad frequency range, so to reduce to detected noise, the EPR absorption signal is modulated by sinusoidal modulation of the magnetic field at 100 kHz (phase sensitive detection). Only this modulation is detected, and as such the EPR signal is the first derivative of the absorption line with an amplitude proportional to the slope of the absorbance signal, which has a Lorentzian lineshape where the area under this line is proportional to the concentration of paramagnetic species. The Lorentzian lineshape is broadened by  $T_1$  and  $T_2$  relaxation but is dominated by  $T_2$  relaxation which is shorter than  $T_1$ .

For protein systems studied using CW-EPR, the lineshape is sensitive to the local steric environment of the spin label, where label dynamics originate from internal bond rotations within the nitroxide, local backbone fluctuations, and the overall molecular tumbling of the molecule [93]. Internal motions within the nitroxide largely originate from torsional oscillations about the dihedral angles  $\chi_4$  and  $\chi_5$  (Figure 2.3) [90]. The overall motion of the nitroxide is coupled to the motion of the protein backbone and the local structure. The degree of steric interaction experienced by the spin label with neighboring residues modulates the dynamic motions conveyed in the lineshape [90]. Only proteins of molecular weights of less than approximately 15 kDa have fast enough rotational tumbling to influence the observed lineshape. Together, the spectra report on motions of the spin label in the pico to nanosecond timescale (fast motions of  $< 2$  ns, intermediate 2-30 ns, and slow  $> 30$  ns) [94, 95].

Motional restrictions result in line broadening of the EPR spectra as anisotropic components become less averaged. Line broadening can be caused by any dynamic process in and around the paramagnetic center, including hindered rotation, tumbling of the molecule, and interactions with other paramagnetic species. Slow conformational changes can be observable as distinct conformer species. The peak to peak amplitude ( $\Delta H_{pp}$ , Figure 2.5) of the derivative line is inversely proportional to the square of linewidth, so small changes in linewidth can cause large changes in relative amplitudes of individual lines in spectrum. As the lines broaden, the amplitudes decrease proportionally.

Attaching a spin label to even an unstructured peptide results in some degree of motional restriction compared to free spin label, which is visually evident in the respective EPR lineshapes. Overall, an EPR spectrum can reflect on a protein's topology and dynamics, where characteristic lineshapes are visible for various secondary structure elements [90, 96]. While deconvolution of the dynamic modes of a spin label requires simulations of the EPR lineshape [97, 98], qualitative spectral differences between spin labels of buried, surface, or loop sites and those in tertiary contact may be sufficient for some protein cases (Figure 2.5). The inverse central linewidth can be used in these cases as a parameter to examine mobility of the protein backbone and the structural identity of that site, where linewidth increases with decreasing motion [90, 98]. Comparing the inverse central linewidth in the presence and absence of ligand can also be useful when identifying conformational changes that occur within a protein upon engaging receptor.



**Figure 2.5: EPR lineshapes reflect the protein environment of the spin label.** *Reprinted (adapted) with permission from [90]. Copyright 1996 American Chemical Society.* The structure of T4 lysozyme is shown, with R1 spin labels at various locations depicted as spheres. The EPR spectra of the black colored spin label is indicative of a buried region of restricted mobility, due to multiple contacts with residues on neighboring helices. Protein regions that are partially mobile, such as the blue label, have restricted backbone motion due to secondary structure elements while having limited contact with neighboring residues, allowing faster label motion with a restriction in amplitude. EPR spectra from spin labels within mobile loop regions, such as the above red label, reflect the degree of backbone mobility at the attached region.  $2A_{zz'}$  and  $\Delta H_{pp}$  report on the mobility of the spin label: anisotropic components are increasingly averaged as the label mobility increases and are visible with linebroadening as mobility is limited.

### 2.5.2 Double electron electron resonance to measure distances

The spin moments of two unpaired electrons can also interact, and because the magnitude of magnetic moments of electrons are approximately 2000 times greater than those of nuclei ( $\beta_e/\beta_n$ ) (due to a smaller electron mass), the dipolar interaction energy between electrons at an interparticle distance of  $r$  is also 2000 times greater [99]. Pulsed EPR is utilized to measure distances between electron spins by probing the strength of the measured dipolar coupling, which is inversely proportional to the interparticle distance cubed ( $r^{-3}$ ). Four-pulse Double Electron Electron Resonance Spectroscopy (DEER) experiments are widely used to measure distances between approximately 1.8 – 6 nm in biological systems [100, 101]. Rather than a single distance value, this technique provides a distance distribution, which in turn can be utilized to characterize conformational distributions of proteins and spatial relationships between domains or secondary structural elements [102].

Typical DEER experiments use a four-pulse sequence to extract the dipolar coupling, and thus the distance, between electron spins (A and B spins) [103] (Figure 2.6). In a four-pulse DEER experiment, a  $\pi/2$  microwave pulse (at a frequency,  $\nu_{\text{obs}}$ , that excites electron spin A, the observer spin) tips the net magnetization of spins A into the  $x$ - $y$  plane. The observer spins A lose phase coherence over time ( $\tau_1$ ) due to  $T_2$  relaxation of the spins in the  $x$ - $y$  plane. A  $\pi$  pulse at  $\nu_{\text{obs}}$  reestablishes phase coherence, which produces a spin echo over time (Hahn echo, Figure 2.7). A fraction of electron spins B that are coupled to observer spin A are excited by a  $\pi$  pump pulse (at a second microwave frequency,  $\nu_{\text{pump}}$ ), which inverts the magnetization of those B spins (at an inversion efficiency fraction  $\lambda$ ).

This inverts the local field imposed by spin B at the site of spin A and changes the frequency of spin A by electron-electron coupling  $\omega_{ee,i}$ :

$$\omega_{ee,i} = \left(\frac{C_i}{r_i^3}\right)(1 - 3 \cos^2 \theta_i) \quad (2.9)$$

where  $C_i$  is the product of the  $g$  values of spins A and B (52.2 MHz nm<sup>-3</sup> for nitroxides) and  $\theta_i$  is the angle between the spin-spin vector and the magnetic field. The inversion of magnetization of pump spin B, and, therefore, the dipole-dipole coupling frequency, creates a phase lag in spins A. The accrued phase lag of spins A is related to the dipolar interaction strength (which is distance-dependent) and the timing of the pump pulse (which affects how long spins A experience the dipolar field), so the pump pulse timing is varied to extract the dipolar coupling. A second  $\pi$  pulse at  $\nu_{\text{obs}}$  reestablishes phase coherence of the observer spins A to generate a refocused echo. The DEER signal is then the modulation of the refocused echo intensity as the timing of the pump pulse ( $t$ ) increases:

$$v(t) = \{1 - \lambda[1 - \cos(\omega_{ee,i} t)]\} \quad (2.10)$$

It is assumed that the spin pair is semi-isolated, where for a given observer spin A, only one spin B on the same protein/complex is presumed to be within the distance range that is measurable with DEER experiments. The time domain data  $v(t)$  contains both inter- and intra- molecular dipolar contributions, where inter-molecular dipolar interactions have a background function:

$$B(t) = e^{-kt^{D/3}} \quad (2.11)$$

where  $k$  is the decay rate constant and  $D$  is the dimensionality of the space (3 for soluble proteins). This intermolecular background is subtracted, generating a signal from only the intramolecular dipolar coupling of interest [104].

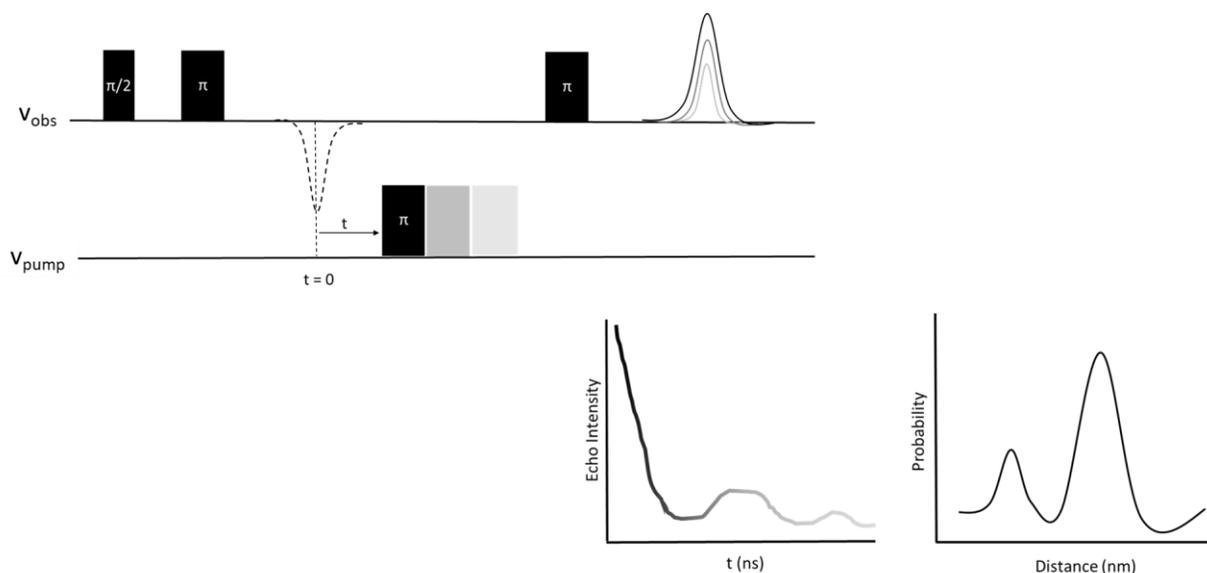
Transforming the intramolecular DEER signal to a distance distribution  $P(r)$  utilizes a fitting approach based on distance models. Fitting the time domain data is an ill-posed mathematical problem, where small changes due to noise or errors in background subtraction can lead to large changes in the output  $P(r)$  [102]. Tikhonov regularization stabilizes the solution by calculating an optimum regularization parameter ( $\alpha$ ) for a sample that is estimated from the corner an L-curve [105] [104, 106]. The L-curve is a plot of the log of the smoothness of the distribution [ $\eta(\alpha)$ ] versus the log of the mean squared deviation [ $\rho(\alpha)$ ] (which incorporates the simulated distribution). The best fit balances smoothness and resolution, which occurs at an  $\alpha$  value at the corner of the L-curve. The resulting distance distribution at the chosen  $\alpha$  yields information on the distances between two spin labels, which is indicative of the conformational distribution of the spin label attached to the protein backbone. The distributions can be characterized by a mean distance and width.

The rate of the initial echo decay is utilized to determine the average distance of a DEER distribution (Figure 2.8). Conformations with short distances below 2.0 nm cannot be reliably measured [107] because the inversion efficiency  $\lambda$  is dependent on the ratio between the excitation bandwidth and dipole-dipole coupling. The upper limit to determine an accurate mean distance is limited by the maximum dipolar evolution time,  $t_{\max}$ , which is dictated by the phase memory time (dependent on transverse relaxation times). This upper mean limit is approximately 6 nm for a membrane protein in detergent micelles with a  $t_{\max}$  of 3.5  $\mu\text{s}$  [108, 109]. Longer distances may be recognized but cannot be quantitatively

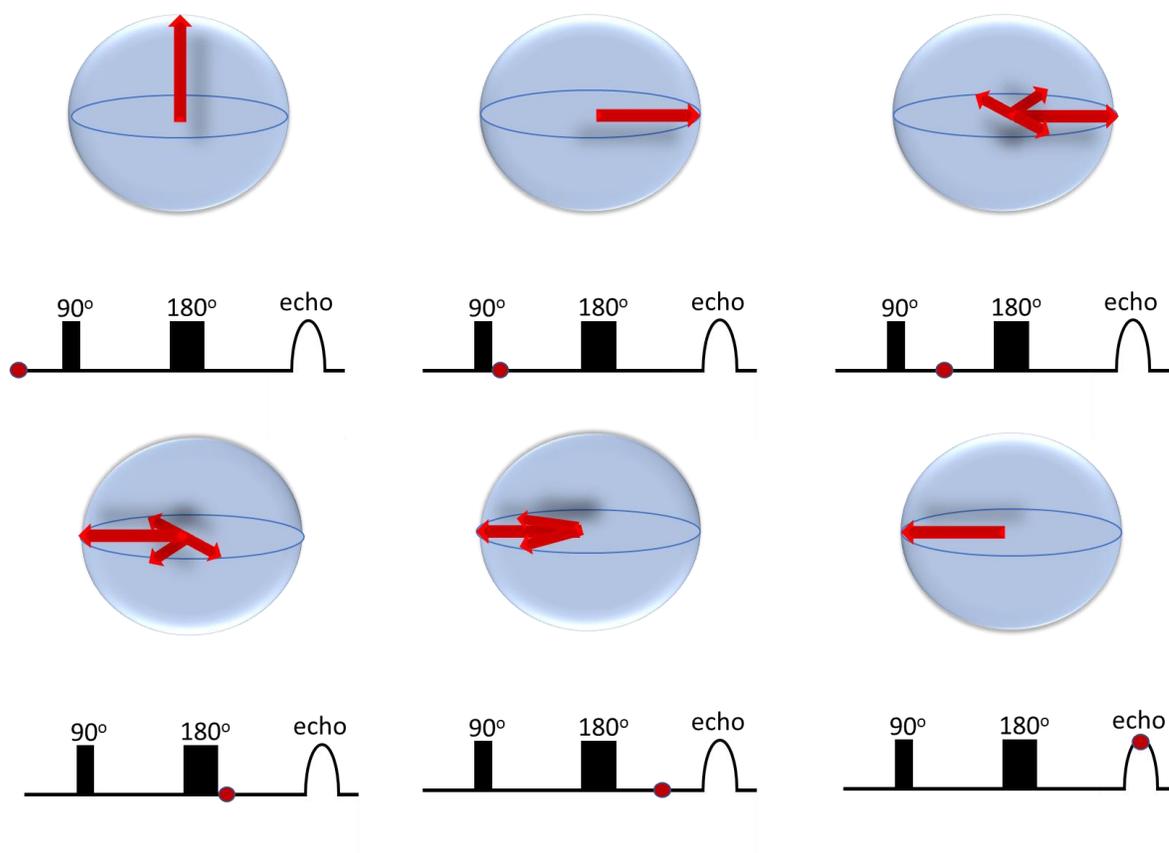
interpreted accurately. Well separated peaks in a distance distribution can be detected and interpreted up to the reliable limit of the mean distance. The decay rate of the dipolar oscillations provides information on the width of the DEER distance distributions. Narrow distance distributions (in rigid molecules) will oscillate before their decay to baseline, whereas wider distributions will have dampened oscillations (Figure 2.8). The limit of determining an accurate distribution width is dependent on  $t_{\max}$  and corresponds to approximately 5 nm for membrane proteins [109].

Ultimately, DEER distributions provide a great deal of information on protein conformational changes, where a shift in conformation may be detected as an altered or added distance in the distribution. Such conformational changes are often measurable due to ligand binding and large-scale domain movements [110-112]. Multiple conformational states are shown as multiple peaks in a DEER distribution. DEER can also be utilized to detect the formation of a protein complex and protein oligomerization [113]. However, generating a sufficiently high number of DEER distances to determine an atomistic structure is a significant limitation which will be further explored in Chapter 3 of this dissertation.

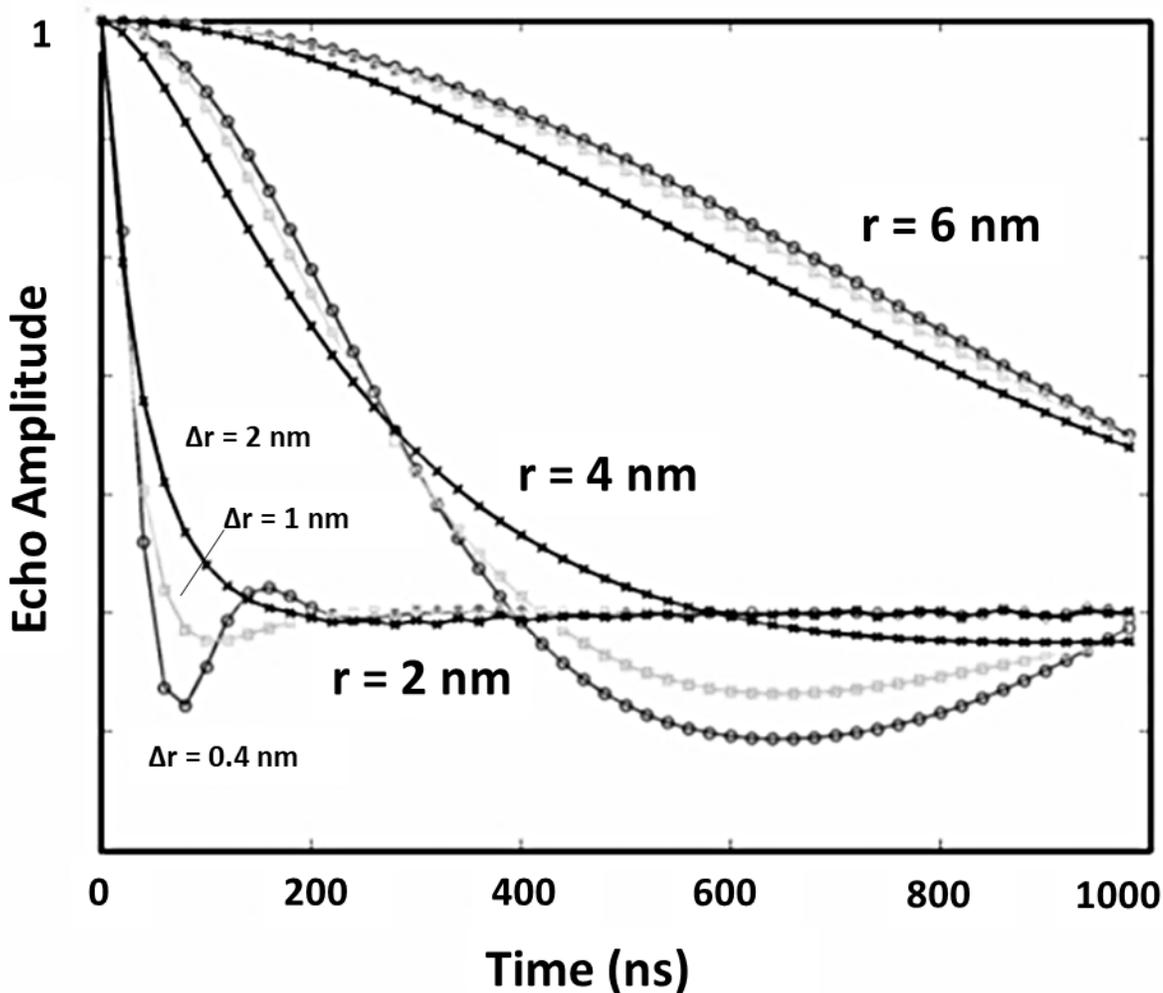
Together, DEER is an ideal technique to measure partially disordered domains [114, 115], structural changes upon ligand binding [112], large-scale domain motions [116], and protein complexes [117]. DEER yields a distance distribution, rather than a single distance value, allowing researchers to examine the flexibility of a protein structure. The applicability of DEER to difficult protein systems, such as membrane proteins, dynamic proteins, and large systems, is a marked advantage of the technique.



**Figure 2.6: Four-pulse DEER experiment setup.** A refocused echo is generated at an observer frequency ( $\nu_{\text{obs}}$ ) (with a  $\pi/2 - \pi - \pi$  sequence) whose intensity is modulated with the application of a  $\pi$  pump pulse at a different frequency ( $\nu_{\text{pump}}$ ). The timing of this pump pulse is varied, which generates a DEER signal that is the intensity of the refocused echo with respect to the timing of the pump pulse. This provides information on the distance-dependent dipolar interaction between spins.



**Figure 2.7: Schematic of a two-pulse sequence to generate a spin echo.** A spin echo is the refocusing of spin magnetization by pulses of electromagnetic radiation. The application of a  $\pi/2$ ,  $90^\circ$  pulse tips the magnetization into the x-y plane, and over time phase coherence is lost due to relaxation effects. The application of a  $\pi$ ,  $180^\circ$  pulse, reestablishes phase coherence over time, producing a spin echo.



**Figure 2.8: DEER signal as a function of the mean distance and width around the mean.** *Figure adapted from [118].* These simulated spectra visually demonstrate the effect of the mean distance and width on the DEER signal. Mean distances ( $r$ ) of 2, 4, and 6 nm are shown with distribution widths around the mean of  $\Delta r = 0.4, 1,$  and 2 nm. The rate of the initial decay is lessened as the mean distance increases, and the oscillations are dampened as the width of the distribution increases, which together generate a distance distribution.

## 2.6 Utilizing EPR to study Opa-CEACAM interactions

The overarching goal of this research is to determine the mechanism by which Neisserial Opa proteins engage host CEACAMs to trigger *Neisseria* phagocytosis into human cells as a means of infection. EPR is an ideal technique to study Opa and CEACAM proteins, as both are natively cys-less. We aim to determine a conformational ensemble of the dynamic Opa loops in isolation and in the presence of CEACAM to identify the mechanism by which Opa proteins engage CEACAM receptors. The dynamic nature of Opa loops will first be examined using CW-EPR, where multiple single site cysteine mutants are spin labeled at various locations along Opa loops. This allows us to compare the timescale of dynamic backbone motions between the loops, as well as identify secondary structure formation in the presence or absence of receptor. DEER will be used to determine the conformational ensemble of the dynamic Opa loops, where Opa proteins are spin labeled along two locations. The resulting DEER distributions of several Opa spin label pairs will be incorporated into MD simulations to identify the Opa loop conformational ensemble in the presence and absence of receptor. CEACAMs will also be single labeled and measured with DEER to better characterize the CEACAM homodimer from the resulting distance distributions. Opa-CEACAM interactions will be measured using DEER, where several single labeled Opa and single labeled CEACAM sites will be probed to generate a structural model of the binding complex from the resulting distance distributions.

1. Kozyrev, S.A.A.t.a.B.M., *Electron Paramagnetic Resonance*. 1964, New York, NY: Academic.
2. Hubbell, W.L., D.S. Cafiso, and C. Altenbach, *Identifying conformational changes with site-directed spin labeling*. *Nature Structural Biology*, 2000. **7**(9): p. 735-739.
3. Berliner, L.J., et al., *A NOVEL REVERSIBLE THIOL-SPECIFIC SPIN LABEL - PAPAINE ACTIVE-SITE LABELING AND INHIBITION*. *Analytical Biochemistry*, 1982. **119**(2): p. 450-455.
4. McHaourab, H.S., et al., *Motion of spin-labeled side chains in T4 lysozyme, correlation with protein structure and dynamics*. *Biochemistry*, 1996. **35**(24): p. 7692-7704.
5. Amadi, S.T., et al., *Structure, Dynamics, and Substrate-induced Conformational Changes of the Multidrug Transporter EmrE in Liposomes*. *Journal of Biological Chemistry*, 2010. **285**(34): p. 26710-26718.
6. McHaourab, Hassane S., P.R. Steed, and K. Kazmier, *Toward the Fourth Dimension of Membrane Protein Structure: Insight into Dynamics from Spin-Labeling EPR Spectroscopy*. *Structure*, 2011. **19**(11): p. 1549-1561.
7. Claxton, D.P., et al., *Chapter Twelve - Navigating Membrane Protein Structure, Dynamics, and Energy Landscapes Using Spin Labeling and EPR Spectroscopy*, in *Methods in Enzymology*, P.Z. Qin and K. Warncke, Editors. 2015, Academic Press. p. 349-387.
8. Atkins, P.W., and Kivelson, D., *ESR Linewidths in Solution. II. Analysis of Spin Rotational Relation Data*. *Journal of Chemical Physics*, 1966. **44**: p. 169-174.
9. Van, S.P., G.B. Birrell, and O.H. Griffith, *RAPID ANISOTROPIC MOTION OF SPIN LABELS - MODELS FOR MOTION AVERAGING OF ESR PARAMETERS*. *Journal of Magnetic Resonance*, 1974. **15**(3): p. 444-459.
10. Kivelson, D., *Theory of ESR linewidths of free radical*. *J. Chem. Phys.*, 1960. **33**.

11. Lietzow, M.A. and W.L. Hubbell, *Motion of spin label side chains in cellular retinol-binding protein: Correlation with structure and nearest-neighbor interactions in an antiparallel beta-sheet*. *Biochemistry*, 2004. **43**(11): p. 3137-3151.
12. Columbus, L. and W.L. Hubbell, *A new spin on protein dynamics*. *Trends in Biochemical Sciences*, 2002. **27**(6): p. 288-295.
13. Columbus, L., et al., *Molecular motion of spin labeled side chains in alpha-helices: Analysis by variation of side chain structure*. *Biochemistry*, 2001. **40**(13): p. 3828-3846.
14. Altenbach, C., et al., *TRANSMEMBRANE PROTEIN-STRUCTURE - SPIN LABELING OF BACTERIORHODOPSIN MUTANTS*. *Science*, 1990. **248**(4959): p. 1088-1092.
15. Jeschke, G. and Y. Polyhach, *Distance measurements on spin-labelled biomacromolecules by pulsed electron paramagnetic resonance*. *Physical Chemistry Chemical Physics*, 2007. **9**(16): p. 1895-1910.
16. Schiemann, O. and T.F. Prisner, *Long-range distance determinations in biomacromolecules by EPR spectroscopy*. *Quarterly Reviews of Biophysics*, 2007. **40**(1): p. 1-53.
17. Jeschke, G., et al., *Direct Conversion of EPR Dipolar Time Evolution Data to Distance Distributions*. *Journal of Magnetic Resonance*, 2002. **155**(1): p. 72-82.
18. Pannier, M., et al., *Dead-Time Free Measurement of Dipole–Dipole Interactions between Electron Spins*. *Journal of Magnetic Resonance*, 2000. **142**(2): p. 331-340.
19. Polyhach, Y., et al., *Spin pair geometry revealed by high-field DEER in the presence of conformational distributions*. *Journal of Magnetic Resonance*, 2007. **185**(1): p. 118-129.
20. Chiang, Y.-W., P.P. Borbat, and J.H. Freed, *The determination of pair distance distributions by pulsed ESR using Tikhonov regularization*. *Journal of Magnetic Resonance*, 2005. **172**(2): p. 279-295.

21. Chiang, Y.-W., P.P. Borbat, and J.H. Freed, *Maximum entropy: A complement to Tikhonov regularization for determination of pair distance distributions by pulsed ESR*. Journal of Magnetic Resonance, 2005. **177**(2): p. 184-196.
22. Banham, J.E., et al., *Distance measurements in the borderline region of applicability of CW EPR and DEER: A model study on a homologous series of spin-labelled peptides*. Journal of Magnetic Resonance, 2008. **191**(2): p. 202-218.
23. Grote, M., et al., *Transmembrane Signaling in the Maltose ABC Transporter MalFGK2-E: PERIPLASMIC MalF-P2 LOOP COMMUNICATES SUBSTRATE AVAILABILITY TO THE ATP-BOUND MalK DIMER*. Journal of Biological Chemistry, 2009. **284**(26): p. 17521-17526.
24. Jeschke, G., *DEER Distance Measurements on Proteins*. Annual Review of Physical Chemistry, 2012. **63**(1): p. 419-446.
25. López, C.J., et al., *Conformational selection and adaptation to ligand binding in T4 lysozyme cavity mutants*. Proceedings of the National Academy of Sciences of the United States of America, 2013. **110**(46): p. E4306-E4315.
26. Braun, P., et al., *Mechanism of Multivalent Carbohydrate-Protein Interactions Studied by EPR Spectroscopy*. Angewandte Chemie International Edition, 2011. **50**(36): p. 8428-8431.
27. Torbeev, V.Y., et al., *Dynamics of "Flap" Structures in Three HIV-1 Protease/Inhibitor Complexes Probed by Total Chemical Synthesis and Pulse-EPR Spectroscopy*. Journal of the American Chemical Society, 2009. **131**(3): p. 884-885.
28. Hilger, D., et al., *Assessing Oligomerization of Membrane Proteins by Four-Pulse DEER: pH-Dependent Dimerization of NhaA Na(+)/H(+) Antiporter of E. coli*. Biophysical Journal, 2005. **89**(2): p. 1328-1338.

29. Bleicken, S., et al., *Structural Model of Active Bax at the Membrane*. *Molecular Cell*, 2014. **56**(4): p. 496-505.
30. Dockter, C., et al., *Rigid Core and Flexible Terminus: STRUCTURE OF SOLUBILIZED LIGHT-HARVESTING CHLOROPHYLL a/b COMPLEX (LHCII) MEASURED BY EPR*. *Journal of Biological Chemistry*, 2012. **287**(4): p. 2915-2925.
31. Georgieva, E.R., et al., *Conformational ensemble of the sodium coupled aspartate transporter*. *Nature structural & molecular biology*, 2013. **20**(2): p. 215-221.
32. Lovett, J.E., et al., *Investigating the structure of the factor B vWF-A domain/CD55 protein-protein complex using DEER spectroscopy: successes and pitfalls*. *Molecular Physics*, 2013. **111**(18-19): p. 2865-2872.
33. Piotr G. Fajer, L.B.a.L.S., *Practical Pulsed Dipolar ESR (DEER)*. Vol. *Biological Magnetic Resonance*. Boston, MA: Springer.

## Chapter 3: Opa loop dynamics and conformational ensemble

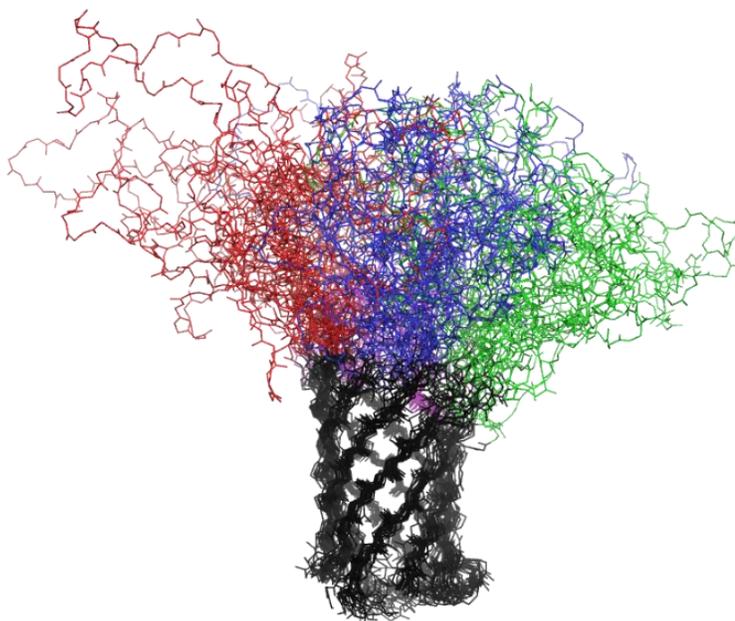
### 3.1 Introduction

Neisserial Opa proteins are a family of outer membrane  $\beta$ -barrel proteins that contain four extracellular loops that are 14 - 47 residues in length. These loops contain regions that are highly diverse in sequence with no apparent receptor binding motif. At least two of the loops are required for receptor engagement, although the binding mechanism to respective receptors remains unclear. To ultimately determine the binding mode of Opa loops, however, we must first understand the nature of these loops in their unbound state. To that end, Opa loop dynamics and their conformational ensemble will be examined in this chapter.

### 3.2 Structure of Opa

This research will focus on one of the most highly studied Opa proteins, Opa<sub>60</sub>, whose cognate receptors have been identified [119]. Opa<sub>60</sub> is found specifically in *N. gonorrhoeae* strain MS11 and is an eight-stranded  $\beta$ -barrel with four extracellular loops. The structure of this protein was determined previously by the Columbus lab (PDB ID: 2MLH) in collaboration with the Kasson lab using a combination of NMR and MD, where distance and dihedral NMR restraints of Opa in detergent micelles were enforced in simulations of Opa in a bilayer [53]. The twenty lowest-energy NMR-generated structures showed Opa<sub>60</sub> loops without a defined secondary structure which occupy a large lateral density over the barrel. However, many of the loop residues (especially in the HV2 region) were not observed and, therefore, unassigned in recorded NMR spectra. The line broadening in these regions is indicative of structural heterogeneity and/or conformational exchange. Thus,

MD was utilized for further structural refinement (in a bilayer). 100-ns simulations were run on the twenty lowest-energy NMR structures [53]. The resulting MD simulations indicated that the loops were highly dynamic and unstructured (Figure 3.1). However, the loop density in these MD-refined structures was localized more closely over the barrel, indicating the loops may interact with each other.



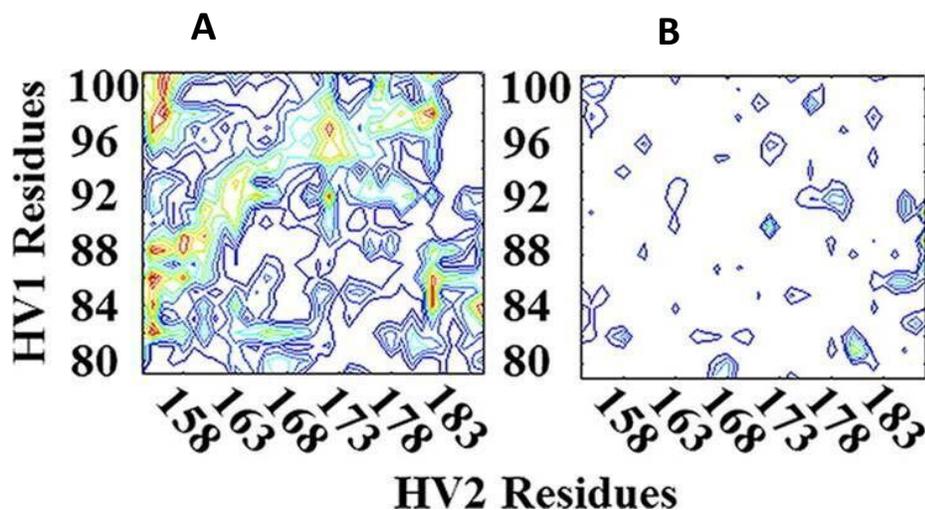
**Figure 3.1: MD-refined NMR structures of Opa<sub>60</sub> proteins.** Overlay of the twenty lowest-energy conformers from MD-refined NMR structure determination of Opa<sub>60</sub> [53]. The loops are disordered and dynamic. Loop 1 containing the SV region is shown in green, loop 2 containing HV1 is colored blue, loop 3 with HV2 is colored red, and the fourth loop is purple.

### 3.2.1 Frequent, short lived contacts between Opa loops

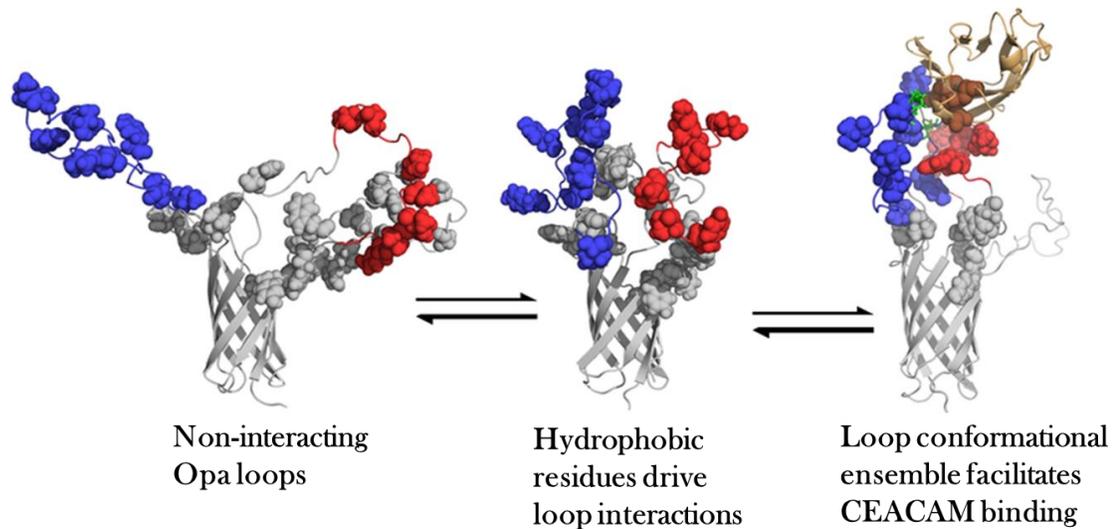
Contact maps were generated between loops containing hypervariable regions 1 and 2 (which are predicted to bind receptors) using MD to further investigate Opa loop interactions [53]. The contact maps were created from the 20 lowest energy conformers, where contacts were defined as residues atoms within 6 Å of each other and each contour line represents a percentile increment in contact probability over the MD snapshots (Figure 3.2 A). The lifetimes of these contacts were also plotted, where each contour line represents a 5-ns lifetime increment (Figure 3.2 B). Most regions that had a high probability of interactions between these two loops (especially contacts between residues 80-88 on HV1 and 155-163 on HV2) lost their contact over 5 ns, which suggests that the loops frequently interact, but these contacts are short lived. Based on the hydrophobic residues within extracellular loops of Opa (Figure 1.8), one hypothesis is that hydrophobic interactions may drive the loops together, forming a flexible and transient loop conformational ensemble that is necessary to promote receptor engagement (Figure 3.3).

Prior to drawing conclusions about how Opa loops selectively engage receptors, dynamic motions throughout loops regions that are not observed with NMR must first be experimentally validated. NMR T1/T2 relaxation and MD experiments indicate that Opa loops are dynamic on the ns timescale, although the timescale of motion could not be determined for the residues that were not experimentally observed (many of which were located along potential receptor binding regions) (Figure 3.4). As such, Opa loop dynamics were further examined via CW-EPR (Section 3.4). The dynamic Opa loops were predicted to interact at hydrophobic residues from MD simulations of Opa<sub>60</sub> (Figure 3.5). To determine if these interactions drive the formation of a loop conformational ensemble that

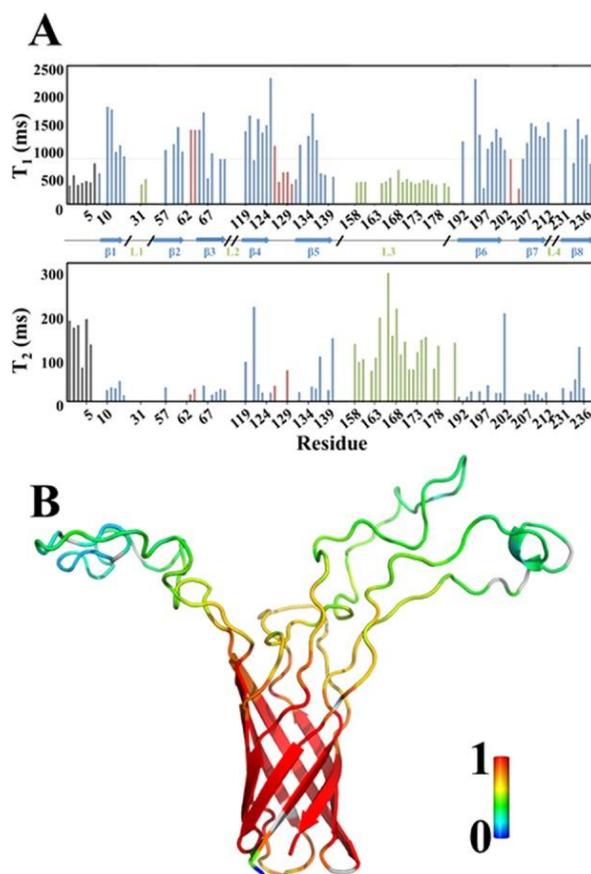
forms in the absence of receptor, a combination of DEER and MD experiments are utilized to identify Opa loop clusters (Section 3.5). Together, CW-EPR and DEER/MD experiments of Opa extracellular loops provide insight into the role of Opa loop dynamics in receptor binding. Studying unbound Opa loops will generate a starting structural model that can be applied to future Opa-CEACAM binding experiments. Understanding Opa loop conformational ensembles will elucidate the mechanism of Opa-receptor interactions (conformational selection, induced folding, etc.) that will ultimately provide insight into the means of *Neisseria* phagocytosis into host human cells.



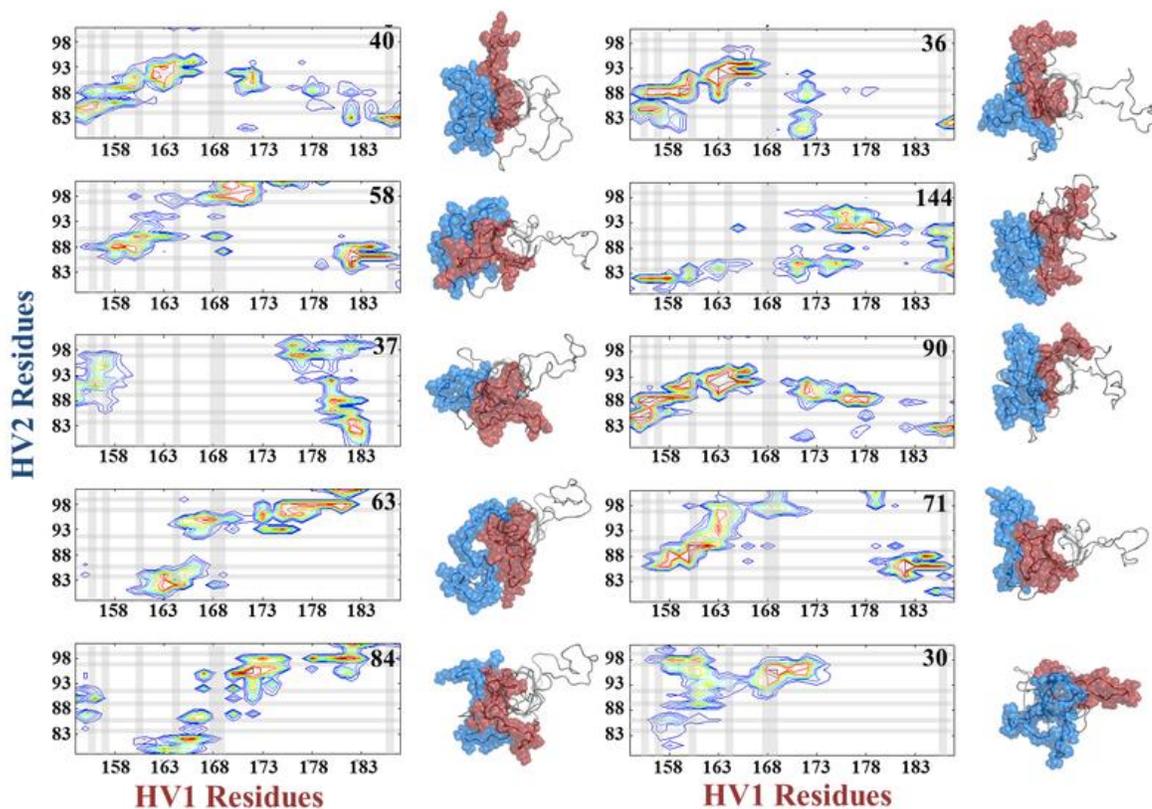
**Figure 3.2: Opa HV regions have a high probability to interact over short periods of time.** *Reprinted with permission from [53].* (A) Contact map generated from snapshots of MD simulations of Opa<sub>60</sub>, where contacts were defined as residues within 6 Å of each other and each contour line represents a percentile increment in contact probability over the MD snapshots. Many residues in the two HV regions had a high probability to interact, especially between residues 80-88 on HV1 and 158-163 on HV2. (B) The lifetimes of these contacts were plotted, where each contour line represents a 5-ns lifetime increment. Most residues lose contact over a 5 ns period, which suggests that the loops frequently interact, but these contacts are short lived.



**Figure 3.3: Hydrophobic residues in Opa loops are hypothesized to promote loop interactions and a receptor-competent conformational ensemble.** Hydrophobic residues span the length of Opa<sub>60</sub> loops (shown here as spheres: HV1 hydrophobic residues are in red and HV2 hydrophobic residues are in blue), and hydrophobic interactions between these residues may promote loop interactions. Such interactions may be required to form a loop conformational ensemble, which may ultimately facilitate receptor engagement and enable receptor selectivity in the highly diverse Opa protein family via conformational selection or induced folding.



**Figure 3.4: NMR and MD predict that Opa extracellular loops are dynamic on the ns-timescale.** Reprinted with permission from [53]. (A)  $T_1$  and  $T_2$  relaxation experiments predict that Opa loops (green) are dynamic on the ns-timescale. Residues with  $T_1$  values that are below the predicted one second  $T_1$  value for a 20 ns overall correlation time (that of an Opa protein detergent complex), shown as a dotted line, are mobile in the ns-timescale. Corresponding loop residues have the highest  $T_2$  values, indicating those residues are the most dynamic.  $T_1$  and  $T_2$  values could not be calculated for residues that were not observed in NMR experiments, namely residues 139-158 and 178-192 that are within HV1 and HV2 regions. (B) MD simulations of Opa support the dynamic nature of Opa loops observed with NMR, where loop backbone residues (green and blue) are not correlated at the end of the 20-ns simulations and the barrel (red) is stationary.



**Figure 3.5: HV1-HV2 contact maps for structural snapshots of Opa loops.** *Reprinted with permission from [53].* Contact maps for structural snapshots of Opa loops that contain a high degree of contact between HV1 and HV2 (greater than 50 residue contacts). Contour lines indicate 10% contact probability increments. Hydrophobic residues are marked with grey lines and the top numbers indicate the number of snapshots in each cluster. A number of loop contact regions occur at hydrophobic residues, and as such, these residues may drive loop interactions to form a loop conformational ensemble.

### 3.3 Folding and site directed spin labeling of Opa proteins

Opa proteins must first be spin labeled and reconstituted into a membrane mimic to measure loop dynamics with EPR techniques. Both micelles and liposomes are utilized to solubilize Opa proteins to determine if the choice of membrane mimic has an effect on Opa loop dynamics. To observe loop dynamics throughout the length of all four loops, a single cysteine was introduced along every ten residues within the loops (as native Opa proteins are cys-less) to compare the backbone mobility at thirteen total sites.

Cysteine residues are introduced via Polymerase Incomplete Primer Extension (PIPE) Mutagenesis, where forward and reverse primers include the mutated residue and are incorporated into the plasmid via the polymerase chain reaction (PCR). Opa proteins are then recombinantly expressed in *E. coli* into insoluble inclusion bodies. The unfolded Opa proteins are extracted from inclusion bodies with 8 M urea and are purified in an unfolded state using  $\text{Co}^{2+}$  Immobilized Metal Affinity Chromatography (IMAC).

When folding Opa proteins into fos-choline-12 (FC-12) detergent micelles, the proteins are first spin labeled using five molar excess MTSL. Unfolded Opa proteins must be spin labeled prior to folding into detergent because unreacted spin label non-specifically associates with the protein detergent complex [120-123]. Following the spin labeling of unfolded Opa, the proteins are rapidly diluted into a buffer containing approximately four molar excess FC-12 micelles, upon which the proteins spontaneously fold into the detergent micelles over the course of several days. While the length of this folding period (five days) was a non-factor in prior NMR studies, the duration of the folding step has a significant impact on Opa EPR studies, as the spin label dissociates from Opa cysteine residues over time. As such, the Opa folding protocol into micelles was optimized to reduce

the protein folding period, thus limiting the dissociation of the spin label from the protein. Incorporating 3M urea into the folding solution has previously been shown to reduce unfolded Opa aggregation when folding into liposomes [124]. Applying this urea protocol towards Opa folding into micelles likewise reduced the duration of Opa folding into detergent to three days (compared to five previously), increasing amount of spin labeled protein and thus the intensity of the EPR signal.

Conversely, Opa proteins are spin labeled following folding into liposomes. The proteins were folded into C10-PC (1,2-didecanoyl-sn-glycero-3-phosphocholine) lipids and were transferred to a more stable lipid mixture [DMPC (1,2-dimyristoyl-sn-glycero-3-phosphocholine), DMPG (1,2-dimyristoyl-sn-glycero-3-phosphoglycerol), cholesterol, and PEG-1000-PE (1,2-distearoyl-sn-glycero-3-phosphoethanolamine-N-[methoxy(polyethylene glycol)-1000])] which is less prone to liposome aggregation prior to Opa labeling [119, 124]. Spin labeling after folding causes only Opa proteins whose loops are facing the exterior of the liposome to be labeled (transferring the proteins to a new lipid mixture with sonication results in roughly half of the Opa proteins to face the interior). Because Opa proteins are folded into liposomes prior to labeling, a small degree of free spin intercalates into the liposome and is visible in the Opa proteoliposome spectra, which causes those spin labeled proteins to appear slightly more mobile than in micelles. However, labeling unfolded Opa proteins prior to folding into liposomes resulted in no EPR signal, as the folding process occurs at 37°C (above the lipid gel-transition temperature) [124], which enhances the rate of MTSL dissociation (unbound label is always removed following spin labeling via a desalting column, so no free spin would be visible in the spectrum).

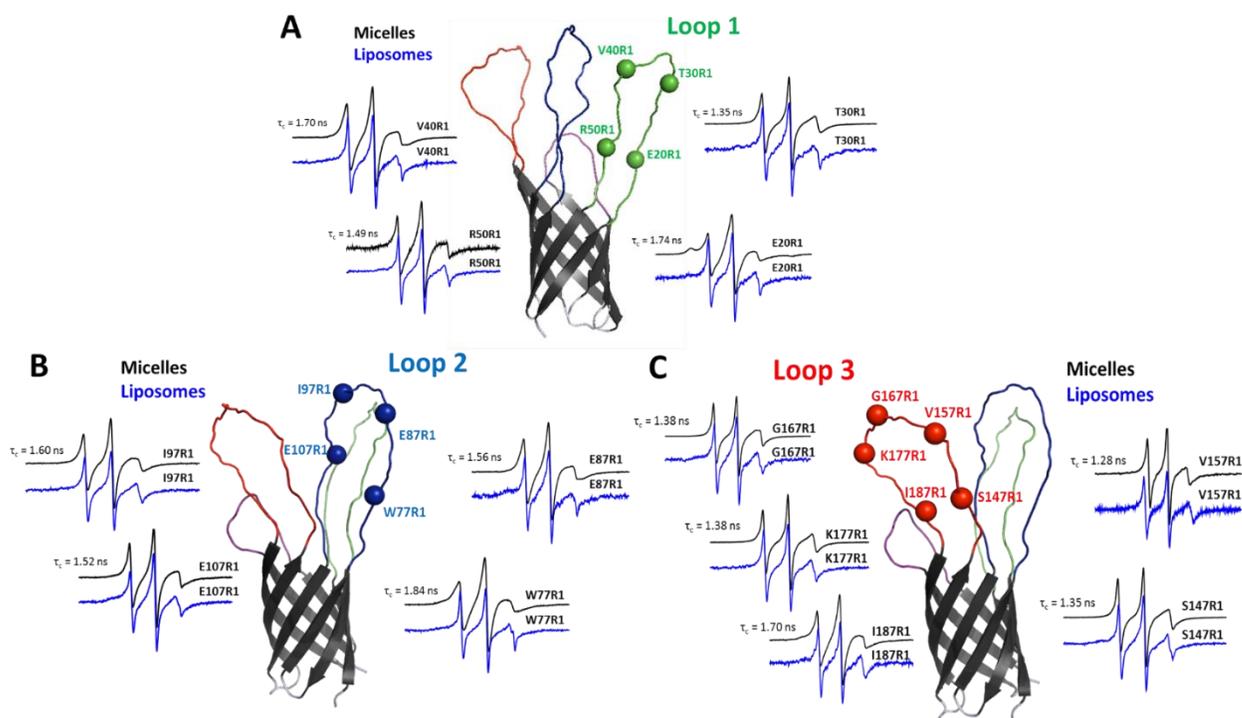
PEGylated lipids serve to roughly sterically mimic LOS moieties found on the outer membrane of *Neisseria* due to their large size, where they extend 7-25 Å from the bilayer [125, 126]. While PEGylated lipids rather than LOS are included in these studies (PEG prevents multilamellar liposome formation and LOS is not commercially available), the impact of LOS on modulating Opa loop dynamics should not be ignored. Many lower residues in Opa loops near the membrane interface are basic and as such may directly bind LOS in addition to the steric impact of the large LOS moieties that populate a significant portion of the outer membrane of *Neisseria*. Although this is beyond the studies explored in this dissertation, LOS interactions with and effects on Opa proteins may be a significant factor in Opa-CEACAM interactions, and thus in *Neisserial* pathogenesis, which will be discussed in Chapter 5 as potential for inclusion in future experiments.

### 3.4 Mapping Opa loop dynamics with CW-EPR

CW-EPR is the ideal technique to measure Opa loop dynamics, as Opa backbone fluctuations were predicted to occur in the ns timescale with NMR and MD. Comparing EPR spectra which span the four loops allows us to determine the rates of spin label motion at various locations. The resulting CW lineshapes may also be indicative of secondary structural elements that may not have been observed in NMR experiments.

The resulting CW-EPR lineshapes of Opa in both micelles and liposomes (Figure 3.6) indicate that each region probed is highly dynamic in the ns-timescale. The lineshapes are indicative of unstructured loop residues, where no local ordering or secondary structure formation was observed for these residues (Figure 3.6). Labels located at regions nearest to the barrel exhibited the most restricted dynamics, where the rate of motion progressively increased towards the top of the loops (as is expected when considering the loops as long

polymers tethered to an immobile barrel). Rotational correlation times ( $\tau_c$ ) for each spin labeled Opa in micelles were calculated via the Redfield theory using the central linewidth [127]. Some of the loop residues closest to the barrel (which are not predicted to be directly involved in receptor binding), specifically 20R1 in Figure 3.6A, show two components of differing mobilities. The two mobilities are likely due to the location of this residue near the barrel, where the hydrogen bonding between  $\beta$ -strands may extend upwards to this residue in one state. The additional immobile component at residue 20R1 is less visible in liposomes due to a small amount of free spin and noise in that sample. Other than this residue, however, the spectra exhibit similar ns loop motions with Opa proteins reconstituted into both micelles and liposomes, indicating that the chosen membrane mimic has a minimal impact on Opa loop dynamics.



**Figure 3.6: Opa loop dynamics via CW-EPR.** CW-EPR lineshapes of Opa<sub>60</sub> proteins that are folded into detergent micelles (black) or liposomes (blue). The proteins were single-labeled, spanning every ten residues of loops 1 (A), 2 (B), and 3 (C). The lineshapes are all indicative of dynamic loop motions in the ns timescale, where the most mobile regions are located near the tops of the loops. Redfield theory was utilized to calculate rotational correlation times ( $\tau_c$ ) from the central linewidth of for each spin labeled site of Opa in detergent.

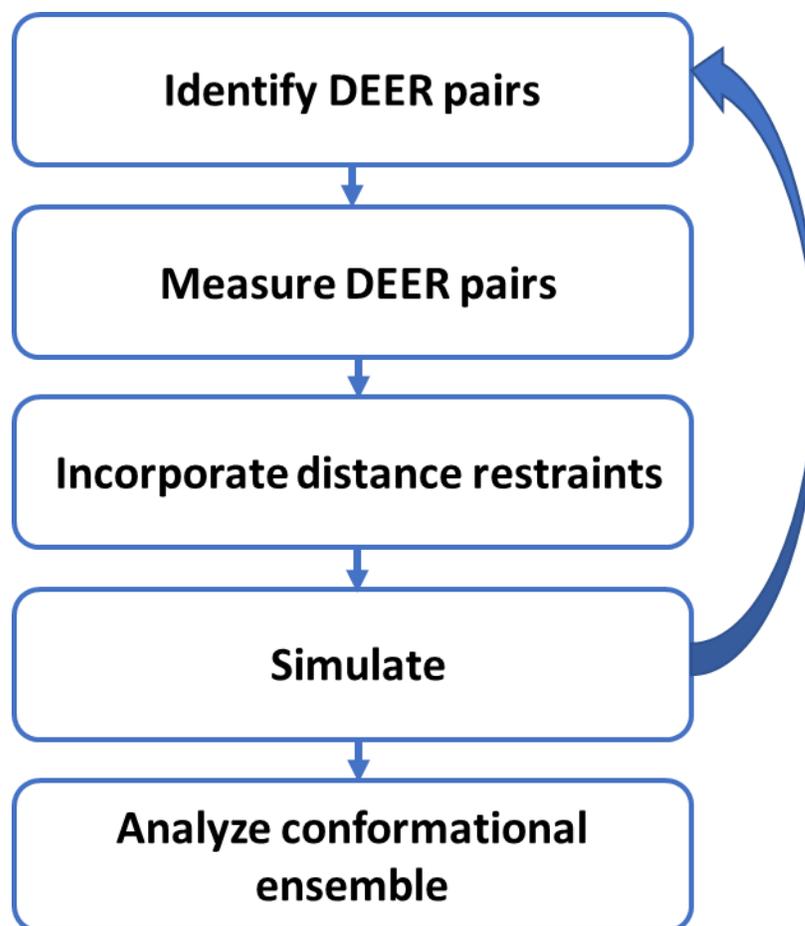
### 3.5 Combining DEER and MD to determine the Opa loop conformational ensemble

Opa extracellular loops are unlikely to be dynamic without any loop interactions due to their hydrophobic nature (as was predicted with previous MD experiments [53]). A loop conformational ensemble may form in the absence of receptor as a prerequisite to binding. Mutagenesis experiments show that a specific combination of Opa HV regions are required for *Neisseria* invasion. Two Opa variants of the same *Neisseria* strain with identical SV and HV2 regions differed in their interactions with CEACAM3 and CEACAM6, suggesting that HV1 is required for receptor recognition [46]. Additionally, generating chimeric Opa proteins, where the HV regions of two CEACAM-binding Opa proteins were switched, eliminated CEACAM interactions [55]. Deleting either HV region also completely abolished *Neisseria* invasion into CEACAM-expressing cells [55]. Thus, a structural ensemble may be required in addition to a specific sequence for receptor engagement. The Opa loop conformational ensemble must first be determined in the absence of receptor to generate loop models in both bound and unbound states, from which a binding mechanism can be identified.

However, determining the conformational ensemble of such a dynamic membrane protein is challenging. MD simulations alone provide a limited solution, because without experimental restraints the potential sampling space of Opa loops is vast and would require prohibitively long (and expensive) computing time. Conversely, experimental methods in isolation, such as DEER, would require a very large number of samples to generate a structural model, as each DEER signal only reports a distance distribution between two labeled sites. Generating such a sufficiently high number (perhaps thousands) of doubly spin labeled Opa proteins that are folded into lipid or detergent membrane mimics would

both be immensely time consuming and expensive. The high conformational heterogeneity of Opa loops also hinders a meaningful structural interpretation of DEER distributions without the use of MD, even if a sufficient number of samples could be measured.

To address these challenges presented by MD and DEER in isolation, an integration of DEER and MD was developed (in collaboration with Jennifer Hays and Peter Kasson, UVA) to generate a model of the Opa loop conformational ensemble that requires as few experimental measurements as possible (Figure 3.7). Our methodology involves three primary steps: (1) predicting DEER pairs to measure using MD, (2) measuring these pairs with DEER, (3) and calculating a multi-conformer model of Opa using DEER restraints to bias MD simulations.



**Figure 3.7: Integrating DEER and MD to generate a conformational ensemble of Opa loops.** In our novel DEER/MD methodology, DEER pairs of maximum relevancy and minimum redundancy (mRMR) are first identified to provide optimal distance restraints for MD simulations of Opa. Following the incorporation of these restraints into simulations, a second round of mRMR pairs are identified and measured to provide further experimental restraints. Ultimately, final atomistic simulations of Opa loops will generate an experimentally-restrained, high-resolution conformational ensemble of Opa loops.

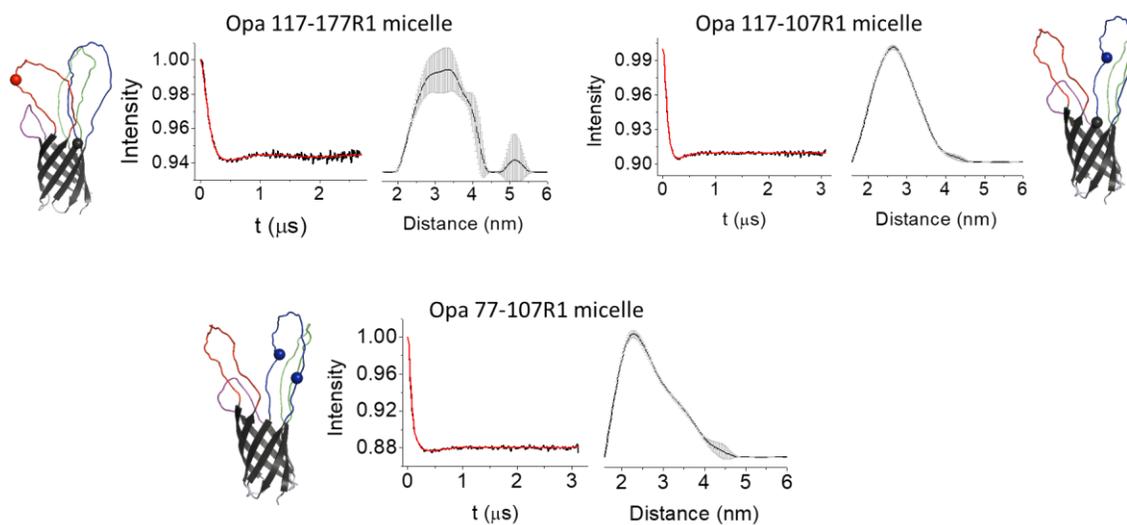
### 3.5.1 Identifying highly informative DEER pairs

While previous groups have integrated DEER and MD experiments, these studies were done on relatively static proteins with defined structures (often solved via X-ray crystallography) -- DEER and MD in these cases are utilized to expand a pre-existing structure to incorporate low range dynamic motions and/or to examine multiple structural states that are static in isolation [128-133]. Docking experiments have also been utilized to incorporate DEER distributions into protein-protein complexes, whose individual structures have been solved [131].

However, Opa loops are very large (roughly 20 – 30 Å extended from the barrel) and have no defined secondary structure, making the applicability of traditional DEER and MD techniques infeasible. In collaboration with Peter Kasson and Jennifer Hays (Biomedical Engineering & Molecular Physiology and Biological Physics, School of Medicine, UVA), we aim to determine the Opa loop conformational ensemble via a novel integration of DEER and MD, where DEER pairs are selected to provide optimal distance restraints in MD simulations. The selected DEER pairs provide experimental restraints for simulations of Opa that are not redundant to distances determined by other measured pairs. The calculation of these maximum relevance and minimum redundancy (mRMR) [134] DEER pairs is the dissertation work of Jennifer Hays in the Kasson lab. Two criteria for pairs were: (1) that the pairs are predicted to be between distances of 25 – 45 Å, as they need to be measurable with DEER, and (2) proline residues were avoided, as these residues (located near the center of loops 2 and 3) restrict local loop dynamics in these regions, which may alter the loop structure and promote a receptor-competent loop conformation.

A ranked list of eleven Opa DEER pairs were identified using the mRMR algorithm, where pairs spanned the length of Opa loops. As a comparison for a validation of our methodology, “spectroscopist-selected” pairs (SSP, identified prior to our collaboration with the Kasson lab) were also measured (Figure 3.8). These pairs were conservative choices, as they are primarily barrel-loop pairs. Without the aid of MD-guided selection of pairs, it is difficult to predict the lateral sampling range of Opa loops, which may extend beyond the distance range measurable with DEER. However, these SSPs provided little information on the relative orientation of the dynamic loops compared to the mRMR-selected pairs. The mRMR method identified primarily loop-loop pairs that incorporate greater conformational heterogeneity and are thus more effective in refining the dynamic loop conformational ensemble. The probability that a spectroscopist chooses the top 25% of the mRMR Opa pairs by chance is only 7%, which highlights the necessity of a guided selection of pairs.

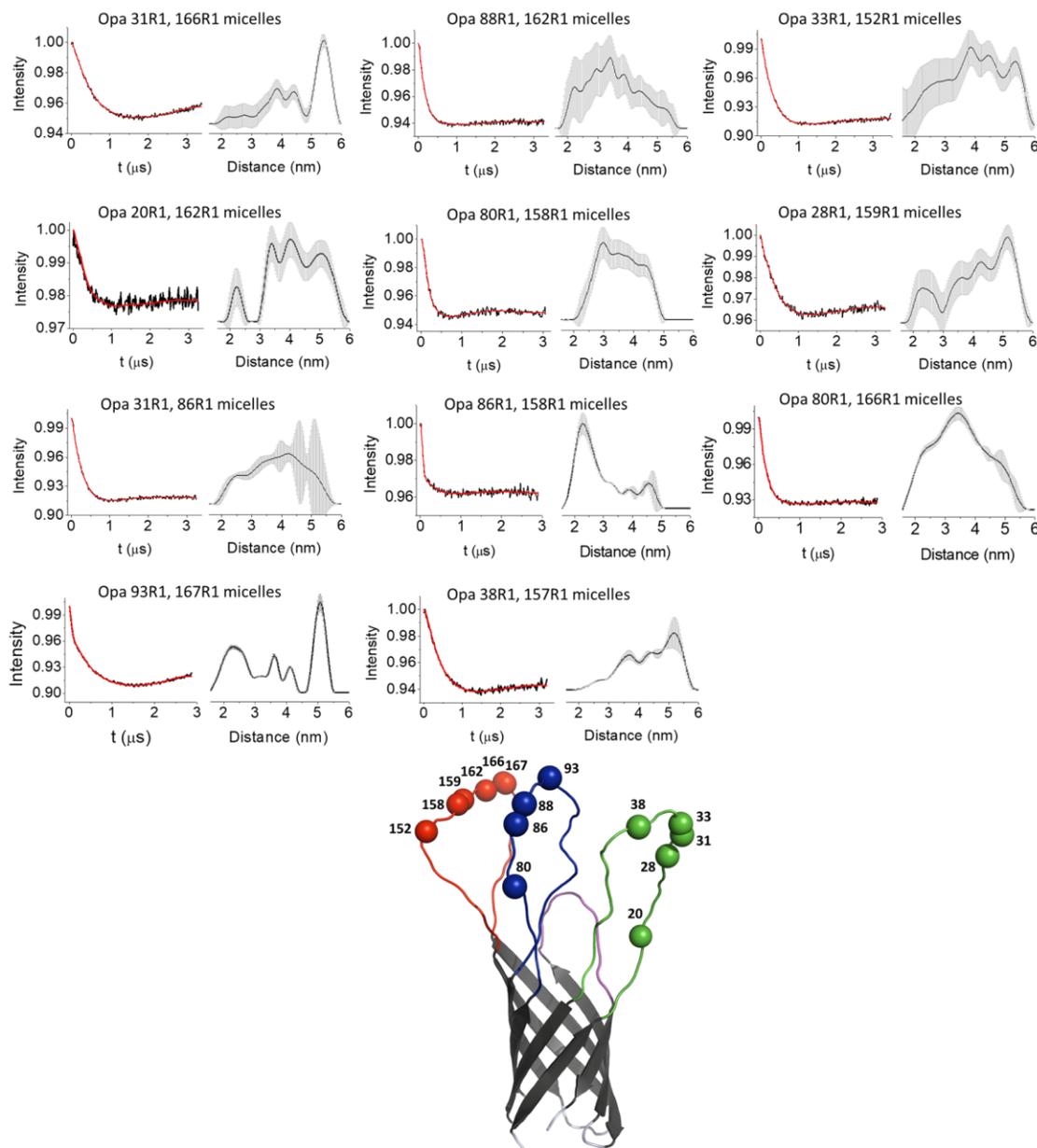
After incorporating this first round of multi-modal DEER distribution restraints into simulations of Opa loops (simulations were performed using Gromacs 5.2 with a CHARMM36 forcefield), a second list of informative DEER pairs was identified and measured. These pairs are incorporated into simulations of Opa loops, from which an experimentally-restrained, atomistic simulation of Opa can be generated to determine the conformational ensemble of unbound Opa loops.



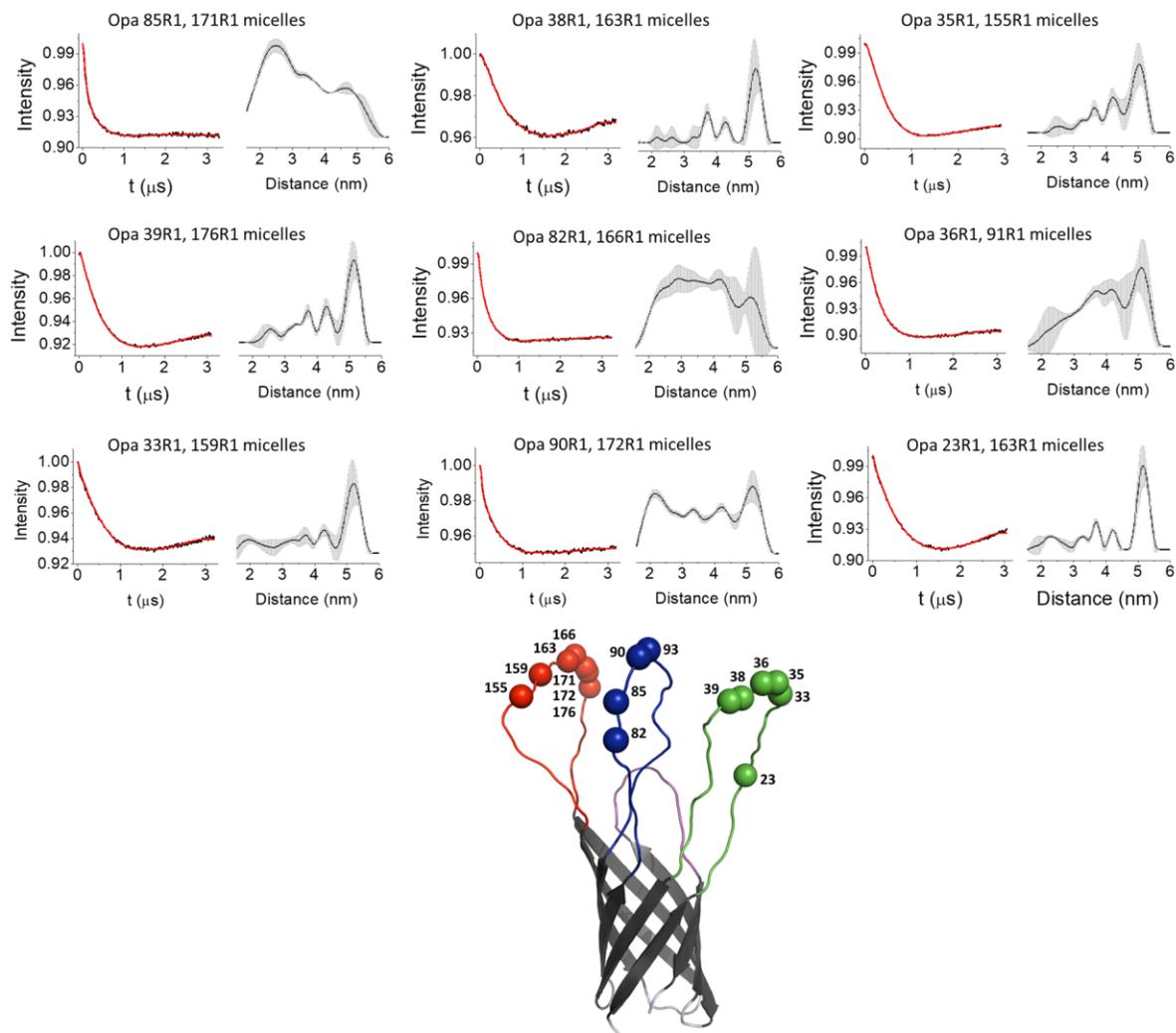
**Figure 3.8: Spectroscopist-selected DEER pairs.** DEER pairs that were generated prior to our collaboration with the Kasson lab were chosen in regions of limited dynamics (loop-barrel or intra-loop) that were comparatively uninformative as experimental restraints for MD simulations of Opa loops.

### 3.5.2 Measuring informative DEER pairs

The mRMR-selected pairs were measured with DEER where Opa proteins were reconstituted into detergent micelles. An initial set of eleven DEER pairs were measured, followed by nine second-round pairs, where the resulting distributions are indicative of high degrees of conformational heterogeneity with broad distance distributions that spanned 30 Å or more (Figures 3.9 and 3.10). Without the application of MD simulations, these distributions are difficult to interpret structurally due to their multi-modal, broad nature. However, a few trends among the DEER distributions were identified. Distance distributions from spin labeled pairs located on 1 and 2 loops all appear very broad with distances ranging from approximately 2-6 nm. Residues pairs located along 2 and 3 generate similarly broad distributions, with the exception of the pairs 93-167 and 86-158. Conversely, most pairs located at the 1 and 3 regions generate distance distributions that are limited below approximately 5 nm, with the exception of 33-152, 28-159, 20-162, and 38-157. Together, these data seem to suggest that there is an interaction between loops 2 & 3 and 1 & 2, but not between loops 1 & 3. However, the pairs that do not follow these trends are particularly interesting and cannot be easily explained by examining DEER distributions independently. These pairs may be suggestive of particular loop arrangements that can best visualized with MD.



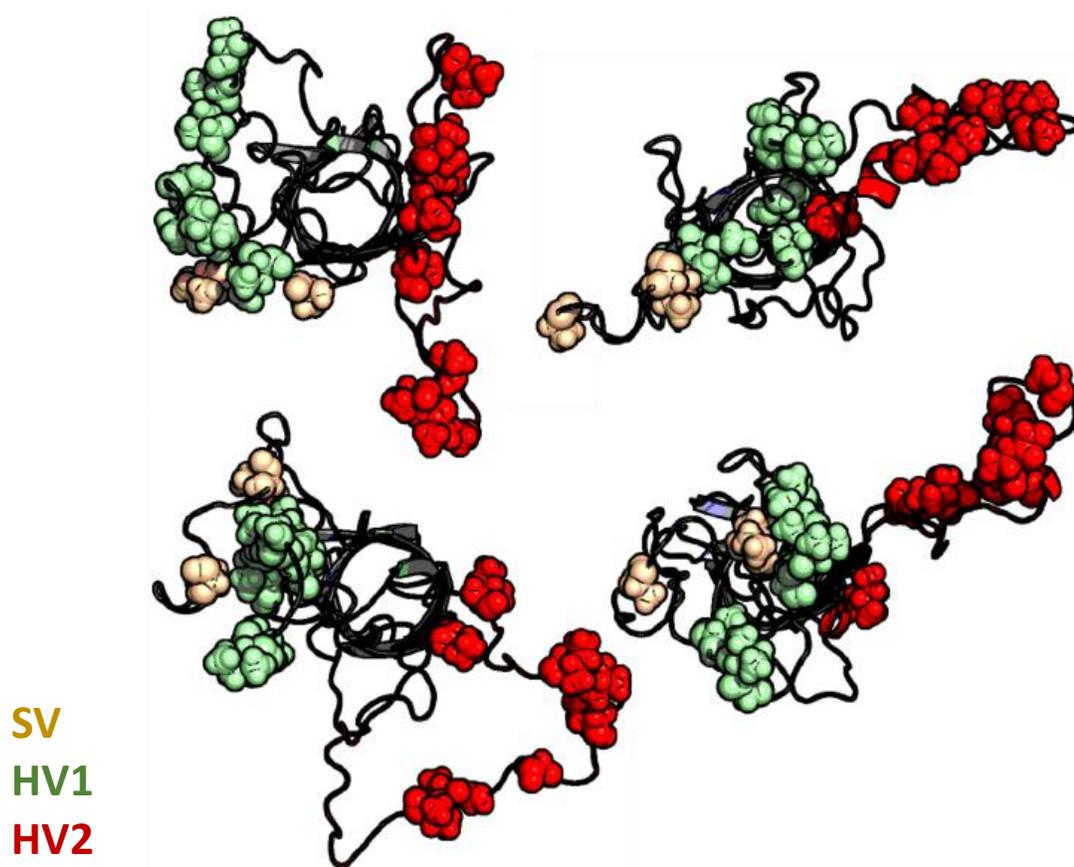
**Figure 3.9: First round of DEER measurements of mRMR pairs.** Eleven DEER pairs were identified using the mRMR method and were measured with Opa proteins in detergent micelles. The resulting distance distributions are all multi-modal and broad, which is indicative of high degrees of conformational heterogeneity. Spin-labeled residues are highlighted as spheres, where loop 1 is shown in green, loop 2 in blue, loop 3 in red, and loop 4 in purple



**Figure 3.10: Second round of DEER measurements of mRMR pairs.** A second list of mRMR DEER pairs were identified from the first round of DEER-restrained Opa simulations. These nine DEER pairs were measured with Opa proteins in detergent micelles. These DEER distributions are currently being incorporated into MD simulation of Opa loops, from which a high-resolution loop conformational ensemble can be generated. Spin-labeled residues are highlighted as spheres, where loop 1 is shown in green, loop 2 in blue, loop 3 in red, and loop 4 in purple.

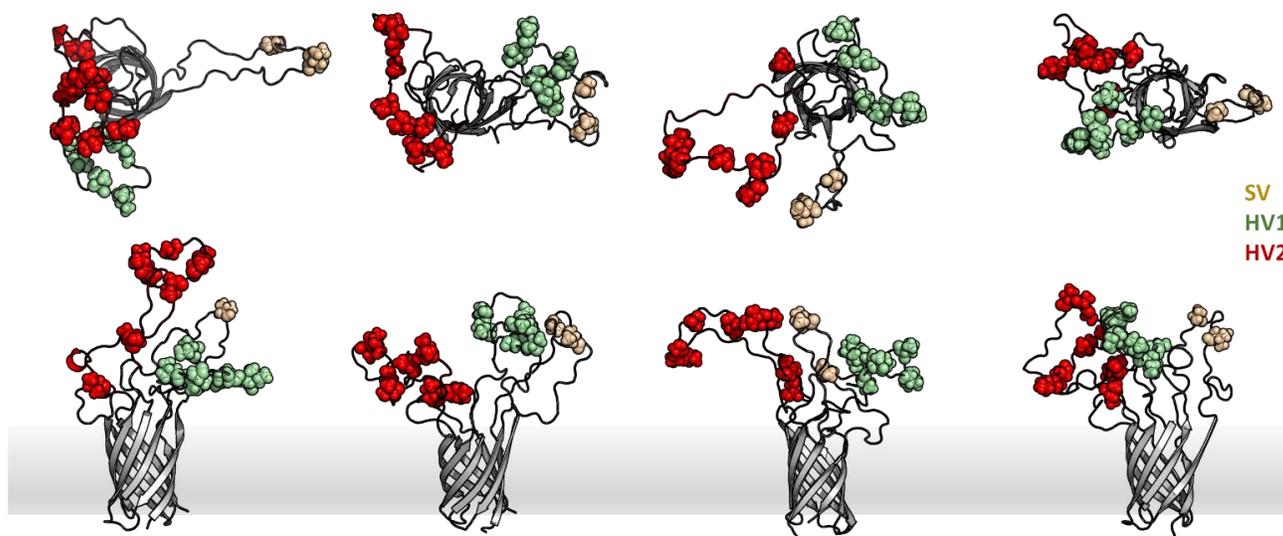
### 3.6 Opa loop conformational ensemble

The first mRMR DEER pairs measured were utilized as multi-modal restraints in MD simulations of Opa loops (with Opa in a DMPC lipid system). Several well separated clusters were identified, from which four of the most populated clusters are shown in Figures 3.11 (from round 1 of mRMR restraints) and 3.12 (from round 2 of mRMR restraints).



**Figure 3.11: Refinement of Opa<sub>60</sub> loops from the first round of mRMR DEER pairs.**

*Figure generated by Jennifer Hays and Peter Kasson.* These structures are from four conformational clusters identified in the DEER-restrained ensemble simulations of Opa loops. Hydrophobic residues are shown as spheres, where residues in the SV region are yellow, HV1 are green, and HV2 are red. Four preliminary Opa loop conformations were identified from round 1 of DEER-restrained MD simulations of Opa loops. SV- and HV1-containing loops interact in two of the clusters, and HV1- and HV-2-containing loops closely interact in the remaining two clusters (where SV and HV1 interact as well).



**Figure 3.12: Refinement of Opa<sub>60</sub> loops from the second-round of mRMR DEER pairs.** Shown are four of the most populated states after incorporating the second-round DEER pairs into restrained-ensemble simulations of Opa loops. The most populated state is shown on the left. Hydrophobic residues are indicated as spheres, where residues in the SV region are yellow, HV1 are green, and HV2 are red.

### 3.7 Conclusions

The preliminary model of the conformational ensemble of Opa loops leads to two potential hypotheses regarding the mechanism of Opa loop interactions with host receptors. Chimeric studies of Opa HV regions between two CEACAM-binding Opa proteins eliminated *Neisseria* internalization into HeLa cells transfected to express CEACAM receptors [55]. This suggests that a specific combination of HV1 and HV2 is required for receptor engagement. As both HV regions were shown to be necessary for internalization, it has long been assumed that interactions between loops 2 and 3 (containing HV1 and HV2 respectively) form the CEACAM-binding site. Our preliminary Opa loop ensemble where HV1 and HV2 closely interact in a subset of the clusters may support a combined loop surface which engages respective receptors. However, these data-driven simulations suggest other hypotheses that are equally feasible and were not previously considered. Other loop conformations observed in simulations of Opa are consistent with the requirement of HV1 and HV2 in chimeric experiments yet suggest the HV1 and HV2 do not need to both form the binding surface. An SV/HV1 interaction that is seen in several clusters of mRMR-restrained ensemble provide support for an alternative binding mechanism, where HV2 is readily available for interaction with receptor because of an interaction between HV1 and SV.

Because no binding motif has been identified in Opa, we can only speculate on how loop dynamics modulate receptor binding and which regions directly interact with receptors. Perhaps induced folding or conformational selection enables one Opa protein to selectively bind several CEACAM receptors (and conversely many Opa proteins to bind the same CEACAM), where a range of loop conformations are sampled and are modulated

by neighboring loops (with only some being receptor competent conformations). The conformational ensemble of Opa loops in the absence of receptor can be compared to the loop ensemble upon binding CEACAM to identify the mechanism of Opa-CEACAM interactions. If the loops sample a receptor-competent state in their unbound form, the mode of binding may be conformational selection, whereas if the loop binding conformation is unobserved in the unbound state, an induced folding or “fuzzy complex” may be the mechanism of Opa-CEACAM interactions. Ultimately, understanding conformation of Opa<sub>60</sub> loops in their bound and unbound forms will elucidate the mechanism of the Opa<sub>60</sub>-CEACAM1 interaction, which could be conserved across the family of Opa and CEACAM proteins.

1. Martin, J.N., et al., *Neisserial Opa Protein-CEACAM Interactions: Competition for Receptors as a Means of Bacterial Invasion and Pathogenesis*. *Biochemistry*, 2016. **55**(31): p. 4286-4294.
2. Fox, D.A., et al., *Structure of the Neisserial Outer Membrane Protein Opa(60): Loop Flexibility Essential to Receptor Recognition and Bacterial Engulfment*. *Journal of the American Chemical Society*, 2014. **136**(28): p. 9938-9946.
3. Kroncke, B.M. and L. Columbus, *Identification and removal of nitroxide spin label contaminant: Impact on PRE studies of  $\alpha$ -helical membrane proteins in detergent*. *Protein Science : A Publication of the Protein Society*, 2012. **21**(4): p. 589-595.
4. Kroncke, B.M., P.S. Horanyi, and L. Columbus, *Structural Origins of Nitroxide Side Chain Dynamics on Membrane Protein  $\alpha$ -Helical Sites*. *Biochemistry*, 2010. **49**(47): p. 10045-10060.
5. Do Cao, M.-A., et al., *Probing the conformation of the resting state of a bacterial multidrug ABC transporter, BmrA, by a site-directed spin labeling approach*. *Protein Science : A Publication of the Protein Society*, 2009. **18**(7): p. 1507-1520.
6. Gross, A., et al., *Structure of the KcsA potassium channel from *Streptomyces lividans*: a site-directed spin labeling study of the second transmembrane segment*. *Biochemistry*, 1999. **38**(32): p. 10324-35.
7. Dewald, Alison H., Jacqueline C. Hodges, and L. Columbus, *Physical Determinants of  $\beta$ -Barrel Membrane Protein Folding in Lipid Vesicles*. *Biophysical Journal*, 2011. **100**(9): p. 2131-2140.
8. Needham, D., T.J. McIntosh, and D.D. Lasic, *Repulsive interactions and mechanical stability of polymer-grafted lipid membranes*. *Biochim Biophys Acta*, 1992. **1108**(1): p. 40-8.

9. Immordino, M.L., F. Dosio, and L. Cattel, *Stealth liposomes: review of the basic science, rationale, and clinical applications, existing and potential*. International Journal of Nanomedicine, 2006. **1**(3): p. 297-315.
10. Redfield, A.G., *The Theory of Relaxation Processes*\* \*This work was started while the author was at Harvard University, and was then partially supported by Joint Services Contract N5ori-76, Project Order I, in *Advances in Magnetic and Optical Resonance*, J.S. Waugh, Editor. 1965, Academic Press. p. 1-32.
11. Virji, M., et al., *Critical determinants of host receptor targeting by Neisseria meningitidis and Neisseria gonorrhoeae: identification of Opa adhesiotopes on the N-domain of CD66 molecules*. Molecular Microbiology, 1999. **34**(3): p. 538-551.
12. Bos, M.P., et al., *Carcinoembryonic antigen family receptor recognition by gonococcal Opa proteins requires distinct combinations of hypervariable Opa protein domains*. Infection and Immunity, 2002. **70**(4): p. 1715-1723.
13. Ding, F., M. Layten, and C. Simmerling, *Solution Structure of HIV-1 Protease Flaps Probed by Comparison of Molecular Dynamics Simulation Ensembles and EPR Experiments*. Journal of the American Chemical Society, 2008. **130**(23): p. 7184-7185.
14. Galiano, L., et al., *Drug Pressure Selected Mutations in HIV-1 Protease Alter Flap Conformations*. Journal of the American Chemical Society, 2009. **131**(2): p. 430-431.
15. Roux, B. and S.M. Islam, *Restrained-Ensemble Molecular Dynamics Simulations Based on Distance Histograms from Double Electron-Electron Resonance Spectroscopy*. The journal of physical chemistry. B, 2013. **117**(17): p. 4733-4739.
16. Tessmer, M.H., et al., *Identification of a ubiquitin-binding interface using Rosetta and DEER*. Proceedings of the National Academy of Sciences, 2018. **115**(3): p. 525-530.

17. Hirst, S.J., et al., *RosettaEPR: An integrated tool for protein structure determination from sparse EPR data*. *Journal of Structural Biology*, 2011. **173**(3): p. 506-514.
18. Islam, S.M., et al., *Structural Refinement from Restrained-Ensemble Simulations Based on EPR/DEER Data: Application to T4 Lysozyme*. *The Journal of Physical Chemistry B*, 2013. **117**(17): p. 4740-4754.
19. Peng, H.C., F.H. Long, and C. Ding, *Feature selection based on mutual information: Criteria of max-dependency, max-relevance, and min-redundancy*. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2005. **27**(8): p. 1226-1238.

## Chapter 4: Homodimerization of nCEACAM1 proteins

### 4.1 CEACAMs are adhesion molecules

Determining the mechanism of Opa-CEACAM interactions also requires a better characterization of CEACAM receptors in the absence of Opa proteins. CEACAMs (carcino embryonic antigen-like cellular adhesion molecules) are membrane proteins that participate in cis and trans homotypic and heterotypic interactions, often through their N-domains (nCEACAM) [86, 135, 136]. The N-terminal Ig domain of CEACAM proteins has two faces – a side that is heavily glycosylated and a face which participates in adhesion interactions (including binding to Opa proteins [46, 86]). Alanine-scanning mutagenesis identified residues V39 and D40 the of non-glycosylated face of nCEACAM1 as being critical in CEACAM homodimerization (Figure 4.1) [136].

The dimerization of cell surface molecules is a common process in the transduction of extracellular signals to the cell interior [137]. The oligomerization of proteins containing several extracellular Ig-like domains in particular are often found in transmembrane signaling events in the cellular immune response [138] and various physiological and pathogenic processes [139]. Likewise, the binding of CEACAM extracellular Ig-like domains stimulates CEACAM-mediated transmembrane signaling. Subtle changes in CEACAM adhesion is utilized to transmit information throughout the length of multiple CEACAM extracellular Ig domains into the cell interior to cytoplasmic ITIM and ITAM motifs (immunoreceptor tyrosine-based inhibitory/activation motif) of certain CEACAMs [140, 141]. Transmembrane signaling of CEACAM1 can be regulated by CEACAM1 trans-homophilic oligomerization, which induces cis-dimerization via an allosteric

mechanism transmitted through the N-domain, and in turn, alters the binding of various signaling molecules [142].

Trans-oligomerization of homotypic nCEACAM interactions with neighboring cells also promotes cell adhesion [136, 143]. These homotypic trans-oligomers are further stabilized by connected Ig-domains [143, 144]. The presence of numerous extracellular Ig domains may also enable CEACAMs to extend a greater distance from the membrane to facilitate binding accessibility. The extracellular domains of CEACAM are flexible with respect to each other, but are stabilized with cis-interactions between Ig-like domains of neighboring CEACAM molecules in the same membrane plane [144]. The oligomeric states of CEACAM1 extracellular domains are also very diverse, where Ig domains aggregate via one or multiple domains on the same cell or neighboring cells (Figure 4.2) [144]. The transmembrane domain of CEACAM1 was also found to facilitate cis-dimerization [145, 146] via a GxxG motif that promotes transmembrane helical association [147]. This clustering of transmembrane domains may be a prerequisite for signaling.

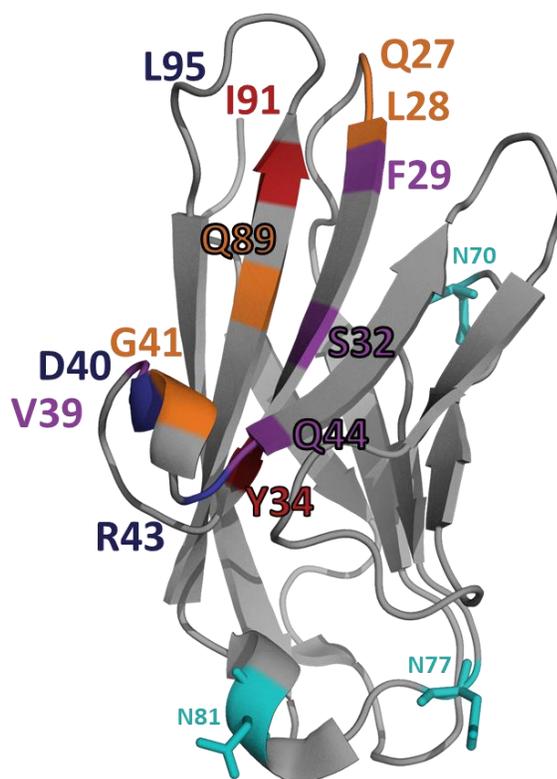
The induction of transmembrane signaling via CEACAM oligomerization requires a connection to the cytosol. CEACAM1 specifically contains a long (L; 71 amino acids) or a short (S; 10 amino acids) cytosolic domain [60], where CEACAM1-L contains an immunoreceptor tyrosine-based inhibitory motif (ITIM). The CEACAM1-L and the CEACAM1-S isoforms are often co-expressed on the same cell, with the expression ratio being specific to cell type [148, 149], as the expression of the short isoform often interferes with CEACAM1-L generated signals [81, 142]. CEACAM1-L (and on occasion CEACAM1-S) interacts with cytoplasmic tyrosine kinases, tyrosine phosphatases, calmodulin, actin, filamin, and tropomyosin, among other signaling molecules to induce

cytoskeletal rearrangements and endocytosis [150]. Bacterial pathogens that bind CEACAMs (e.g. *Neisseria*) exploit these native CEACAM functions to enhance their colonization of the host.

Analysis of the dimerization of recombinant, non-glycosylated, isolated Ig domains of CEACAM proteins was examined via sedimentation equilibrium analysis using analytical ultracentrifugation [151]. nCEACAM1 exhibited an average molecular weight of  $23.0 \pm 0.9$  kDa (theoretical monomer MW = 12.2 kDa) indicating the non-glycosylated nCEACAM1 proteins are primarily dimer, and a dimerization constant of  $450 \text{ nM}$  ( $+370/-230 \text{ nM}$ ) was determined [151]. Conversely nCEACAM6 appears to exhibit a weak monomer-dimer association with an average molecular weight of  $15.7 \pm$  kDa and a theoretical affinity of  $60 \pm 15 \mu\text{M}$  [151]. nCEACAM5 has an average molecular weight of  $23 \pm 0.8$  kDa and exhibits a homodimer binding affinity between that of nCEACAM1 and nCEACAM6 at  $1.3 \pm 0.6 \mu\text{M}$  [151]. As such, the oligomerization propensities of different CEACAMs may be varied. Similar experiments found the dissociation constant of the recombinant, non-glycosylated nCEACAM1 homodimer to be  $100 \pm 17 \text{ nM}$  and nCEACAM1 interactions with the Ig-like A3 domain to yield a  $K_D$  of  $18 \pm 3 \text{ nM}$  using ELISA binding assays [152].

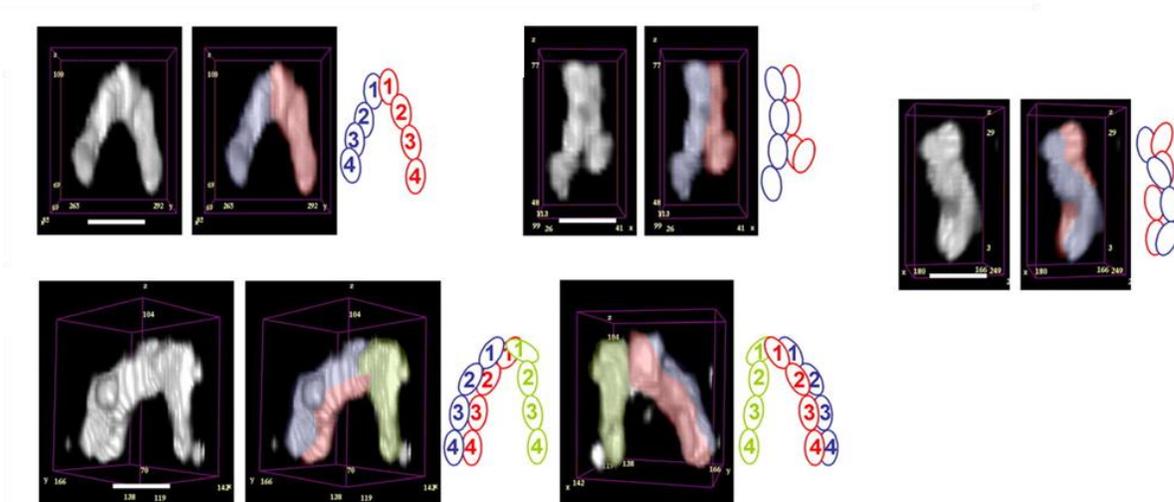
Two crystal structures of nCEACAM1 (PDB IDs: 2gk2 and 5dzl, Figure 4.3) have been suggested as models of the N-domain CEACAM dimer [86, 153]. 2gk2 is an unlikely candidate, as the CEACAM oligomerization face in this crystal structure would be occluded *in vivo* by the heavy glycosylation of that region [86]. The subunits of the 2gk2 crystal structure are also coordinated via  $\text{Ni}^{2+}$  to H105 residues, where such divalent interactions may mediate the formation of artificial dimers. The 5dzl dimer structure is

more probable, as the CEACAM contact face includes residues known to be involved in homodimerization [153]. However, neither of these structures contain proteins that are glycosylated, which may ultimately impact the dimerization of nCEACAM1 *in vivo*. Likewise, crystal packing interactions and crystallization conditions (especially large PEG molecules [154]) may impact the orientation of CEACAMs within the crystal units, and consequently the dimer structure [155]. As such, we aim to study glycosylated and non-glycosylated nCEACAM1 homodimerization, as well as Opa proteins, in solution to further investigate nCEACAM interactions.

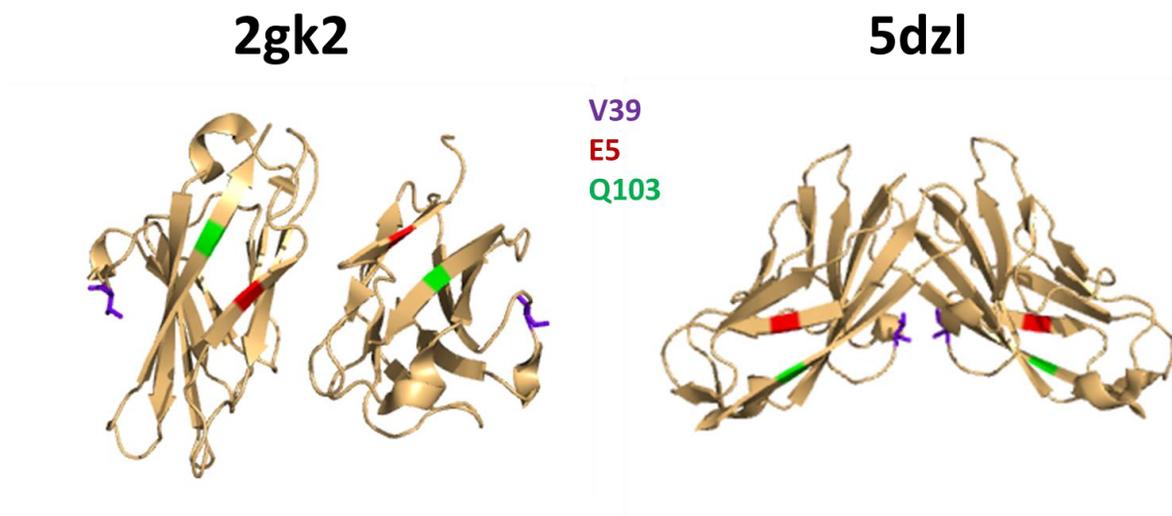


**Figure 4.1: Residues involved in nCEACAM1 binding interactions (2gk2) [86, 136].**

Residues in red are necessary for all interactions with CEACAM-binding Opa proteins. Orange residues are necessary to bind only certain Opa variants. Residues in blue are involved in hetero- and homo-typic CEACAM interactions. Purple residues are necessary to interact with Opa proteins as well as other CEACAMs. N-linked glycosylation sites (N70, N77, N81) are shown in cyan.



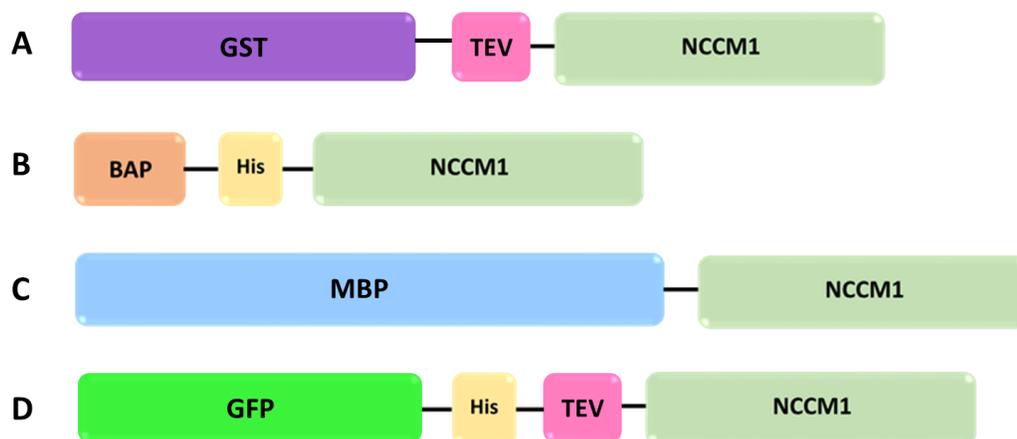
**Figure 4.2: Arrangements of CEACAM1 ectodomains.** *Reprinted with permission from [142].* Selected arrangements of soluble CEACAM1 ectodomains (domains 1 - 4) on the same cell that were imaged via molecular electron tomography are shown at a resolution of approximately 20 Å. The interactions of these domains are complex and diverse, where the N-domain alone or multiple domains adhere.



**Figure 4.3: Potential nCEACAM1 homodimer crystal structures.** [86, 153] The structures 2gk2 and 5dzl have been proposed as potential structures of the nCEACAM1 homodimer. Residue V39 (purple) is known to be necessary for CEACAM adhesion. Two spin-labeled sites used in DEER experiments, E5 (red) and Q103 (green), are also highlighted. 2gk2 is an unlikely candidate as the homodimer structure because the contact face is glycosylated *in vivo*, unlike the crystal contacts of 5dzl.

#### 4.2 Developing methods to disrupt CEACAM homodimerization

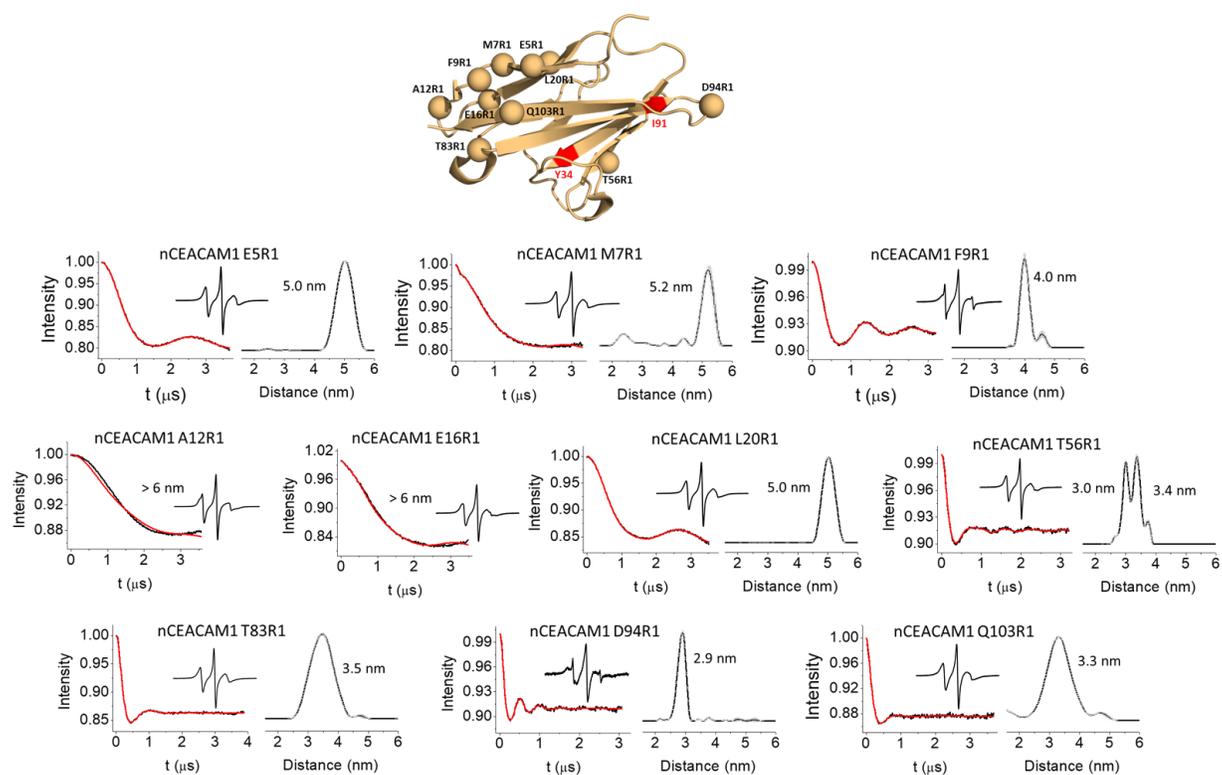
Prior to performing Opa-CEACAM binding experiments with EPR, CEACAM must be spin labeled and soluble at concentrations of roughly 200  $\mu\text{M}$ . Unfortunately, the nCEACAM1 construct that was previously used in Opa-binding studies [119] is not applicable to EPR experiments. The nCEACAM1 construct used in these binding studies is only monomeric at low concentrations ( $\sim 50 \mu\text{M}$ ) or with an attached glutathione S-transferase (GST) affinity tag [119]. The GST tag contains several critical cysteine residues, and therefore cannot remain attached to nCEACAM1 for spin labeling. As such, the dimerization propensity of several different CEACAM constructs were studied prior to performing Opa-CEACAM experiments. Four CEACAM variants were examined in this work: (1) nCEACAM with no attached affinity tag or glycosylation, (2) MBP-nCEACAM1, (3) BAP-nCEACAM1, and (4) N-linked glycosylated nCEACAM1 produced in HEK293 cells (Figure 4.4).



**Figure 4.4: Schematic of nCEACAM1 constructs used to study nCEACAM1 homodimerization.** (A) The GST-TEV-nCEACAM1 construct was previously used in Opa binding studies, where the GST tag remained attached [119]. However, this tag must be removed using TEV protease for DEER experiments, leaving non-glycosylated nCEACAM1. (B) A BAP-His<sub>6</sub>-nCEACAM1 construct was synthesized using the Gibson Assembly protocol (New England BioLabs, [156]). This protein was biotinylated, spin labeled, and immobilized onto avidin resin to limit CEACAM diffusion and dimerization for DEER experiments. (C) A MBP-nCEACAM1 construct was synthesized by General Biosystems. The large MBP tag may sterically limit CEACAM-CEACAM interactions. (D) A GFP-His<sub>6</sub>-TEV-nCEACAM1 construct was used to generate glycosylated nCEACAM1 proteins in mammalian HEK293 cell cultures [157]. The GFP is cleaved using TEV protease to leave purified, glycosylated nCEACAM1 proteins.

#### 4.2.1 nCEACAM1 with no glycosylation or attached affinity tag

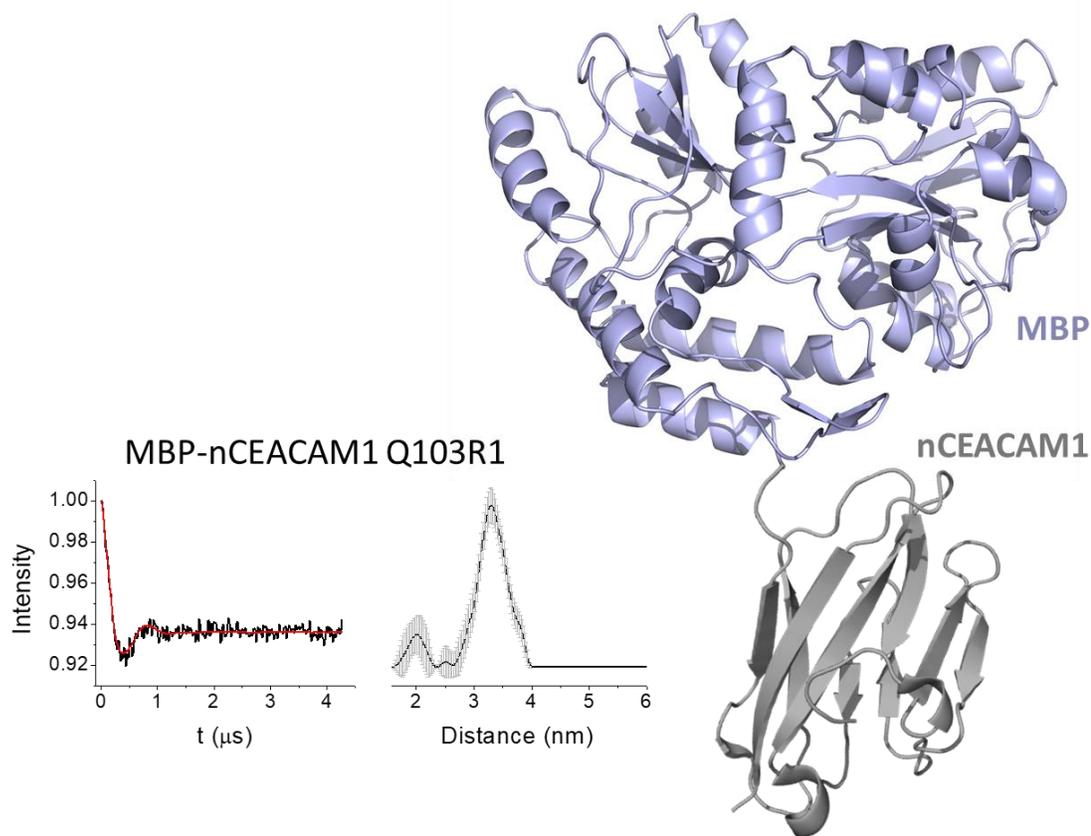
The non-glycosylated nCEACAM1 construct that was previously utilized in Opa-CEACAM binding experiments [119] (with glutathione S-transferase (GST) cleaved) was first spin-labeled for DEER experiments. Ten cys-mutants of this non-glycosylated nCEACAM1 were spin labeled and measured using DEER. Each DEER distribution contained a single distance value that is indicative of CEACAM homodimerization (Figure 4.5). Binding was not observed using DEER, as the addition of Opa proteins (in liposomes or micelles) showed only the dimer peak in the DEER distribution. Additional studies of the nCEACAM1 homodimer structure will utilize these DEER distributions and are discussed later in this chapter.



**Figure 4.5: DEER distributions of non-glycosylated nCEACAM1.** Ten spin labeled locations along non-glycosylated nCEACAM1 were probed using DEER. These distances are used as a comparison to predicted distances in the nCEACAM1 homodimer 5dzl structure. Labeled residues 12R1 and 16R1 contain long distances that are above the measurable mean limit of approximately 6 nm. The distance distribution of 56R1 contains two peaks of similar mean distances, which is likely indicative of two spin label conformers rather than two dimer conformations. CW-EPR lineshapes of the spin labeled CEACAMs are shown as inserts.

#### 4.2.2 MBP-nCEACAM1

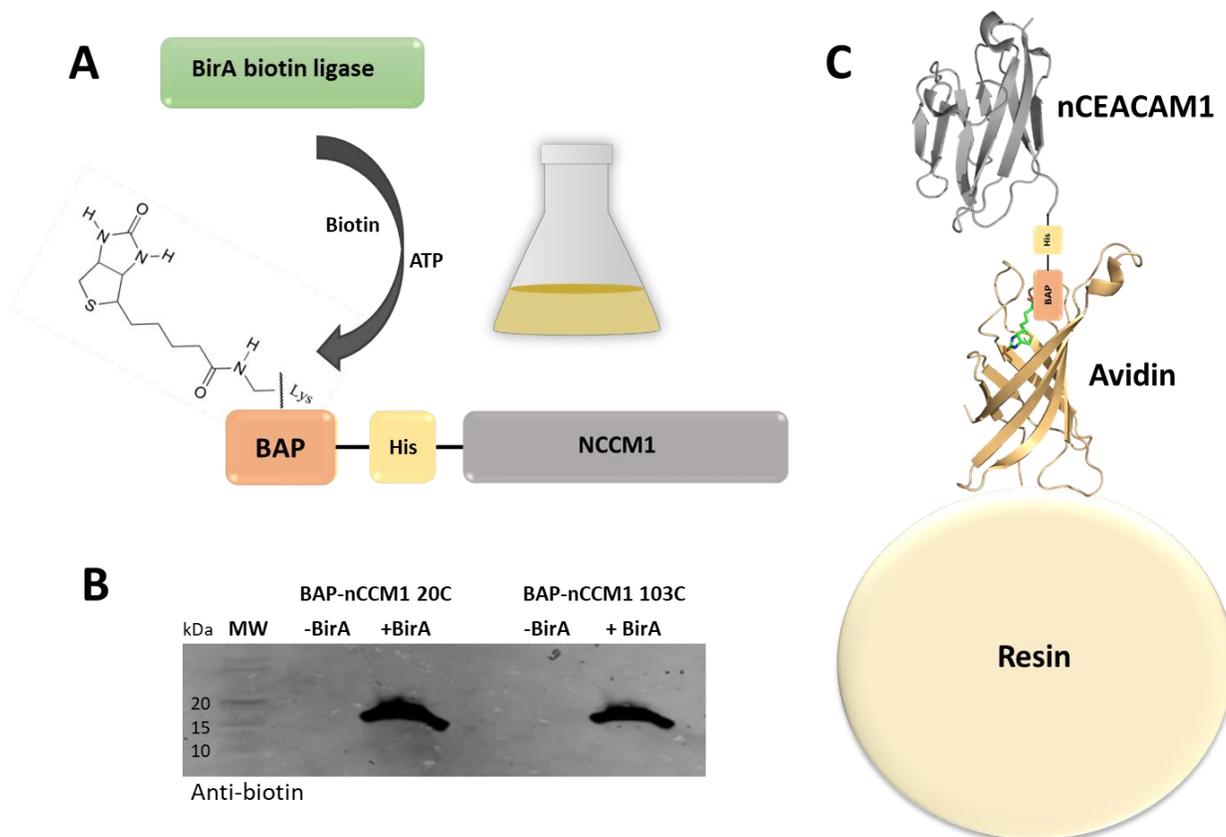
Maltose-binding protein (MBP) is a large 43 kDa protein that is often used as an affinity tag for recombinant protein purification. MBP binds to amylose with high affinity, which can be utilized for protein purification via amylose immobilized resin. MBP tags often prevent protein aggregation due to its large size [158, 159]. An MBP-nCEACAM1 construct was made that was hypothesized to prevent CEACAM dimerization as a result of steric restrictions from the large size of the attached MBP. While the expression, purification, and spin labeling of this MBP-nCEACAM1 construct were successful, the resulting DEER signal of spin labeled CEACAM proteins (103R1) in isolation was still indicative of a CEACAM dimer. Binding was not observed using DEER with Opa in liposomes or micelles (Figure 4.6).



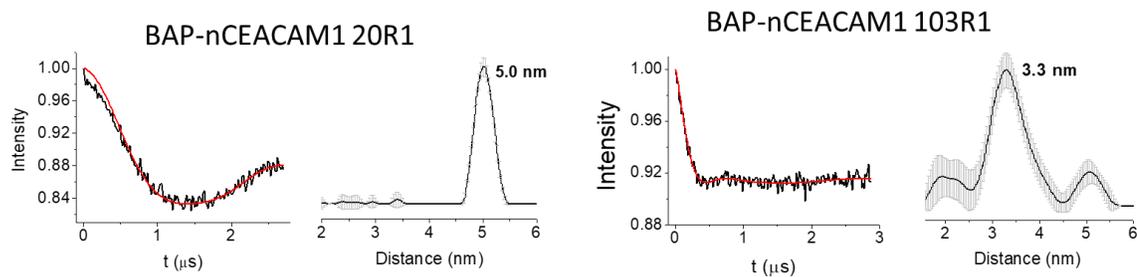
**Figure 4.6: DEER distributions of MBP-nCEACAM1.** MBP-nCEACAM1 proteins were spin labeled with the MBP affinity tag attached. The resulting Q103R1 DEER distribution is indicative of nCEACAM1 homodimerization. MBP is shown in blue and nCEACAM1 in grey.

### 4.2.3 BAP-nCEACAM1

Next, a nCEACAM1 construct was generated with a biotinylatable BAP (biotin acceptor peptide) tag (Figure 4.7). We hypothesized that immobilizing nCEACAM1 directly from the cell lysate onto avidin agarose beads (via biotin-avidin binding) would limit nCEACAM1 dimerization, as freely diffusing nCEACAM1 in solution would be limited. BAP-CEACAM proteins were successfully expressed and biotinylated in *E. coli* via a co-transformation of the desired cysteine-containing CEACAM plasmid and a pBirAcm plasmid that expresses a BirA biotin ligase protein. Dual antibiotic selection was utilized, and IPTG induced the expression of both proteins. BirA biotinylates the BAP tag upon the incorporation of biotin into the expression media (a Western Blot validating this in-cell biotinylation technique is shown in Figure 4.7). The cell lysate was added directly to monomeric avidin resin, which both purified and immobilized nCEACAM1, after which the proteins were spin labeled. However, like the two previously studied CEACAM constructs, the spin labeled, immobilized BAP-nCEACAM1 proteins generated a strong DEER signal that is indicative of CEACAM homodimerization. Binding was not observed using DEER with the addition of Opa proteins in liposomes or micelles (Figure 4.8).



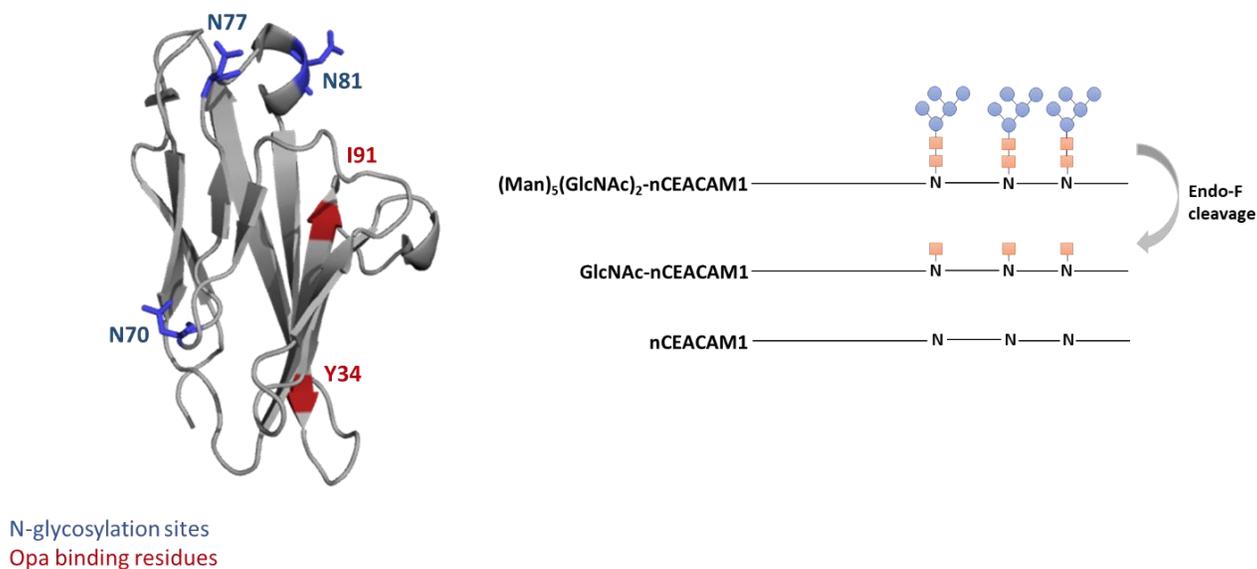
**Figure 4.7: Generating BAP-nCEACAM1 proteins that are immobilized on avidin resin using in-cell biotinylation with BirA.** (A) *E. coli* are co-transformed with a pET9a vector containing the BAP-nCEACAM1 sequence and a pBirAcm plasmid encoding a gene for the BirA enzyme. IPTG induces the expression of both proteins, and the addition of biotin to the growth media results in the biotinylation of an attached BAP tag. (B) This in-cell biotinylation technique was verified using an anti-biotin western blot. Only cells expressing both BirA and BAP-nCEACAM1 proteins resulted in BAP biotinylation. (C) These biotinylated BAP-nCEACAM1 proteins were then immobilized onto monomeric avidin beads. The CEACAM1 proteins are spin-labeled following immobilization for DEER studies.



**Figure 4.8: Immobilized BAP-nCEACAM1 DEER distributions.** Biotinylated BAP-nCEACAM1 proteins are immobilized onto monomeric avidin resin and are spin labeled. The resulting DEER distance distributions are indicative of nCEACAM1 homodimerization.

### 4.3 N-linked glycosylation may disrupt nCEACAM1 homodimerization

In 2016 the Moremen and Prestegard labs at the University of Georgia utilized N-linked glycosylated nCEACAM1 proteins and NMR residual dipolar coupling (RDC) techniques to study CEACAM homodimerization [157], as nCEACAM1 contains three N-glycosylation sites - N70, N77, and N81 (Figure 4.9). These labs compared two glycosylated nCEACAM1 variants: one CEACAM contains a high degree of glycosylation, and a second contains a single GlcNAc at each of the three N-glycosylation sites. Their results indicated that non-glycosylated nCEACAM1 proteins exist primarily as dimers, whereas minimal glycosylation inhibits homodimerization (with nCEACAM GlcNAc being the most monomeric form of CEACAM1). Thus, we utilized their approach (with the guidance of Jeong Yeh in the Moremen lab at the University of Georgia) to generate N-glycosylated nCEACAM1 proteins to compare the dimerization propensity and examine the binding of glycosylated nCEACAM1 to Opa<sub>60</sub> proteins. Our findings on the dimerization of glycosylated nCEACAM1 will be compared to the results of the Moremen and Prestegard labs and are discussed later in this chapter.



**Figure 4.9: nCEACAM1 contains three N-linked glycosylation sites.** nCEACAM1 proteins are characterized by two sides – a glycosylated face and an adhesion face. Three N-glycosylation sites (blue) are on strands opposite those containing Opa and CEACAM binding residues (red). An *in vitro* nCEACAM1 N-glycosylation method utilizes HEK293S GnT1<sup>-</sup> cells to attach (Man)<sub>5</sub>(GlcNAc)<sub>2</sub> glycans to the three N-glycosylation motifs. EndoF1 enzymes cleave the glycans to leave a single GlcNAc unit at each of the three sites.

#### 4.3.1 N-linked glycosylation

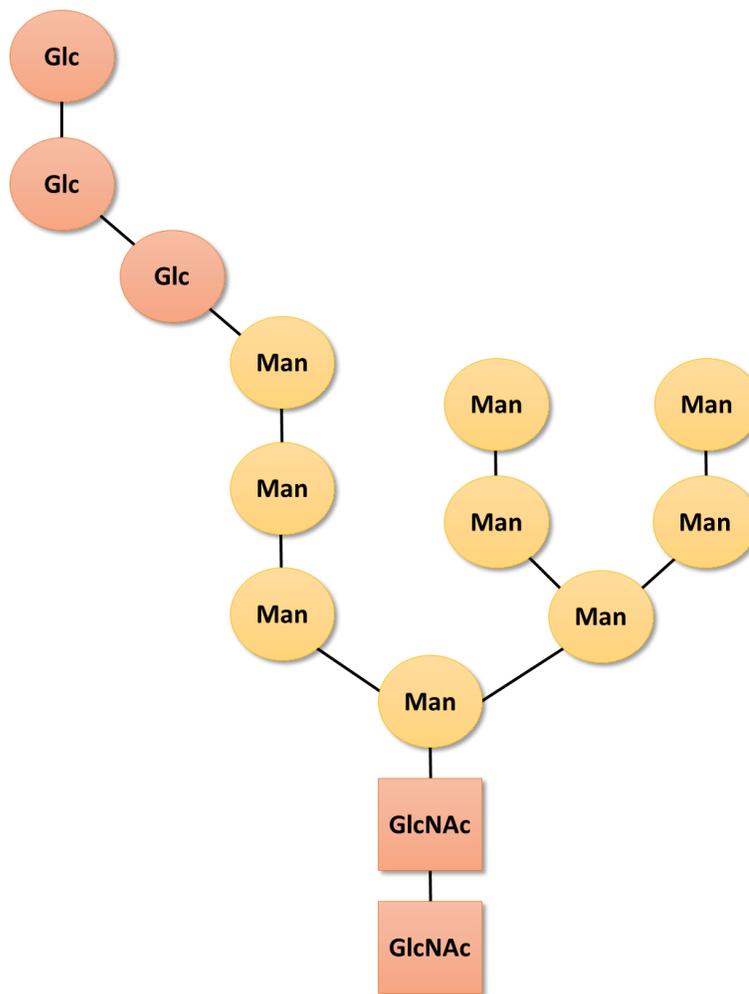
Asparagine (N)-linked glycosylation is a post-translational modification mechanism found in prokaryotes and eukaryotes that results in the covalent attachment of an oligosaccharide onto specific N residues. The oligosaccharide chain length and composition vary; however, the basic mechanism of N-glycosylation is conserved among species [160]:

- (1) A glycan is assembled on a lipid anchor (dolichylpyrophosphate in eukaryotes) from nucleotide-activated monosaccharide building blocks (GDP-Man, UDP-GlcNAc, and UDP-Glc) via the stepwise incorporation of monosaccharides by specific glycosyltransferases to form a lipid-linked oligosaccharide (LLO) (Figure 4.10). The LLO is then re-oriented from the cytosolic to the luminal side of the eukaryotic ER membrane (or plasma membrane of prokaryotes) where it acts as a glycosylation donor and the glycans can be extended further in eukaryotes.
- (2) The protein consensus sequence for N-linked glycosylation (N-x-S/T) serves as an acceptor when translocated to the ER lumen (or periplasm) [161].
- (3) Oligosaccharyltransferase (OST) catalyzes the formation of an N-glycosidic linkage of the oligosaccharide to the asparagine of the acceptor protein.

The LLO has three principal carbohydrate components: (1) two N-acetylglucosamine (GlcNAc), (2) nine mannose (Man), and (3) three glucose (Glc) residues (Figure 4.1). This base oligosaccharide structure can be further modified in eukaryotes. In a later stage of eukaryotic glycoprotein processing (located in the Golgi), N-linked glycans are remodeled

in a protein, cell-type, and species-specific manner, which generates a high degree of glycan structural diversity.

When attached to a protein, the hydrophilic carbohydrates alter the biophysical properties of the protein, which can affect their folding, solubility, and activity [162-165]. Glycosylation serves as a quality control of protein folding in the ER, where the glycan signals whether a protein is correctly folded and can leave the ER to continue maturation in the Golgi [166, 167]. This quality control is mediated by the specificity of OST enzymes. However, the role of these important protein glycans is often unstudied due to their complexity and difficulty of glycosylated protein production.



**Figure 4.10: Lipid-linked oligosaccharide (LLO) glycan core assembled in N-linked glycosylation.** During N-linked glycosylation, two GlcNAc units, nine mannoses, and three glucose sugars are incrementally incorporated by glycosyltransferases onto a lipid anchor, forming a lipid-linked oligosaccharide. These glycans are transferred to a specific N-glycosylation consensus sequence.

#### 4.3.2 Mammalian recombinant expression to generate glycosylated nCEACAM1

Traditional recombinant protein expression techniques in bacterial systems do not incorporate post-translational modifications that are often found *in vivo* in eukaryotes. As CEACAM1 is a mammalian protein, recombinant expression in human cell cultures is required to generate glycosylated nCEACAM1 proteins with glycosylation patterns that model those found *in vivo* (as opposed to the less complex glycosylation patterns from insect cell cultures, for example [168]).

HEK293F (human embryonic kidney) cells are utilized to prepare a highly glycosylated nCEACAM1 isoform with N-linked glycans that are large and complex (with additional glycans such as GalNAc and galactose, for example [168]). These cells are well suited for suspension culture, where large volumes (often several liters) of HEK293F cells are transfected to express nCEACAM1 proteins, which are then collected in soluble mg quantities in the serum [157]. nCEACAM1 constructs contain an attached GFP (green fluorescent protein) tag to aid in protein solubilization, a His<sub>6</sub> tag used for IMAC affinity purification, a TEV cleavage site, and nCEACAM1 (Figure 4.4). The GFP-His-TEV residues are removed, leaving an isolated, purified nCEACAM1 protein for dimerization studies.

HEK293S GnT1<sup>-</sup> cells are transfected to express nCEACAM1 proteins that contain a high-mannose glycan variety with three (Man)<sub>5</sub>-(GlcNAc)<sub>2</sub> glycans bound. The GnTI gene is knocked out in this cell type, therefore the native expression of the protein GlcNAcT-I (alpha-1,3-mannosyl-glycoprotein 2-beta-N-acetylglucosaminyltransferase) is eliminated [169]. GlcNAcT1 is located within the Golgi membrane and initiates complex N-linked carbohydrate formation, as it converts high-mannose base glycans to hybrid and complex

N-glycans. [170, 171]. Therefore, nCEACAM1 proteins generated using this cell type have a monodisperse (Man)<sub>5</sub>-(GlcNAc)<sub>2</sub> glycan pattern at the three potential N-glycosylation sites. The complexity of these glycans can be further reduced using the enzyme Endoglycosidase F1 (EndoF1), which truncates the sugars to leave a lone, core GlcNAc unit at each site (but cannot cleave complex glycans that are generated with HEK293F cells) (Figure 4.9). Such nCEACAM1 GlcNAc proteins were also used to study CEACAM1 homodimerization.

#### 4.4 Characterizing nCEACAM1 homodimerization

While the signaling cascades induced by CEACAM1 homodimerization have been studied [150] and the residues involved in homodimerization have been identified [136], the dimer structure itself continues to be debated. While the 5dzl crystal structure of nCEACAM1 yields a potential homodimer structure with the appropriate binding interface, our non-glycosylated nCEACAM1 DEER distances do not match what would be expected in this crystal structure (Table 4.1, Figure 4.5). This indicates that crystallization conditions or the crystal contacts themselves may have an impact on the orientations of CEACAM within the crystal lattice. Thus, we aim to characterize nCEACAM1 homodimerization in its glycosylated and non-glycosylated forms in solution using a variety of techniques, namely size-exclusion chromatography, DEER, and, in the future, small angle X-ray scattering (SAXS). X-ray crystallography will also be utilized in future experiments to further examine the structure of the nCEACAM1 GlcNAc protein (and possibly its homodimer) to determine the potential impact of N-glycosylation on the structure of CEACAM1.

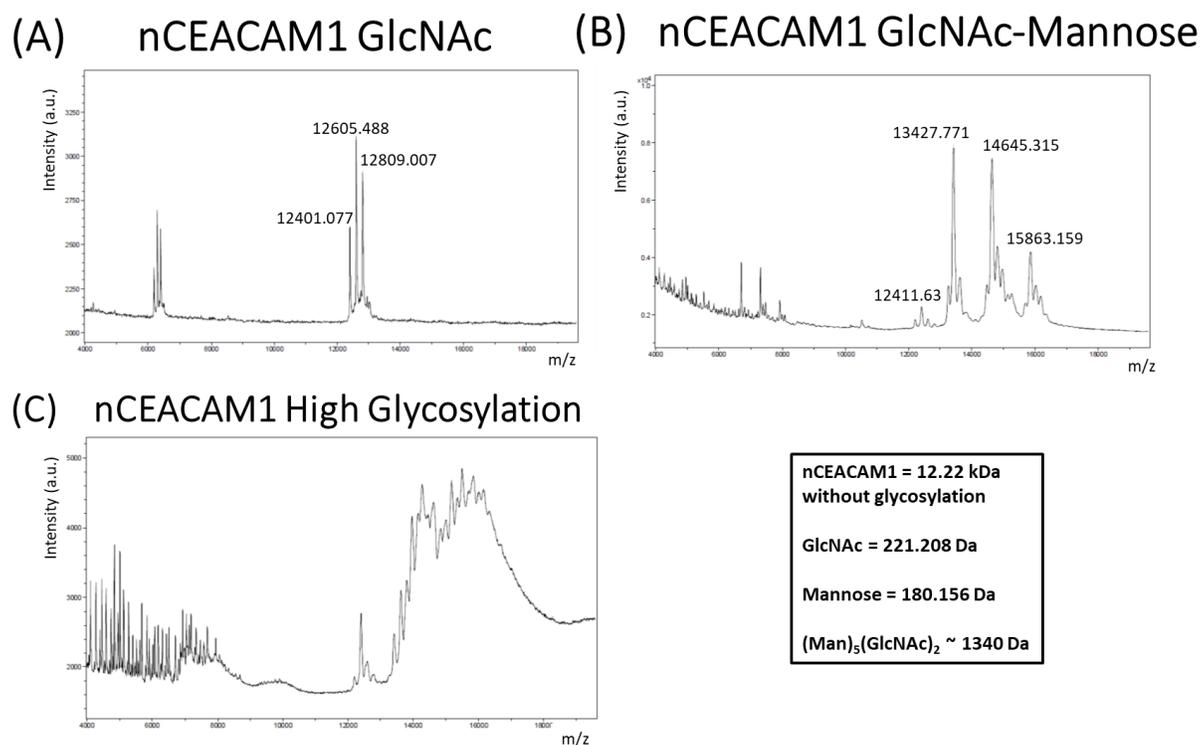
We aim to characterize the homodimerization of three protein constructs: non-glycosylated nCEACAM, nCEACAM1 GlcNAc, and highly glycosylated nCEACAM1. Studying both glycosylated and non-glycosylated nCEACAM1 variants will enable an evaluation of the effect of N-glycosylation on CEACAM homodimerization, which may in turn impact *in vivo* CEACAM adhesion functions. Such CEACAM homo- and heterodimerization interactions potentially play a key role in various diseases, including *Neisseria* pathogenesis and tumor progression [41, 69, 70]. Thus, the characterization of CEACAM oligomeric states in various glycosylation constructs prior to drawing conclusions on *in vivo* effects of CEACAM glycosylation is critical. Likewise, determining an accurate model of the CEACAM homodimer in solution is necessary to appropriately interpret CEACAM adhesion experiments, as the overall shape of the dimer and an identification of the binding surface may impact such data interpretations, and in turn the biological effect of CEACAM dimerization.

MALDI-TOF (matrix assisted laser desorption/ionization-time of flight mass spectrometry) was utilized to examine the glycosylation extent of nCEACAM1 along the three potential N-glycosylation sites (Figure 4.11). MALDI results confirmed that the techniques used to generate glycosylated CEACAM (adapted slightly from the protocol in [157]) were successful. nCEACAM1 GlcNAc proteins contain three possible N-glycosylation sites, and three MALDI peaks are observed with a mass separation equal to the molecular weight of a single GlcNAc (Figure 4.11 A). This indicates that the recombinant EndoF1 proteins used (which were generated in-house) successfully cleaved (Man)<sub>5</sub>-(GlcNAc) from the core GlcNAc glycans. The non-EndoF1-cleaved nCEACAM1 construct yields three primary MALDI peaks that are indicative of the three N-

glycosylation sites that differ in mass by a  $(\text{Man})_5(\text{GlcNAc})_2$  unit (Figure 4.11 B). The highly glycosylated nCEACAM1 construct shows heterogeneous MALDI peaks, as would be expected of complex glycans (Figure 4.11 C). As such, the N-glycosylated nCEACAM1 constructs studied previously [157] were successfully generated, which can be used in further nCEACAM1 homodimerization and Opa-CEACAM studies.

Spin labeled site	CEACAM dimer in 2gk2 crystal structure	CEACAM dimer in 5dzl crystal structure	CEACAM dimer in DEER
E5R1*	1.6 nm	3.8 nm	5.0 nm
M7R1	0.6 nm	4.8 nm	5.2 nm
F9R1	1.2 nm	4.7 nm	4.0 nm
A12R1	2.4 nm	5.8 nm	> 6 nm
E16R1*	1.7 nm	5.3 nm	> 6 nm
L20R1*	1.3 nm	3.9 nm	5.0 nm
T56R1	2.8 nm	2.7 nm	3.2 nm
T83R1	3.0 nm	3.7 nm	3.5 nm
D94R1*	4.8 nm	1.5 nm	2.9 nm
Q103R1	2.4 nm	3.1 nm	3.3 nm

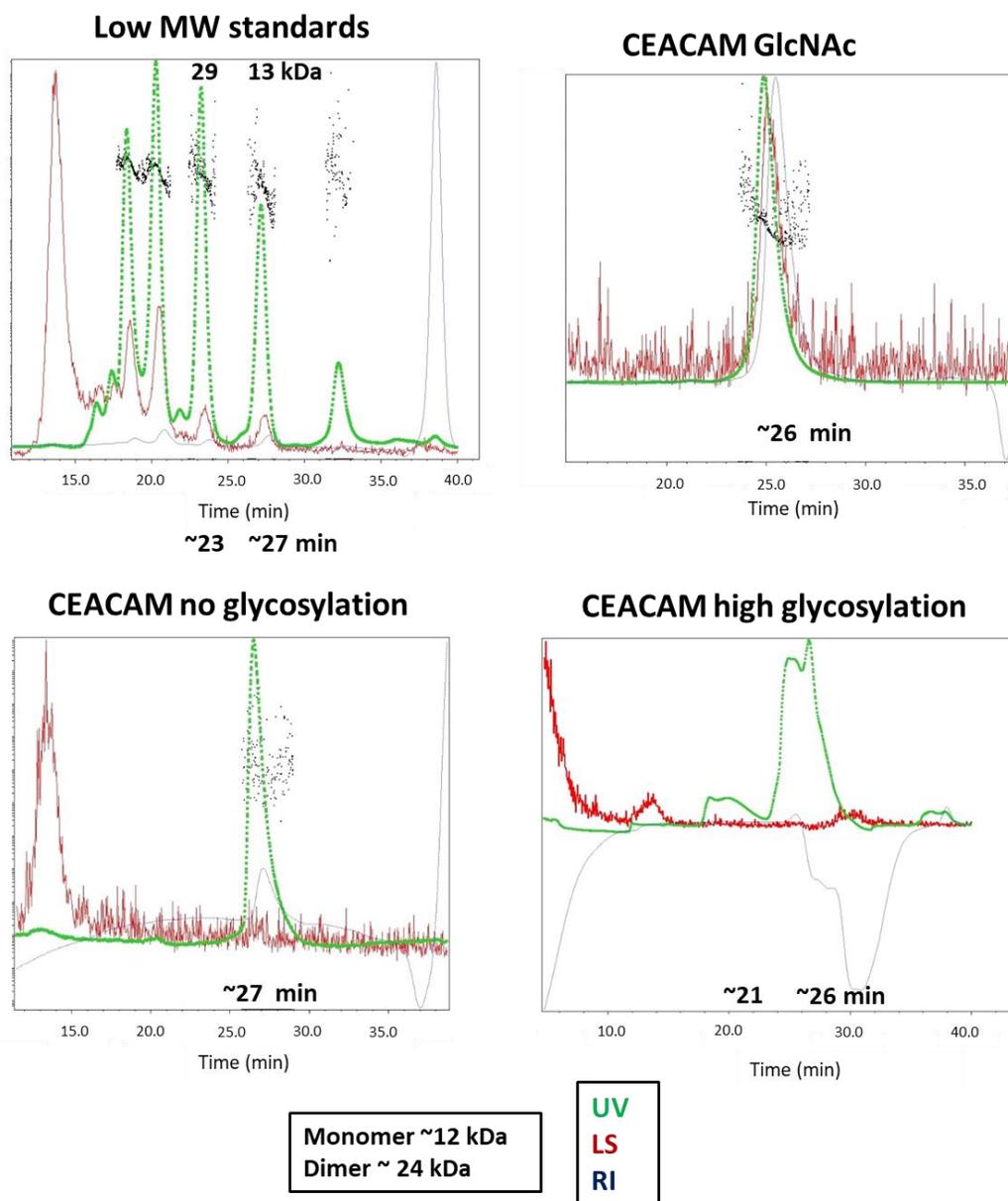
**Table 4.1: Average distance between non-glycosylated nCEACAM1 residues in DEER vs. homodimer crystal structures.** Two potential crystal structures, 2gk2 and 5dzl, have been proposed to model the nCEACAM1 homodimer. However, non-glycosylated nCEACAM1 DEER distances between spin labeled sites do not match those of either crystal structure (distances were calculated in Pymol between C $\beta$ ). Distances between residues in 5dzl most closely match those measured with DEER. However, several inter-residue distances differ by greater than 1 nm (\* residues), which indicates that the overall shape or orientation of chains within the homodimer crystal structure may be impacted by crystallization conditions.



**Figure 4.11: MALDI-TOF of nCEACAM1 GlcNAc, nCEACAM1 high mannose, and nCEACAM1 with complex glycosylation.** Three nCEACAM1 glycosylation products utilized in this research were validated using MALDI-TOF. (A) nCEACAM1 GlcNAc proteins contain the base GlcNAc unit (221 Da) at the three potential N-glycosylation sites. (B) nCEACAM1 GlcNAc proteins are generated using HEK293 GnT1<sup>-</sup> cells that lack the enzyme required to form complex glycans. A (Man)<sub>5</sub>(GlcNAc)<sub>2</sub> glycan unit (1340 Da) is bound to three potential N-glycosylation sites to form nCEACAM1 high mannose. (Man)<sub>5</sub>(GlcNAc) glycans are cleaved from each of the three sites with EndoF1 to yield nCEACAM1 GlcNAc in (A). (C) nCEACAM1 proteins that are glycosylated with complex glycans are generated using HEK293F cells which contain the GnT1 enzyme. The mass of this CEACAM variant is highly heterogeneous.

#### 4.4.1 SEC-MALS to identify oligomeric state of nCEACAM1 proteins

Size-exclusion chromatography with multi-angle light scattering detection (SEC-MALS) was used to identify the oligomeric state of nCEACAM1 (both glycosylated and non-glycosylated) at various concentrations. A size-exclusion chromatography column (Superdex 75 Increase 10/300, GE) that is optimal for the separation of monomeric and dimeric proteins the size of nCEACAM1 was used to separate oligomers for each CEACAM variant. UV detection at 280 nm identifies protein in elution timepoints and multi-angle light scattering is used to calculate the molecular weight of the proteins in such fractions. However, only one eluent peak is observed for each of the CEACAM constructs at all concentrations tested (Figure 4.12). The elution profile of low molecular weight standards shows a clear separation of proteins near the molecular weights of nCEACAM1 monomer and dimer (carbonic anhydrase, 29 kDa; ribonuclease A, 13.7 kDa) (Figure 4.12). Because only one peak is observed, non-glycosylated nCEACAM1, nCEACAM1 GlcNAc, and highly glycosylated nCEACAM1 are primarily dimer. This dimeric state is supported by the observation that nCEACAM1 GlcNAc proteins elute in nearly identical fractions to cleaved GFP (MW 27 kDa) during previous size-exclusion purification steps, as a nCEACAM1 dimer would have an approximately MW of 24 kDa. Likewise, DEER results indicate that there is a significant amount of nCEACAM1 dimer in each construct studied.



**Figure 4.12: SEC-MALS elution profiles of various nCEACAM1 constructs.** Low molecular weight standards of MWs 13.7 kDa and 29 kDa are well separated, which indicates that the chosen size-exclusion column can resolve nCEACAM1 monomer and dimer fractions at similar concentrations to the standards. However, only one peak is observed with UV detection at 280 nm for the three nCEACAM1 constructs (no

glycosylation, GlcNAc, and high glycosylation), which suggests that these proteins are in one oligomeric state.

#### 4.4.2 DEER distances of nCEACAM1 homodimer

DEER was initially utilized as a method to screen CEACAM constructs for CEACAM homodimerization and Opa binding. Non-glycosylated nCEACAM1 with a cleaved affinity tag, BAP-nCEACAM1, MBP-nCEACAM1, nCEACAM1 GlcNAc, and highly glycosylated nCEACAM1 were spin labeled and analyzed using DEER (Figures 4.5, 4.6, 4.8, and 4.13). While the original aim of this research, to generate monomeric nCEACAM1 to study interactions with Opa proteins, was somewhat unsuccessful (binding experiments will be examined in Chapter 5), the DEER results of these CEACAM constructs alone are useful in studying nCEACAM1 homodimerization, which is largely uncharacterized.

The DEER distributions of nCEACAM1 GlcNAc, highly glycosylated nCEACAM1, and non-glycosylated nCEACAM1 are nearly identical (Figure 4.13). The similar DEER distributions indicate that the overall structure of the CEACAM1 homodimer is unaffected by the presence of N-glycosylation in either form. These DEER experiments are contrary to the results of nCEACAM1 homodimerization found in the Moremen and Prestegard labs. Their NMR RDC experiments indicated that at CEACAM concentrations of approximately 150  $\mu\text{M}$  non-glycosylated nCEACAM1 is almost entirely dimer, nCEACAM1 GlcNAc is primarily monomer, and highly glycosylated nCEACAM1 is both monomeric and dimeric [157]. Conversely, all of the SEC and DEER results presented in this dissertation indicate that the three CEACAM constructs are homodimers (at concentrations of approximately 100-200  $\mu\text{M}$  in identical buffer conditions to [157]).

The contradictory results observed pose an interesting question of how the results of certain in-solution techniques may be biased by potentially inaccurate structural models. An estimation of CEACAM monomer and dimer populations was determined in [157] by estimating the expected rotational correlation times for a monomer and dimer and comparing these predicted numbers to rotational correlation times ( $\tau_c$ ) that are determined experimentally using RDC NMR. nCEACAM1 samples with GlcNAc residues resulted in an average correlation time of 8.5 ns, (Man)<sub>5</sub>-(GlcNAc)<sub>2</sub> a  $\tau_c$  of 9.3 ns, complex glycans an average  $\tau_c$  of 14.2 ns, and non-glycosylated a  $\tau_c$  of 13.3 ns [157]. The estimated rotational correlation times (in ns) are approximately one half the effective molecular mass (in kDa) when measured at room temperature in an aqueous solution. A nCEACAM1 monomer would be expected to have a correlation time of approximately 6.5 ns and a nCEACAM1 dimer 13 ns. Thus, RDC NMR results predict nCEACAM1 GlcNAc to be primarily monomer, (Man)<sub>5</sub>-(GlcNAc)<sub>2</sub> nCEACAM1 a mixture of monomer and dimer, and highly glycosylated and non-glycosylated nCEACAM1 primarily dimer [157]. Such a relationship between molecular mass and rotational correlation time is derived from Stokes formula for rotational Brownian diffusion ( $D_r$ ):

$$D_r = \frac{kT}{8\pi\eta r^3} = \frac{1}{\tau_c} \quad (4.1)$$

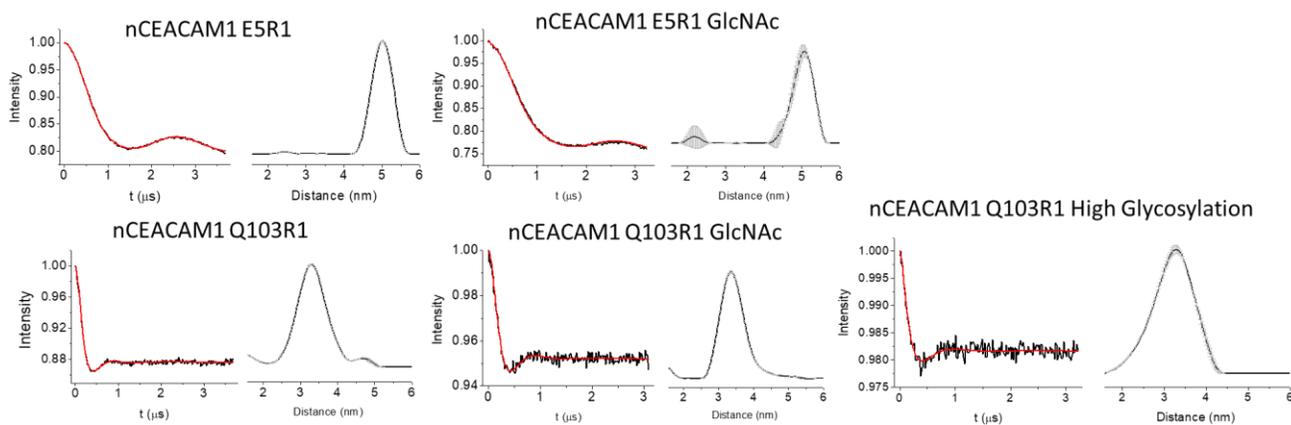
where T is the temperature, r is the radius of the spherical solute, k is the Boltzmann constant, and  $\eta$  is the shear viscosity of the solvent. This estimation is only applicable for spherical particles, thus, an accurate interpretation of the overall shape of a CEACAM homodimer is critical. When these authors compared rotational correlation times of different nCEACAM1 constructs, they assumed the CEACAM1 homodimer was roughly

spherical. However, DEER distance distributions of non-glycosylated nCEACAM1 proteins yielded average distances that are varied from what would be predicted in the 5dzl structure (Table 4.1). The measurement of several distances that are  $\geq 1$  nm longer than in the 5dzl structure (namely 5R1, 12R1, 16R1, 20R1, 56R1, and 94R1) suggests that the homodimer chains may be twisted and/or elongated compared to the crystal structure. Such an alteration in the overall shape of the homodimer would impact the expected calculations of CEACAM1 correlation times in a monomer and dimer state, and consequently the interpretation of experimental calculations.

Conversely, the limited number of experimental measurements may have impacted the average calculations of  $\tau_c$ . Only six  $^{15}\text{N}$ -labeled alanine residues were used to calculate  $\tau_c$ , from which an average value was used to compare overall protein rotational correlation times. Such a limited sample size may influence the calculation of average correlation times: the alanine  $\tau_c$  calculations varied by a range of approximately 5 ns for each nCEACAM1 construct. As this error is similar to the predicted  $\tau_c$  of a CEACAM monomer, it is difficult to accurately draw conclusions regarding differences in oligomeric states among various CEACAM constructs. Likewise, the added weight of the glycans complicates the ability to predict rotational correlation times based on molecular weight. Approximately 8000 Da may be added to the protein molecular mass with wild type glycans (at roughly 12 sugar residues with 100% N-glycosylation site occupancy) [157].

Because such discrepancies may be due to inaccuracies in the overall shape of the 5dzl homodimer structure, we aim to refine this basic structure using non-glycosylated nCEACAM1 homodimer DEER distances. XPLOR-NIH is a program that is typically used to generate structural models from NMR experimental restraints. We are currently using

this program to refine the orientation of CEACAM chains within the 5dzl homodimer structure using experimental DEER restraints. This will elucidate the overall shape of the homodimer, which may be elongated or twisted compared to the crystal structure. Gaining insight into the in-solution shape of the nCEACAM1 homodimer will provide a model that more closely mimics the *in vivo* system, from which a more accurate assessment of the binding interface can also be drawn. Together, such information can have a profound impact on experimental interpretations, as was shown with NMR RDC experiments of nCEACAM1. Likewise, generating an accurate model of nCEACAM1 homodimerization may be critical to Opa-CEACAM binding experiments.



**Figure 4.13: DEER distributions of nCEACAM1 in its glycosylated and non-glycosylated forms.** The DEER distributions of nCEACAM1 5R1 and 103R1 are not changed with the addition of GlcNAc or complex glycans at three N-glycosylation sites. This indicates that the presence of glycosylation on nCEACAM1 does not alter the structure of the nCEACAM1 homodimer.

#### 4.5 Future research directions

Several additional experiments must be performed to characterize nCEACAM1 homodimerization in its glycosylated and non-glycosylated forms, which may ultimately impact Opa binding. Currently, preliminary experiments of nCEACAM1 using SAXS and X-ray crystallography are in progress and are highlighted in the below sections. Additional potential future experiments of CEACAM proteins are discussed further in Chapter 5.

##### 4.5.1 SAXS to determine the shape of the nCEACAM1 homodimer

In future experiments, small angle X-ray scattering (SAXS) will be utilized to determine the overall shape of the nCEACAM1 homodimer. nCEACAM1 non-glycosylated, nCEACAM1 GlcNAc, and nCEACAM1 highly glycosylated samples have been sent to the Argonne National Laboratory for analysis. SAXS is a technique that quantifies nanoscale density differences in a sample via a scattering profile, from which the size and shape of a protein can be determined. Thus, SAXS will be used to confirm the observation derived from the DEER data that the nCEACAM1 homodimer may be extended, rather than compact as is seen in the 5dzl dimer crystal structure. SAXS will provide another in solution method to examine the structure of the homodimer, without the necessity of crystal formation or added crystal components, both of which may impact a dimer structure. For example, the nCEACAM1 samples sent to Argonne were in HEPES and NaCl buffer conditions, compared to the addition of the large PEG-2000 and the partially hydrophobic octyl glucoside in the two crystal structure solutions.

Together, a combination of SAXS and DEER data of the three CEACAM constructs will enable conclusions to be drawn about the overall shape and structure of the

CEACAM homodimer. A greater understanding of the nature of the CEACAM1 homodimer is necessary to accurately interpret experimental results regarding CEACAM adhesion functions, which play a critical role in tumor progression and *Neisseria* pathogenesis [41, 69, 70]. Likewise, a comparison of DEER and SAXS data between glycosylated and non-glycosylated CEACAM constructs will provide insight into the role of N-glycosylation on the formation of CEACAM homodimers (or the lack thereof).

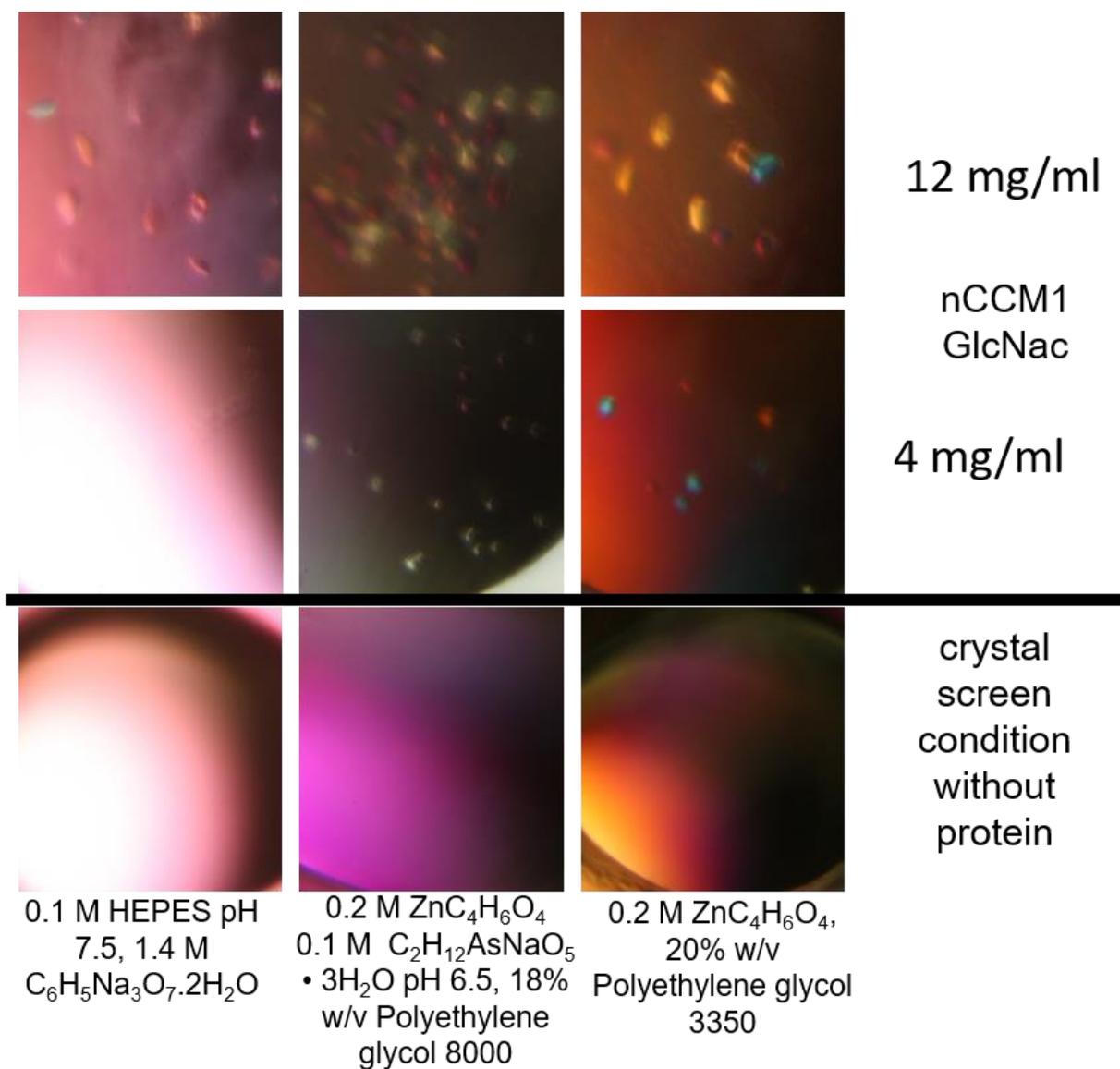
#### 4.5.2 X-ray crystallography of nCEACAM GlcNAc

The manner by which N-linked glycosylation may impact CEACAM dimerization remains unclear. The Moremen and Prestegard labs hypothesize that local steric effects of glycosylation may propagate throughout the CEACAM structure, perhaps altering the formation of salt bridges between  $\beta$ -strands and thus the fold itself [157]. Slight variations in CEACAM fold may be significant enough to impact its adhesion function. Thus, a complicated interplay between the degree/pattern of glycosylation and protein fold/function may be present *in vivo* to regulate CEACAM homotypic and heterotypic adhesion (which may also involve additional glycosylated domains of CEACAM). However, the nCEACAM1 oligomeric states observed in DEER and SEC experiments indicate that nCEACAM1 is primarily dimer in all of its glycosylated forms. As such, N-linked glycosylation on the Ig-like domains of CEACAMs may enhance adhesion directly or slightly alter the structure of nCEACAM1 itself to that end.

To gain further insight into how N-linked glycosylation may alter the fold of nCEACAM1 and in turn affect its function, we are working towards the crystallization of nCEACAM1 GlcNAc. Crystallization conditions have been screened, and several conditions generated crystal hits (Figure 4.14). These X-ray crystallography experiments

are currently ongoing. A nCEACAM1 GlcNAc crystal structure will provide insight into the mechanism by which N-linked glycosylation affects protein structure (if at all) via a comparison to previously solved structures of non-glycosylated nCEACAM1.

X-ray crystallography of glycosylated proteins is rarely successful due to the heterogeneous and flexible nature of the glycans (although a few glycan-containing crystal structures exist [172-174]). However, because roughly 50% of eukaryotic proteins are glycosylated, the importance of such post-translational modifications cannot be understated; they increase protein solubility, protect from proteolysis, influence protein localization and cell-cell interactions [175]. Thus, determining the crystal structure of nCEACAM1 GlcNAc will also be of a broader importance, where the incorporation of small and homogeneous GlcNAc units provides an alternative method for the structural determination of N-glycosylated proteins, from which the effects of such glycosylation on structure can be explored.



**Figure 4.14: Initial nCEACAM1 GlcNAc crystal screens.** Several nCEACAM1 GlcNAc crystallization conditions yielded numerous crystals. Determining the structure of nCEACAM1 containing three GlcNAc units will provide insight into the effect (if any) of glycosylation on the structure of nCEACAM1.

## 4.6 Conclusions

The CEACAM family of proteins have long been classified as adhesion molecules that are associated with numerous cellular processes including angiogenesis [56], cell proliferation [57], cell motility [58], invasion [59], infection and inflammation [60], and tumor growth [176]. CEACAMs contain a variable number of Ig-like ectodomains that participate in cis and trans homo- and heterotypic interactions, which trigger transmembrane signaling events. Numerous studies have investigated CEACAM oligomerization, however, the results are often contradictory and difficult to interpret in or apply to an *in vivo* setting.

Fully characterizing the mechanism and effect of CEACAM dimerization is complicated by different CEACAMs often yielding dissimilar experimental results. For example, the crystal structure of the nCEACAM6 homodimer (PDB ID: 4Y8A) is distinct from that of the nCEACAM5 homodimer (PDB ID: 2QSQ) and the nCEACAM1 homodimer structures [151]. While the chain of one CEACAM unit overlays completely in each case, the second chain of nCEACAM1, nCEACAM5, and nCEACAM6 are twisted compared to the aligned chain (Figure 4.15) [151]. This indicates that crystallization conditions may impact the overall shape of the homodimers. This gives credence to our observation that in-solution DEER measurements of nCEACAM1 provide a different and more accurate model of the homodimer than the nCEACAM1 crystal structure. Applying in-solution experimental DEER measurements to the core 5dzl homodimer crystal structure provides a way to bridge the two techniques to fine-tune the model of a nCEACAM1 homodimer. SAXS is also currently being utilized to further define the overall shape of the nCEACAM1 homodimer, from which a more accurate model of the dimer structure can be

generated. However, such techniques will need to be applied to other CEACAM variants prior to drawing any comprehensive biological conclusions on the mode of CEACAM homo and heterodimerization.

The effect of glycosylation on the ectodomains of various CEACAMs will also need to be examined to determine its impact on dimerization. In this dissertation, the DEER distributions of non-glycosylated nCEACAM1, nCEACAM1 GlcNAc, and highly glycosylated nCEACAM1 proteins were nearly identical, indicating that glycosylation does not affect the overall orientation of the CEACAM1 homodimer. Likewise, size exclusion experiments indicate that the glycosylation on nCEACAM1 does not affect the oligomeric state of the proteins. This is contrary to what was observed in the Moremen and Prestegard labs, as they suggest that a slight amount of glycosylation on nCEACAM1 decreases the dimerization propensity significantly. Such a discrepancy in results highlights the necessity of an accurate structural model of the homodimer to interpret experimental results. SAXS and X-ray crystallography are currently being utilized to further examine the homodimer structure of nCEACAM1 GlcNAc proteins to determine the effect of minimal glycosylation on nCEACAM1 structure. Due to the structural heterogeneity of the highly glycosylated nCEACAM1 construct, it will be difficult to examine such effects with this variant. However, such large and complex glycosylation, as well as the presence of additional domains, may have a significant impact on CEACAM function *in vivo* that cannot be easily probed experimentally.

Ultimately, each CEACAM variant must be examined individually to fully characterize hetero- and homo-dimerization across the CEACAM protein family. Subtle variations in dimerization propensities may play a significant role in the regulation of CEACAM

signaling events and cell adhesion. Because CEACAM proteins are widely and precisely expressed among many cell types in mammals, these dimerization functions may be highly regulated and have a significant impact on disease states involving aberrant CEACAM function, including tumor progression and metastasis. The homodimerization of various CEACAMs and their glycosylation likely also impacts the binding of pathogenic proteins, such as Opa, to host CEACAMs.



**Figure 4.15: Overlay of nCEACAM1, nCEACAM5, and nCEACAM6 homodimer structures.** The structures of nCEACAM1 (gray, PDB ID: 5dzl), nCEACAM5 (tan, PDB ID: 2qsq), nCEACAM6 (red, PDB ID: 4y8a) are aligned by a single chain (right). The second chain of each of the three structures are somewhat well aligned (left), although the chains may be twisted with respect to each other. This indicates that the crystallization conditions of the respective structures may impact the arrangement of the CEACAM proteins within the dimer.

1. Fedarovich, A., et al., *Structure of the N-terminal domain of human CEACAM1: binding target of the opacity proteins during invasion of Neisseria meningitidis and Neisseria gonorrhoeae*. Acta Crystallographica Section D-Biological Crystallography, 2006. **62**: p. 971-979.
2. Oikawa, S., et al., *A specific heterotypic cell adhesion activity between members of carcinoembryonic antigen family, W272 and NCA, is mediated by N-domains*. Journal of Biological Chemistry, 1991. **266**(13): p. 7995-8001.
3. Watt, S.M., et al., *Homophilic adhesion of human CEACAM1 involves N-terminal domain interactions: structural analysis of the binding site*. Blood, 2001. **98**(5): p. 1469-1479.
4. Virji, M., et al., *Critical determinants of host receptor targeting by Neisseria meningitidis and Neisseria gonorrhoeae: identification of Opa adhesiotopes on the N-domain of CD66 molecules*. Molecular Microbiology, 1999. **34**(3): p. 538-551.
5. Heldin, C.-H., *Dimerization of cell surface receptors in signal transduction*. Cell, 1995. **80**(2): p. 213-223.
6. Tchoupa, A.K., T. Schuhmacher, and C.R. Hauck, *Signaling by epithelial members of the CEACAM family – mucosal docking sites for pathogenic bacteria*. Cell Communication and Signaling, 2014. **12**(1): p. 27.
7. Huber, M., et al., *The Carboxyl-terminal Region of Biliary Glycoprotein Controls Its Tyrosine Phosphorylation and Association with Protein-tyrosine Phosphatases SHP-1 and SHP-2 in Epithelial Cells*. Journal of Biological Chemistry, 1999. **274**(1): p. 335-344.
8. Müller, M.M., et al., *Homophilic adhesion and CEACAM1-S regulate dimerization of CEACAM1-L and recruitment of SHP-2 and c-Src*. The Journal of Cell Biology, 2009. **187**(4): p. 569-581.

9. Zhou, H., et al., *Homophilic adhesion between Ig superfamily carcinoembryonic antigen molecules involves double reciprocal bonds*. The Journal of Cell Biology, 1993. **122**(4): p. 951-960.
10. Klaile, E., et al., *The CEACAM1 N-terminal Ig domain mediates cis- and trans-binding and is essential for allosteric rearrangements of CEACAM1 microclusters*. The Journal of Cell Biology, 2009. **187**(4): p. 553-567.
11. Patel, P.C., et al., *Inside-out Signaling Promotes Dynamic Changes in the Carcinoembryonic Antigen-related Cellular Adhesion Molecule 1 (CEACAM1) Oligomeric State to Control Its Cell Adhesion Properties*. Journal of Biological Chemistry, 2013. **288**(41): p. 29654-29669.
12. Lawson, E.L., et al., *The Transmembrane Domain of CEACAM1-4S Is a Determinant of Anchorage Independent Growth and Tumorigenicity*. PLOS ONE, 2012. **7**(1): p. e29606.
13. Russ, W.P. and D.M. Engelman, *The GxxxG motif: A framework for transmembrane helix-helix association*<sup>11</sup>Edited by G. von Heijne. Journal of Molecular Biology, 2000. **296**(3): p. 911-919.
14. Gray-Owen, S.D. and R.S. Blumberg, *CEACAM1: contact-dependent control of immunity*. Nature Reviews Immunology, 2006. **6**: p. 433.
15. Singer, B.B., et al., *Deregulation of the CEACAM Expression Pattern Causes Undifferentiated Cell Growth in Human Lung Adenocarcinoma Cells*. PLOS ONE, 2010. **5**(1): p. e8747.
16. Singer, B.B., I. Scheffrahn, and B. Öbrink, *The Tumor Growth-inhibiting Cell Adhesion Molecule CEACAM1 (C-CAM) Is Differently Expressed in Proliferating and Quiescent Epithelial Cells and Regulates Cell Proliferation*. Cancer Research, 2000. **60**(5): p. 1236-1244.

17. Turbide, C., et al., *Optimal ratios of biliary glycoprotein isoforms required for inhibition of colonic tumor cell growth*. *Cancer Res*, 1997. **57**(13): p. 2781-8.
18. Beauchemin, N. and A. Arabzadeh, *Carcinoembryonic antigen-related cell adhesion molecules (CEACAMs) in cancer progression and metastasis*. *Cancer and Metastasis Reviews*, 2013. **32**(3): p. 643-671.
19. Bonsor, D.A., et al., *Diverse oligomeric states of CEACAM IgV domains*. *Proceedings of the National Academy of Sciences*, 2015. **112**(44): p. 13561-13566.
20. Abdul-Wahid, A., et al., *The carcinoembryonic antigen IgV-like N domain plays a critical role in the implantation of metastatic tumor cells*. *Molecular Oncology*, 2014. **8**(2): p. 337-350.
21. Huang, Y.-H., et al., *CEACAM1 regulates TIM-3-mediated tolerance and exhaustion*. *Nature*, 2014. **517**: p. 386.
22. Dym, O., et al., *The impact of crystallization conditions on structure-based drug design: A case study on the methylene blue/acetylcholinesterase complex*. *Protein Sci*, 2016. **25**(6): p. 1096-114.
23. Zhang, X.-j., J.A. Wozniak, and B.W. Matthews, *Protein Flexibility and Adaptability Seen in 25 Crystal Forms of T4 Lysozyme*. *Journal of Molecular Biology*, 1995. **250**(4): p. 527-552.
24. Martin, J.N., et al., *Neisserial Opa Protein-CEACAM Interactions: Competition for Receptors as a Means of Bacterial Invasion and Pathogenesis*. *Biochemistry*, 2016. **55**(31): p. 4286-4294.
25. Gibson, D.G., et al., *Enzymatic assembly of DNA molecules up to several hundred kilobases*. *Nature Methods*, 2009. **6**: p. 343.

26. Zhuo, Y., et al., *Glycosylation Alters Dimerization Properties of a Cell-surface Signaling Protein, Carcinoembryonic Antigen-related Cell Adhesion Molecule 1 (CEACAM1)*. Journal of Biological Chemistry, 2016. **291**(38): p. 20085-20095.
27. Kapust, R.B. and D.S. Waugh, *Escherichia coli maltose-binding protein is uncommonly effective at promoting the solubility of polypeptides to which it is fused*. Protein Science : A Publication of the Protein Society, 1999. **8**(8): p. 1668-1674.
28. Bach, H., et al., *Escherichia coli maltose-binding protein as a molecular chaperone for recombinant intracellular cytoplasmic single-chain antibodies*. J Mol Biol, 2001. **312**(1): p. 79-93.
29. Schwarz, F. and M. Aebi, *Mechanisms and principles of N-linked protein glycosylation*. Current Opinion in Structural Biology, 2011. **21**(5): p. 576-582.
30. Marshall, R.D., *Glycoproteins*. Annu Rev Biochem, 1972. **41**: p. 673-702.
31. Shental-Bechor, D. and Y. Levy, *Effect of glycosylation on protein folding: a close look at thermodynamic stabilization*. Proc Natl Acad Sci U S A, 2008. **105**(24): p. 8256-61.
32. Hanson, S.R., et al., *The core trisaccharide of an N-linked glycoprotein intrinsically accelerates folding and enhances stability*. Proceedings of the National Academy of Sciences of the United States of America, 2009. **106**(9): p. 3131-3136.
33. Culyba, E.K., et al., *Protein native-state stabilization by placing aromatic side chains in N-glycosylated reverse turns*. Science, 2011. **331**(6017): p. 571-5.
34. Skropeta, D., *The effect of individual N-glycans on enzyme activity*. Bioorg Med Chem, 2009. **17**(7): p. 2645-53.
35. Helenius, A. and M. Aebi, *Roles of N-linked glycans in the endoplasmic reticulum*. Annu Rev Biochem, 2004. **73**: p. 1019-49.

36. Aebi, M., et al., *N-glycan structures: recognition and processing in the ER*. Trends Biochem Sci, 2010. **35**(2): p. 74-82.
37. Shi, X.Z. and D.L. Jarvis, *Protein N-glycosylation in the baculovirus-insect cell system*. Current Drug Targets, 2007. **8**(10): p. 1116-1125.
38. Reeves, P.J., et al., *Structure and function in rhodopsin: high-level expression of rhodopsin with restricted and homogeneous N-glycosylation by a tetracycline-inducible N-acetylglucosaminyltransferase I-negative HEK293S stable mammalian cell line*. Proc Natl Acad Sci U S A, 2002. **99**(21): p. 13419-24.
39. von Schaewen, A., et al., *Isolation of a mutant Arabidopsis plant that lacks N-acetylglucosaminyl transferase I and is unable to synthesize Golgi-modified complex N-linked glycans*. Plant Physiol, 1993. **102**(4): p. 1109-18.
40. Gomez, L. and M.J. Chrispeels, *Complementation of an Arabidopsis thaliana mutant that lacks complex asparagine-linked glycans with the human cDNA encoding N-acetylglucosaminyltransferase I*. Proc Natl Acad Sci U S A, 1994. **91**(5): p. 1829-33.
41. Hauck, C.R. and T.F. Meyer, *'Small' talk: Opa proteins as mediators of Neisseria-host-cell communication*. Current Opinion in Microbiology, 2003. **6**(1): p. 43-49.
42. Neumaier, M., et al., *Biliary glycoprotein, a potential human cell adhesion molecule, is down-regulated in colorectal carcinomas*. Proc Natl Acad Sci U S A, 1993. **90**(22): p. 10744-8.
43. Rosenberg, M., et al., *The expression of mouse biliary glycoprotein, a carcinoembryonic antigen-related gene, is down-regulated in malignant mouse tissues*. Cancer Res, 1993. **53**(20): p. 4938-45.

44. Behrens, A.J., et al., *Integrity of Glycosylation Processing of a Glycan-Depleted Trimeric HIV-1 Immunogen Targeting Key B-Cell Lineages*. Journal of Proteome Research, 2018. **17**(3): p. 987-999.
45. Medina-Ramirez, M., et al., *Design and crystal structure of a native-like HIV-1 envelope trimer that engages multiple broadly neutralizing antibody precursors in vivo*. Journal of Experimental Medicine, 2017. **214**(9): p. 2573-2590.
46. Ward, A.B. and I.A. Wilson, *The HIV-1 envelope glycoprotein structure: nailing down a moving target*. Immunological Reviews, 2017. **275**(1): p. 21-32.
47. Apweiler, R., H. Hermjakob, and N. Sharon, *On the frequency of protein glycosylation, as deduced from analysis of the SWISS-PROT database*. Biochim Biophys Acta, 1999. **1473**(1): p. 4-8.
48. Horst, A.K., et al., *Carcinoembryonic antigen-related cell adhesion molecule 1 modulates vascular remodeling in vitro and in vivo*. The Journal of Clinical Investigation, 2006. **116**(6): p. 1596-1605.
49. Scheffrahn, I., et al., *Control of density-dependent, cell state-specific signal transduction by the cell adhesion molecule CEACAM1, and its influence on cell cycle regulation*. Experimental Cell Research, 2005. **307**(2): p. 427-435.
50. Klaile, E., et al., *CEACAM1 functionally interacts with filamin A and exerts a dual role in the regulation of cell migration*. Journal of Cell Science, 2005. **118**(23): p. 5513-5524.
51. Ebrahimnejad, A., et al., *CEACAM1 Enhances Invasion and Migration of Melanocytic and Melanoma Cells*. The American Journal of Pathology, 2004. **165**(5): p. 1781-1787.
52. Leung, N., et al., *Intestinal tumor progression is promoted by decreased apoptosis and dysregulated Wnt signaling in Ceacam1<sup>-/-</sup> mice*. Oncogene, 2008. **27**: p. 4943.

## Chapter 5: Prospects for future research into Opa-CEACAM interactions

### 5.1 Introduction to Opa-CEACAM interactions

The overarching goal of the research presented in this dissertation was to better characterize the mechanism of Opa protein interactions with CEACAMs to gain insight into the mode of *Neisseria* phagocytosis into host human cells. To that end, Opa and CEACAM proteins needed to first be studied in isolation, the results of which are described in previous chapters. Much remains unknown about the mechanism of Opa-CEACAM binding, specifically the manner by which highly diverse and dynamic Opa proteins engage a subset of CEACAMs. While portions of the loops of Opa proteins are variable in sequence, the N-terminal domains of CEACAMs are largely conserved and the binding patterns observed *in vivo* are highly selective (Figure 5.1).

Previous Opa-CEACAM binding studies indicate that there is little receptor selectivity between Opa proteins for various nCEACAM receptors *in vitro*, despite trends observed via invasion assays with Opa in *Neisseria gonorrhoeae* and CEACAM in HeLa cells [50, 119]. This may be explained by several factors. Reconstituting Opa proteins into a liposome system isolates the proteins from additional membrane components of *Neisseria*, such as LOS, that may interact directly with Opa loops or impact Opa dynamics in such a way that the binding mechanism of Opa is altered. The lack of *in vitro* selectivity may also indicate that the specificity of host receptor-induced *Neisseria* internalization may instead be the result of differential downstream signaling events that may or may not trigger phagocytosis rather than an altered binding mechanism between Opa and CEACAM at the

cell surface. Additionally, the presence of heavy glycosylation on CEACAMs may have a significant impact on Opa binding *in vivo* that was not captured in previous *in vitro* binding assays that utilized non-glycosylated nCEACAM constructs [119].

As such, several questions remain regarding the mechanism of Opa-CEACAM binding that must be answered using a variety of experimental techniques. First, how are the dynamics of Opa loops altered upon receptor engagement? Second, what is the impact of LOS on Opa loop dynamics and CEACAM binding? Third, what is the effect of CEACAM glycosylation on Opa binding? Finally, which residues of Opa are directly involved in interactions with CEACAMs and how is the mechanism of binding altered between various Opa and CEACAM proteins? This chapter highlights several experiments aimed at answering these questions as well as preliminary results and future research directions.

```

      40          50          60          70          80          90
CEACAM1  QLTTE SMPFNVAEGKEV LLLVHNL PQQ LFGYSWYKGERV DGNRQIVGYAIGT-QQATPGP
CEACAM5  KLTIE ST.FNV...E...VH.LPQH LFG.S.Y...ERVDGNRQ.I..VIGT-.QAT..P
CEACAM3  KLTIE SM.LSV...E...VH.LPQH LFG.S.Y...ERVDGNSL.V..VIGT-.QAT..A
CEACAM6  KLTIE ST.FNV...E...AH.LPQN RIG.S.Y...ERVDGNSL.V..VIGT-.QAT..P
CEACAM8  QLTIE AV.SNA...E...VH.LPQD PRG.N.Y...ETVDANRR.I..VISN-.QIT..P
CEACAM7  QTNID VV.FNV...E...VH.ESQN LYG.N.Y...ERVHANYR.I..VKNIS.ENA..P
CEACAM4  QFTIE AL.SSA...D...AC.ISET IQA.Y.H...KTAEGSPL.A..ITDI-.ANI..A

      100       110       120       130       140
CEACAM1  ANSGRE TIYPNASLLIQNV TQN DTFYTLQVIKSDLVNEEATGQFHVYP
CEACAM5  .YSG..II.P.AS..IQ.IIQN.T.F...HV.KSDLVNEEA.G.FR.YP
CEACAM3  .YSG..TI.T.AS..IQ.VTQN.I.F...QV.KSDLVNEEA.G.FH.YQ
CEACAM6  .YSG..TI.P.AS..IQ.VTQN.T.F...QV.KSDLVNEEA.G.FH.YP
CEACAM8  .YSN..TI.P.AS..MR.VTRN.T.S...QV.KLNLMS EEV.G.FS.HP
CEACAM7  .HNG..TI.P.GT..IQ.VTHN.A.F...HV.KENLVNEEV.R.FY.FS
CEACAM4  .YSG..TV.P.GS..FQ.ITLE.A.S...RT.NASYDSDQA.G.LH.HQ

```

**Figure 5.1: Sequence alignment of human CEACAM N-terminal domains.** Residues required for nCEACAM1 binding to Opa are shown in red, and residues necessary to bind only certain Opa proteins are shown in orange. Residue numbering corresponds to the full nCEACAM1 domain (containing a linker region), UniProt ID P13688-1. Dots indicate conserved residues. CEACAMs 1, 3, 5, and 6 bind Opa proteins.

## 5.2 Effect of CEACAM binding on Opa loop dynamics

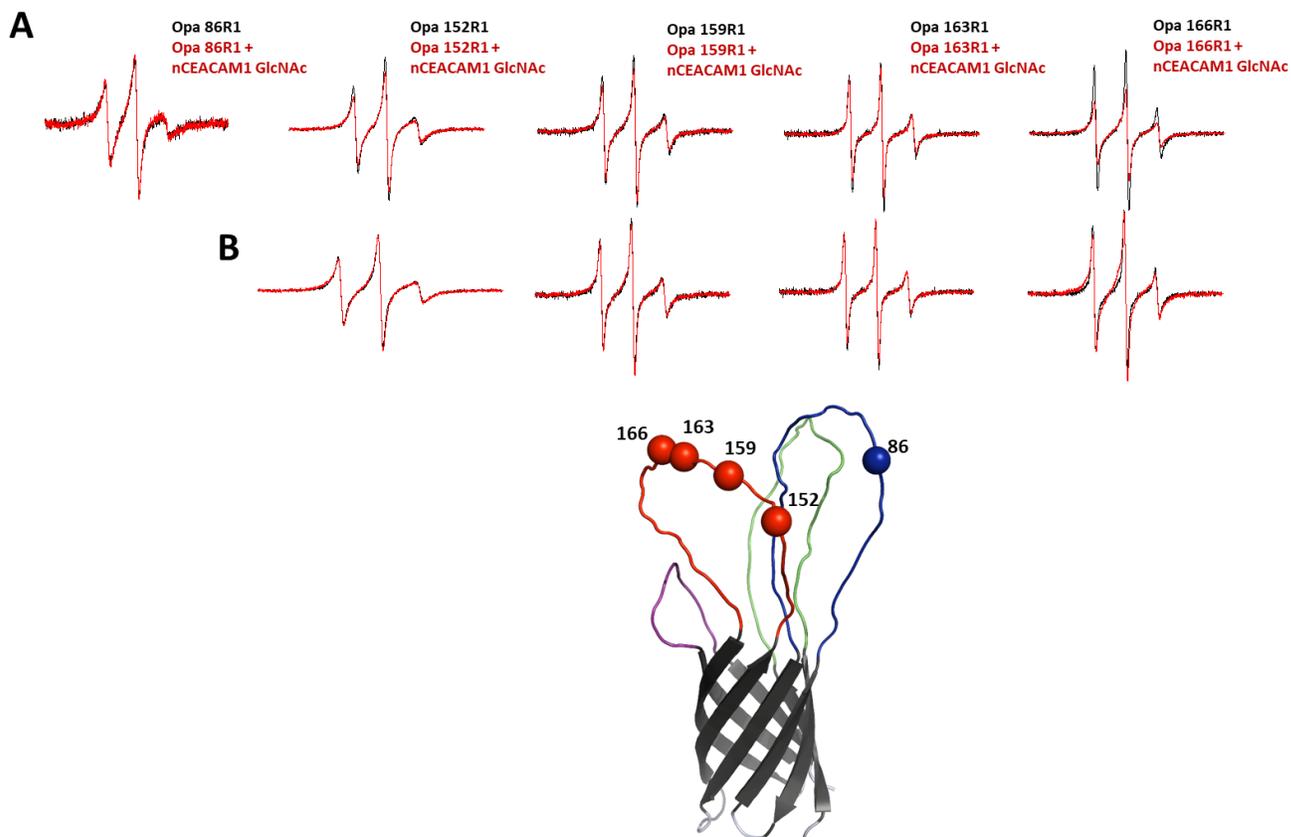
It is important to understand how Opa loop dynamics provide a biological advantage to *Neisseria*, and therefore how such dynamics are involved in receptor engagement. While Opa loop dynamics in isolation have been extensively investigated and discussed in this dissertation, complementary future CW-EPR and DEER experiments must be performed in the presence of nCEACAM to determine the effect of receptor binding on loop dynamics. Unfortunately, the dimerization of nCEACAM1 proteins has complicated such measurements, as the addition of Opa proteins does not appear to disrupt the nCEACAM1 homodimer. As such, this work is ongoing, where preliminary results and future directions are presented in this chapter.

### 5.2.1 Opa loop dynamics upon CEACAM binding with CW-EPR

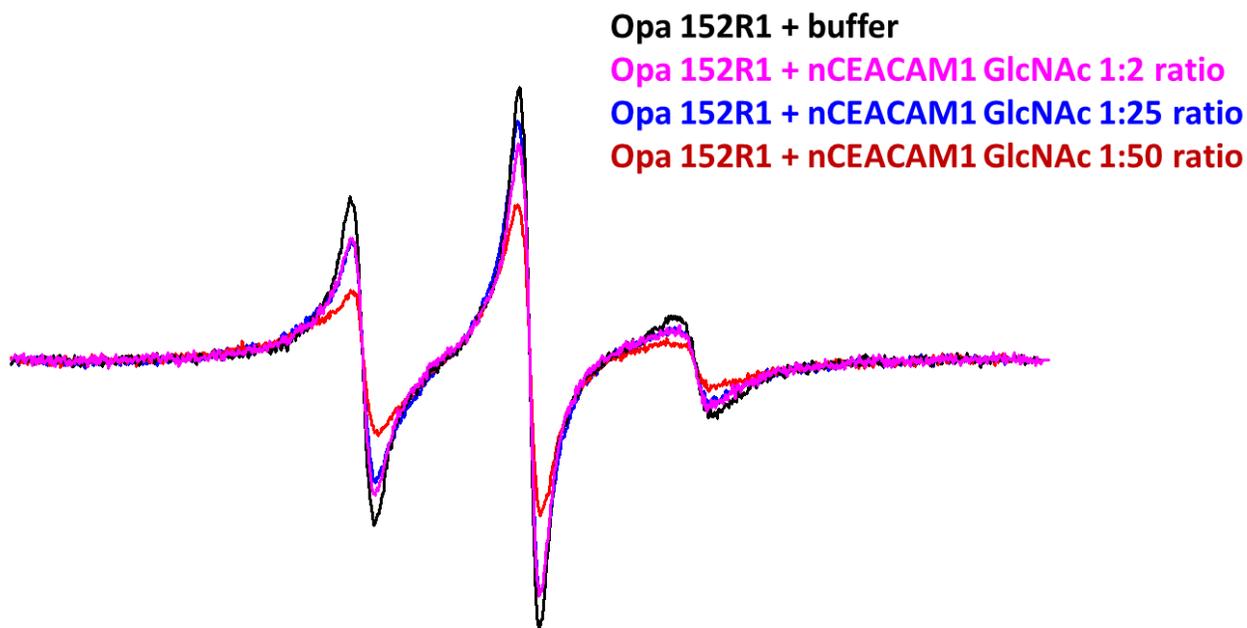
Single labeled Opa proteins in detergent are added to an excess of nCEACAM1 proteins in non-glycosylated or glycosylated forms and are measured using CW-EPR. The Opa EPR lineshape should broaden as the labeled loop region becomes locally ordered upon CEACAM binding. Line broadening may only be observed for labeled loop regions that are directly adjacent to the CEACAM binding site, as it is expected that the remaining loop regions remain largely mobile (a full loop ordering upon binding would be entropically unfavorable). Loops 1, 2, and 3 will all be probed for CEACAM binding, where many residues near and within the HV regions (predicted to be receptor-binding regions) will be spin labeled in particular. Examining Opa loop dynamics in the presence of CEACAMs will report on alterations in Opa loop mobility upon binding and highlight the relative location of CEACAM binding to Opa (via a comparison of lineshapes that

display broadening) that is currently unknown. Together, examining Opa loop dynamics upon engaging CEACAMs will provide insight into the mechanism of Opa-receptor binding and the role of loop motion in such interactions.

Preliminary CW-EPR results of single labeled Opa proteins in detergent display line broadening with the addition of an excess of nCEACAM1 GlcNAc proteins for Opa loop sites 152R1 and 166R1 (Figure 5.2). Sites that do not exhibit line broadening near those that do (159R1 and 163R1) may indicate that the incorporation of the spin label at that site disrupts CEACAM binding. The Opa 152R1 sample was also exposed to increasing concentrations of nCEACAM1 GlcNAc and the resulting EPR lineshapes were correspondingly broadened (Figure 5.3). Additional single labeled Opa sites to fully characterize Opa loop dynamics upon CEACAM binding will be investigated in the future.



**Figure 5.2: CW-EPR lineshapes of single labeled Opa proteins in detergent with excess nCEACAM1 GlcNAc proteins.** (A) Lineshapes of Opa proteins with (red) or without (black) nCEACAM1 GlcNAc proteins at a 2:1 Opa:CEACAM molar ratio. (B) Spectra that display an intensity decrease upon the addition of CEACAM were normalized to the intensity of the central manifold to better visualize altered lineshapes. Spectra in (A) that perfectly overlay when normalized in (B) only display a decrease in intensity in (A) due to noise rather than a decreased mobility upon binding, as the data processing normalizes the area under the signal to remove intensity changes due to concentration differences. Thus, only residues 152R1 and 166R1 appear to display linebroadening and CEACAM binding. Neighboring regions that do not exhibit binding (159R1 and 163R1) may be impacted by mutations and the incorporation of R1.

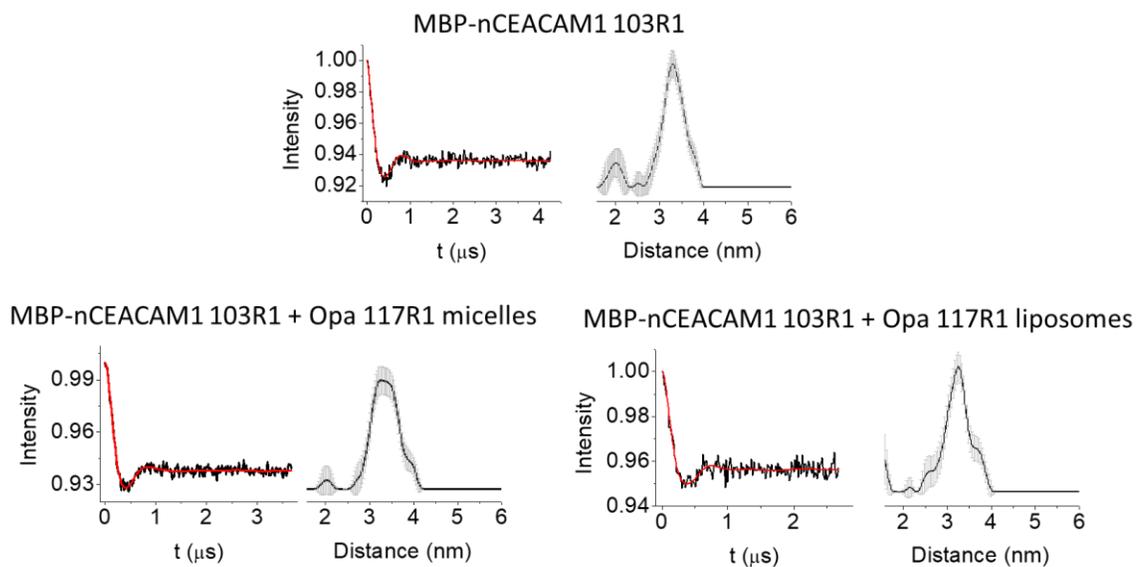


**Figure 5.3: CW-EPR lineshapes of Opa 152R1 micelles upon the addition of varying concentrations of nCEACAM1 GlcNAc.** Opa 152R1 proteins are maintained at the same concentration and volume for samples containing buffer (black), nCEACAM1 GlcNAc at a 1:2 Opa:CEACAM ratio (pink), at a 1:25 ratio (blue), and at a 1:50 ratio (red). As increasing concentrations of nCEACAM1 GlcNAc are added, EPR lineshapes broaden, which is indicative of Opa-CEACAM binding.

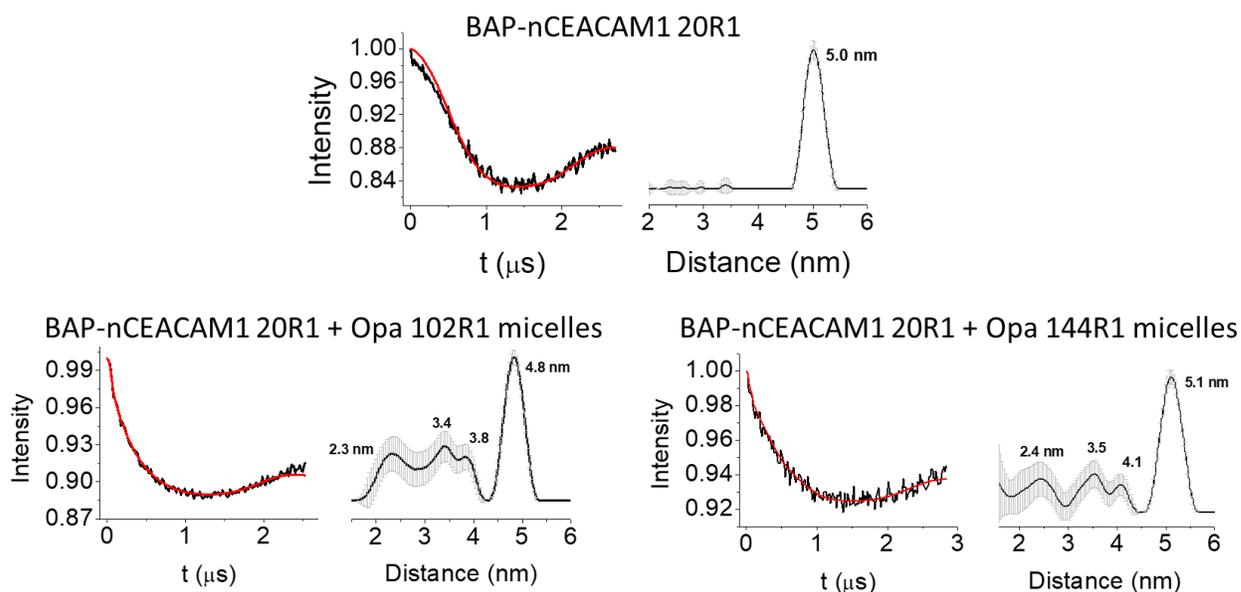
### 5.2.2 Modeling the Opa-CEACAM complex with DEER

DEER will also be utilized to examine the impact of CEACAM binding on Opa loop dynamics and conformational ensemble. We aim to map the Opa-CEACAM binding complex using single labeled Opa and CEACAM proteins together to generate a model of the bound complex that could be refined using MD. However, as mentioned previously, the addition of labeled Opa proteins to any of the CEACAM constructs generated did not result in additional “binding peaks” in the DEER distance distributions (Figures 5.4, 5.5, and 5.6). This lack of observed binding could be the result of nCEACAM1 being primarily homodimer that cannot be disrupted with the addition of Opa proteins or because Opa proteins are reconstituted into micelles or liposomes.

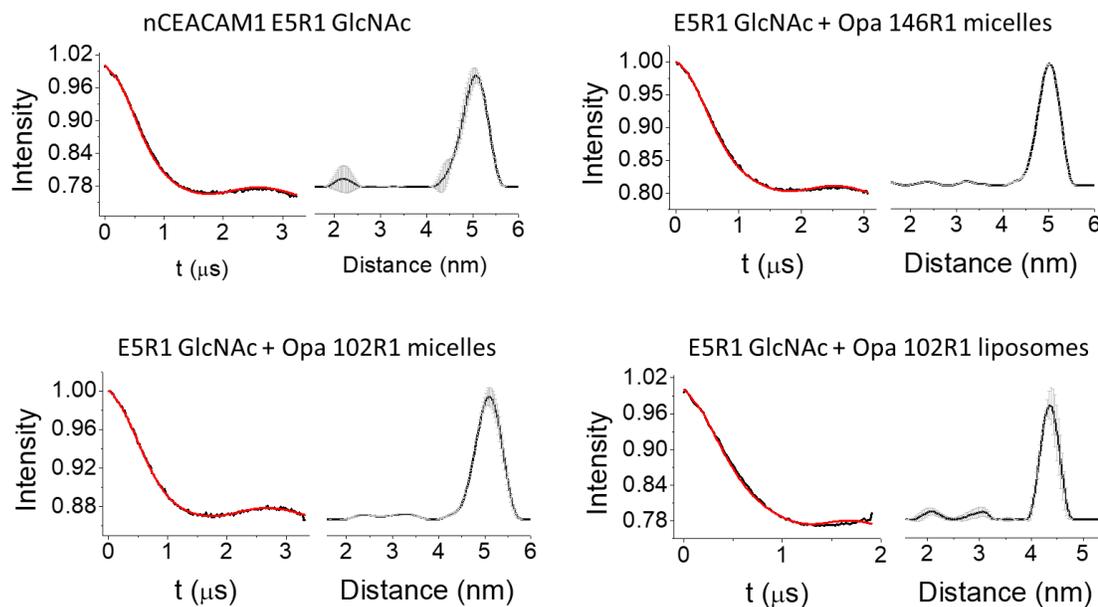
When detergent micelles form above the CMC, a portion of free detergent monomers remain in solution. Because the loops of Opa contain a large number of hydrophobic residues that may engage corresponding non-polar residues of CEACAM (I91 is critical to Opa-binding, among others), the hydrophobic hydrocarbon chain of detergent monomers in solution may directly interact with such hydrophobic residues and disrupt binding. While reconstituting Opa proteins into liposomes does not produce similar free lipid monomers, Opa proteins (or the liposomes themselves) often aggregate at the concentrations required to run DEER. As such, future experiments measuring Opa-CEACAM binding with DEER may need to be performed in different detergent or lipid types or in another membrane mimic altogether, such as nanodics. Altering the membrane mimic used to reconstitute Opa may still not result in CEACAM binding, as additional *in vivo* components, such as LOS and CEACAM glycosylation, may have a significant impact on Opa-CEACAM interactions *in vitro*.



**Figure 5.4: DEER distributions of MBP-nCEACAM1 Q103R1 + Opa 117R1.** MBP-nCEACAM1 proteins were added to Opa 117R1 micelles or liposomes, where the resulting DEER distributions did not result in additional binding peak(s).



**Figure 5.5: DEER distributions of BAP-nCEACAM1 L20R1 proteins immobilized onto avidin resin + Opa R1 micelles.** Biotinylated BAP-nCEACAM1 20R1 proteins are immobilized onto monomeric avidin resin, where the resulting DEER distributions upon the addition of Opa 102R1 micelles or Opa 144R1 micelles are largely the same as BAP-nCEACAM1 20R1 proteins alone. While three additional peaks are visible upon the addition of Opa in these representative cases, many other Opa proteins spin labeled at a variety of sites were probed (in liposomes and micelles). The DEER distributions for each Opa site probed were identical to 102R1 and 144R1, thus these interactions are likely some homodimer artifact. As such, even if a “binding peak” is present in these samples it is impossible to identify between different labeled Opa samples.



**Figure 5.6: DEER distributions of nCEACAM1 GlcNAc proteins + Opa R1 micelles or liposomes.** nCEACAM1 GlcNAc 5R1 proteins are added to an excess of Opa 146R1 micelles, Opa 102R1 micelles, or Opa 102R1 liposomes. The resulting DEER distributions are identical, which indicates Opa-CEACAM binding is not observed with N-glycosylation of nCEACAM1.

## Selected Opa-CEACAM DEER Distances

Opa + BAP-nCEACAM1		
Opa site	CEACAM site	Mean Distance (nm)
-	20R1	5.0
102R1 (micelle)	20R1	4.8
144R1 (micelle)	20R1	5.1
Opa + MBP-nCEACAM1		
Opa site	CEACAM site	Mean Distance (nm)
-	103R1	3.3
117R1 (micelle)	103R1	3.3
117R1 (lipid)	103R1	3.3
Opa + nCEACAM1 GlcNAc		
Opa site	CEACAM site	Mean Distance (nm)
-	5R1	5.0
146R1 (micelle)	5R1	5.0
102R1 (micelle)	5R1	5.0
102R1 (lipid)	5R1	4.5

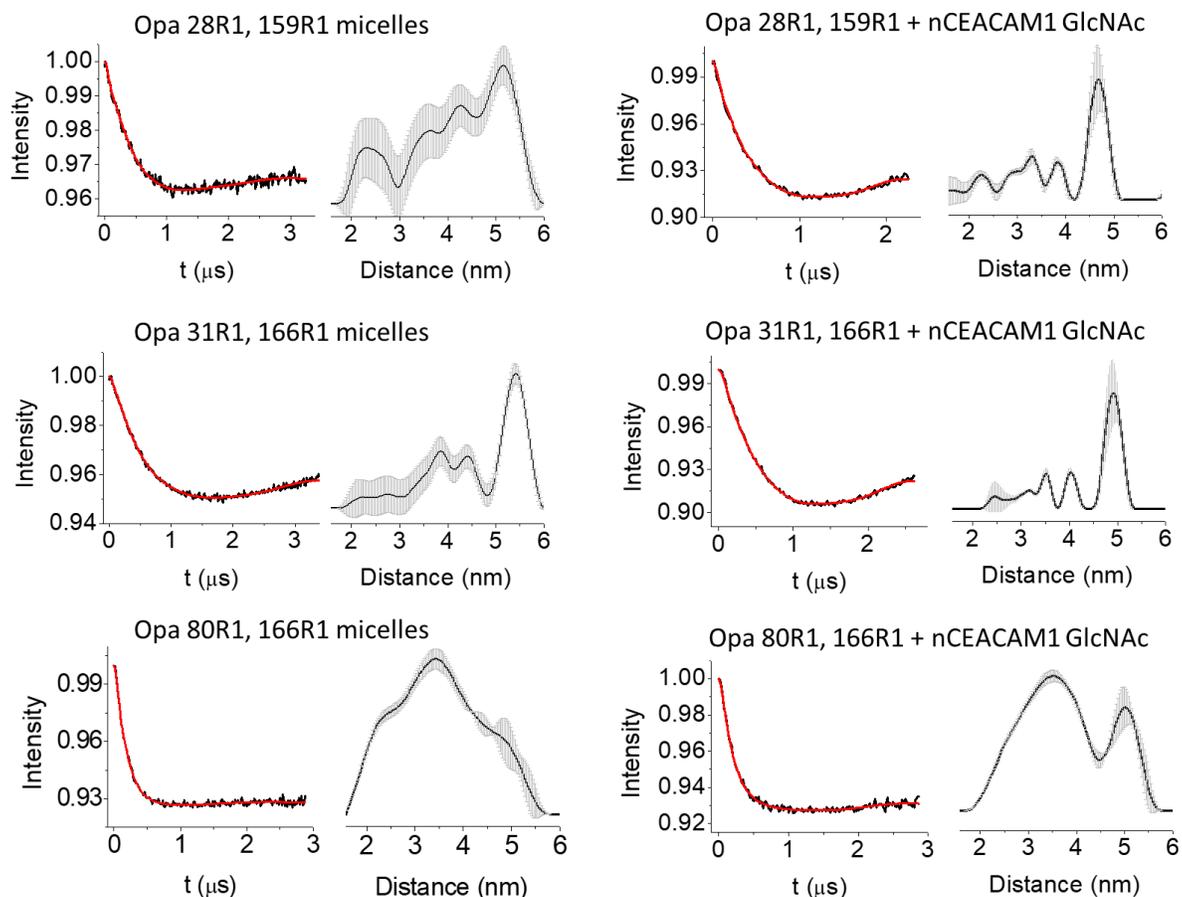
**Table 5.1: Selected mean DEER distances between single labeled Opa and CEACAM proteins.** The various CEACAM constructs (BAP, MBP, and GlcNAc) did not result in an additional “binding peak” upon the addition of Opa proteins in liposomes or micelles.

### 5.2.3 Opa loop conformational ensemble upon binding CEACAMs using DEER

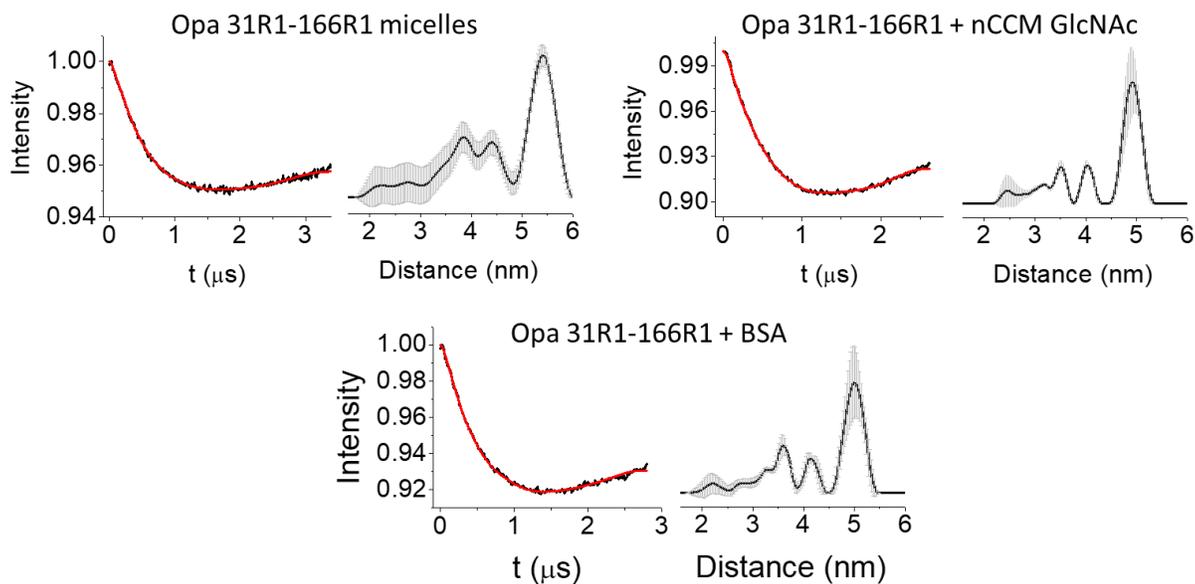
To further probe Opa loop dynamics upon engaging CEACAMs, double spin labeled Opa proteins will be added to an excess of CEACAMs and measured with DEER. A comparison of the resulting DEER distributions to those without CEACAM will allow us to examine alterations in the Opa loop conformational ensemble upon binding CEACAMs. It is expected that CEACAM binding will result in altered distances or a narrowing of the distribution, as the overall lateral loop mobility is predicted to decrease upon binding receptor if multiple Opa loops are involved. Preliminary DEER measurements to this end result in a narrowing of the DEER distribution for three double-labeled sites probed upon the addition of excess nCEACAM1 GlcNAc (Figure 5.7). However, a similar distribution is generated upon the addition of bovine serum albumin (BSA) to the double labeled Opa proteins for one loop pair (Figure 5.8). This indicates that a decrease in mobility observed for that Opa loop pair (via a narrowing of the distribution) may be due to a steric effect of highly concentrated BSA/CEACAM proteins in solution obstructing the motion of Opa loops rather than a binding event causing such ordering. However, the slope of the background signals of the +CEACAM samples are significantly different from those of the -CEACAM samples, which maybe suggestive of binding with a narrowing of the loop sampling range (a great deal of the background signal of the -CEACAM samples comes from real long-distance loop conformations). Thus, additional sites must be probed to determine if Opa-CEACAM binding can be measured in this manner.

If binding is observed with DEER, many labeled locations can be probed to generate a model of the Opa-CEACAM binding complex. Our DEER/MD methodology

(described in Chapter 3) can be applied to this system, where Opa and CEACAM spin label sites (or double labeled Opa pairs with unlabeled CEACAM) can be identified and measured. An incorporation of DEER distance restraints into MD simulations of the complex can be compared to those of Opa loops in isolation to determine the shift in the Opa loop conformational ensemble in the presence of receptor. This comparison will enable us to determine the mechanism by which Opa loop dynamics mediate CEACAM binding (conformational selection, fuzzy complex, etc.), and thus *Neisseria* pathogenesis. When measuring single or double labeled Opa proteins with DEER in the presence of CEACAM, Opa proteins must be saturated with receptor to ensure unbound Opa loops are not biasing the DEER distribution significantly. Thus far, the nCEACAM1 homodimer has not been disrupted sufficiently to be confident Opa-CEACAM binding is being observed with such techniques; however, preliminary results suggest that a binding event is being detected, albeit minimally.



**Figure 5.7: DEER distributions of double spin labeled Opa proteins + unlabeled nCEACAM1 GlcNAc proteins.** DEER distributions of Opa 28R1-159R1, Opa 31R1-166R1, and Opa 80R1-166R1 micelles with and without an excess of nCEACAM1 GlcNAc proteins. Upon the addition of CEACAM, the DEER distributions of each of these double labeled Opa proteins become narrowed, which may indicate decreased Opa loop mobility upon binding. The slope of the background signals of the +CEACAM samples are significantly different from those of the -CEACAM samples, which maybe suggestive of binding with a narrowing of the loop sampling range (a great deal of the background signal of the -CEACAM samples comes from real long-distance loop conformations).



**Figure 5.8: DEER distributions of Opa 31-166R1 micelles with nCEACAM1 GlcNAc or BSA.** The DEER distribution of Opa 31-166R1 micelles becomes narrowed upon the addition of nCEACAM1 GlcNAc which could be indicative of binding. Adding similar amounts of BSA to Opa 31-166R1 micelles resulted in a similar DEER distribution, which suggests that the altered distribution upon the addition of CEACAM is a result of steric effects from proteins in solution, rather than a binding event.

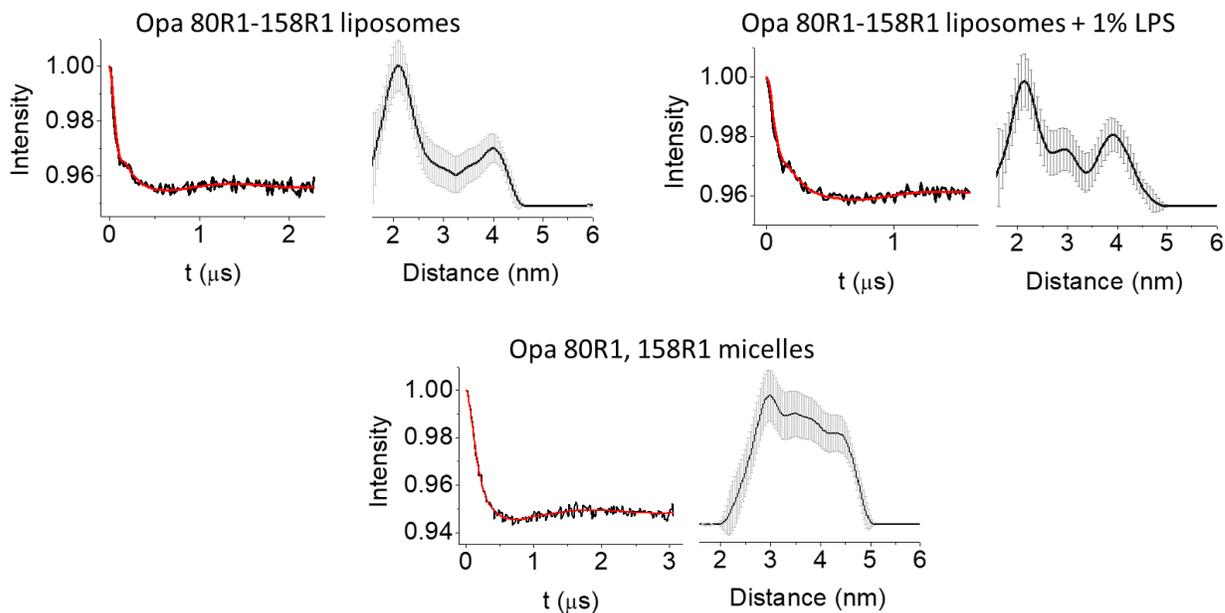
### 5.3 Impact of LOS on Opa loop dynamics and CEACAM binding

The observation of minimal Opa-CEACAM binding with EPR techniques could indicate that an additional factor found in *Neisseria* or human epithelial cells may not be captured in our experimental design. One of these factors could be LOS found in the outer membrane of *Neisseria*. Several lower loop residues of Opa are basic and directly interact with the negatively charged LOS moieties [177]. LOS could serve to sterically modulate Opa loop dynamics, as the LOS lipids occupy a large portion of the outer membrane and extend vertically from the membrane. Mutations in the structure of *Neisseria meningitidis* LOS resulted in diminished Nm internalization into host cells [178], which may be related to LOS interactions with Opa (and in turn affect CEACAM binding).

To model LOS in future Opa dynamics or binding experiments, commercially available short-chain LPS lipids can be incorporated into the membrane mimic. CW-EPR and DEER experiments using spin labeled Opa proteins reconstituted into liposomes containing LPS will be used to examine the impact of LPS on Opa loop dynamics and conformational ensemble. Ultimately, comparable DEER experiments will be performed with the addition of nCEACAM1 proteins (in glycosylated and non-glycosylated forms) to determine the effect of LPS on Opa interactions with CEACAM. Likewise, SPR experiments (discussed in section 5.1.4) can be performed using Opa proteoliposomes containing LPS to determine the effect of LPS on Opa-CEACAM interactions.

Preliminary DEER experiments of Opa 80-158R1 with 1% LPS incorporated into liposomes provide data similar to samples without LPS, but different from the DEER distribution with Opa reconstituted into micelles (Figure 5.9). Distances observed for Opa 80-158R1 liposomes with or without LPS are near 2 nm, which is approaching the lower

observable limit of DEER and as such minor changes in the conformational ensemble that occur from the addition of LPS may not be observed. Additionally, only 1% LPS was incorporated in this experiment; increased amounts of LPS may be required to observe changes in the DEER signal. The DEER distances of Opa 80-158R1 liposomes are shorter than the distances observed for Opa 80-158R1 in micelles, which may indicate that the addition of the large PEG-ylated lipids as well as LPS to the liposomes decreases the lateral mobility of Opa loops. Additional Opa pairs must be measured in liposomes (with and without LPS) to gain a further understanding of the effect of the membrane mimic on the Opa loop conformational ensemble as well as the impact of LPS on Opa loop dynamics.



**Figure 5.9: DEER distributions of Opa 80-158R1 liposomes with and without LPS.**

The addition of 1% LPS to Opa proteoliposomes did not result in an altered DEER distribution of Opa 80-158R1. The primary distance observed in both cases is near 2 nm, which is approaching the lower observable limit of DEER. As such, any alterations in the Opa loop conformations that may occur with the addition of LPS may be too small to be observed by this technique. Likewise, higher amounts of LPS may be required to observe an impact on Opa loop dynamics. The DEER distributions of both Opa 80-158R1 liposome samples contain shorter distances than in micelles, which indicates that the large PEGylated and LPS lipids affect the lateral mobility of Opa loops.

#### 5.4 Effect of CEACAM glycosylation on Opa binding

As previously mentioned, the ectodomains of the CEACAM family are heavily glycosylated *in vivo*. The effect of N-glycosylation on the homodimerization of the N-terminal domain of nCEACAM1 was examined in Chapter 4 of this dissertation, although the full impact of glycosylation on modulating CEACAM oligomerization is still being examined. It is likely that such CEACAM glycosylation impacts Opa binding as well, which poses several interesting hypotheses regarding the mode of glycosylated CEACAM interactions with Opa proteins. Glycosylation on the ectodomains of CEACAMs may serve to modify the dimerization propensity of CEACAMs, perhaps by sterically limiting the formation of non-native dimers or by subtly altering the structure of CEACAMs themselves to impact the dimer structure and binding affinity. Such a modification of CEACAM dimerization via glycosylation may serve to promote (or limit) Opa binding. It is also possible that Opa proteins may bind glycosylated CEACAMs as dimers *in vivo*, where monomeric CEACAMs are not required for binding.

Likewise, certain Opa proteins may directly engage the glycans of CEACAM in conjunction with other Opa proteins that bind CEACAM residues. While mutagenesis studies have shown that several CEACAM residues are required for Opa binding, only three Opa proteins of Nm strain C751 were examined in that work [46]. As such, the binding partners of the majority of Opa proteins (which have hundreds of potential HV sequences) have not been identified, and some may engage CEACAM glycans directly. The role of CEACAM glycans in Opa binding or *Neisseria* invasion has not been studied previously. Other bacteria that bind CEACAMs to promote invasion and colonization of the host, such as *E. coli* [179], have been shown to interact directly with the glycans of

CEACAMs on host cell membranes, as de-glycosylation of the receptors results in decreased bacterial adhesion to host cells [180]. This CEACAM binding is mediated via the Dr family of adhesions of *E. coli* [181]. Likewise, certain Opa proteins, or perhaps other *Neisseria* outer membrane adhesions such as Opc, may engage CEACAM glycans directly to promote bacterial colonization. As such, the possible impact of CEACAM glycosylation on receptor engagement and *Neisseria* phagocytosis should not be ignored.

Additional Ig-like domains of CEACAMs may also impact Opa binding, as these ectodomains promote CEACAM1 oligomerization [144]. The domains of CEACAMs directly associate, where recombinant, non-glycosylated N-terminal domain of CEACAM1 binds the Ig-like A3 domain of CEACAM1 with high affinity [152], for example. The promotion of CEACAM interactions with multiple domains may affect Opa binding by occluding Opa binding residues located on the same binding surface as CEACAM dimerization [86, 136]. Thus, the outlined CEACAM homodimerization and Opa binding experiments must also be performed using soluble CEACAMs containing various numbers of extracellular domains (ideally in both glycosylated and non-glycosylated forms). For example, CEACAM1 contains four Ig-like domains that can be varied in number (and glycosylation) to identify the impact of these domains on CEACAM oligomerization and as a result Opa interactions.

### 5.5 Opa residues involved in CEACAM interactions

Although the Opa-binding residues on CEACAM have been identified, the specific residues and regions on Opa that engage CEACAM are unknown (although HV1 and HV2 are likely required). This information is critical to determine the molecular mechanism by which Opa proteins engage host receptors to promote bacterial phagocytosis, and as such

several future experiments must be performed to identify the CEACAM-binding motif of Opa. CW and DEER experiments mentioned above will provide preliminary insight into this binding motif from regions that display the greatest lineshape/distribution changes upon binding. However, it is possible that the incorporation of spin labels near or at the CEACAM interaction site disrupts binding, and as such, additional methods must be utilized to fully characterize the CEACAM binding motif of Opa.

Surface plasmon resonance (SPR) is currently being used to determine the binding affinity and kinetics of the Opa<sub>60</sub>-nCEACAM1 GlcNAc interaction (previous BLI binding experiments necessitated immobilized CEACAMs that could not be applied to glycosylated CEACAM). For SPR experiments Opa proteoliposomes are immobilized onto a gold sensor chip coated with streptavidin by incorporating biotinylated PEG lipids into the liposomes. CEACAM proteins flow over this sensor chip at varying concentrations, which are significantly lower than those required for EPR (approximately 0.2 – 3  $\mu$ M compared to roughly 150  $\mu$ M with DEER), so CEACAM homodimerization may be reduced. Preliminary SPR experiments are promising, as the association signal for Opa HV<sup>-</sup> proteins (Opa<sub>60</sub> with the HV regions deleted) is less than that of Opa<sub>60</sub> proteoliposomes. Conditions are currently being optimized to diminish non-specific binding and improve chip regeneration between CEACAM injections. Upon optimization of binding conditions, we can compare the interactions of Opa proteins with non-glycosylated and glycosylated CEACAMs. Mutagenesis studies of the Opa loops (alanine-scanning or deletions of selective HV regions), as well as competition experiments where selective HV peptides are added to compete for CEACAM binding, can be used to identify loop regions and residues that are involved in receptor interactions. Experiments that oblate

CEACAM binding (without affecting the solubility or overall structure of Opa) would allow a determination of the CEACAM-binding region. SPR will also be utilized to measure the binding of spin labeled Opa proteins and CEACAMs to ensure that the introduction of R1 does not disrupt Opa-CEACAM interactions.

#### 5.6 Mechanism of Opa-CEACAM binding among different protein variants

Because of the high degree of sequence variability of Opa proteins, it is currently impossible to design a *Neisseria* antibiotic that targets the Opa-CEACAM interaction to disrupt *Neisseria* colonization. Nevertheless, a common Opa loop structure at the binding interface may be identified using the above methods, from which an antibiotic can be modeled. While the research in this dissertation was limited to Opa<sub>60</sub> and nCEACAM1 specifically, additional Opa and CEACAM proteins must be also studied to fully characterize the mechanism of Opa-CEACAM interactions and generate a molecular model of Opa-mediated *Neisseria* phagocytosis into host cells. Future experiments will need to take into account a variety of Opa proteins that bind different CEACAMs (or none at all), as well as various CEACAMs. Measuring Opa-CEACAM interactions using proteins of different sequences and binding partners will allow us to determine the molecular mechanism of Opa-CEACAM interactions (fuzzy complex, conformational selection, etc.) and identify a universal binding motif or mechanism that could be applied to *Neisseria* targeted antibiotics.

#### 5.7 Concluding remarks

Neisserial Opa protein interactions with human CEACAMs triggers the uptake of *Neisseria gonorrhoeae* (Gc) and *N. meningitides* (Nm) into non-phagocytic and phagocytic

cells as a means of infection. The internalization into host mucosal tissue enhances the survivability of Gc and Nm, which cause the diseases gonorrhea and meningococcal meningitis, respectively. Gc is of particular interest due to its increasing antibiotic resistance, and as such the need for alternative treatments is critical. Understanding the molecular mechanism of Opa-CEACAM interactions will provide insight into the mode of *Neisseria* internalization and could in turn be utilized as an antibiotic target.

The extracellular loops of Opa<sub>60</sub> were found to be highly mobile in the ns-timescale with no observable secondary structure. Because Opa proteins are highly dynamic, a hybrid DEER and MD technique was developed in collaboration with Peter Kasson and Jennifer Hays to identify DEER pairs that provide optimal experimental restraints for simulations of the Opa loop conformational ensemble. The preliminary conformational ensemble indicates that the extracellular loops interact in absence of receptor. Together, the highly dynamic nature of Opa loops may enable diverse Opa proteins to selectively engage various CEACAM receptors, where a flexible binding surface allows Opa proteins to recognize CEACAM proteins with different sequences.

CEACAMs participate in homophilic dimerization events that are largely uncharacterized. Crystal structures of nCEACAM1 dimers are different from what we observe in solution. As such, a model of the nCEACAM1 homodimer from DEER data is currently being generated. Likewise, the addition of N-glycosylation to nCEACAM1 proteins did not disrupt homodimerization, as was previously predicted. Future experiments will be performed to characterize the nCEACAM1 homodimer in its glycosylated and non-glycosylated forms.

Studying Opa and CEACAM proteins in isolation allows us to gain insight into the mechanism of Opa-CEACAM interactions. If Opa loops sample a receptor-competent state in their unbound form, the mode of binding may be conformational selection, whereas if the loop binding conformation is unobserved in the unbound state, an induced folding or fuzzy complex may be the mechanism of Opa-CEACAM interactions. Understanding the molecular mechanism of Opa-CEACAM binding will ultimately elucidate the manner by which pathogenic bacteria induce phagocytosis into human cells.

1. Kendrew, J.C., et al., *A Three-Dimensional Model of the Myoglobin Molecule Obtained by X-Ray Analysis*. Nature, 1958. **181**: p. 662.
2. Blake, C.C.F., et al., *Structure of Hen Egg-White Lysozyme: A Three-dimensional Fourier Synthesis at 2 Å Resolution*. Nature, 1965. **206**: p. 757.
3. Braun, W., et al., *COMBINED USE OF PROTON-PROTON OVERHAUSER ENHANCEMENTS AND A DISTANCE GEOMETRY ALGORITHM FOR DETERMINATION OF POLYPEPTIDE CONFORMATIONS - APPLICATION TO MICELLE-BOUND GLUCAGON*. Biochimica Et Biophysica Acta, 1981. **667**(2): p. 377-396.
4. Deisenhofer, J., et al., *Structure of the protein subunits in the photosynthetic reaction centre of Rhodospseudomonas viridis at 3Å resolution*. Nature, 1985. **318**(6047): p. 618-24.
5. Fischer, E., *Einfluss der Configuration auf die Wirkung der Enzyme*. Berichte der deutschen chemischen Gesellschaft, 1894. **27**(3): p. 2985-2993.
6. Oldfield, C.J., et al., *Comparing and combining predictors of mostly disordered proteins*. Biochemistry, 2005. **44**(6): p. 1989-2000.
7. Austin, R.H., et al., *Dynamics of ligand binding to myoglobin*. Biochemistry, 1975. **14**(24): p. 5355-5373.
8. Henzler-Wildman, K. and D. Kern, *Dynamic personalities of proteins*. Nature, 2007. **450**(7172): p. 964-972.
9. Claxton, D.P., et al., *Chapter Twelve - Navigating Membrane Protein Structure, Dynamics, and Energy Landscapes Using Spin Labeling and EPR Spectroscopy*, in *Methods in Enzymology*, P.Z. Qin and K. Warncke, Editors. 2015, Academic Press. p. 349-387.

10. Koshland, D.E., *Application of a Theory of Enzyme Specificity to Protein Synthesis*. Proceedings of the National Academy of Sciences of the United States of America, 1958. **44**(2): p. 98-104.
11. Lindorff-Larsen, K., et al., *Simultaneous determination of protein structure and dynamics*. Nature, 2005. **433**(7022): p. 128-32.
12. Tompa, P. and M. Fuxreiter, *Fuzzy complexes: polymorphism and structural disorder in protein-protein interactions*. Trends Biochem Sci, 2008. **33**(1): p. 2-8.
13. Zimmermann, J., et al., *Antibody evolution constrains conformational heterogeneity by tailoring protein dynamics*. Proceedings of the National Academy of Sciences of the United States of America, 2006. **103**(37): p. 13722-13727.
14. Thorpe, I.F. and C.L. Brooks, *Molecular evolution of affinity and flexibility in the immune system*. Proceedings of the National Academy of Sciences of the United States of America, 2007. **104**(21): p. 8821-8826.
15. Keskin, O., *Binding induced conformational changes of proteins correlate with their intrinsic fluctuations: a case study of antibodies*. BMC Structural Biology, 2007. **7**: p. 31-31.
16. James, L.C., P. Roversi, and D.S. Tawfik, *Antibody multispecificity mediated by conformational diversity*. Science, 2003. **299**(5611): p. 1362-7.
17. O'Brien, P.J. and D. Herschlag, *Catalytic promiscuity and the evolution of new enzymatic activities*. Chem Biol, 1999. **6**(4): p. R91-r105.
18. James, L.C. and D.S. Tawfik, *Conformational diversity and protein evolution--a 60-year-old hypothesis revisited*. Trends Biochem Sci, 2003. **28**(7): p. 361-8.
19. Mannige, R.V., *Dynamic New World: Refining Our View of Protein Structure, Function and Evolution*. Proteomes, 2014. **2**(1): p. 128-153.

20. Wallin, E. and G. von Heijne, *Genome-wide analysis of integral membrane proteins from eubacterial, archaean, and eukaryotic organisms*. Protein Science, 1998. **7**(4): p. 1029-1038.
21. Fagerberg, L., et al., *Prediction of the human membrane proteome*. Proteomics, 2010. **10**(6): p. 1141-1149.
22. Overington, J.P., B. Al-Lazikani, and A.L. Hopkins, *Opinion - How many drug targets are there?* Nature Reviews Drug Discovery, 2006. **5**(12): p. 993-996.
23. Patching, S.G., *NMR structures of polytopic integral membrane proteins*. Molecular Membrane Biology, 2011. **28**(6): p. 370-397.
24. Dutta, R.C., *Drug carriers in pharmaceutical design: Promises and progress*. Current Pharmaceutical Design, 2007. **13**(7): p. 761-769.
25. Kellenberger, E. and A. Ryter, *Cell Wall and Cytoplasmic Membrane of Escherichia coli*. The Journal of Biophysical and Biochemical Cytology, 1958. **4**(3): p. 323-326.
26. Zhai, Y.F. and M.H. Saier, *The beta-barrel finder (BBF) program, allowing identification of outer membrane beta-barrel proteins encoded within prokaryotic genomes*. Protein Science, 2002. **11**(9): p. 2196-2207.
27. Wimley, W.C., *Toward genomic identification of beta-barrel membrane proteins: Composition and architecture of known structures*. Protein Science, 2002. **11**(2): p. 301-312.
28. Wimley, W.C., *The versatile beta-barrel membrane protein*. Current Opinion in Structural Biology, 2003. **13**(4): p. 404-411.
29. Chou, K.C. and L. Carlacci, *ENERGETIC APPROACH TO THE FOLDING OF ALPHA-BETA BARRELS*. Proteins-Structure Function and Genetics, 1991. **9**(4): p. 280-295.

30. van den Berg, B., *The FadL family: unusual transporters for unusual substrates*. Current Opinion in Structural Biology, 2005. **15**(4): p. 401-407.
31. Choutko, A., et al., *Membrane protein dynamics in different environments: simulation study of the outer membrane protein X in a lipid bilayer and in a micelle*. European Biophysics Journal with Biophysics Letters, 2011. **40**(1): p. 39-58.
32. Tamaki, S., T. Sato, and M. Matsuhashi, *Role of Lipopolysaccharides in Antibiotic Resistance and Bacteriophage Adsorption of Escherichia coli K-12*. Journal of Bacteriology, 1971. **105**(3): p. 968-975.
33. Sanderson, K.E., et al., *Permeability of lipopolysaccharide-deficient (rough) mutants of Salmonella typhimurium to antibiotics, lysozyme, and other agents*. Can J Microbiol, 1974. **20**(8): p. 1135-45.
34. Ried, G., I. Hindennach, and U. Henning, *Role of lipopolysaccharide in assembly of Escherichia coli outer membrane proteins OmpA, OmpC, and OmpF*. Journal of Bacteriology, 1990. **172**(10): p. 6048-6053.
35. Sen, K. and H. Nikaido, *In vitro trimerization of OmpF porin secreted by spheroplasts of Escherichia coli*. Proceedings of the National Academy of Sciences of the United States of America, 1990. **87**(2): p. 743-747.
36. Sen, K. and H. Nikaido, *Lipopolysaccharide structure required for in vitro trimerization of Escherichia coli OmpF porin*. Journal of Bacteriology, 1991. **173**(2): p. 926-928.
37. Kramer, R.A., et al., *Lipopolysaccharide regions involved in the activation of Escherichia coli outer membrane protease OmpT*. Eur J Biochem, 2002. **269**(6): p. 1746-52.
38. Merz, A.J. and M. So, *Interactions of pathogenic Neisseriae with epithelial cell membranes*. Annual Review of Cell and Developmental Biology, 2000. **16**: p. 423-+.

39. Diaz Romero, J. and I.M. Outshoorn, *Current status of meningococcal group B vaccine candidates: capsular or noncapsular?* Clinical Microbiology Reviews, 1994. **7**(4): p. 559-575.
40. Ohnishi, M., et al., *THE NEW SUPERBUG NEISSERIA GONORRHOEAE MAKES GONORRHOEA UNTREATABLE?-FIRST HIGH-LEVEL CEFTRIAXONE RESISTANCE WORLDWIDE AND PUBLIC HEALTH IMPORTANCE.* Sexually Transmitted Infections, 2011. **87**: p. A76-A76.
41. Hauck, C.R. and T.F. Meyer, *'Small' talk: Opa proteins as mediators of Neisseria-host-cell communication.* Current Opinion in Microbiology, 2003. **6**(1): p. 43-49.
42. Segal, E., et al., *Antigenic variation of gonococcal pilus involves assembly of separated silent gene segments.* Proceedings of the National Academy of Sciences of the United States of America, 1986. **83**(7): p. 2177-2181.
43. Stern, A., et al., *Opacity genes in Neisseria gonorrhoeae: control of phase and antigenic variation.* Cell, 1986. **47**(1): p. 61-71.
44. Billker, O., et al., *Distinct mechanisms of internalization of Neisseria gonorrhoeae by members of the CEACAM receptor family involving Rac1- and Cdc42-dependent and -independent pathways.* Embo Journal, 2002. **21**(4): p. 560-571.
45. Dehio, C., S.D. Gray-Owen, and T.F. Meyer, *The role of neisserial Opa proteins in interactions with host cells.* Trends in Microbiology, 1998. **6**(12): p. 489-495.
46. Virji, M., et al., *Critical determinants of host receptor targeting by Neisseria meningitidis and Neisseria gonorrhoeae: identification of Opa adhesiotopes on the N-domain of CD66 molecules.* Molecular Microbiology, 1999. **34**(3): p. 538-551.

47. Ball, L.M. and A.K. Criss, *Constitutively Opa-Expressing and Opa-Deficient Neisseria gonorrhoeae Strains Differentially Stimulate and Survive Exposure to Human Neutrophils*. Journal of Bacteriology, 2013. **195**(13): p. 2982-2990.
48. Criss, A.K. and H.S. Seifert, *A bacterial siren song: intimate interactions between Neisseria and neutrophils*. Nature Reviews Microbiology, 2012. **10**(3): p. 178-190.
49. Sadarangani, M., A.J. Pollard, and S.D. Gray-Owen, *Opa proteins and CEACAMs: pathways of immune engagement for pathogenic Neisseria*. Fems Microbiology Reviews, 2011. **35**(3): p. 498-514.
50. Bos, M.P., F. Grunert, and R.J. Belland, *Differential recognition of members of the carcinoembryonic antigen family by Opa variants of Neisseria gonorrhoeae*. Infection and Immunity, 1997. **65**(6): p. 2353-2361.
51. Bilek, N., C.A. Ison, and B.G. Spratt, *Relative Contributions of Recombination and Mutation to the Diversification of the opa Gene Repertoire of Neisseria gonorrhoeae*. Journal of Bacteriology, 2009. **191**(6): p. 1878-1890.
52. Callaghan, M.J., K.A. Jolley, and M.C.J. Maiden, *Opacity-Associated Adhesin Repertoire in Hyperinvasive Neisseria meningitidis*. Infection and Immunity, 2006. **74**(9): p. 5085-5094.
53. Fox, D.A., et al., *Structure of the Neisserial Outer Membrane Protein Opa(60): Loop Flexibility Essential to Receptor Recognition and Bacterial Engulfment*. Journal of the American Chemical Society, 2014. **136**(28): p. 9938-9946.
54. Malorny, B., et al., *Sequence diversity, predicted two-dimensional protein structure, and epitope mapping of neisserial Opa proteins*. Journal of Bacteriology, 1998. **180**(5): p. 1323-1330.

55. Bos, M.P., et al., *Carcinoembryonic antigen family receptor recognition by gonococcal Opa proteins requires distinct combinations of hypervariable Opa protein domains*. *Infection and Immunity*, 2002. **70**(4): p. 1715-1723.
56. Horst, A.K., et al., *Carcinoembryonic antigen-related cell adhesion molecule 1 modulates vascular remodeling in vitro and in vivo*. *The Journal of Clinical Investigation*, 2006. **116**(6): p. 1596-1605.
57. Scheffrahn, I., et al., *Control of density-dependent, cell state-specific signal transduction by the cell adhesion molecule CEACAM1, and its influence on cell cycle regulation*. *Experimental Cell Research*, 2005. **307**(2): p. 427-435.
58. Klaile, E., et al., *CEACAM1 functionally interacts with filamin A and exerts a dual role in the regulation of cell migration*. *Journal of Cell Science*, 2005. **118**(23): p. 5513-5524.
59. Ebrahimnejad, A., et al., *CEACAM1 Enhances Invasion and Migration of Melanocytic and Melanoma Cells*. *The American Journal of Pathology*, 2004. **165**(5): p. 1781-1787.
60. Gray-Owen, S.D. and R.S. Blumberg, *CEACAM1: contact-dependent control of immunity*. *Nature Reviews Immunology*, 2006. **6**: p. 433.
61. Beauchemin, N., et al., *Redefined nomenclature for members of the carcinoembryonic antigen family*. *Exp Cell Res*, 1999. **252**(2): p. 243-9.
62. Hefta, S.A., et al., *Carcinoembryonic antigen is anchored to membranes by covalent attachment to a glycosylphosphatidylinositol moiety: identification of the ethanolamine linkage site*. *Proc Natl Acad Sci U S A*, 1988. **85**(13): p. 4648-52.
63. Obrink, B., *CEA adhesion molecules: multifunctional proteins with signal-regulatory properties*. *Curr Opin Cell Biol*, 1997. **9**(5): p. 616-26.
64. McCaw, S.E., E.H. Liao, and S.D. Gray-Owen, *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): p. 2742-2752.
65. Voges, M., et al., *Extracellular IgC2 constant domains of CEACAMs mediate PI3K sensitivity during uptake of pathogens*. PLoS One, 2012. **7**(6): p. e39908.
66. Hammarstrom, S., *The carcinoembryonic antigen (CEA) family: structures, suggested functions and expression in normal and malignant tissues*. Semin Cancer Biol, 1999. **9**(2): p. 67-81.
67. Frangsmyr, L., et al., *Cell- and region-specific expression of biliary glycoprotein and its messenger RNA in normal human colonic mucosa*. Cancer Res, 1995. **55**(14): p. 2963-7.
68. Prall, F., et al., *CD66a (BGP), an adhesion molecule of the carcinoembryonic antigen family, is expressed in epithelium, endothelium, and myeloid cells in a wide range of normal human tissues*. J Histochem Cytochem, 1996. **44**(1): p. 35-41.
69. Neumaier, M., et al., *Biliary glycoprotein, a potential human cell adhesion molecule, is down-regulated in colorectal carcinomas*. Proc Natl Acad Sci U S A, 1993. **90**(22): p. 10744-8.
70. Rosenberg, M., et al., *The expression of mouse biliary glycoprotein, a carcinoembryonic antigen-related gene, is down-regulated in malignant mouse tissues*. Cancer Res, 1993. **53**(20): p. 4938-45.
71. Busch, C., et al., *Down-regulation of CEACAM1 in human prostate cancer: correlation with loss of cell polarity, increased proliferation rate, and Gleason grade 3 to 4 transition*. Hum Pathol, 2002. **33**(3): p. 290-8.
72. Huang, J., et al., *Expression of biliary glycoprotein (CD66a) in normal and malignant breast epithelial cells*. Anticancer Res, 1998. **18**(5a): p. 3203-12.

73. Oliveira-Ferrer, L., et al., *Dual role of carcinoembryonic antigen-related cell adhesion molecule 1 in angiogenesis and invasion of human urinary bladder cancer*. *Cancer Res*, 2004. **64**(24): p. 8932-8.
74. Lu, R., et al., *Tumor Angiogenesis Mediated by Myeloid Cells Is Negatively Regulated by CEACAM1* (vol 72, pg 2239, 2012). *Cancer Research*, 2016. **76**(4): p. 984-985.
75. Jothy, S., S.Y. Yuan, and K. Shiota, *Transcription of carcinoembryonic antigen in normal colon and colon carcinoma. In situ hybridization study and implication for a new in vivo functional model*. *Am J Pathol*, 1993. **143**(1): p. 250-7.
76. Kodera, Y., et al., *Expression of carcinoembryonic antigen (CEA) and nonspecific crossreacting antigen (NCA) in gastrointestinal cancer; the correlation with degree of differentiation*. *Br J Cancer*, 1993. **68**(1): p. 130-6.
77. Tiernan, J.P., et al., *Carcinoembryonic antigen is the preferred biomarker for in vivo colorectal cancer targeting*. *Br J Cancer*, 2013. **108**(3): p. 662-7.
78. Scholzel, S., et al., *Carcinoembryonic antigen family members CEACAM6 and CEACAM7 are differentially expressed in normal tissues and oppositely deregulated in hyperplastic colorectal polyps and early adenomas*. *Am J Pathol*, 2000. **156**(2): p. 595-605.
79. Kim, K.S., et al., *Overexpression and clinical significance of carcinoembryonic antigen-related cell adhesion molecule 6 in colorectal cancer*. *Clinica Chimica Acta*, 2013. **415**: p. 12-19.
80. Kammerer, R. and W. Zimmermann, *Coevolution of activating and inhibitory receptors within mammalian carcinoembryonic antigen families*. *BMC Biol*, 2010. **8**: p. 12.
81. Turbide, C., et al., *Optimal ratios of biliary glycoprotein isoforms required for inhibition of colonic tumor cell growth*. *Cancer Res*, 1997. **57**(13): p. 2781-8.

82. Wang, L., et al., *C-CAM1, a candidate tumor suppressor gene, is abnormally expressed in primary lung cancers*. Clin Cancer Res, 2000. **6**(8): p. 2988-93.
83. Hill, D.J. and M. Virji, *A novel cell-binding mechanism of Moraxella catarrhalis ubiquitous surface protein UspA: specific targeting of the N-domain of carcinoembryonic antigen-related cell adhesion molecules by UspA1*. Mol Microbiol, 2003. **48**(1): p. 117-29.
84. Hill, D.J., et al., *The variable P5 proteins of typeable and non-typeable Haemophilus influenzae target human CEACAM1*. Molecular Microbiology, 2001. **39**(4): p. 850-862.
85. Sauter, S.L., et al., *Identification of the specific oligosaccharide sites recognized by type 1 fimbriae from Escherichia coli on nonspecific cross-reacting antigen, a CD66 cluster granulocyte glycoprotein*. J Biol Chem, 1993. **268**(21): p. 15510-6.
86. Fedarovich, A., et al., *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**: p. 971-979.
87. Kozyrev, S.A.A.t.a.B.M., *Electron Paramagnetic Resonance*. 1964, New York, NY: Academic.
88. Hubbell, W.L., D.S. Cafiso, and C. Altenbach, *Identifying conformational changes with site-directed spin labeling*. Nature Structural Biology, 2000. **7**(9): p. 735-739.
89. Berliner, L.J., et al., *A NOVEL REVERSIBLE THIOL-SPECIFIC SPIN LABEL - PAPAINE ACTIVE-SITE LABELING AND INHIBITION*. Analytical Biochemistry, 1982. **119**(2): p. 450-455.
90. McHaourab, H.S., et al., *Motion of spin-labeled side chains in T4 lysozyme, correlation with protein structure and dynamics*. Biochemistry, 1996. **35**(24): p. 7692-7704.

91. Amadi, S.T., et al., *Structure, Dynamics, and Substrate-induced Conformational Changes of the Multidrug Transporter EmrE in Liposomes*. Journal of Biological Chemistry, 2010. **285**(34): p. 26710-26718.
92. McHaourab, Hassane S., P.R. Steed, and K. Kazmier, *Toward the Fourth Dimension of Membrane Protein Structure: Insight into Dynamics from Spin-Labeling EPR Spectroscopy*. Structure, 2011. **19**(11): p. 1549-1561.
93. Atkins, P.W., and Kivelson, D., *ESR Linewidths in Solution. II. Analysis of Spin Rotational Relation Data*. Journal of Chemical Physics, 1966. **44**: p. 169-174.
94. Van, S.P., G.B. Birrell, and O.H. Griffith, *RAPID ANISOTROPIC MOTION OF SPIN LABELS - MODELS FOR MOTION AVERAGING OF ESR PARAMETERS*. Journal of Magnetic Resonance, 1974. **15**(3): p. 444-459.
95. Kivelson, D., *Theory of ESR linewidths of free radical*. J. Chem. Phys., 1960. **33**.
96. Lietzow, M.A. and W.L. Hubbell, *Motion of spin label side chains in cellular retinol-binding protein: Correlation with structure and nearest-neighbor interactions in an antiparallel beta-sheet*. Biochemistry, 2004. **43**(11): p. 3137-3151.
97. Columbus, L. and W.L. Hubbell, *A new spin on protein dynamics*. Trends in Biochemical Sciences, 2002. **27**(6): p. 288-295.
98. Columbus, L., et al., *Molecular motion of spin labeled side chains in alpha-helices: Analysis by variation of side chain structure*. Biochemistry, 2001. **40**(13): p. 3828-3846.
99. Altenbach, C., et al., *TRANSMEMBRANE PROTEIN-STRUCTURE - SPIN LABELING OF BACTERIORHODOPSIN MUTANTS*. Science, 1990. **248**(4959): p. 1088-1092.
100. Jeschke, G. and Y. Polyhach, *Distance measurements on spin-labelled biomacromolecules by pulsed electron paramagnetic resonance*. Physical Chemistry Chemical Physics, 2007. **9**(16): p. 1895-1910.

101. Schiemann, O. and T.F. Prisner, *Long-range distance determinations in biomacromolecules by EPR spectroscopy*. Quarterly Reviews of Biophysics, 2007. **40**(1): p. 1-53.
102. Jeschke, G., et al., *Direct Conversion of EPR Dipolar Time Evolution Data to Distance Distributions*. Journal of Magnetic Resonance, 2002. **155**(1): p. 72-82.
103. Pannier, M., et al., *Dead-Time Free Measurement of Dipole–Dipole Interactions between Electron Spins*. Journal of Magnetic Resonance, 2000. **142**(2): p. 331-340.
104. Polyhach, Y., et al., *Spin pair geometry revealed by high-field DEER in the presence of conformational distributions*. Journal of Magnetic Resonance, 2007. **185**(1): p. 118-129.
105. Chiang, Y.-W., P.P. Borbat, and J.H. Freed, *The determination of pair distance distributions by pulsed ESR using Tikhonov regularization*. Journal of Magnetic Resonance, 2005. **172**(2): p. 279-295.
106. Chiang, Y.-W., P.P. Borbat, and J.H. Freed, *Maximum entropy: A complement to Tikhonov regularization for determination of pair distance distributions by pulsed ESR*. Journal of Magnetic Resonance, 2005. **177**(2): p. 184-196.
107. Banham, J.E., et al., *Distance measurements in the borderline region of applicability of CW EPR and DEER: A model study on a homologous series of spin-labelled peptides*. Journal of Magnetic Resonance, 2008. **191**(2): p. 202-218.
108. Grote, M., et al., *Transmembrane Signaling in the Maltose ABC Transporter MalFGK2-E: PERIPLASMIC MalF-P2 LOOP COMMUNICATES SUBSTRATE AVAILABILITY TO THE ATP-BOUND MalK DIMER*. Journal of Biological Chemistry, 2009. **284**(26): p. 17521-17526.
109. Jeschke, G., *DEER Distance Measurements on Proteins*. Annual Review of Physical Chemistry, 2012. **63**(1): p. 419-446.

110. López, C.J., et al., *Conformational selection and adaptation to ligand binding in T4 lysozyme cavity mutants*. Proceedings of the National Academy of Sciences of the United States of America, 2013. **110**(46): p. E4306-E4315.
111. Braun, P., et al., *Mechanism of Multivalent Carbohydrate–Protein Interactions Studied by EPR Spectroscopy*. Angewandte Chemie International Edition, 2011. **50**(36): p. 8428-8431.
112. Torbeev, V.Y., et al., *Dynamics of “Flap” Structures in Three HIV-1 Protease/Inhibitor Complexes Probed by Total Chemical Synthesis and Pulse-EPR Spectroscopy*. Journal of the American Chemical Society, 2009. **131**(3): p. 884-885.
113. Hilger, D., et al., *Assessing Oligomerization of Membrane Proteins by Four-Pulse DEER: pH-Dependent Dimerization of NhaA Na(+)/H(+) Antiporter of E. coli*. Biophysical Journal, 2005. **89**(2): p. 1328-1338.
114. Bleicken, S., et al., *Structural Model of Active Bax at the Membrane*. Molecular Cell, 2014. **56**(4): p. 496-505.
115. Dockter, C., et al., *Rigid Core and Flexible Terminus: STRUCTURE OF SOLUBILIZED LIGHT-HARVESTING CHLOROPHYLL a/b COMPLEX (LHCII) MEASURED BY EPR*. Journal of Biological Chemistry, 2012. **287**(4): p. 2915-2925.
116. Georgieva, E.R., et al., *Conformational ensemble of the sodium coupled aspartate transporter*. Nature structural & molecular biology, 2013. **20**(2): p. 215-221.
117. Lovett, J.E., et al., *Investigating the structure of the factor B vWF-A domain/CD55 protein–protein complex using DEER spectroscopy: successes and pitfalls*. Molecular Physics, 2013. **111**(18-19): p. 2865-2872.
118. Piotr G. Fajer, L.B.a.L.S., *Practical Pulsed Dipolar ESR (DEER)*. Vol. Biological Magnetic Resonance. Boston, MA: Springer.

119. Martin, J.N., et al., *Neisserial Opa Protein-CEACAM Interactions: Competition for Receptors as a Means of Bacterial Invasion and Pathogenesis*. *Biochemistry*, 2016. **55**(31): p. 4286-4294.
120. Kroncke, B.M. and L. Columbus, *Identification and removal of nitroxide spin label contaminant: Impact on PRE studies of  $\alpha$ -helical membrane proteins in detergent*. *Protein Science : A Publication of the Protein Society*, 2012. **21**(4): p. 589-595.
121. Kroncke, B.M., P.S. Horanyi, and L. Columbus, *Structural Origins of Nitroxide Side Chain Dynamics on Membrane Protein  $\alpha$ -Helical Sites*. *Biochemistry*, 2010. **49**(47): p. 10045-10060.
122. Do Cao, M.-A., et al., *Probing the conformation of the resting state of a bacterial multidrug ABC transporter, BmrA, by a site-directed spin labeling approach*. *Protein Science : A Publication of the Protein Society*, 2009. **18**(7): p. 1507-1520.
123. Gross, A., et al., *Structure of the KcsA potassium channel from *Streptomyces lividans*: a site-directed spin labeling study of the second transmembrane segment*. *Biochemistry*, 1999. **38**(32): p. 10324-35.
124. Dewald, Alison H., Jacqueline C. Hodges, and L. Columbus, *Physical Determinants of  $\beta$ -Barrel Membrane Protein Folding in Lipid Vesicles*. *Biophysical Journal*, 2011. **100**(9): p. 2131-2140.
125. Needham, D., T.J. McIntosh, and D.D. Lasic, *Repulsive interactions and mechanical stability of polymer-grafted lipid membranes*. *Biochim Biophys Acta*, 1992. **1108**(1): p. 40-8.
126. Immordino, M.L., F. Dosio, and L. Cattel, *Stealth liposomes: review of the basic science, rationale, and clinical applications, existing and potential*. *International Journal of Nanomedicine*, 2006. **1**(3): p. 297-315.

127. Redfield, A.G., *The Theory of Relaxation Processes*\* \*This work was started while the author was at Harvard University, and was then partially supported by Joint Services Contract N5ori-76, Project Order I, in *Advances in Magnetic and Optical Resonance*, J.S. Waugh, Editor. 1965, Academic Press. p. 1-32.
128. Ding, F., M. Layten, and C. Simmerling, *Solution Structure of HIV-1 Protease Flaps Probed by Comparison of Molecular Dynamics Simulation Ensembles and EPR Experiments*. *Journal of the American Chemical Society*, 2008. **130**(23): p. 7184-7185.
129. Galiano, L., et al., *Drug Pressure Selected Mutations in HIV-1 Protease Alter Flap Conformations*. *Journal of the American Chemical Society*, 2009. **131**(2): p. 430-431.
130. Roux, B. and S.M. Islam, *Restrained-Ensemble Molecular Dynamics Simulations Based on Distance Histograms from Double Electron-Electron Resonance Spectroscopy*. *The journal of physical chemistry. B*, 2013. **117**(17): p. 4733-4739.
131. Tessmer, M.H., et al., *Identification of a ubiquitin-binding interface using Rosetta and DEER*. *Proceedings of the National Academy of Sciences*, 2018. **115**(3): p. 525-530.
132. Hirst, S.J., et al., *RosettaEPR: An integrated tool for protein structure determination from sparse EPR data*. *Journal of Structural Biology*, 2011. **173**(3): p. 506-514.
133. Islam, S.M., et al., *Structural Refinement from Restrained-Ensemble Simulations Based on EPR/DEER Data: Application to T4 Lysozyme*. *The Journal of Physical Chemistry B*, 2013. **117**(17): p. 4740-4754.
134. Peng, H.C., F.H. Long, and C. Ding, *Feature selection based on mutual information: Criteria of max-dependency, max-relevance, and min-redundancy*. *Ieee Transactions on Pattern Analysis and Machine Intelligence*, 2005. **27**(8): p. 1226-1238.

135. Oikawa, S., et al., *A specific heterotypic cell adhesion activity between members of carcinoembryonic antigen family, W272 and NCA, is mediated by N-domains*. Journal of Biological Chemistry, 1991. **266**(13): p. 7995-8001.
136. Watt, S.M., et al., *Homophilic adhesion of human CEACAM1 involves N-terminal domain interactions: structural analysis of the binding site*. Blood, 2001. **98**(5): p. 1469-1479.
137. Heldin, C.-H., *Dimerization of cell surface receptors in signal transduction*. Cell, 1995. **80**(2): p. 213-223.
138. Wu, H., P.D. Kwong, and W.A. Hendrickson, *Dimeric association and segmental variability in the structure of human CD4*. Nature, 1997. **387**: p. 527.
139. Yu, X.-L., et al., *Crystal Structure of HAb18G/CD147: IMPLICATIONS FOR IMMUNOGLOBULIN SUPERFAMILY HOMOPHILIC ADHESION*. Journal of Biological Chemistry, 2008. **283**(26): p. 18056-18065.
140. Tchoupa, A.K., T. Schuhmacher, and C.R. Hauck, *Signaling by epithelial members of the CEACAM family – mucosal docking sites for pathogenic bacteria*. Cell Communication and Signaling, 2014. **12**(1): p. 27.
141. Huber, M., et al., *The Carboxyl-terminal Region of Biliary Glycoprotein Controls Its Tyrosine Phosphorylation and Association with Protein-tyrosine Phosphatases SHP-1 and SHP-2 in Epithelial Cells*. Journal of Biological Chemistry, 1999. **274**(1): p. 335-344.
142. Müller, M.M., et al., *Homophilic adhesion and CEACAM1-S regulate dimerization of CEACAM1-L and recruitment of SHP-2 and c-Src*. The Journal of Cell Biology, 2009. **187**(4): p. 569-581.
143. Zhou, H., et al., *Homophilic adhesion between Ig superfamily carcinoembryonic antigen molecules involves double reciprocal bonds*. The Journal of Cell Biology, 1993. **122**(4): p. 951-960.

144. Klaile, E., et al., *The CEACAM1 N-terminal Ig domain mediates cis- and trans-binding and is essential for allosteric rearrangements of CEACAM1 microclusters*. *The Journal of Cell Biology*, 2009. **187**(4): p. 553-567.
145. Patel, P.C., et al., *Inside-out Signaling Promotes Dynamic Changes in the Carcinoembryonic Antigen-related Cellular Adhesion Molecule 1 (CEACAM1) Oligomeric State to Control Its Cell Adhesion Properties*. *Journal of Biological Chemistry*, 2013. **288**(41): p. 29654-29669.
146. Lawson, E.L., et al., *The Transmembrane Domain of CEACAM1-4S Is a Determinant of Anchorage Independent Growth and Tumorigenicity*. *PLOS ONE*, 2012. **7**(1): p. e29606.
147. Russ, W.P. and D.M. Engelman, *The GxxxG motif: A framework for transmembrane helix-helix association*<sup>11</sup>Edited by G. von Heijne. *Journal of Molecular Biology*, 2000. **296**(3): p. 911-919.
148. Singer, B.B., et al., *Deregulation of the CEACAM Expression Pattern Causes Undifferentiated Cell Growth in Human Lung Adenocarcinoma Cells*. *PLOS ONE*, 2010. **5**(1): p. e8747.
149. Singer, B.B., I. Scheffrahn, and B. Öbrink, *The Tumor Growth-inhibiting Cell Adhesion Molecule CEACAM1 (C-CAM) Is Differently Expressed in Proliferating and Quiescent Epithelial Cells and Regulates Cell Proliferation*. *Cancer Research*, 2000. **60**(5): p. 1236-1244.
150. Beauchemin, N. and A. Arabzadeh, *Carcinoembryonic antigen-related cell adhesion molecules (CEACAMs) in cancer progression and metastasis*. *Cancer and Metastasis Reviews*, 2013. **32**(3): p. 643-671.
151. Bonsor, D.A., et al., *Diverse oligomeric states of CEACAM IgV domains*. *Proceedings of the National Academy of Sciences*, 2015. **112**(44): p. 13561-13566.

152. Abdul-Wahid, A., et al., *The carcinoembryonic antigen IgV-like N domain plays a critical role in the implantation of metastatic tumor cells*. *Molecular Oncology*, 2014. **8**(2): p. 337-350.
153. Huang, Y.-H., et al., *CEACAM1 regulates TIM-3-mediated tolerance and exhaustion*. *Nature*, 2014. **517**: p. 386.
154. Dym, O., et al., *The impact of crystallization conditions on structure-based drug design: A case study on the methylene blue/acetylcholinesterase complex*. *Protein Sci*, 2016. **25**(6): p. 1096-114.
155. Zhang, X.-j., J.A. Wozniak, and B.W. Matthews, *Protein Flexibility and Adaptability Seen in 25 Crystal Forms of T4 Lysozyme*. *Journal of Molecular Biology*, 1995. **250**(4): p. 527-552.
156. Gibson, D.G., et al., *Enzymatic assembly of DNA molecules up to several hundred kilobases*. *Nature Methods*, 2009. **6**: p. 343.
157. Zhuo, Y., et al., *Glycosylation Alters Dimerization Properties of a Cell-surface Signaling Protein, Carcinoembryonic Antigen-related Cell Adhesion Molecule 1 (CEACAM1)*. *Journal of Biological Chemistry*, 2016. **291**(38): p. 20085-20095.
158. Kapust, R.B. and D.S. Waugh, *Escherichia coli maltose-binding protein is uncommonly effective at promoting the solubility of polypeptides to which it is fused*. *Protein Science : A Publication of the Protein Society*, 1999. **8**(8): p. 1668-1674.
159. Bach, H., et al., *Escherichia coli maltose-binding protein as a molecular chaperone for recombinant intracellular cytoplasmic single-chain antibodies*. *J Mol Biol*, 2001. **312**(1): p. 79-93.
160. Schwarz, F. and M. Aebi, *Mechanisms and principles of N-linked protein glycosylation*. *Current Opinion in Structural Biology*, 2011. **21**(5): p. 576-582.

161. Marshall, R.D., *Glycoproteins*. Annu Rev Biochem, 1972. **41**: p. 673-702.
162. Shental-Bechor, D. and Y. Levy, *Effect of glycosylation on protein folding: a close look at thermodynamic stabilization*. Proc Natl Acad Sci U S A, 2008. **105**(24): p. 8256-61.
163. Hanson, S.R., et al., *The core trisaccharide of an N-linked glycoprotein intrinsically accelerates folding and enhances stability*. Proceedings of the National Academy of Sciences of the United States of America, 2009. **106**(9): p. 3131-3136.
164. Culyba, E.K., et al., *Protein native-state stabilization by placing aromatic side chains in N-glycosylated reverse turns*. Science, 2011. **331**(6017): p. 571-5.
165. Skropeta, D., *The effect of individual N-glycans on enzyme activity*. Bioorg Med Chem, 2009. **17**(7): p. 2645-53.
166. Helenius, A. and M. Aebi, *Roles of N-linked glycans in the endoplasmic reticulum*. Annu Rev Biochem, 2004. **73**: p. 1019-49.
167. Aebi, M., et al., *N-glycan structures: recognition and processing in the ER*. Trends Biochem Sci, 2010. **35**(2): p. 74-82.
168. Shi, X.Z. and D.L. Jarvis, *Protein N-glycosylation in the baculovirus-insect cell system*. Current Drug Targets, 2007. **8**(10): p. 1116-1125.
169. Reeves, P.J., et al., *Structure and function in rhodopsin: high-level expression of rhodopsin with restricted and homogeneous N-glycosylation by a tetracycline-inducible N-acetylglucosaminyltransferase I-negative HEK293S stable mammalian cell line*. Proc Natl Acad Sci U S A, 2002. **99**(21): p. 13419-24.
170. von Schaewen, A., et al., *Isolation of a mutant Arabidopsis plant that lacks N-acetylglucosaminyl transferase I and is unable to synthesize Golgi-modified complex N-linked glycans*. Plant Physiol, 1993. **102**(4): p. 1109-18.

171. Gomez, L. and M.J. Chrispeels, *Complementation of an Arabidopsis thaliana mutant that lacks complex asparagine-linked glycans with the human cDNA encoding N-acetylglucosaminyltransferase I*. Proc Natl Acad Sci U S A, 1994. **91**(5): p. 1829-33.
172. Behrens, A.J., et al., *Integrity of Glycosylation Processing of a Glycan-Depleted Trimeric HIV-1 Immunogen Targeting Key B-Cell Lineages*. Journal of Proteome Research, 2018. **17**(3): p. 987-999.
173. Medina-Ramirez, M., et al., *Design and crystal structure of a native-like HIV-1 envelope trimer that engages multiple broadly neutralizing antibody precursors in vivo*. Journal of Experimental Medicine, 2017. **214**(9): p. 2573-2590.
174. Ward, A.B. and I.A. Wilson, *The HIV-1 envelope glycoprotein structure: nailing down a moving target*. Immunological Reviews, 2017. **275**(1): p. 21-32.
175. Apweiler, R., H. Hermjakob, and N. Sharon, *On the frequency of protein glycosylation, as deduced from analysis of the SWISS-PROT database*. Biochim Biophys Acta, 1999. **1473**(1): p. 4-8.
176. Leung, N., et al., *Intestinal tumor progression is promoted by decreased apoptosis and dysregulated Wnt signaling in Ceacam1<sup>-/-</sup> mice*. Oncogene, 2008. **27**: p. 4943.
177. Blake, M.S., et al., *Gonococcal opacity: lectin-like interactions between Opa proteins and lipooligosaccharide*. Infect Immun, 1995. **63**(4): p. 1434-9.
178. Plant, L., et al., *Lipooligosaccharide Structure Contributes to Multiple Steps in the Virulence of Neisseria meningitidis*. Infection and Immunity, 2006. **74**(2): p. 1360-1367.
179. Mulvey, M.A., *Adhesion and entry of uropathogenic Escherichia coli*. Cellular Microbiology, 2002. **4**(5): p. 257-271.
180. Leusch, H.G., et al., *Escherichia coli of human origin binds to carcinoembryonic antigen (CEA) and non-specific crossreacting antigen (NCA)*. FEBS Lett, 1990. **261**(2): p. 405-9.

181. Korotkova, N., et al., *The Dr Family of E.coli Adhesins bind independently to the Decay-Accelerating Factor and the N-domain of Carcinoembryonic Antigen*. The Journal of biological chemistry, 2006. **281**(39): p. 29120-29130.
182. Dewald, A.H., J.C. Hodges, and L. Columbus, *Physical Determinants of beta-Barrel Membrane Protein Folding in Lipid Vesicles*. Biophysical Journal, 2011. **100**(9): p. 2131-2140.
183. Fox, D.A. and L. Columbus, *Solution NMR resonance assignment strategies for -barrel membrane proteins*. Protein Science, 2013. **22**(8): p. 1133-1140.
184. Pannier, M., et al., *Dead-time free measurement of dipole-dipole interactions between electron spins*. Journal of Magnetic Resonance, 2000. **142**(2): p. 331-340.
185. Jeschke, G., et al., *DeerAnalysis2006 - a comprehensive software package for analyzing pulsed ELDOR data*. Applied Magnetic Resonance, 2006. **30**(3-4): p. 473-498.

## Materials and Methods

### *Expression, purification, spin-labeling, and refolding of Opa<sub>60</sub> proteins into micelles*

Opa<sub>60</sub> expression, purification, and refolding protocols were previously published [53, 119, 182, 183]. Briefly, the *opa60* gene was sub-cloned into a pET28b vector (EMD chemicals, Gibbstown, NJ) containing N and C terminal His<sub>6</sub>- tags. Cysteine residues were introduced at regions of interest on Opa using PIPE Mutagenesis, and gene sequencing confirmed the mutations (Genewiz Inc., South Plainfield, NJ). The pET28b vectors containing a mutated *opa60* gene were transformed into BL21(DE3) *E. coli* cells, and cultures were grown in Luria-Burtani (LB) media. Opa protein expression to inclusion bodies was induced with 1 mM isopropyl- $\beta$ -thio-D-galactoside (IPTG). Cells were harvested and resuspended in lysis buffer [50 mM Tris-HCl, pH 8.0, 150 mM NaCl, and 1 mM TCEP-HCl (tris(2-carboxyethyl)phosphine hydrochloride)]. Following cell lysis, insoluble fractions were pelleted and resuspended overnight with lysis buffer containing 8 M urea. Cell debris was removed via centrifugation and unfolded Opa proteins in the soluble fraction were purified using Co<sup>2+</sup> immobilized metal affinity chromatography, eluting in 20 mM sodium phosphate, pH 7.0, 150 mM NaCl, 680 mM imidazole, 8 M urea, and 1 mM TCEP. Purified Opa proteins were loaded on a PD-10 column (GE Healthcare Biosciences, Pittsburg, PA) to remove TCEP. Opa proteins were eluted with buffer (20 mM sodium phosphate, pH 7.0, 150 mM NaCl, and 8M urea) directly into five molar excess MTSL/R1 spin label [S-(2, 2, 5, 5-tetramethyl-2,5-dihydro-1H-pyrrol-3-yl)methyl methanesulfonothioate, Toronto Research Chemicals Inc., Toronto, Canada, stored as 100 mM stock in acetonitrile] for proteins containing a single cysteine and ten molar excess MTSL for Opa proteins with two cysteine residues. The proteins were spin labeled overnight at room temperature.

Excess spin label was removed using a second PD-10 column, and the eluted protein was concentrated to approximately 150 - 200  $\mu\text{M}$ . The labeled proteins were rapidly diluted 20-fold into 20 mM Tris-HCl, pH 8.0, 500 mM NaCl, 3 M urea, and 4.6 mM n-dodecylphosphocholine (FC-12, Anatrace), upon which Opa proteins fold into the detergent micelles over the course of three days at room temperature [53, 183]. Folding efficiency was assessed with SDS-PAGE. Samples were dialyzed against 3 x 4L of 20 mM sodium phosphate, 150 mM NaCl for an hour each, removing any free spin. Opa proteins were concentrated to approximately 200 – 400  $\mu\text{M}$ .

*Spin labeling and folding Opa proteins into liposomes*

Opa<sub>60</sub> protein refolding into liposomes was adapted from Dewald, 2011. Lipid chloroform stocks of 1,2-didecanoyl-*sn*-glycero-3-phosphocholine (diC<sub>10</sub>PC, Avanti Polar Lipids) were dried under a continuous stream of nitrogen and resuspended in borate buffer (10 mM sodium borate, pH 12, 1 mM EDTA). Liposomes were formed via 30 minutes of sonication (Q Sonica model Q500, Newtown, CT) with an 1/8 microtip at 40% amplitude. 4 M urea was added to the lipids, followed by Opa proteins in 20  $\mu\text{L}$  aliquots with constant stirring, yielding a final protein:lipid ratio of 1:1160. Proteoliposomes were incubated at 37°C for three days and folding was assessed via SDS-PAGE. Opa proteoliposomes are harvested using ultracentrifugation (142,400 x g, 2 hr, 10°C), resuspended in a new lipid mixture, and pulse sonicated (30 seconds on, 30 seconds off, 20 min). The new lipid composition contained 63 mol% 1,2-dimyristoyl-*sn*-glycero-3-phosphocholine (DMPC), 16 mol% 1,2-dimyristoyl-*sn*-glycero-3-phospho-(1'-*rac*-glycerol) sodium salt (DMPG), 16 mol% cholesterol, 5 mol% 1,2-dimyristoyl-*sn*-glycero-3-phosphoethanolamine-N-[methoxy(polyethylene glycol)-1000] ammonium salt (DMPE-PEG 1000), with a

protein:lipid ratio of 1:234. Folded Opa proteins were then spin labeled using the same protocol used above.

*Expression, purification, and spin-labeling of the non-glycosylated N-terminal domain CEACAM1 proteins*

The expression and purification protocol of the N-terminal domain of CEACAM1 has been published previously (Martin, 2016) and was adapted from Fedarovich et al. A pGEX-2T vector containing the N-terminal D1 domain of the human *ceacam1* gene (amino acids 35-141 of the full length protein, NCEACAM1) was generously provided by Rob Nicholas (University of North Carolina at Chapel Hill). Cysteine residues were introduced at specific locations on NCCM1 using PIPE Mutagenesis. Mutations were confirmed using gene sequencing (Genewiz Inc, South Plainfield, NJ). Plasmids containing the mutated *ceacam1* gene were transformed into MC1061 *E. coli* cells, which were grown in LB media to an OD<sub>600</sub> of 0.6. Protein expression was induced with 1 mM IPTG overnight. Cells were harvested, resuspended in lysis buffer (20 mM Tris, pH 8.0, 150 mM NaCl, 2 mM ethylenediaminetetraacetic acid (EDTA), 2 mM dithiothreitol (DTT), and 10% glycerol), and lysed. Cell debris was removed via centrifugation, and the addition of ammonium sulfate to 55% precipitated proteins within the supernatant. These proteins were pelleted and resuspended in 20 mM Tris pH 7.3, 150 mM NaCl, 2 mM DTT, and 10% glycerol at 4°C. NCCM1 was purified using a glutathione resin column at 4°C, eluting in 20 mM Tris, pH 7.3, 150 mM NaCl, 2 mM DTT, 10% glycerol, and 10 mM reduced glutathione. The GST tag was cleaved from the N-terminal domain of CEACAM using tobacco etch virus (TEV) protease (at approximately 3.5 μM), which was added to the eluent and dialyzed overnight at 4°C against 20 mM Tris pH 7.3, 150 mM NaCl, 10% glycerol, and 2 mM DTT. NCCM1 was isolated from TEV and cleaved GST using a HR Sephacryl S-200 Gel

Filtration column (26/60 mm, GE Healthcare) equilibrated with 20 mM Tris, pH 8.0, 500 mM NaCl, 10% glycerol, and 2 mM DTT. SDS-PAGE was utilized to analyze NCCM1 purity within the eluted fractions. Purified NCEACAM1 fractions were combined and passed through a PD10 column at 4°C to remove DTT. NCEACAM1 was eluted with buffer (20 mM sodium phosphate, 150 mM NaCl, and 10% glycerol) directly into five molar excess MTSL/R1 [S-(2, 2, 5, 5-tetramethyl-2,5-dihydro-1H-pyrrol-3-yl)methyl methanesulfonothioate, Toronto Research Chemicals Inc., Toronto, Canada, stored as 100 mM stock in acetonitrile], and NCEACAM1 was spin labeled overnight in the dark at 4°C. Excess spin label was removed with a second PD10 column, and NCEACAM1 was concentrated to approximately 100 µM-200 µM.

*Expression, purification, and spin-labeling of the N-terminal domain CEACAM1-GlcNAc proteins*

An expression and purification protocol for glycosylated N-CEACAM1 proteins has been adapted from Zhuo, 2016. HEK293S GnT1<sup>-</sup> cells (which express primarily (Man)<sub>5</sub> – (GlcNAc)<sub>2</sub> glycans) were generously provided by Kelley Mormen (University of Georgia) and were maintained using FreeStyle™ 293 expression medium (Thermo Fisher Scientific) in a humidified CO<sub>2</sub> platform shaker incubator at 37°C. Kelley Moremen also provided a pGen2 expression vector containing a N-terminal sequence that signals protein secretion into the medium, followed by a His<sub>8</sub> tag, an AviTag, green fluorescent protein (GFP), a TEV cleavage sequence, and the human *ceacam1* gene encoding the N-terminal domain of CEACAM1 (residues 34-141, UniProt P13688). Cysteine residues were introduced into *ceacam1* using PIPE Mutagenesis, and gene sequencing confirmed the mutations (Genewiz Inc., South Plainfield, NJ). A 250 mL suspension culture of HEK293S (GnT1<sup>-</sup>) cells in a 9:1 ratio of FreeStyle<sup>T</sup> 293 media and EX-CELL media (Sigma) was transfected with the

NCEACAM1-pGEn2 plasmid using polyethyleneimine (Polysciences, Inc., Warrington, PA) as described previously (Zhuo, 2016). After incubating for 24 hours, 250 mL of a 9:1 ratio of FreeStyle<sup>T</sup> 293 EX-CELL medias and 2.2 mM valproic acid (Sigma) were added to the suspension culture. Glycosylated NCEACAM1 proteins were produced over the course of five days at 37°C, after which cell debris was removed via centrifugation (20 min, 150 x g, 4°C). Glycosylated NCEACAM1 proteins were purified from the supernatant via Co<sup>2+</sup> immobilized metal affinity chromatography (IMAC), eluting in ten column volumes of elution buffer [25 mM HEPES, 300 mM NaCl, 680 mM imidazole, pH 7.0] at 4°C. The eluent was dialyzed into 4L of 25 mM HEPES, 300 mM NaCl, 10% glycerol, pH 7.0 containing approximately 3.5 μM TEV and endoglycosidase F1 (both enzymes were expressed in BL21(DE3) *E. coli* and purified via Co<sup>2+</sup> IMAC), removing the GFP tag from NCEACAM1 and truncating the glycans to single GlcNAc residues on NCEACAM1. Excess GFP is removed using Co<sup>2+</sup> IMAC, where the flow-through containing CEACAM proteins is collected. NCEACAM1 is further purified from GFP, TEV, and EndoF1 using a HR Sephacryl S-200 Gel Filtration column (26/60 mm, GE Healthcare) equilibrated with 20 mM HEPES, pH 7.0, 150 mM NaCl, 10% glycerol, and 2 mM DTT. Fractions containing pure NCEACAM1 (assessed via SDS-PAGE) are concentrated and run through a PD-10 column to remove DTT. NCEACAM1 proteins were eluted using buffer [20 mM sodium phosphate, 150 mM NaCl, and 10% glycerol] directly into five molar excess MTSL. NCEACAM1 was spin labeled overnight in the dark at 4°C, after which excess spin label was removed with a second PD10 column. The glycosylated NCEACAM1 proteins were concentrated to approximately 100 μM-200 μM.

*Expression, purification, and spin-labeling of the N-terminal domain CEACAM1 highly glycosylated proteins*

An expression and purification protocol for highly glycosylated N-CEACAM1 proteins has been adapted from Zhuo, 2016. HEK293F cells (which express complex glycans) were generously provided by Kelley Mormen (University of Georgia) and were maintained using FreeStyle™ 293 expression medium in a humidified CO<sub>2</sub> platform shaker incubator at 37°C. A 250 mL suspension culture of HEK293F cells in a 9:1 ratio of FreeStyle<sup>T</sup> 293 media and EX-CELL media was transfected with the NCEACAM1-pGen2 plasmid using polyethyleneimine. After incubating for 24 hours, 250 mL of a 9:1 ratio of FreeStyle<sup>T</sup> 293 EX-CELL medias and 2.2 mM valproic acid were added to the suspension culture. Glycosylated NCEACAM1 proteins were produced over the course of five days at 37°C, after which cell debris was removed via centrifugation (20 min, 150 x g, 4°C ). Glycosylated NCEACAM1 proteins were purified from the supernatant via Co<sup>2+</sup> immobilized metal affinity chromatography (IMAC), eluting in ten column volumes of elution buffer [25 mM HEPES, 300 mM NaCl, 680 mM imidazole, pH 7.0] at 4°C. The eluent was dialyzed into 4L of 25 mM HEPES, 300 mM NaCl, 10% glycerol, pH 7.0 containing approximately 3.5 μM TEV, removing the GFP tag from NCEACAM1. Excess GFP is removed using Co<sup>2+</sup> IMAC, where the flow-through containing CEACAM proteins is collected. NCEACAM1 is further purified from GFP and TEV using a HR Sephacryl S-200 Gel Filtration column (26/60 mm, GE Healthcare) equilibrated with 20 mM HEPES, pH 7.0, 150 mM NaCl, 10% glycerol, and 2 mM DTT. Fractions containing pure NCEACAM1 (assessed via SDS-PAGE) are concentrated and run through a PD-10 column to remove DTT. NCEACAM1 proteins were eluted using buffer [20 mM sodium phosphate, 150 mM NaCl, and 10% glycerol] directly into five molar excess MTSL. NCEACAM1 was spin labeled overnight in the dark at 4°C, after which excess spin label

was removed with a second PD10 column. The glycosylated NCEACAM1 proteins were concentrated to approximately 100  $\mu$ M-200  $\mu$ M.

#### *Continuous wave EPR*

CW-EPR experiments were measured using an X-band Bruker EMX continuous wave spectrometer with an ER4123D dielectric resonator (Bruker Biospin, Billerica, MA) at room temperature. Five  $\mu$ L of each sample (100 – 300  $\mu$ M) was measured in pyrex capillaries (0.6 mm id x 0.84 od, Vitrocom, Mountain Lakes, NJ). Spectra were baseline corrected and normalized using Lab-VIEW software generously provided by C. Altenbach (University of California at Los Angeles).

#### *Double-electron electron resonance spectroscopy*

Double-labeled samples were measured using pulsed EPR with a Q-band Bruker E580 Spectrometer fitted with an ER5106-QT flexline resonator (Bruker Biospin) at 80 K. All samples were prepared to a final protein concentration between approximately 100 and 200  $\mu$ M with 10% deuterated glycerol. The samples were loaded into quartz capillaries with a 1.6 mm od x 1.1 mm id (Vitrocom) and were flash frozen in liquid nitrogen. A four pulse DEER sequence was used with one 16 ns  $\pi/2$ , two 32 ns  $\pi$  observed pulses (at an observed frequency  $\nu_1$ ), and a  $\pi$  pump pulse (at a frequency  $\nu_2$ ) optimized at approximately 32 ns [184]. A pump frequency ( $\nu_2$ ) is set at the maximum of the nitroxide spectrum and the observed frequency ( $\nu_1$ ) is set to 75 MHz lower. Increasing inter-pulse delays at 16 ns increments were utilized with a 16-step phase cycle during data collection. Accumulation times were typically between 18 and 24 hours, with a dipolar evolution time between 2 and 3  $\mu$ s.

Dipolar evolution data were processed using DEERAnalysis2016 software using Tikhonov regularization to generate distance distributions [185]. Automatic zero-time determination is used in DEERAnalysis2016, which is approximated by the time at which the real part of the signal is maximum. Automatic phase correction ensures the signal is entirely from the real portion of the dataset. Points near the end of the dataset are cut, because at longer times the pump pulse begins to interfere with the observe pulse. Background subtraction of the distance distribution yields error at each distance which is plotted as ranges that represent fits that are within 15% root-mean-square-deviation of the best fit. This homogeneous background subtraction is validated by varying the starting time point of background subtraction and the dimensionality of the space (3 for soluble proteins, around 2 for membrane proteins). This intermolecular background is subtracted, generating a signal from only the intramolecular dipolar coupling of interest