Spectroscopic Studies of Glycosylated CEACAM1 Ectodomains and Membrane Protein Folding in Lipid-Detergent Mixtures

Nicole Kay Swope
North Branch, MI

B.S. Saginaw Valley State University, 2015

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, 2020

Abstract

Biomolecular nuclear magnetic resonance (NMR) spectroscopy, small-angle X-ray scattering (SAXS), and electron paramagnetic resonance (EPR) spectroscopy are established methods for determining the structure and dynamics of macromolecules. The work in this dissertation utilized magnetic resonance and SAXS methods to investigate two major research aims that are outlined in the following sections.

The first study describes the human carcinoembryonic antigen-like cellular adhesion molecule 1 (CEACAM1), which facilitates cell-cell adhesion in healthy tissue. Impaired CEACAM1-mediated cellular adhesions result in tumor formation, and CEACAM1 is also targeted by surface proteins of pathogenic bacteria. CEACAM1 is localized to the plasma membrane and contains four highly-glycosylated ectodomains, of which the N-terminal domain is both essential for normal function and specific to pathogen interactions. Investigating the CEACAM1 N-domain is therefore of broad interest for relevant disease states. However, the majority of studies with isolated CEACAM1 N-domain proteins lack glycan modifications that may be important for CEACAM1 homotypic and heterotypic interactions. In this work, EPR spectroscopy was used to investigate four CEACAM1 N-domain glycoforms. All CEACAM1 N-domain proteins form a homodimer in solution, and, contrary to previous literature, glycosylation does not perturb CEACAM1 dimerization. Furthermore, the interaction interface between CEACAM1 N-domains is maintained across all glycoforms. Collectively, glycosylated

CEACAM1 N-domain proteins recapitulate native homotypic interactions, and these results provide a foundation for understanding the role of CEACAM1 glycans *in vivo*.

The second major aim of this dissertation investigated a subclass of membrane mimics used to solubilize membrane proteins. Approximately 50% of drugs on the market target membrane proteins, and they are often isolated outside of the native membrane for structural and functional studies. Membrane proteins are commonly solubilized in detergent micelles, but lipid-detergent mixtures, or bicelles, represent another water-soluble amphipathic assembly. The ideal bicelle morphology contains a phase-separated detergent rim and lipid core, which is meant to provide a bilayer-like environment for an embedded membrane protein. This work described the morphologies of eleven lipid-detergent compositions without protein using SAXS. For each investigated mixture, evidence of internal lipid organization occurred when the ratio of lipid to detergent was at least 1:2. Results suggest that bicelles undergo a composition-independent phase transition that requires a minimum number of lipids, and the detergent component modulates bicelle size. To assess the impact of bicelle phase transitions and size in a system containing protein, a model, polytopic α-helical transmembrane protein was prepared in bicelles characterized with SAXS. The overall protein fold and tertiary helical contacts, evaluated with NMR and EPR spectroscopy, demonstrated that bicelles stabilize the protein fold in a unique manner from detergent micelles. These results suggest that bicelle lipid and detergent molecules may rearrange to stabilize protein folds through preferential lipid solvation and/or alleviating hydrophobic mismatch.

Dedication Page

To my advisor, Linda Columbus, for helping me become a confident, capable scientist and allowing me to grow as a person. I would not be the woman I am today without her.

To all current and past members of the Columbus laboratory for tremendous support and friendship over the years. All of them are wonderful both at the bench and outside of the laboratory, and I could not have asked for a better group of people to work alongside every day. For mentorship, I would especially like to thank Dr. Meagan Belcher Dufrisne, Dr. Marissa Kieber, Dr. Jason Kuhn, Dr. Steven Keller, Dr. Jennifer Martin, and Dr. Ashton Brock. I would also like to thank my undergraduate mentee, Katherine Lake, for her fantastic work on the phosphoglucose isomerase project. She amazes and impresses me, always. Finally, I would like to recognize Tracy Caldwell for her constant encouragement and being a solid support system throughout my graduate career. I would not have survived without her.

To the Department of Chemistry staff, especially Dr. Carol Price, for being both a mentor and friend. I would also like to thank Dr. Jeff Ellena for his training and continued guidance.

To my partner, friends, family, and kitties, for being a constant source of love and laughter. I am so lucky to have them in my life.

Table of Contents

List of Figures

List of Tables

Chapter 1: Introduction

1.1 Protein structure, dynamics, and function

When hearing the word "protein" as a child, I always thought of the stuff that makes our body strong. Protein equaled meat, mostly. My little self could have never fathomed how ubiquitous proteins are, nor the exquisite processes that require proteins. Functions of proteins are vast, reflecting the many biological processes that occur within or between cells. Proteins are involved in cell adhesion to form new tissue, transport of ions and vitamins across the cell membrane, catalyzing chemical reactions, triggering cell death, protecting the body against pathogens, and more. The central dogma of molecular biology – transcription of DNA into RNA, followed by translation of RNA by the ribosome for protein synthesis – is only the beginning.

1.1.1 Protein structure

The ribosome translates RNA to build sequences of amino acids that form a polypeptide chain. This sequence of amino acids comprises a protein's primary structure. The chemical structure of amino acids allows the ribosome to catalyze a condensation reaction between the carboxyl group of one amino acid and the amine group of another, forming a peptide bond (Figure 1.1A). Together, N, C_α, and C´ form the backbone atoms of the polypeptide chain. Torsion angles about the C_α-N (φ) and C_α-C´ (ψ) bonds define the polypeptide backbone conformation, while the C´-N peptide bond is rigid. Amino acids possess side chains with functional groups that allow hydrogen bonding, van der Waals

interactions, electrostatic interactions, acid/base chemistry, and varying degrees of flexibility, all of which play roles in protein folding, dynamics, and function.

The linear sequence of amino acids ultimately gives rise to hydrogen bonding networks along the polypeptide backbone. The spatial arrangement of the backbone resulting from hydrogen bonds, excluding contributions from amino acid side chains, defines a protein's secondary structure. Secondary structural elements common to proteins include α-helices and β-sheets (Figure 1.1B), which are defined within specific distributions of φ and ψ backbone torsion angles that are dictated by van der Waals radii. Allowable configurations of φ and ψ angles that lead to α-helical and β-sheet formation are summarized well by a Ramachandran plot (Figure 1.2) [1]. Loops, although not organized structural protein elements, often serve to connect β-strands and α-helices as well as provide flexibility within the protein backbone.

Beyond α-helices, β-sheets, and loops, a protein adopts a unique folded structure in order to carry out its intended function. Organized secondary structural elements, such as α-helical bundles and β-barrels, often define the tertiary structure of a protein. The amino sequence of a soluble, globular protein often contains residues with both hydrophobic and hydrophilic side chains. A soluble protein must adopt a folded structure such that its core contains hydrophobic amino acid residues, to minimize interactions with water, while the surface of the protein contains hydrophilic residues. The partitioning of water from the hydrophobic core of the protein while maintaining favorable surface interactions with aqueous solvent contributes stability to the overall protein fold and spatial arrangement of secondary structural elements [2]. Finally, quaternary structure is defined by the assembly

of more than one polypeptide chain, or domain, that arises from a combination of covalent (disulfide bonds) and noncovalent interactions. The majority of this thesis is dedicated to understanding the stability of a final protein's fold, how a protein's fold dictates assembly with identical subunits, and the functional elements within a protein that perform catalytic reactions.

1.1.2 Protein dynamics

Proteins do not possess static structures; they undergo fluctuations in motion that can generally be categorized as "local" or "collective" (Figure 1.3) [3]. Local protein fluctuations occur approximately on the femtosecond to nanosecond timescales, while collective motions occur on the microsecond to second timescales. The fastest fluctuations occurring within a protein, as within any chemical structure, are vibrations between chemical bonds (femtosecond). Backbone fluctuations about C_α-N and C_α- C´ bonds and side-chain rotamer fluctuations occur in the picosecond to nanosecond regime. Additional examples of motions that occur on the nanosecond timescale are unstructured protein elements, with conformational rearrangement of a protein's secondary structural elements occurring on the micro- to millisecond timescales.

Protein fluctuations can be probed by numerous, complementary biophysical techniques such as nuclear magnetic resonance spectroscopy (NMR), electron paramagnetic resonance spectroscopy (EPR), small-angle X-ray scattering (SAXS), and molecular dynamics (MD) simulations. Although successful X-ray crystallography experiments rely on a homogenous crystal that "traps" a protein in a specific conformation,

evidence of protein dynamics can also be observed from crystal structures. For example, protein crystal structures obtained with and without a ligand often show that binding events facilitate protein structural rearrangements. Advances in time-resolved crystallography methods have facilitated direct protein measurements of protein dynamics on the order of hundreds of picoseconds [4], though the technique has not been widely adopted. While routine crystallography methods cannot capture protein motion as it occurs, structures of protein elements that undergo large-scale, dynamic motions offers a high-resolution complementary approach to solution spectroscopy and scattering methods [5].

1.1.3 Protein function

Protein structure and dynamics are intimately related with protein function, as the former elements determine the fate of binding events, catalysis, transport, and more. Logic may lead the reader to conclude that each protein is manufactured by the cell to carry out a unique function. However, a single protein may possess multiple functions, as is the case with "moonlighting" proteins [6]. Chapter 5 of this dissertation is dedicated to characterization of a heat-stable phosphoglucose isomerase, whose homologues serve additional, localization-dependent functions such as neurotrophic, motility, and differentiation factors in eukaryotic organisms [7]. These moonlighting proteins provide a mechanism for efficiency in cells, as one protein is able to carry out a variety of necessary functions.

Due to the important relationship between protein structure and function, especially for structural perturbations that cause disease, determining protein structures remains a

priority in the biophysical community. To date, there are approximately 150,000 protein structures deposited in the Protein Data Bank (PDB), obtained by X-ray crystallography, cryo-electron microscopy, and NMR methods. Most protein structures deposited in the PDB are globular and soluble in aqueous solvent. However, an entire class of proteins has struggled to see the same boon as globular protein structure determination – those that are peripheral to or embedded in cell membranes. Aptly called “membrane proteins”, these proteins, encoded by approximately 30% of the genome and targeted by 50% of drugs on the market, represent less than 1% of deposited protein structures in the Protein Data Bank [8]. The slow progress of membrane protein structure determination (Figure 1.4) and study is largely due to challenges associated with requirement of a membrane environment.

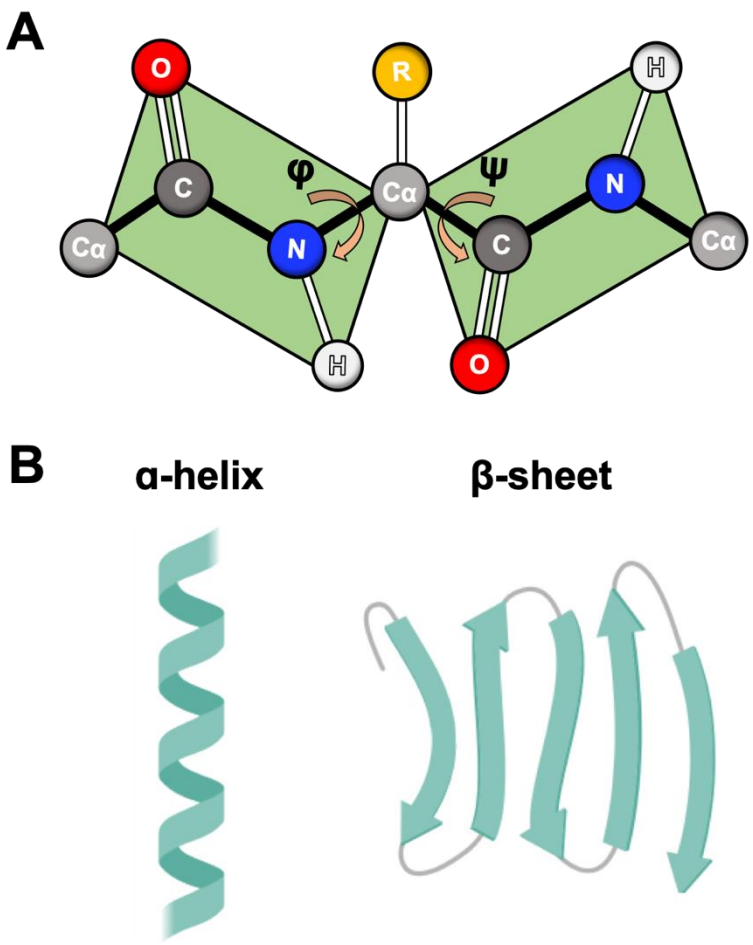


Figure 1.1: Primary and secondary protein structures. (A) Protein primary structure includes the polypeptide chain, formed by amino acids covalently linked with peptide bonds. Green areas represent the amide plane. Allowable free movements within the polypeptide backbone occur about the Cα-N (φ) and Cα-C (ψ) bonds. (B) Alpha-helix and β-sheet secondary protein structures (created with BioRender).

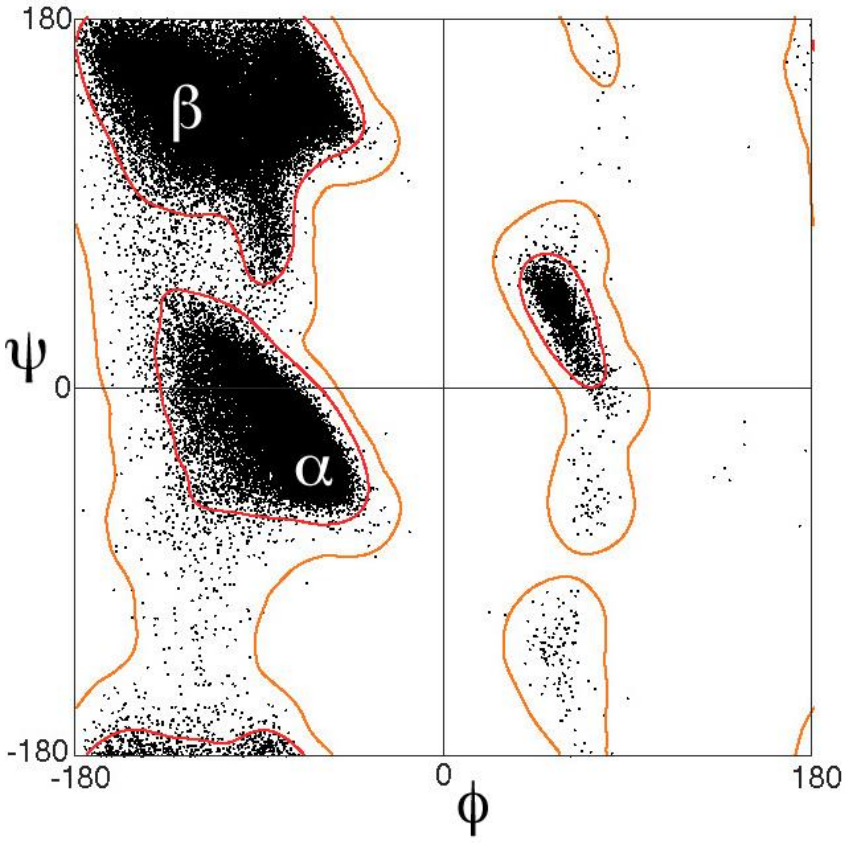


Figure 1.2: General Ramachandran plot determined from high-quality protein crystal structures. *Reprinted from Jane S. Richardson with permission*. Data points represent φ and ψ angles of 100,000 residues from 500 structures (resolution > 1.8 Å). Contour regions of α-helix and β-sheet structures are labeled. Points outside the orange contour lines are considered outliers.

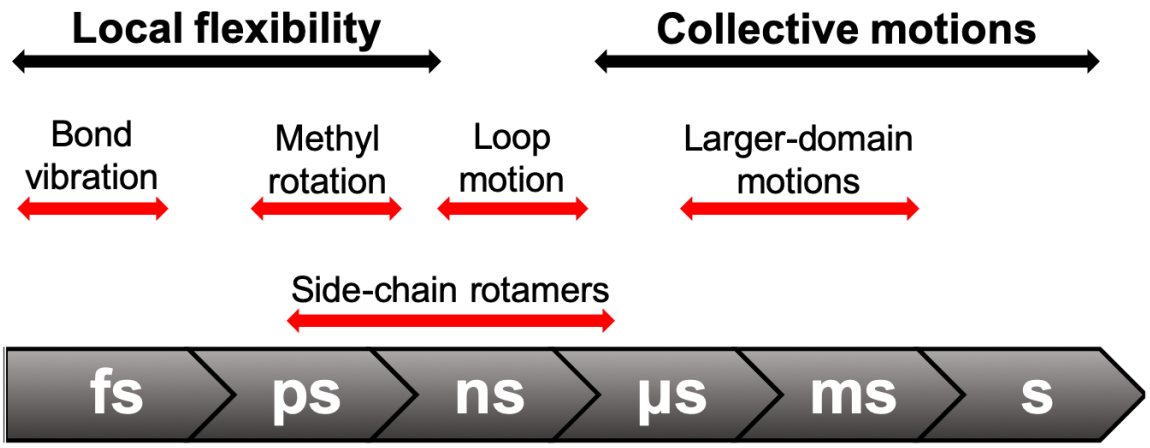


Figure 1.3: Dynamic processes in proteins occur over a range of timescales. *Adapted from [3] with permission*. General categories of motion are considered "local" and "collective." Examples of fluctuations that occur within proteins and their respective timescales are included.

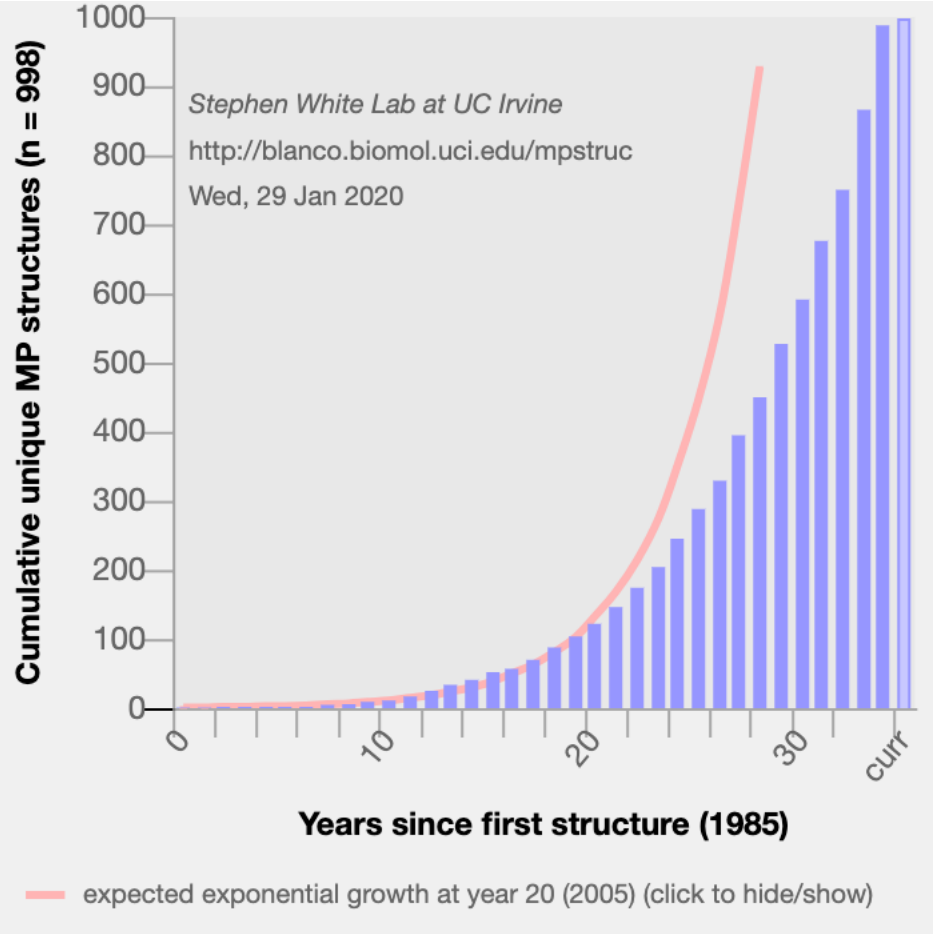


Figure 1.4: Membrane protein structure determination trends are significantly slower than expected. *Reprinted from Stephen White with permission*. Expected exponential growth in the rate of membrane protein structure determination far exceeds the current rate. At the time of writing, 998 unique membrane protein structures are deposited in the PDB.

1.2 Cell membrane environments

A cell membrane is a semi-permeable barrier that resides both on the outside of a cell and surrounds various cellular compartments, which facilitates organelle organization. The cell membrane is effectively the gate into and out of the cytoplasm as well as maintaining a cell's shape and morphology. Eukaryotic cells typically have a plasma membrane that forms the cell and organelles, each with their own membranes. In contrast, bacterial membranes differ based on the presence of an additional outer membrane (Gram-negative bacteria) vs. inner membrane alone (Gram-positive bacteria). Similar to Gram-negative bacteria, several eukaryotic organelles such as mitochondria and chloroplasts contain an outer membrane, likely due to bacterial origins [9]. Finally, archaeal species show evidence of both Gram-positive and Gram-negative-like membrane architectures with adapted chemical components [10].

The membrane itself is composed of lipids and proteins that are either embedded within the membrane or associated with it. Sterols, such as cholesterol (Figure 1.5A), are an additional common feature of eukaryotic membranes, though they are rarely found in prokaryotes [11]. A lipid is an amphipathic biomolecule composed of two components: a polar headgroup and nonpolar acyl chains. Lipids found in biological membranes are further generally classified as glycerophospholipids, with glycerol-derived headgroups and two acyl chains (Figure 1.5B), or sphingolipids, with a sphingosine-derived headgroup and one acyl chain (Figure 1.5C) [12]. Lipids spontaneously assemble into bilayers, with their polar head groups facing solvent while sequestering their nonpolar acyl chains, due to the hydrophobic effect [13]. The hydrophobic effect driving bilayer formation offsets the large

entropic cost of ordered water structures called clathrates that form hydrogen-bonded cages around exposed lipid acyl chains [14]. Lipid association into bilayers therefore minimizes the exposed surface area of acyl chains, and by extension ordered water, in the system. Van der Waals interactions between the hydrocarbon chains provide an enthalpic contribution to lipid bilayer formation and stability [15].

Our understanding of biological membranes has drastically evolved since the first lipid bilayer model for the cell membrane was proposed by Gorter & Grendel in 1925 [16]. Gorter & Grendel's simplistic membrane model failed to account for the diverse functionality of biological membranes [17]. A biological membrane model that included proteins was first described a decade later by Danielli & Davson. Danielli & Davson's membrane model limited the description of membrane proteins as those adsorbed to the surface of bilayer [18]. In 1972, experiments from Singer & Nicolson demonstrated that proteins may also span through biological membranes [19]. The Singer-Nicolson biological membrane model, or "fluid mosaic model," is famously taught in biochemistry textbooks and describes a dynamic system of lipids, peripherally-adsorbed proteins, and integral membrane proteins, all with the ability to freely diffuse within the bilayer. Various modifications to the fluid mosaic model have been proposed as new discoveries are made. Nicolson recently published a revised fluid mosaic model that takes into account lipid rafts, membrane domains, and cytoskeletal fencing (Figure 1.6) [20]. The revised fluid mosaic model depicts a crowded membrane environment and components that restrict lateral diffusion. Notably, this updated model acknowledges that biological membranes possess

non-uniform, non-random interacting elements that undergo both compositional fluctuations and phases of thermodynamic equilibrium.

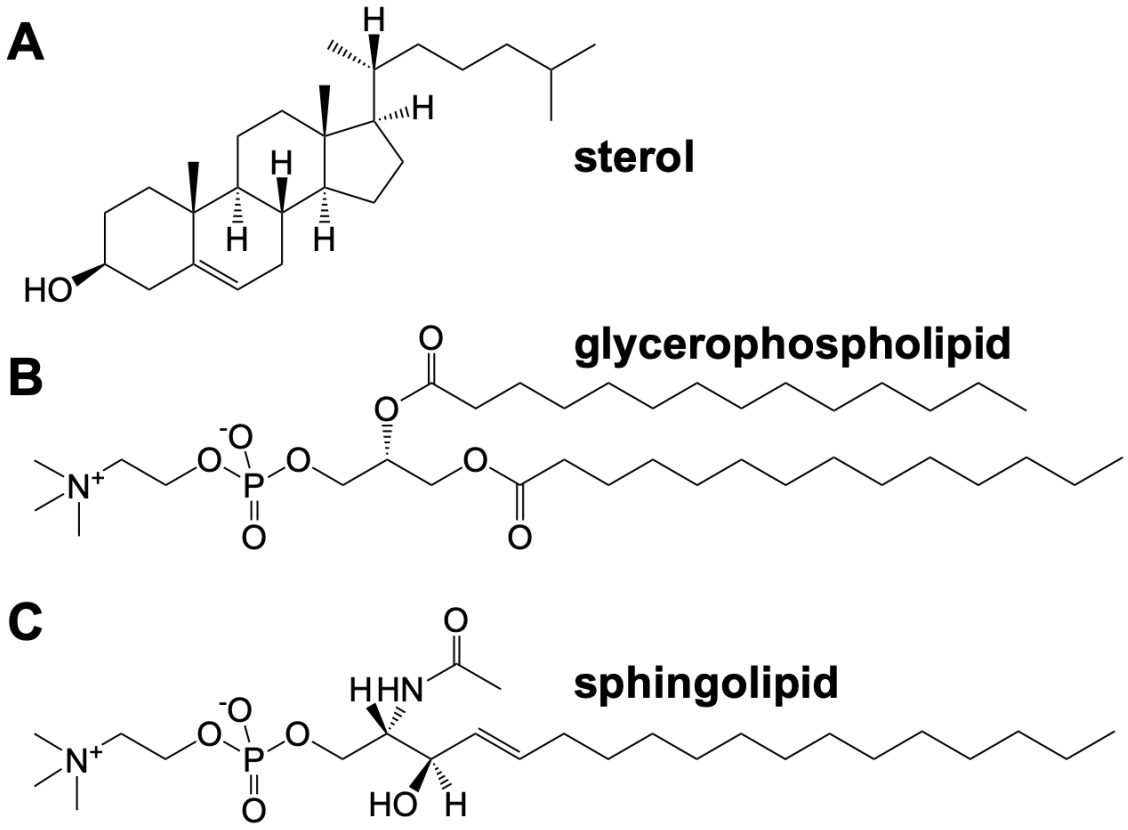


Figure 1.5: Chemical structures of common biological membrane components. Sterols (A), glycerophospholipids (B), and sphingolipids (C) are featured in a variety of cell membranes and possess amphipathic characteristics.

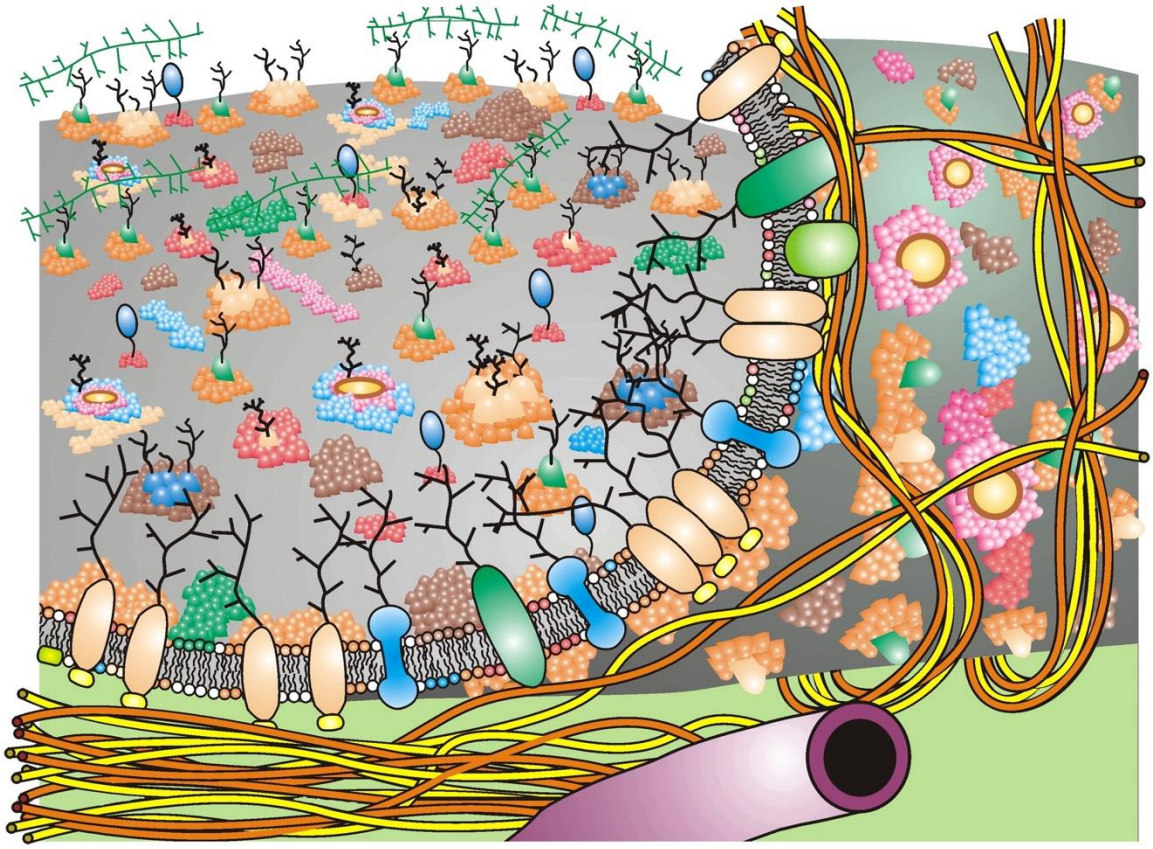


Figure 1.6: Revised Fluid-Mosaic Membrane Model. *Reprinted from [20] with permission.* Major updates to the original fluid-mosaic model [19] primarily include membrane domains, extracellular domains, and membrane-associated cytoskeletal structures. Different colors represent integral membrane proteins, glycoproteins, lipids, and oligosaccharides. Surface cytoskeletal fencing is rendered in the peeled view of the membrane.

1.3 Membrane protein structure, folding, and stability

As introduced in section 1.2, biological membranes are crucially composed of both lipids and proteins. Membrane-embedded or associated proteins represent a unique class of proteins, appropriately called “membrane proteins.” Membrane proteins typically function as transporters, receptors, enzymes, or anchors. Transporters may actively or passively transfer a molecule from one side of the membrane to the other. The sodium potassium pump, for example, utilizes ATP hydrolysis to facilitate the exchange of potassium and sodium ions across the membrane [21]. Transport through membrane channels, on the other hand, represents a passive process. Classic examples of membrane receptors are the G-protein coupled receptor family of proteins that, upon binding to an external signaling molecule, triggers cell signaling events [22]. Membrane enzymes carry out a myriad of catalytic reactions, ranging from hydrolases [23] to oxidoreductases [24]. A membrane enzyme popular in the biophysics field is the ATP-dependent diacylglycerol kinase (DgkA) that converts diacylglycerol to phosphatidic acid for membrane-derived oligosaccharide synthesis [25]. Finally, membrane anchors link extracellular and intracellular structures, or may be involved in cell adhesion processes [26]. Chapter 2 of this dissertation will focus on the carcinoembryonic antigen-like cell adhesion molecule (CEACAM) protein family that is expressed in mammalian membranes and plays a role in cell-cell adhesion and immune responses [27].

1.3.1 General membrane protein structural features

In relation to the bilayer, membrane proteins are classified as integral membrane proteins, which span the lipid bilayer, and peripheral membrane proteins, which are adsorbed on the surface of the bilayer. Integral membrane proteins may contain one or more hydrophobic transmembrane domains. Structurally, membrane proteins with more than one transmembrane element fall into two categories: α-helical bundles and β-barrels (Figure 1.7). Most secondary structural elements crossing the membrane are longer than those in their soluble protein counterparts, reflecting the ~ 30 Å hydrophobic lipid core. Within the membrane, α-helices are typically longer than 20 amino acids (1.5 Å/residue), and β-strands are longer than 10 amino acids (3.3 Å/residue) [28]. Due to the characteristic length and hydrophobicity of transmembrane helices, their topologies can generally be predicted using hydropathy plots based on amino acid sequence [29]. Additionally, certain amino acids have preferred localizations with respect to the membrane and aid protein topology prediction. For example, positively-charged amino acids are abundant on the cytoplasmic side of the membrane, a phenomenon known as the "positive-inside" rule [30]. Aromatic amino acids such as tryptophan and tyrosine have a strong preference to be localized at the membrane interface, forming characteristic "belts" that interact with lipid head groups [31].

It was previously thought that membrane proteins were "inside-out" proteins with a core of polar residues and nonpolar residues exposed to the surrounding lipids [32]. A number of authors have challenged the "inside-out" hypothesis, contradicting the early notion that membrane protein structures were mostly stabilized by interior polar

interactions [28, 32, 33]. Their results suggested that for α-helical membrane proteins, enthalpic contributions dominate helix packing based on maximized van der Waals contacts rather than entropic contributions from lipid exclusion of polar surface area [32, 33]. Similarly, β-barrel membrane proteins may appear "inside-out" with hydrophobic residues facing lipid and a polar interior, but β-barrel stability in the membrane depends on hydrogen bonds between strands [34]. Membrane protein structure, of course, plays a critical role in understanding their functions and ultimately serves as the starting point for drug development and targeting. Both structural classes of membrane proteins present interesting case studies in physical properties of membrane protein folding and stability, and each merit further discussion.

1.3.2 α-helical membrane protein folding and stability

Helical transmembrane proteins typically reside in the plasma membrane, specifically the inner membrane for bacteria or organelles that have additional outer membranes. The majority of α-helical transmembrane proteins are co-translationally inserted and folded into the plasma membrane via the Sec translocon pathway [35]. Membrane insertion occurs via attachment of a ribosome to a membrane-embedded translocon complex, which guides the secreted nascent chain from the ribosome into the membrane (Figure 1.8). In eukaryotes, the translocon complex is Sec61αβγ; bacteria use the homologous SecYEG complex [36]. Protein-conducting channel subunits Sec61α and SecY, for eukaryotes and bacteria respectively, contain a pore with exits for the translocating peptide chain to either be secreted to the aqueous environment or partitioned into the lipid bilayer [37, 38]. Translocon selection for the peptide chain to be secreted or

inserted into the membrane is attributed to hydrophobicity of transmembrane segments and resulting interactions with both the translocon and lipid bilayer [39].

A four-step thermodynamic cycle describing partitioning, folding, insertion, and association of α-helical transmembrane segments was introduced by White & Wimley [28], combining earlier models from Jacobs & White [40] and Popot & Engelman [41] (Figure 1.9). Importantly, helix formation occurs first, followed by helix packing and formation of the tertiary structure. The stability of folded protein elements in the membrane depends on free energy contributions of side chains, hydrogen bonds, and interactions with the lipid bilayer itself. Secondary structural elements of membrane proteins notoriously resist complete denaturation due to the free energy cost of disrupting hydrogen bonds in a membrane (~80-100 kcal mol^{-1} for a 20 amino acid transmembrane helix [28]). Denaturation of transmembrane helical bundles, on the other hand, occurs more readily than disruption of the helix itself [42]. Reversible folding from partially-denatured α-helical transmembrane proteins has been successful [43], likely owing to residual helical packing and/or interactions that are critical to the final folded state.

Focke *et al.* recently argued that *in vitro* folding of α-helical membrane proteins with no translocon present supports that protein sequence alone determines helical bundle folding [44]. It should be noted that the authors utilize a detergent for intermediate steps between synthesis and folding, rather than a chaotrope such as urea. A number of authors have suggested that establishing a helical core may be crucial for correct folding, and detergents such as sodium dodecyl sulfate (SDS) may facilitate such an intermediate structure *in vitro* [45-47]. Furthermore, unlike urea, SDS will partition into lipid vesicles

[48] with an encapsulated protein, complicating the assumption that sequence alone determines membrane protein folding based on results from *in vitro* folding protocols to date.

Bilayer adaptations during partitioning, folding, insertion, and association of α-helical membrane proteins play an additional, critical role in minimizing the free energy of the system [49]. For example, bilayer thickness may be perturbed in the presence of helical segments to reduce hydrophobic mismatch [28]. Chapter 4 will further explore the relationship between α-helical transmembrane protein fold and surrounding environment effects based on *in vitro* systems. Compared to β-barrel membrane proteins, the physical principles that drive α-helical bundles into a folded, free-energy minimum after insertion into the bilayer, as well as forces that maintain structural equilibrium, are not well understood.

1.3.3 β-barrel membrane protein folding and stability

The process of β-barrel membrane protein folding is fundamentally different than α-helical bundles. Differences are evident in the topology requirements for β-barrel membrane proteins, which may be considered more complex than transmembrane α-helices. Antiparallel β-strands form the barrel itself and are a hallmark structural feature of outer membrane proteins [50]. Unlike α-helices, in which backbone hydrogen bonding is fulfilled within the helix, β-strands in the membrane must form hydrogen bonds with neighboring strands, ultimately creating a barrel-like configuration. While soluble β-barrels can be found in the cytoplasm, the limit of strands in soluble barrels is eight. This

limit is the result of tight packing that is necessary to exclude water from the hydrophobic interior [34]. However, the minimum number of strands found in β-barrel membrane proteins begins at eight and may reach up to 36 strands [51]. The size of the barrel in the membrane is limited by stability of the hydrogen bonding network rather than solvent exclusion, allowing barrels to be much larger in the membrane compared to barrels found in the cytoplasm. Finally, residues alternate facing outward toward the lipid environment and facing inward toward the core of the barrel. Outward-facing β-barrel residues are characteristically nonpolar, while those facing inward may be either nonpolar or polar [52]. Because hydrophobicity of the interior may vary, it is typically not possible to define residue orientation of β-barrel membrane proteins based on sequence alone [34].

To correctly fold in outer membranes, β-barrel termini must start and end on the periplasmic side of the membrane, with short turns additionally facing the periplasm and long loops exposed to the extracellular space [53]. To achieve this topology, β-barrel membrane proteins must have an even number of strands, which is another defining characteristic of this protein class. Like α-helical transmembrane proteins, a translocon complex guides the nascent peptide chain from the ribosome. However, β-barrel membrane proteins remarkably must avoid insertion into the plasma membrane and instead be secreted from the translocon into the periplasmic space, where the protein is insoluble [54]. Traversing the periplasmic space is facilitated by chaperones that bind to unfolded outer membrane proteins, such as DegP, Skp, and SurA [55]. Chaperone proteins then deliver the protein to the outer membrane Bam complex that guides β-barrel insertion and folding [56].

Fleming recently presented an elegant "kinetic push, thermodynamic pull" mechanism for outer membrane β-barrel sorting across the periplasm and folding (Figure 1.10) [57]. The "thermodynamic pull" arises from characteristically high stabilities of β-barrel proteins in the membrane. Several free energies of outer membrane proteins folding into lipid vesicles are reported in the range of -18 to -32 kcal mol^{-1}, which suggests that the sorting and folding trajectory of β-barrel membrane proteins is governed by a thermodynamic "pull" toward its most energetically-favorable state in the outer membrane. It follows that the binding free energies of unfolded outer membrane proteins with periplasmic chaperones would be thermodynamically less favorable to facilitate insertion and folding into the outer membrane. Experimental values of outer membrane protein aggregation and binding to DegP, Skp, and SurA fall in the range of -8.5 – 11 kcal mol^{-1}, supporting the "thermodynamic pull" hypothesis [57].

Kinetic barriers to folding play an additional crucial role in outer membrane protein sorting, beginning with secretion from the translocon. PE and PG lipids in the inner membrane as well as bilayer thickness slow the folding kinetics of the outer membrane protein, which "pushes" the unfolded β-barrel from folding into the incorrect membrane [58]. Crucially, the association kinetics of unfolded outer membrane proteins with chaperones is at least an order of magnitude faster than spontaneous folding into bilayers containing PE and PG [59]. Addition of lipopolysaccharide, which is present in the outer membrane, does not substantially improve spontaneous folding rates [60]. The final kinetic "push" for β-barrel membrane protein folding *in vivo* occurs with the Bam complex, acting as a catalyst to overcome the kinetic barrier for spontaneous folding in bilayers. A proposed

mechanism for the Bam complex involves membrane thinning [61], which accelerates outer membrane protein folding *in vitro* [62, 63].

Although the kinetics of spontaneous β-barrel folding into membranes with *E. coli* lipid compositions is slow, a key observation is that they *will* spontaneously refold into bilayers. The presence of short-chain lipids, akin to membrane thinning *in vivo*, and zwitterionic lipid headgroups accelerate outer membrane protein folding [64]. Outer membrane proteins may be reversibly folded from a denatured state (e.g. in urea) into lipid vesicles, which is a common practice in β-barrel membrane protein preparations. Within highly denaturing conditions (>8 M urea), β-barrel folding may occur via a two-state reaction [65]. The transition state observed for PagP supports tilted insertion mechanisms that have been proposed for β-barrel membrane protein folding [66]. Similar denaturing conditions with helical membrane proteins minimally disrupts secondary structure, as described in section 1.3.2. Combined, understanding the apparent differences between α-helical and β-barrel membrane protein folding and stability is fundamental to isolating each protein class from the membrane for *in vitro* characterization.

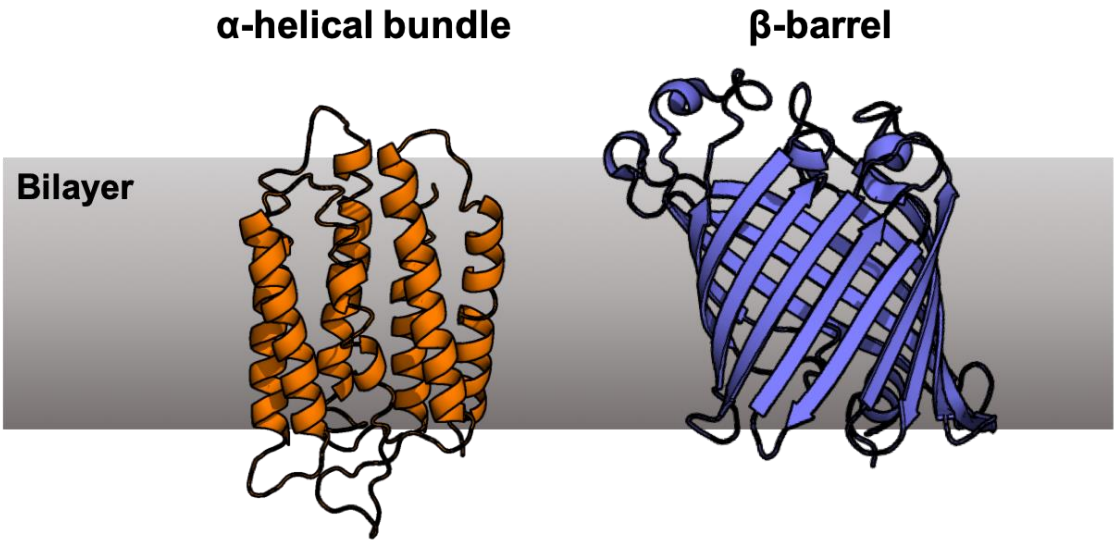


Figure 1.7: General folds adopted by transmembrane proteins. Alpha-helical bundles are typically localized to the plasma membrane (eukarya) and inner membrane (bacteria). Beta-barrel membrane proteins are found in the outer membranes of Gram-negative bacteria, mitochondria, and chloroplasts. One lipid bilayer is shown for simplicity. Selected structures are bacteriorhodopsin (α-helical bundle, PDB ID: 1AP9 [67]) and outer membrane phospholipase A (β-barrel, PDB ID: 1QD6 [68]).

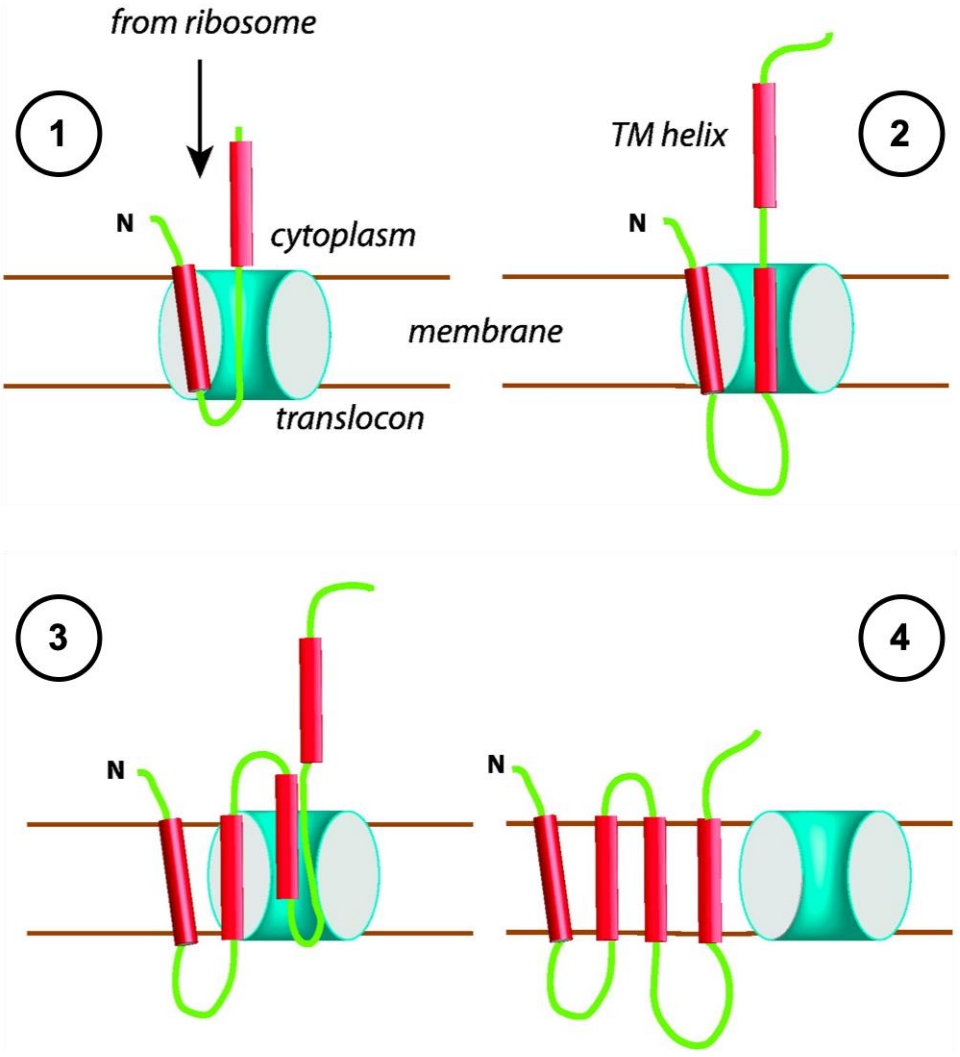


Figure 1.8: Translocon pathway for alpha-helical membrane protein insertion and folding. *Adapted from [69] with permission.* Transmembrane segments are co-translationally passed from the ribosome into the translocon complex (Steps 1-2). The translocon facilitates segment partitioning into the membrane (Step 3) for each helix to achieve the final protein fold (Step 4).

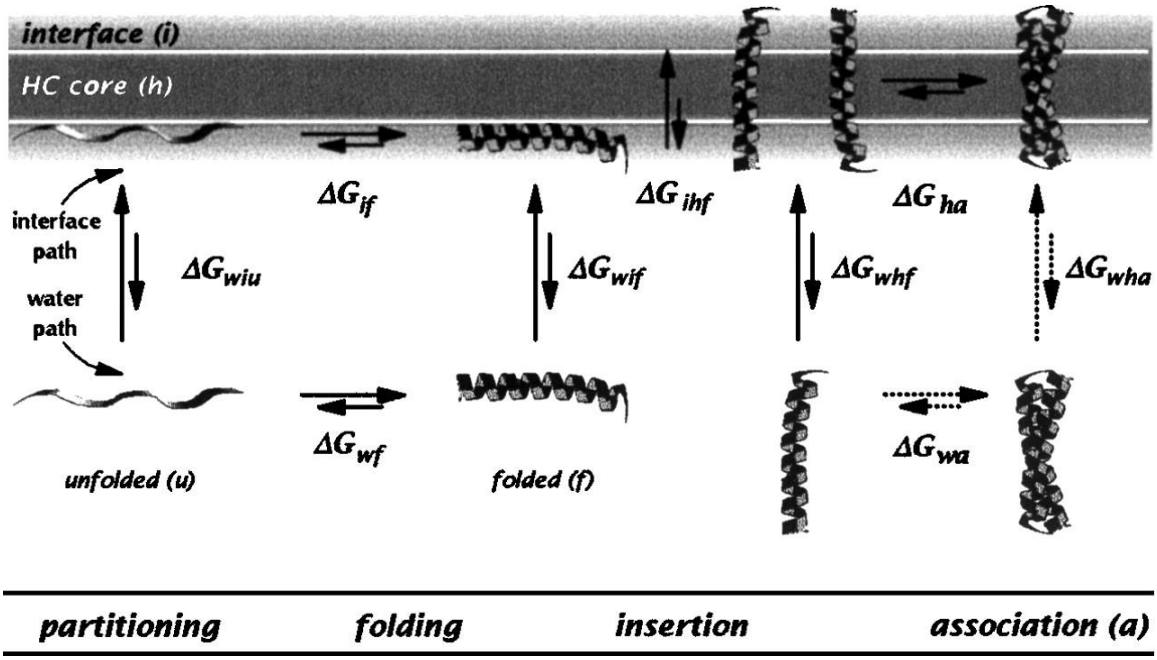


Figure 1.9: A four-step thermodynamic cycle of alpha-helix membrane partitioning, folding, insertion, and association. *Reprinted from [28] with permission*. Subscripts denote specific steps in the cycle: interface (i), hydrocarbon core (h), unfolded (u), folded (f), and association (a). The combination "wif," for example, indicates a process that sequentially follows water, interface, and fold.

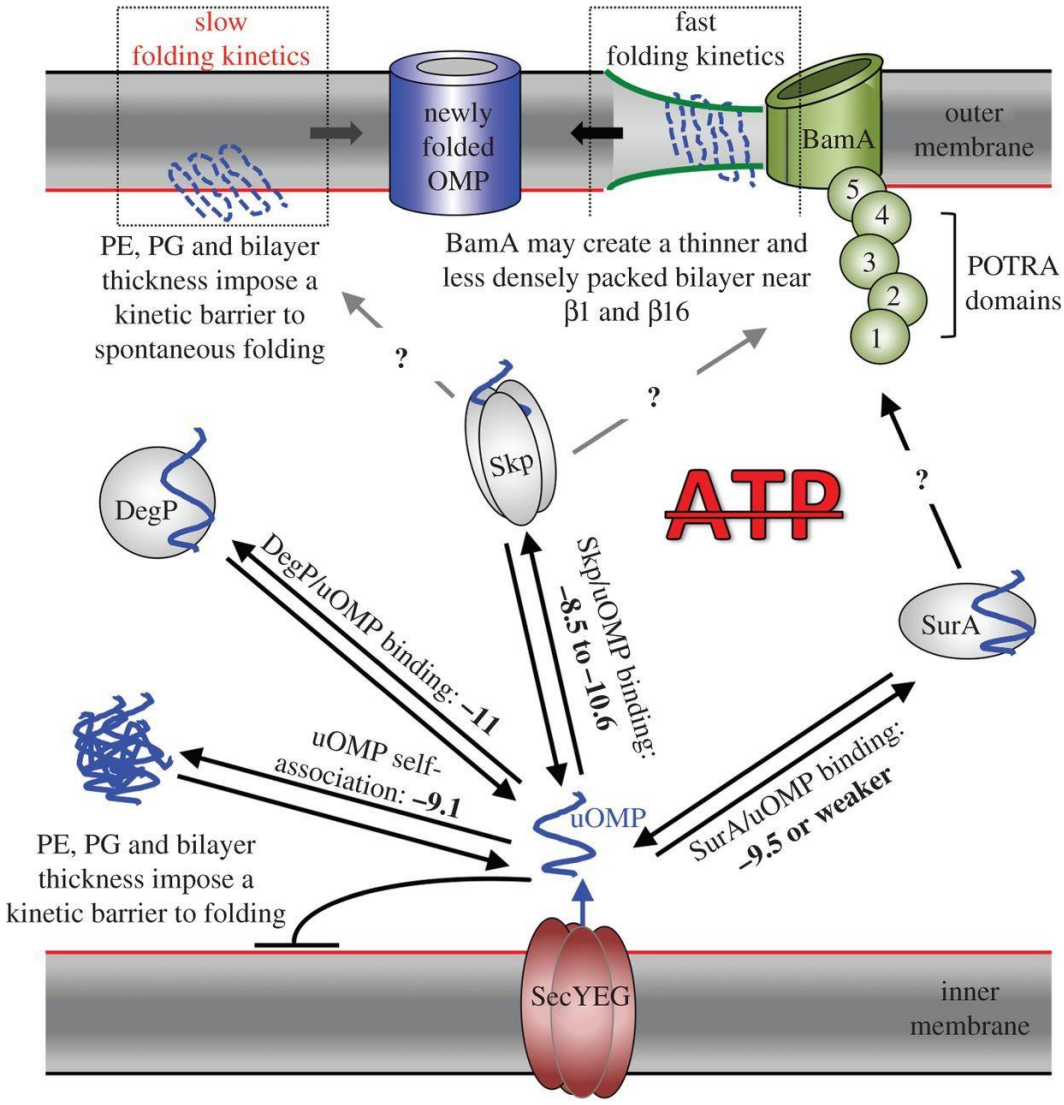


Figure 1.10: Outer membrane protein "kinetic push, thermodynamic pull" sorting and folding pathway. *Reprinted from [57] with permission.* An energy source such as ATP is notably absent from the process, suggesting that kinetic and thermodynamic contributions of protein-protein interactions and folding drives an outer membrane protein from the inner to the outer membrane. Known thermodynamic parameters are included for each step.

1.4 Membrane proteins in solution

To obtain atomic-level structures of membrane proteins, they typically must be isolated from their native cell membranes. The native membrane environment is often not amenable to biophysical techniques used to elucidate membrane protein structures, such as x-ray crystallography, cryo-electron microscopy, and nuclear magnetic resonance (NMR) spectroscopy. Thus, the majority of membrane protein structures have been solved using bilayer mimics that will be discussed in section 1.4.2.

1.4.1 Recombinant membrane protein expression and solubilization methods

Structural and functional studies of membrane proteins typically require milligram quantities of protein, which is not feasible to achieve using the natural abundance of proteins in the membrane. Recombinant systems to overexpress membrane proteins are therefore critical for biophysical characterization. A major pitfall in overexpressing membrane protein targets to the host is that any perturbation to the native membrane, particularly over-crowding with additional proteins, is often toxic [70]. This leads to the low yields typical of membrane protein expression systems. Additionally, there are insufficient populations of translocon complexes and chaperone proteins available from the host to accommodate proper sorting, folding, and insertion into the membrane, leading to aggregate formation [71]. To overcome the burden on the host cell membrane and inefficient expression, membrane proteins may instead be expressed to inclusion bodies, isolated, and refolded *in vitro* [72]. Overexpression into inclusion bodies is a common practice for β-barrel membrane proteins because they possess the ability to spontaneously

refold into bilayers (section 1.3.3). However, α-helical membrane protein refolding after isolation from inclusion bodies is rarely successful [43]. An alternative approach for increasing helical bundle yields involves tighter transcription control to improve cell viability [73]. Successful expression often relies on optimization of growth conditions and inducer concentrations, with variable results across protein targets. Consideration for the native environment of the target membrane protein is crucial in selecting an appropriate expression system.

Choice of an *in situ* expression system for a given membrane protein depends on the folding and membrane insertion requirements. For example, specific lipid compositions, translocon complexes, and chaperone proteins may be required for proper sorting and insertion into host membranes [74]. Specific post-translational modifications may be an additional requirement to maintain native membrane protein structure and function, which limits viable expression systems for a target. Depending on the membrane protein, the most common cell systems used for expression are derived from bacteria, yeast, and insects [75]. *E. coli* is the most widely-employed cell expression system for recombinant membrane proteins and is cost-effective, easy to use, and amenable to a variety of membrane localization and *in situ* labeling techniques. Insect and yeast cell lines are highly useful expression systems for eukaryotic membrane proteins, as they possess the necessary machinery for post-translational modifications such as glycosylation and disulfide-bond formation. Though multiple cell lines are available for recombinant membrane protein expression, each system has drawbacks. Major limitations include

aggregation, mis-targeting, and misfolding in the membrane with *E. coli*, and limited sterol populations crucial for efficient eukaryotic expression with insect/yeast cell lines [75].

In the event of successful expression of a target protein to the membrane, membrane protein isolation is typically required for structural and functional studies. Detergents are frequently used for membrane protein extraction into aqueous solvents. The detergent must be "harsh" in the sense that replacement of native lipids with detergent molecules is efficient and allows membrane protein solubilization (Figure 1.11). However, efficient membrane solubilization and subsequent purification steps in detergent often results in stripping native lipids from the protein interface that may be necessary to maintain an active conformation and/or stability [76]. Furthermore, solubilizing detergents may not be compatible with the size, surface properties, or native tertiary structure of a target membrane protein [77], forcing experimentalists to empirically determine appropriate conditions on a protein-by-protein basis. Compared with direct solubilization of the lipid bilayer, unfolded membrane proteins expressed into inclusion bodies are typically extracted in highly denaturing conditions (>8 M urea) for initial solubilization followed by rapid dilution into a detergent solution or short-chain lipid vesicles [72].

An alternative to *in situ* membrane protein expression and subsequent extraction from the cell membrane is utilizing *in vitro* expression systems. Within the past decade, expression of membrane proteins *in vitro* (cell-free) has improved dramatically. In principle, cell-free expression circumvents pitfalls of *in situ* membrane protein expression, chiefly with respect to cytotoxicity and challenges with protein isolation from cell membranes. For cell-free synthesis of soluble proteins, basic requirements for the reaction

mixture are (1) mRNA or DNA for the target protein and (2) necessary transcriptional and translational components. Cell extracts are commonly prepared for transcription and translation machinery, which includes RNA polymerase, ribosomes, tRNA, amino acids, enzymatic cofactors and energy sources, and chaperones for protein folding [44]. Membrane protein cell-free synthesis ultimately requires that a membrane-like environment is provided for proper folding and stabilization. For α-helical membrane proteins, correct co-translational folding is challenging in cell-free systems. Lipid vesicles may be added during cell-free expression, but without translocon complexes and membrane-associated chaperones, polypeptide chains must spontaneously insert into the bilayer. Focke *et al.* recently published a protocol for coupling cell-free membrane protein expression and folding *in vitro*, separating protein synthesis and folding steps [44]. The utility of this two-step process was demonstrated with the amino acid transporter LeuT and K^+ channels KcsA and MVP. A two-step approach for cell-free membrane protein production draws from earlier studies of *in vitro* KscA folding in which unfolded KscA was solubilized in SDS, followed by rapid dilution into lipid vesicles [78]. Advancements in cell-free expression will be important for identifying principles that lead to correct co-translational membrane protein folding, and screening approaches for solubilizing conditions may be improved with *in vitro* expression systems.

1.4.2 Soluble bilayer mimics for membrane protein applications

Regardless of the expression and solubilization method selected for membrane protein studies, an environment that maintains membrane protein structure and function in the absence of its native bilayer is a prerequisite for any protein target. The majority of *in*

vitro membrane protein experiments to date have utilized molecules with "lipid-like" properties that form amphipathic assemblies in aqueous solvent. A wide range of bilayer mimics have been introduced for membrane protein applications, which will be explored in this section: detergent micelles, liposomes, lipid-detergent bicelles, nanodiscs, and styrene maleic acid lipid nanoparticles (Figure 1.12). Detergent micelles and lipid-detergent bicelles were primarily used in this research, and bicelle physical properties will be discussed at length in Chapter 3.

1.4.2.1 Detergent micelles

Detergents are similar to lipids in that they have amphipathic character – a water-soluble head, and a hydrophobic tail. This physical property of detergents suits membrane protein solubilization well, as the hydrophobic or transmembrane portion of the membrane protein can be surrounded by the alkyl tails of detergent molecules, with headgroups able to provide similar polar interactions and hydrogen bonding that would be present in the native membrane environment. A key difference between lipids and detergents is that detergents commonly contain one alkyl tail, as opposed to the two tails belonging to lipids. Detergent molecules exist in solution as monomers until the detergent is at or exceeds a concentration required to form aggregate assemblies, referred to as the critical micelle concentration (CMC). Once the CMC is reached, detergent monomers spontaneously assemble into micelle aggregates in solution (Figure 1.12A), and both components are in equilibrium [79]. The CMC of each detergent is unique because head group repulsive forces and alkyl tail interactions contribute to aggregation propensity, which varies based on the detergent physical properties. The micelle aggregation number (n) represents the

number of detergent monomers that form a micelle and is also dependent on detergent character.

Detergent micelles form oblate (frisbee-shaped), prolate (football-shaped), disk-shaped, and aggregate structures, and both micelle size and shape depend on detergent monomer size and shape. For a given detergent, the packing parameter (p) describes the resulting assembly shape and is given by:

$$\mathrm{p} = \frac{v}{la} \tag{1.1}$$

where *v* is the volume occupied by the detergent tail, *l* is the tail length, and *a* defines the head group surface area [80]. Detergent tail length and volume contributions to the packing parameter are typically smaller than the headgroup surface area, owing to the single tail. Micelle oblate, prolate, or spherical shapes result when $0 < p < 1/3$, and cylinders form when $1/3 < p < 1/2$ (Figure 13). Compared with detergents, the two alkyl chains of lipid monomers increase the resulting volume and length contributions in the tail region, increasing the packing parameter such that bilayer or disc-like shapes form. Discs and bilayers result when $1/2 < p < 1$ (Figure 1.13).

Detergents are generally classified based on head group charge properties and include non-ionic, zwitterionic, and ionic categories (Figure 1.14). Non-ionic detergents such as n-decyl-β-D-maltopyranoside (DM) and n-octyl-β-D-glucopyranoside (OG) are commonly used for membrane solubilization and protein extraction to prevent protein denaturation [81]. Zwitterionic detergents contain both positive and negative charges that result in a net zero charge, and the head group properties of zwitterionic detergents are

diverse. Similar to non-ionic detergents, zwitterionic detergents such as the foscholine class are widely used in membrane protein structural biology applications. Ionic detergents possess head groups with net positive (cationic) or negative (anionic) charges. SDS and sarkosyl are popular ionic detergents. While ionic detergents are efficient in solubilizing membranes, they are often denaturing and not compatible with many membrane proteins [82]. Steroid-based detergents, or bile salts, represent a unique category with a hydrophilic head group and both nonpolar and polar surfaces from the sterol (Figure 14). The nonpolar sterol surface is sequestered to the interior of the micelle, exposing the polar sterol and headgroup to solvent. The head group may additionally possess charge properties, similar to head groups among single-chain detergents. Sodium cholate (anionic) and 3-((3-cholamidopropyl)dimethylammonio)-1-propanesulfonate (CHAPS) are typical bile salt detergents. Non-ionic, zwitterionic, ionic, and bile salt detergents were all used to probe lipid-detergent bicelle phase transitions described in Chapter 3.

Compared to lipid bilayers, detergent micelles are water-soluble assemblies and are suitable for use in a wide range of biophysical techniques. Detergent micelles are furthermore advantageous when one protein per assembly is desirable, as lipid bilayers and large discs can accommodate multiple proteins. As described in section 1.4.1, detergents are commonly used in extraction of proteins from the membrane. Once the membrane protein is extracted from the membrane with detergent, the same detergent may be used for purification. However, the solubilizing detergent is often not suitable for stabilizing a functional fold of a target membrane protein and may be exchanged for another detergent during purification. To maintain stable protein-detergent complexes, the detergent

concentration in solution must remain above the CMC. Practically speaking, this procedure may be expensive and challenging, as detergent solutions are required throughout the protein purification process. When dialysis is required for protein purification, the exchange buffer must also contain detergent, or dialysis times must be carefully controlled to avoid excess detergent loss.

Pitfalls of detergent use for membrane protein functional assays are primarily concerned with free detergent monomers and empty micelles present in solution. Detergent monomers may interfere with ligand binding, convolute light scattering or absorbance measurements, and affect calorimetry measurements [82, 83]. Furthermore, membrane transport function is not compatible with detergent micelles because they lack an inside and outside compartment necessary to establish gradients across the interface. A structural pitfall of using detergent micelles that ultimately impacts function occurs when detergent micelles stabilize non-functional folds of membrane proteins [79]. The physical principles that bias membrane protein functionally-folded states in lipid and detergent environments is poorly understood at present and remains a challenge to the field.

1.4.2.2 Liposomes

Liposomes are spherical in shape and composed of one or more lipid bilayers (Figure 1.12B). This class of membrane mimics is well-suited for many functional assays, including transport, and is preferred over detergent micelles whenever compatible with specific biophysical techniques. Liposome size and curvature depends both on the lipid composition and formation method. Liposomes may be formed by a variety of methods such as lipid sonication, extrusion, and initial lipid solubilization with detergent followed

by dialysis for detergent removal [84]. Preparation methods additionally determine whether a liposome is unilamellar (single bilayer) or multilamellar (multiple bilayers). Multilamellar vesicles (MLVs) may form with sonication employed at low shear rates [85]. Small unilamellar vesicles (SUVs) are typically prepared from extensive sonication of MLVs and possess diameters of 15 – 50 nm. Large unilamellar vesicle (LUV) diameters range from 100 nm – 5 μm and may be formed by freeze-thawing SUVs or extrusion using membranes with specific pore sizes [86]. Finally, cell-sized vesicles with 5 – 300 μm diameters are known as giant unilamellar vesicles (GUVs), which are gaining popularity in the membrane protein field but are fragile and require separate techniques to form [87].

For membrane protein incorporation, denatured membrane proteins may be directly refolded into a liposome solution. This method is most often successful with β-barrel outer membrane proteins and liposomes with short-chain lipid compositions [88]. Alternatively, prepared protein-detergent complexes may be diluted into a liposome solution followed by detergent removal with biobeads and/or extensive dialysis. Dilution of the protein-detergent complex causes the detergent concentration to fall below the CMC, shifting equilibrium from micelle to detergent monomer. During this process, liposomes are available for protein incorporation upon micelle destabilization. The liposome solution may additionally contain a small amount of solubilizing detergent such as sodium cholate or β-octyl-D-maltopyranoside (OG) to facilitate protein incorporation. As is the case with all membrane mimetics, liposome composition and size must be screened for compatibility with each membrane protein, especially if the protein requires specific lipids or sterol content for function.

1.4.2.3 Bicelles

Lipid and detergent molecules may be combined and spontaneously form a "hybrid" bilayer mimic called a bicelle (Figure 1.12C). The "ideal" bicelle possesses a lipid-bilayer core surrounded by a rim of detergent molecules, forming a disc-like morphology [89]. The detergent rim "shields" lipids in the bilayer phase from water and stabilizes the edge of the assembly. Bicelle size depends on both the mole ratio of lipid to detergent (q-ratio, or "q") and total phospholipid concentration (C_L). When $0.2 < q < 2.0$, bicelles tumble isotropically in solution. Bicelles undergo a phase transition for $q > 2.0$, exhibiting a liquid-crystalline phase when $2.0 < q < 5.0$. For bicelle q-ratios above 7.0, multilamellar sheets or vesicles form with perforations edge-stabilized by detergent molecules, resembling Swiss cheese-like structures [90]. Bicelle diameter (long-dimension, *d*) is directly proportional to the q-ratio and may exceed several hundred angstroms [91]. The thickness or short-dimension of the bicelle is instead limited by the headgroup-to-headgroup distance (*L*) of the lipid in the bilayer phase.

DMPC/DHPC bicelles, formed with 14:0 (long-chain) and 6:0 (short-chain) glycerophospholipids, are the most well-characterized bicelles [92]. Although DHPC possesses two acyl chains, DHPC monomers form micelles due to the short chain-length and tail volume relative to the lipid head group. Bicelle formation is not limited to combining long-chain lipids and short-chain detergents. Bile salts (CHAPS and CHAPSO), long-chain fos-cholines (FC12), and detergents with acyl chains that terminate in cyclohexane moieties (Cyclofos-6), all form soluble assemblies with DMPC lipid [93]. However, bicelle assembly appears to be limited to specific combinations of lipids and

detergents, and use of a lipid in one bicelle composition does not guarantee success with another. For example, DMPC may be edge-stabilized by DHPC (6:0 and 7:0), CHAPS, FC12, and Cyclofos-6, all possessing unique physicochemical properties, but it may not be compatible with shorter-chain fos-cholines, for example [93]. Similarly, DHPC (7:0) may edge-stabilize DMPC but not POPC (16:0-18:1). The principles that govern bicelle formation and allow prediction of assembly are not well-understood. Furthermore, q-ratios and C_L values required to achieve a core of segregated lipids in the bilayer phase when $0.2 < q < 1.0$ is heavily contested in the literature [94]. Lipid-detergent bicelle assemblies and physical properties are the focus of Chapter 3.

Membrane protein incorporation into bicelles may be performed by several methods described in the literature. If the membrane protein is purified with the detergent used in the bicelle system, dried lipid may be added directly to the protein-detergent sample. To achieve the correct q-ratio in the final bicelle sample, detergent concentrations must be determined before lipid addition and is typically measured with ^{1}H-NMR. If the membrane protein is purified in a different detergent, the initial protein-detergent sample may be immobilized onto an affinity column, and a bicelle solution is used to exchange micelles for bicelles. This technique utilizes the same principles as detergent exchange methods for membrane proteins. Certain membrane proteins require initial incorporation into liposomes and nanodiscs, followed by addition of detergent to form protein-containing bicelles [95, 96]. Prior to detergent addition, it is necessary to measure the lipid concentration in the proteoliposome sample if a specific q-ratio is desired. For phospholipid systems, a phosphate assay is the preferred lipid quantitation method. Authors who

describe the liposome to bicelle method suggest that the pre-formed lipid core is not disrupted by addition of detergent [95]. Preference of direct purification in bicelles compared with initial liposome incorporation for a given membrane protein is at present limited to empirical observation. Current hypotheses and evidence from membrane proteins EmrE and LspA prepared in bicelles support that these membrane proteins, among others, may form specific contacts with the surrounding lipid that stabilizes the protein fold [95, 96]. Membrane proteins may also be purified and lyophilized, and bicelle incorporation occurs upon addition of solution directly to the protein [97].

Bicelles are attractive for membrane protein applications that require fast, isotropically-tumbling protein complexes such as solution NMR spectroscopy. The lipid core of bicelles maintains the bilayer-like architecture of the membrane, which offers an advantage over detergent micelles. When $2.0 < q < 3.5$, bicelles align in a magnetic field; this property has been exploited for both solution and solid-state NMR applications for soluble and membrane protein structure determination [89]. For several decades, isotropic bicelles were considered the biologically-relevant “cousin” of detergent micelles in the solution NMR field. The Columbus lab has recently demonstrated that q-ratios most often used in solution NMR studies do not contain a lipid core and instead possess mixed lipid and detergent molecules [94]. Whether a mixed lipid-detergent assembly containing protein (1) functions as a mixed micelle and/or (2) facilitates lipid-protein interactions stabilizing to a folded membrane protein is the motivation for work presented in Chapter 4.

1.4.2.4 Nanodiscs

Lipid nanodiscs are similar to bicelles in that they are disc-like particles with a lipid core. Rather than containing a detergent rim, lipid nanodiscs are encircled by membrane scaffold proteins (MSPs) (Figure 1.12D). MSPs are amphipathic helical proteins, and two MSPs form a bundle around the lipid disc per assembly [98]. The size of lipid nanodiscs is controlled by the MSP length, and MSP constructs have been engineered for specific nanodisc sizes. Membrane proteins are typically incorporated into nanodiscs by combining the purified protein (in detergent), lipids, and prepared MSPs, followed by detergent removal with biobeads [99]. Optimization of the protein:MSP:lipid ratio is required for efficient nanodisc assembly and yield, which is often a challenging bottleneck in membrane protein preparations with nanodiscs. When detergent-free systems are desired, nanodiscs are a clear advantage over detergent micelles and bicelles, although detergent is involved in initial purification of the protein. Once protein is successfully incorporated into nanodiscs, the sample can undergo further purification steps and dilution without addition of nanodisc components, which offers another advantage compared with detergent use. While minimally explored in this work, nanodiscs in particular have come of age in the membrane protein field over the past decade. Advancements in nanodisc technology have especially facilitated membrane protein structure determination with solution NMR spectroscopy and cryo-electron microscopy. Nanodiscs are furthermore amenable to many biophysical techniques that probe membrane protein function [98, 99].

1.4.2.5 Styrene maleic acid lipid nanoparticles

Styrene maleic acid (SMA) lipid nanoparticles, or SMALPs (Figure 1.12E), are a relatively new addition to the suite of membrane protein solubilization options. SMA is an amphipathic polymer capable of adsorbing to and destabilizing biological membranes [100]. Membrane proteins may be extracted directly from the membrane with SMA, which encapsulates the protein in small discs that partially retain the surrounding lipid environment [76]. In simple terms, SMALP formation is analogous to "hole-punching" a membrane to obtain lipid patches with the protein of interest. Because SMA "hole-punches" the entire membrane, lipid disc patches are not specific to the target protein, and affinity tags are typically employed for purification with conventional methods. SMALPs containing protein have been reported as monodisperse lipid discs with diameters in the range of 10-11 nm [101]. Similar to nanodiscs, SMALPs are water-soluble and preserve the integrity of folded transmembrane proteins. SMA, however, offers the additional advantage of membrane solubilization properties, which make SMALPs an attractive detergent-free bilayer system. SMA has been used to successfully extract and maintain function of a variety of membrane proteins [102]. Notable limitations of SMALP technology include sensitivity to divalent cations and low pH (< 6.5). The former property is a challenge to a number of proteins that rely on cations for function and/or stability, such as $MgCl_2$-dependent ATPases. Development of novel polymers for membrane encapsulation is a growing field, and recent work has been dedicated to overcoming challenges associated with SMALPs [103].

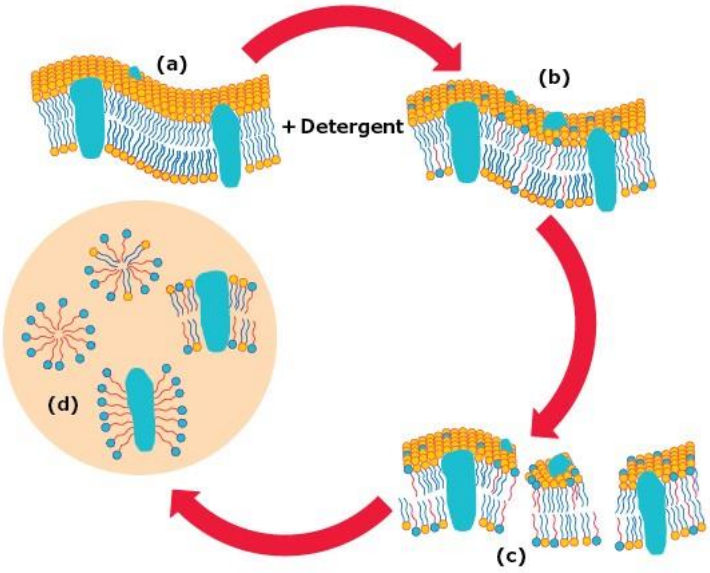


Figure 1.11: Process of membrane solubilization with detergent. *Reprinted from Sigma-Aldrich*. Lipids are represented with yellow head groups and blue tails, and membrane-spanning proteins are teal (a). Addition of low detergent concentrations (blue headgroups and red tails) causes detergent insertion into the lipid bilayer without destabilization (b). As more detergent is added to the system, membrane destabilization occurs (c) such that the majority of lipids surrounding the membrane protein are replaced with detergent molecules. A water-soluble protein-detergent complex forms in addition to lipid/detergent complexes without protein and unsolubilized protein (d).

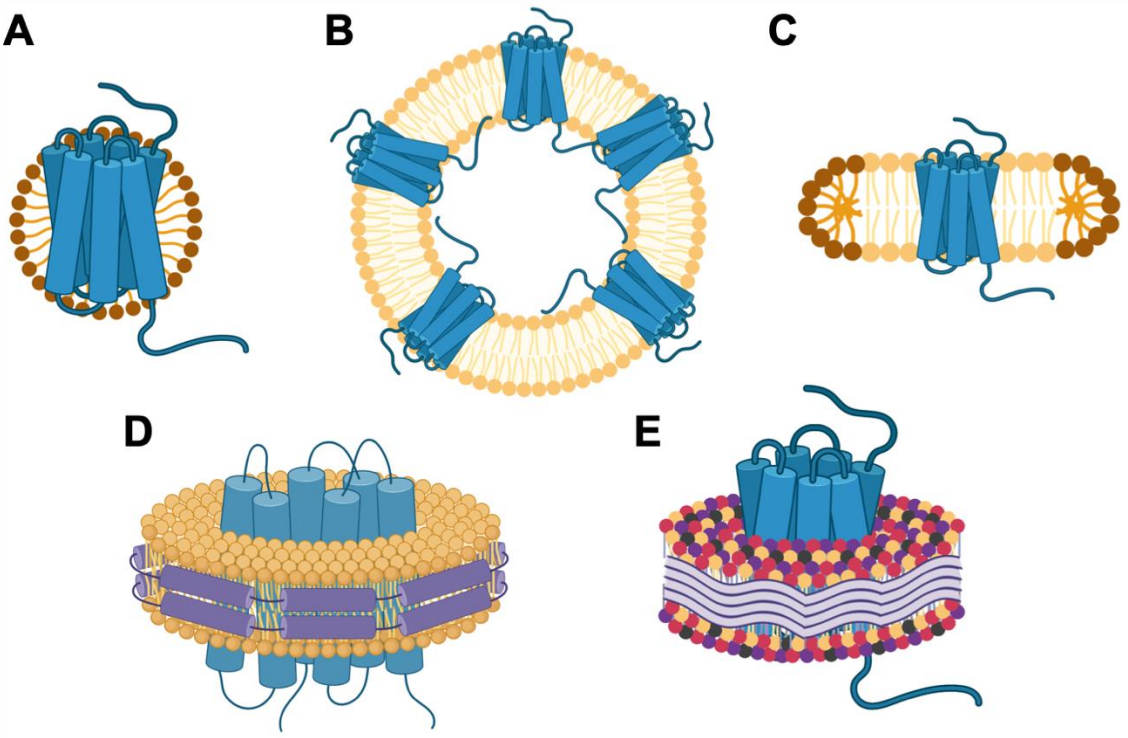


Figure 1.12: Soluble membrane mimics incorporated with protein. *Created with BioRender*. General α-helical bundle encapsulated in a (A) detergent micelle, (B) liposome, (C) lipid-detergent bicelle, (D) nanodisc, and (E) Styrene maleic acid lipid nanoparticle (SMALP). Purple cylinders and squiggles represent membrane scaffolding proteins (MSP) and SMA polymers, respectively. Panels (A-D) contain yellow lipid and brown detergent molecules. Multicolor lipids in panel E demonstrate the extraction of native lipids with SMA.

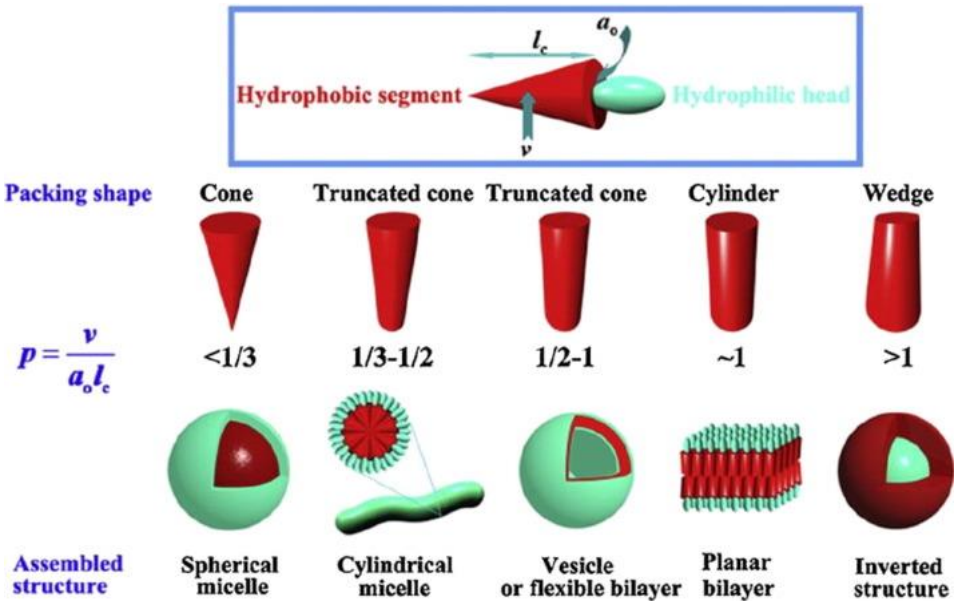


Figure 1.13: Morphology of amphiphile assembly depends on the packing parameter. *Reprinted from [104] with permission.* The volume of the hydrophobic segment with respect to its length and surface area of the hydrophilic component defines the amphiphile packing parameter. Packing parameter values can generally predict assembly structures, ranging from spherical micelles to inverted micelles.

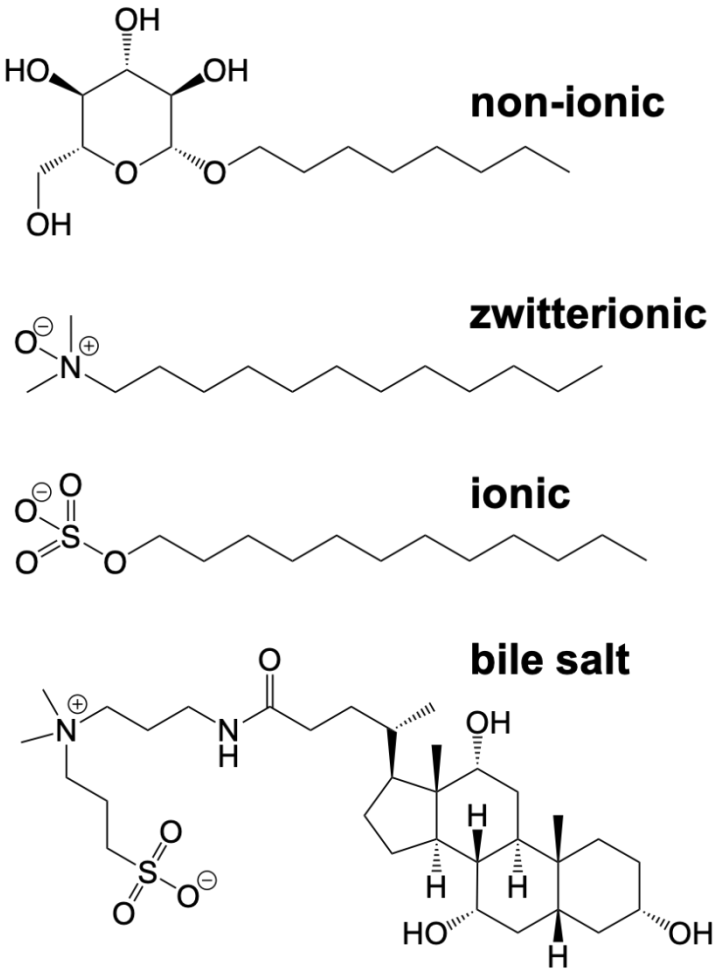


Figure 1.14: Chemical structures of detergent classes. Non-ionic, zwitterionic, ionic, and bile salt detergents represent the four major categories of detergents. Representative structures are n-octyl-β-D-glucopyranoside (OG, non-ionic), n-dodecyl-N,N-dimethylamine-N-oxide (LDAO, zwitterionic), sodium dodecyl sulfate (SDS, ionic), and 3-((3-cholamidopropyl)dimethylammonio)-1-propanesulfonate (CHAPS, bile salt).

1.5 Dissertation overview

The majority of this dissertation focuses on solution-based studies of lipid bilayer mimics, membrane proteins, and receptors with small-angle scattering and magnetic resonance spectroscopy techniques. Chapter 2 describes the homodimer structure of the carcinoembryonic antigen-like cellular adhesion molecule 1 (CEACAM1) N-terminal domain in solution with and without glycan modifications, contradicting recent work suggesting that glycan modifications reduce CEACAM1 N-domain dimerization. Physical properties and phase transitions of lipid-detergent bicelles commonly used for membrane protein structure determination are presented in Chapter 3, with emphasis on establishing the presence of a bilayer phase in isotropically-tumbling bicelles. Drawing from bicelle characterization described in this work, Chapter 4 compares α-helical transmembrane protein fold and interactions between helical segments of a model membrane protein, TM0026, in pure detergent and bicelle environments. Bicelle properties stabilizing to α-helical membrane protein folds are discussed, as well as future directions for assessing α-helical membrane protein complexes formed with bicelles. Finally, a comprehensive study of the function and mechanism of phosphoglucose isomerase from *Thermotoga maritima* is presented in Chapter 5, which may be beneficial for bioengineering thermostable enzymes in industrial settings.

1.6 References

1. Hardianto, A., et al., *Structure-Based Drug Design Workflow*, in *Encyclopedia of Bioinformatics and Computational Biology*, S. Ranganathan, et al., Editors. 2019, Academic Press: Oxford. p. 273-282.
2. Baldwin, R.L. and G.D. Rose, *How the hydrophobic factor drives protein folding.* Proceedings of the National Academy of Sciences, 2016. **113**(44): p. 12462.
3. Henzler-Wildman, K. and D. Kern, *Dynamic personalities of proteins.* Nature, 2007. **450**(7172): p. 964-972.
4. Doerr, A., *Protein structure through time.* Nature Methods, 2016. **13**(1): p. 34-34.
5. Brünger, A.T., *X-ray crystallography and NMR reveal complementary views of structure and dynamics.* Nature structural biology, 1997. **4 Suppl**: p. 862-865.
6. Jeffery, C.J., *Why study moonlighting proteins?* Frontiers in genetics, 2015. **6**: p. 211-211.
7. Lee, J.H. and C.J. Jeffery, *The crystal structure of rabbit phosphoglucose isomerase complexed with D-sorbitol-6-phosphate, an analog of the open chain form of D-glucose-6-phosphate.* Protein science : a publication of the Protein Society, 2005. **14**(3): p. 727-734.
8. Wang, J.-h. and M.J. Eck, *Chapter 6 - Protein Architecture: Relationship of Form and Function*, in *Hematology (Seventh Edition)*, R. Hoffman, et al., Editors. 2018, Elsevier. p. 59-67.
9. Martin, W.F., S. Garg, and V. Zimorski, *Endosymbiotic theories for eukaryote origin.* Philosophical transactions of the Royal Society of London. Series B, Biological sciences, 2015. **370**(1678): p. 20140330-20140330.
10. Holst, O., A.P. Moran, and P.J. Brennan, *Chapter 1 - Overview of the glycosylated components of the bacterial cell envelope*, in *Microbial Glycobiology*, O. Holst, et al., Editors. 2010, Academic Press: San Diego. p. 1-13.
11. Stillwell, W., *Chapter 5 - Membrane Polar Lipids*, in *An Introduction to Biological Membranes (Second Edition)*, W. Stillwell, Editor. 2016, Elsevier. p. 63-87.
12. Ouellette, R.J. and J.D. Rawn, *31 - Lipids and Biological Membranes*, in *Organic Chemistry (Second Edition)*, R.J. Ouellette and J.D. Rawn, Editors. 2018, Academic Press. p. 1001-1032.
13. Tanford, C., *Interfacial free energy and the hydrophobic effect.* Proceedings of the National Academy of Sciences, 1979. **76**(9): p. 4175.
14. Sharp, K.A., *The hydrophobic effect.* Current Opinion in Structural Biology, 1991. **1**(2): p. 171-174.
15. Chen, L., M.L. Johnson, and R.L. Biltonen, *A Macroscopic Description of Lipid Bilayer Phase Transitions of Mixed-Chain Phosphatidylcholines: Chain-Length and Chain-Asymmetry Dependence.* Biophysical Journal, 2001. **80**(1): p. 254-270.
16. Gorter, E. and F. Grendel, *ON BIMOLECULAR LAYERS OF LIPOIDS ON THE CHROMOCYTES OF THE BLOOD.* The Journal of Experimental Medicine, 1925. **41**(4): p. 439-443.
17. Robertson, J.D., *The ultrastructure of cell membranes and their derivatives.* Biochemical Society symposium, 1959. **16**: p. 3-43.

18. Danielli, J.F. and H. Davson, *A contribution to the theory of permeability of thin films.* Journal of Cellular and Comparative Physiology, 1935. **5**(4): p. 495-508.
19. Singer, S.J. and G.L. Nicolson, *The Fluid Mosaic Model of the Structure of Cell Membranes.* Science, 1972. **175**(4023): p. 720.
20. Nicolson, G.L., *The Fluid—Mosaic Model of Membrane Structure: Still relevant to understanding the structure, function and dynamics of biological membranes after more than 40years.* Biochimica et Biophysica Acta (BBA) - Biomembranes, 2014. **1838**(6): p. 1451-1466.
21. Suhail, M., *Na, K-ATPase: Ubiquitous Multifunctional Transmembrane Protein and its Relevance to Various Pathophysiological Conditions.* Journal of clinical medicine research, 2010. **2**(1): p. 1-17.
22. Hilger, D., M. Masureel, and B.K. Kobilka, *Structure and dynamics of GPCR signaling complexes.* Nature Structural & Molecular Biology, 2018. **25**(1): p. 4-12.
23. Chang, J.W., R.E. Moellering, and B.F. Cravatt, *An Activity-Based Imaging Probe for the Integral Membrane Hydrolase KIAA1363.* Angewandte Chemie International Edition, 2012. **51**(4): p. 966-970.
24. Lu, P., et al., *Structure and mechanism of a eukaryotic transmembrane ascorbate-dependent oxidoreductase.* Proceedings of the National Academy of Sciences, 2014. **111**(5): p. 1813.
25. Li, D. and M. Caffrey, *Lipid cubic phase as a membrane mimetic for integral membrane protein enzymes.* Proceedings of the National Academy of Sciences, 2011. **108**(21): p. 8639.
26. Denker, S.P. and D.L. Barber, *Ion transport proteins anchor and regulate the cytoskeleton.* Current Opinion in Cell Biology, 2002. **14**(2): p. 214-220.
27. Gray-Owen, S.D. and R.S. Blumberg, *CEACAM1: contact-dependent control of immunity.* Nature Reviews Immunology, 2006. **6**(6): p. 433-446.
28. White, S.H. and W.C. Wimley, *MEMBRANE PROTEIN FOLDING AND STABILITY: Physical Principles.* Annual Review of Biophysics and Biomolecular Structure, 1999. **28**(1): p. 319-365.
29. White, S.H., *Hydropathy Plots and the Prediction of Membrane Protein Topology*, in *Membrane Protein Structure: Experimental Approaches*, S.H. White, Editor. 1994, Springer New York: New York, NY. p. 97-124.
30. von Heijne, G., *Membrane protein structure prediction: Hydrophobicity analysis and the positive-inside rule.* Journal of Molecular Biology, 1992. **225**(2): p. 487-494.
31. Yau, W.-M., et al., *The Preference of Tryptophan for Membrane Interfaces.* Biochemistry, 1998. **37**(42): p. 14713-14718.
32. Stevens, T.J. and I.T. Arkin, *Are membrane proteins "inside-out" proteins?* Proteins: Structure, Function, and Bioinformatics, 1999. **36**(1): p. 135-143.
33. Rees, D.C., L. DeAntonio, and D. Eisenberg, *Hydrophobic organization of membrane proteins.* Science, 1989. **245**(4917): p. 510.
34. Schulz, G.E., *The structure of bacterial outer membrane proteins.* Biochimica et Biophysica Acta (BBA) - Biomembranes, 2002. **1565**(2): p. 308-317.

35. Mori, H. and K. Ito, *The Sec protein-translocation pathway.* Trends in Microbiology, 2001. **9**(10): p. 494-500.
36. Mandon, E.C., S.F. Trueman, and R. Gilmore, *Translocation of proteins through the Sec61 and SecYEG channels.* Current opinion in cell biology, 2009. **21**(4): p. 501-507.
37. White, S.H. and G. von Heijne, *How Translocons Select Transmembrane Helices.* Annual Review of Biophysics, 2008. **37**(1): p. 23-42.
38. Park, E. and T.A. Rapoport, *Mechanisms of Sec61/SecY-Mediated Protein Translocation Across Membranes.* Annual Review of Biophysics, 2012. **41**(1): p. 21-40.
39. Hessa, T., et al., *Recognition of transmembrane helices by the endoplasmic reticulum translocon.* Nature, 2005. **433**(7024): p. 377-381.
40. Jacobs, R.E. and S.H. White, *The nature of the hydrophobic binding of small peptides at the bilayer interface: implications for the insertion of transbilayer helices.* Biochemistry, 1989. **28**(8): p. 3421-3437.
41. Popot, J.L. and D.M. Engelman, *Membrane protein folding and oligomerization: the two-stage model.* Biochemistry, 1990. **29**(17): p. 4031-4037.
42. Vogel, R. and F. Siebert, *Conformation and Stability of α-Helical Membrane Proteins. 2. Influence of pH and Salts on Stability and Unfolding of Rhodopsin.* Biochemistry, 2002. **41**(11): p. 3536-3545.
43. Booth, P.J., *The trials and tribulations of membrane protein folding in vitro.* Biochimica et Biophysica Acta (BBA) - Biomembranes, 2003. **1610**(1): p. 51-56.
44. Focke, P.J., et al., *Combining in Vitro Folding with Cell Free Protein Synthesis for Membrane Protein Expression.* Biochemistry, 2016. **55**(30): p. 4212-4219.
45. Skach, W.R., *Cellular mechanisms of membrane protein folding.* Nature structural & molecular biology, 2009. **16**(6): p. 606-612.
46. Eilers, M., et al., *Internal packing of helical membrane proteins.* Proceedings of the National Academy of Sciences, 2000. **97**(11): p. 5796.
47. Booth, P.J. and J. Clarke, *Membrane protein folding makes the transition.* Proceedings of the National Academy of Sciences, 2010. **107**(9): p. 3947.
48. Tan, A., et al., *Thermodynamics of sodium dodecyl sulfate partitioning into lipid membranes.* Biophysical journal, 2002. **83**(3): p. 1547-1556.
49. Booth, P.J., *Sane in the membrane: designing systems to modulate membrane proteins.* Current Opinion in Structural Biology, 2005. **15**(4): p. 435-440.
50. Schulz, G.E., *β-Barrel membrane proteins.* Current Opinion in Structural Biology, 2000. **10**(4): p. 443-447.
51. Zhang, X.C. and L. Han, *How does a β-barrel integral membrane protein insert into the membrane?* Protein & cell, 2016. **7**(7): p. 471-477.
52. Taylor, P.D., et al., *Beta barrel trans-membrane proteins: Enhanced prediction using a Bayesian approach.* Bioinformation, 2006. **1**(6): p. 231-233.
53. Bagos, P.G., T.D. Liakopoulos, and S.J. Hamodrakas, *Evaluation of methods for predicting the topology of beta-barrel outer membrane proteins and a consensus prediction method.* BMC bioinformatics, 2005. **6**: p. 7-7.

54. Ricci, D.P. and T.J. Silhavy, *Outer Membrane Protein Insertion by the β-barrel Assembly Machine.* EcoSal Plus, 2019. **8**(2): p. 10.1128/ecosalplus.ESP-0035-2018.
55. Sklar, J.G., et al., *Defining the roles of the periplasmic chaperones SurA, Skp, and DegP in Escherichia coli.* Genes & development, 2007. **21**(19): p. 2473-2484.
56. Hagan, C.L., T.J. Silhavy, and D. Kahne, *β-Barrel Membrane Protein Assembly by the Bam Complex.* Annual Review of Biochemistry, 2011. **80**(1): p. 189-210.
57. Fleming, K.G., *A combined kinetic push and thermodynamic pull as driving forces for outer membrane protein sorting and folding in bacteria.* Philosophical Transactions of the Royal Society B: Biological Sciences, 2015. **370**(1679): p. 20150026.
58. Danoff, E.J. and K.G. Fleming, *The soluble, periplasmic domain of OmpA folds as an independent unit and displays chaperone activity by reducing the self-association propensity of the unfolded OmpA transmembrane β-barrel.* Biophysical chemistry, 2011. **159**(1): p. 194-204.
59. Gessmann, D., et al., *Outer membrane β-barrel protein folding is physically controlled by periplasmic lipid head groups and BamA.* Proceedings of the National Academy of Sciences of the United States of America, 2014. **111**(16): p. 5878-5883.
60. Bulieris, P.V., et al., *Folding and Insertion of the Outer Membrane Protein OmpA Is Assisted by the Chaperone Skp and by Lipopolysaccharide.* Journal of Biological Chemistry, 2003. **278**(11): p. 9092-9099.
61. Noinaj, N., et al., *Structural insight into the biogenesis of β-barrel membrane proteins.* Nature, 2013. **501**(7467): p. 385-390.
62. Burgess, N.K., et al., *β-Barrel Proteins That Reside in the Escherichia coli Outer Membrane in Vivo Demonstrate Varied Folding Behavior in Vitro.* Journal of Biological Chemistry, 2008. **283**(39): p. 26748-26758.
63. Kleinschmidt, J.H. and L.K. Tamm, *Secondary and Tertiary Structure Formation of the β-Barrel Membrane Protein OmpA is Synchronized and Depends on Membrane Thickness.* Journal of Molecular Biology, 2002. **324**(2): p. 319-330.
64. Hussain, S. and H.D. Bernstein, *The Bam complex catalyzes efficient insertion of bacterial outer membrane proteins into membrane vesicles of variable lipid composition.* The Journal of biological chemistry, 2018. **293**(8): p. 2959-2973.
65. Huysmans, G.H.M., et al., *The transition state for folding of an outer membrane protein.* Proceedings of the National Academy of Sciences, 2010. **107**(9): p. 4099.
66. Bond, P.J. and M.S.P. Sansom, *Insertion and assembly of membrane proteins via simulation.* Journal of the American Chemical Society, 2006. **128**(8): p. 2697-2704.
67. Pebay-Peyroula, E., et al., *X-ray Structure of Bacteriorhodopsin at 2.5 Angstroms from Microcrystals Grown in Lipidic Cubic Phases.* Science, 1997. **277**(5332): p. 1676.
68. Snijder, H.J., et al., *Structural evidence for dimerization-regulated activation of an integral membrane phospholipase.* Nature, 1999. **401**(6754): p. 717-721.

69. Cymer, F., G. von Heijne, and S.H. White, *Mechanisms of integral membrane protein insertion and folding.* Journal of molecular biology, 2015. **427**(5): p. 999-1022.
70. Gubellini, F., et al., *Physiological response to membrane protein overexpression in E. coli.* Molecular & cellular proteomics : MCP, 2011. **10**(10): p. M111.007930-M111.007930.
71. Wuu, J.J. and J.R. Swartz, *High yield cell-free production of integral membrane proteins without refolding or detergents.* Biochimica et Biophysica Acta (BBA) - Biomembranes, 2008. **1778**(5): p. 1237-1250.
72. Bannwarth, M. and G.E. Schulz, *The expression of outer membrane proteins for crystallization.* Biochimica et Biophysica Acta (BBA) - Biomembranes, 2003. **1610**(1): p. 37-45.
73. Jensen, H.M., et al., *Improving membrane protein expression and function using genomic edits.* Scientific Reports, 2017. **7**(1): p. 13030.
74. Bogdanov, M., W. Dowhan, and H. Vitrac, *Lipids and topological rules governing membrane protein assembly.* Biochimica et biophysica acta, 2014. **1843**(8): p. 1475-1488.
75. He, Y., K. Wang, and N. Yan, *The recombinant expression systems for structure determination of eukaryotic membrane proteins.* Protein & cell, 2014. **5**(9): p. 658-672.
76. Teo, A.C.K., et al., *Analysis of SMALP co-extracted phospholipids shows distinct membrane environments for three classes of bacterial membrane protein.* Scientific Reports, 2019. **9**(1): p. 1813.
77. Columbus, L., et al., *Mixing and Matching Detergents for Membrane Protein NMR Structure Determination.* Journal of the American Chemical Society, 2009. **131**(21): p. 7320-7326.
78. Valiyaveetil, F.I., Y. Zhou, and R. MacKinnon, *Lipids in the Structure, Folding, and Function of the KcsA K+ Channel.* Biochemistry, 2002. **41**(35): p. 10771-10777.
79. Seddon, A.M., P. Curnow, and P.J. Booth, *Membrane proteins, lipids and detergents: not just a soap opera.* Biochimica et Biophysica Acta (BBA) - Biomembranes, 2004. **1666**(1): p. 105-117.
80. Stuart, M.C.A. and E.J. Boekema, *Two distinct mechanisms of vesicle-to-micelle and micelle-to-vesicle transition are mediated by the packing parameter of phospholipid–detergent systems.* Biochimica et Biophysica Acta (BBA) - Biomembranes, 2007. **1768**(11): p. 2681-2689.
81. Arachea, B.T., et al., *Detergent selection for enhanced extraction of membrane proteins.* Protein Expression and Purification, 2012. **86**(1): p. 12-20.
82. Yang, Z., et al., *Membrane protein stability can be compromised by detergent interactions with the extramembranous soluble domains.* Protein science : a publication of the Protein Society, 2014. **23**(6): p. 769-789.
83. Gimpl, K., J. Klement, and S. Keller, *Characterising protein/detergent complexes by triple-detection size-exclusion chromatography.* Biological procedures online, 2016. **18**: p. 4-4.

84. Has, C. and P. Sunthar, *A comprehensive review on recent preparation techniques of liposomes.* Journal of Liposome Research, 2019: p. 1-30.
85. Khadke, S., et al., *Point of use production of liposomal solubilised products.* International journal of pharmaceutics, 2018. **537**(1-2): p. 1-8.
86. Cheng, H.-T. and E. London, *Preparation and properties of asymmetric large unilamellar vesicles: interleaflet coupling in asymmetric vesicles is dependent on temperature but not curvature.* Biophysical journal, 2011. **100**(11): p. 2671-2678.
87. Vallejo, D., et al., *Cell-sized lipid vesicles for cell-cell synaptic therapies.* Technology, 2017. **5**(4): p. 201-213.
88. Dewald, A.H., J.C. Hodges, and L. Columbus, *Physical determinants of β-barrel membrane protein folding in lipid vesicles.* Biophysical journal, 2011. **100**(9): p. 2131-2140.
89. Dürr, U.H.N., M. Gildenberg, and A. Ramamoorthy, *The magic of bicelles lights up membrane protein structure.* Chemical reviews, 2012. **112**(11): p. 6054-6074.
90. Dürr, U.H.N., R. Soong, and A. Ramamoorthy, *When detergent meets bilayer: birth and coming of age of lipid bicelles.* Progress in nuclear magnetic resonance spectroscopy, 2013. **69**: p. 1-22.
91. Yamamoto, K., et al., *Probing the Transmembrane Structure and Topology of Microsomal Cytochrome-P450 by Solid-State NMR on Temperature-Resistant Bicelles.* Scientific Reports, 2013. **3**(1): p. 2556.
92. Wu, H., et al., *Assessing the size, stability, and utility of isotropically tumbling bicelle systems for structural biology.* Biochimica et biophysica acta, 2010. **1798**(3): p. 482-488.
93. Lu, Z., et al., *Bicelles at low concentrations.* Molecular pharmaceutics, 2012. **9**(4): p. 752-761.
94. Caldwell, T.A., et al., *Low-q Bicelles Are Mixed Micelles.* The Journal of Physical Chemistry Letters, 2018. **9**(15): p. 4469-4473.
95. Morrison, E.A. and K.A. Henzler-Wildman, *Reconstitution of integral membrane proteins into isotropic bicelles with improved sample stability and expanded lipid composition profile.* Biochimica et biophysica acta, 2012. **1818**(3): p. 814-820.
96. Laguerre, A., et al., *From Nanodiscs to Isotropic Bicelles: A Procedure for Solution Nuclear Magnetic Resonance Studies of Detergent-Sensitive Integral Membrane Proteins.* Structure (London, England : 1993), 2016. **24**(10): p. 1830-1841.
97. Piai, A., et al., *Optimal Bicelle Size q for Solution NMR Studies of the Protein Transmembrane Partition.* Chemistry (Weinheim an der Bergstrasse, Germany), 2017. **23**(6): p. 1361-1367.
98. McLean, M.A., M.C. Gregory, and S.G. Sligar, *Nanodiscs: A Controlled Bilayer Surface for the Study of Membrane Proteins.* Annual Review of Biophysics, 2018. **47**(1): p. 107-124.
99. Rouck, J.E., et al., *Recent advances in nanodisc technology for membrane protein studies (2012-2017).* FEBS letters, 2017. **591**(14): p. 2057-2088.
100. Swainsbury, D.J.K., et al., *The effectiveness of styrene-maleic acid (SMA) copolymers for solubilisation of integral membrane proteins from SMA-accessible*

and SMA-resistant membranes. Biochimica et biophysica acta. Biomembranes, 2017. **1859**(10): p. 2133-2143.

101. Postis, V., et al., *The use of SMALPs as a novel membrane protein scaffold for structure study by negative stain electron microscopy.* Biochimica et biophysica acta, 2015. **1848**(2): p. 496-501.
102. Pollock, N.L., et al., *Structure and function of membrane proteins encapsulated in a polymer-bound lipid bilayer.* Biochimica et Biophysica Acta (BBA) - Biomembranes, 2018. **1860**(4): p. 809-817.
103. Simon, K.S., N.L. Pollock, and S.C. Lee, *Membrane protein nanoparticles: the shape of things to come.* Biochemical Society transactions, 2018. **46**(6): p. 1495-1504.
104. Zhang, J., X. Li, and X. Li, *Stimuli-triggered structural engineering of synthetic and biological polymeric assemblies.* Progress in Polymer Science, 2012. **37**(8): p. 1130-1176.

Chapter 2: Glycan-(in)dependent Human CEACAM1 N-Domain Dimerization

2.1 Overview

This chapter is adapted from the forthcoming manuscript, "Human CEACAM1 N-domain Dimerization is Independent from Glycan Modifications." The CEACAM protein family is localized to the plasma membrane of mammalian cells, and their major biological function is to facilitate cell-cell adhesion in healthy tissue [1]. Modes of membrane attachment and number of extracellular domains vary among CEACAM family members, but the conserved N-terminal domain is required for adhesion functionality. The Columbus laboratory specifically focuses on studying CEACAM1 adhesion at the N-terminal domain due to the prevalence of CEACAM1 in a number of late-stage cancers [2] and pathogen-associated diseases [3-5]. Prof. Kelley Moremen and Prof. James Prestegard at the University of Georgia helped us establish mammalian cell culture in the laboratory to generate CEACAM1 N-domain (CCM1) proteins with glycan modifications, which they proposed disrupted N-domain dimerization [6]. At the time of publication, the mechanism for glycan-dependent CCM1 dimer disruption was unclear because the glycosylation sites are located opposite from one of the proposed dimer interfaces. Additionally, distinguishing between oligomeric states within CCM1 glycoforms was underdetermined from a limited number of residue-specific rotational correlation times within the protein (< 5% of total residues). Through continued collaboration with the Moremen and Prestegard laboratories, we have determined that CEACAM1 N-domain dimerization appears to be

independent from glycan modifications based on site-directed spin labeling coupled to EPR and DEER spectroscopy. I have established the preferred dimer configuration of CEACAM1 N-domain glycoforms in solution. In this chapter, I additionally explore the structural, functional, and evolutionary basis for CEACAM1 N-domains maintaining a consistent dimer interface with or without glycan modifications. From results of this work, I have generated a hypothesized model for glycan interactions at the CEACAM1 N-domain during cell-cell adhesion based on *in vivo* experiments described by Patel *et al* [7].

2.2 Introduction

Carcinoembryonic antigen cellular adhesion molecules (CEACAMs) are cell surface adhesion proteins with immunoglobin (Ig)-like ectodomains. CEACAMs mediate many different functions in the cell such as proliferation, signaling, survival, and tumor suppression [1]. The diverse functional roles of CEACAMs may be attributed to structural variation within the protein family, which are encoded by 12 genes (CEACAM1, 3-8, 16, 18-21) and can exist in multiple isoforms [8] (Figure 2.1). Isoforms can be secreted by the cell or tethered to the membrane through a transmembrane helix that may include cytoplasmic signaling domains, or a glycosylphosphatidylinositol (GPI) anchor [8]. The mode of membrane attachment affects CEACAM functionality, as those with transmembrane helices and cytoplasmic domains often contribute to signaling pathways dependent on immunoreceptor tyrosine-based inhibition/activation motif (ITIM/ITAM) motifs [9]. Among the CEACAMs, CEACAM1 distribution varies extensively across tissue types, with expression on epithelial cells, endothelial cells, and immune cell types [1].

Some CEACAMs, including CEACAM1, are able to form homo- and hetero-oligomers. For example, equilibrium between monomeric and dimeric CEACAM1 on the apical cell surface is crucial for tight control of cell-cell adhesion, which is facilitated by activation of its cytoplasmic domain [7] (Figure 2.2). Loss of CEACAM1 monomer/dimer equilibrium control by the cell promotes tumor formation [10] and may be disrupted by pathogen receptors for adhesion and/or immune suppression [5, 11, 12]. The extracellular N-terminal domain (CCM) is required for the adhesion properties of CEACAM1, and it

mediates interactions with itself [13], other CEACAM isoforms [14], and pathogens [4, 5, 15, 16]. Molecular interactions that occur between N-terminal CEACAM1 domains (CCM1) are, therefore, important to understand CEACAM1 recognition by pathogens, tumor formation, and the broader physiological role of CEACAMs.

The structures of CCMs have been well-established (Table 2.1) and are remarkably similar between human isoforms (Table 2.2 and Table 2.3) [5, 17-20]. Most CCMs crystallize as dimers, even if the concentration of dimer in solution is small [19]. CCM1 in particular has a strong propensity to dimerize (homodimer K_D ~ 0.5 [19] to 2.0 μM [5]) (Table 2.1). Two distinct CCM1 dimer interfaces are reported by crystal structures to date: one occurring at the ABED β-strand interface (PDB ID: 2GK2 [17]), and the other occurring at the GFCC′C″ interface (PDB ID: 4QXW [20]) (Figure 2.3). While studies have reported CCM1 interactions at the ABED interface, the asymmetric unit can be expanded to identify GFCC′C″ interface dimers [6] (Table 2.3). Furthermore, CCM1 homotypic and heterotypic interactions are typically attributed to the GFCC′C″ interface [3, 5, 11, 19, 21-23]. Electron tomography data with soluble rat CEACAM1 ectodomains containing all four Ig domains supports dimers of the GFCC′C″ interface and shows that a distinct, minor population of dimers can occur between N-terminal domains of CEACAM1 molecules [24]. Additionally, a recent study of recombinantly-expressed CCM1 in partially-aligned media with nuclear magnetic resonance (NMR) spectroscopy favors the GFCC′C″ dimer interface. Collectively, these data suggest the dimer observed in the 4QXW structure is the dominant domain orientation in solution [6].

Understanding oligomerization behavior of CCMs is further complicated by post-translational modifications. All reported human CCM crystal structures to date lack post-translational modifications, so the atomic-level understanding of CCM dimerization in this context is limited. Three Asn residues can be glycosylated on the ABED face (N70, N77, and N81) [6], and these residues are conserved among CCM1, 3, 5, 6, 7, and 8 (Figure 2.4). Previous studies suggest that CCM glycosylation would obstruct formation of ABED-interface dimers (Figure 2.3A), although this hypothesis is not definitive due to the dynamic nature of glycans [6, 17, 24]. Cleavage of glycosylation moieties with endoglycosidase F1 to produce human CCM1 constructs with single GlcNAc modifications (gCCM1, Figure 2.5) is reported to minimize dimerization. However, CCM1 with heterogeneous complex glycan modifications (cgCCM1, Figure 2.5) does not appear to perturb oligomerization [6]. From these studies, the rationale for CCM1 dimer inhibition with small GlcNAc modifications compared to CCM1 with large, complex glycans is unclear. Furthermore, the basis for glycan-modulated CCM1 dimerization was not clear, as Asn modification sites are clustered opposite the GFCC'C'' dimeric interface that is known to occur for non-glycosylated CCM1 (Figure 2.3B). Finally, it is not known if dimer interfaces are consistent between CCM1 glycoforms.

In this work, the impact of glycosylation on CCM1 dimer orientation and glycan-modulated protein association was investigated with continuous-wave (CW) and double electron-electron resonance (DEER) electron paramagnetic resonance (EPR) spectroscopy. The solution CCM1 structure in the absence of glycosylation was determined through coupling experimental DEER data with structure refinement using the crystal structure as

a template and supports CCM1 GFCC′C″ interface dimers. Contrary to glycan-dependent disruption of CCM1 oligomerization that has been previously reported [6], all CCM1 glycoforms studied in this work are dimeric, including those with single GlcNAc modifications. Additionally, CCM1 glycoform variants possess identical domain orientations matching the GFCC′C″ interface. The persistence of CCM1 dimers with post-translational modifications coupled to known structural information across the CCM protein family supports that there may be a strong evolutionary bias for maintaining GFCC′C″ contacts. Conserved residue pairs that co-evolve along the CCM1 dimer interface are consequently described in the context of homotypic and heterotypic interactions. Interface contacts that have been evolutionarily-selected for can provide insight not only for their role in CCM-dependent cell-cell adhesion, but also interactions with proteins on other cells and pathogens. Combined, these results suggest that GFCC′C″ interface contacts are not disrupted by CCM1 glycan modifications, which may instead play alternative roles in immune recognition and/or strengthen adhesion between cells.

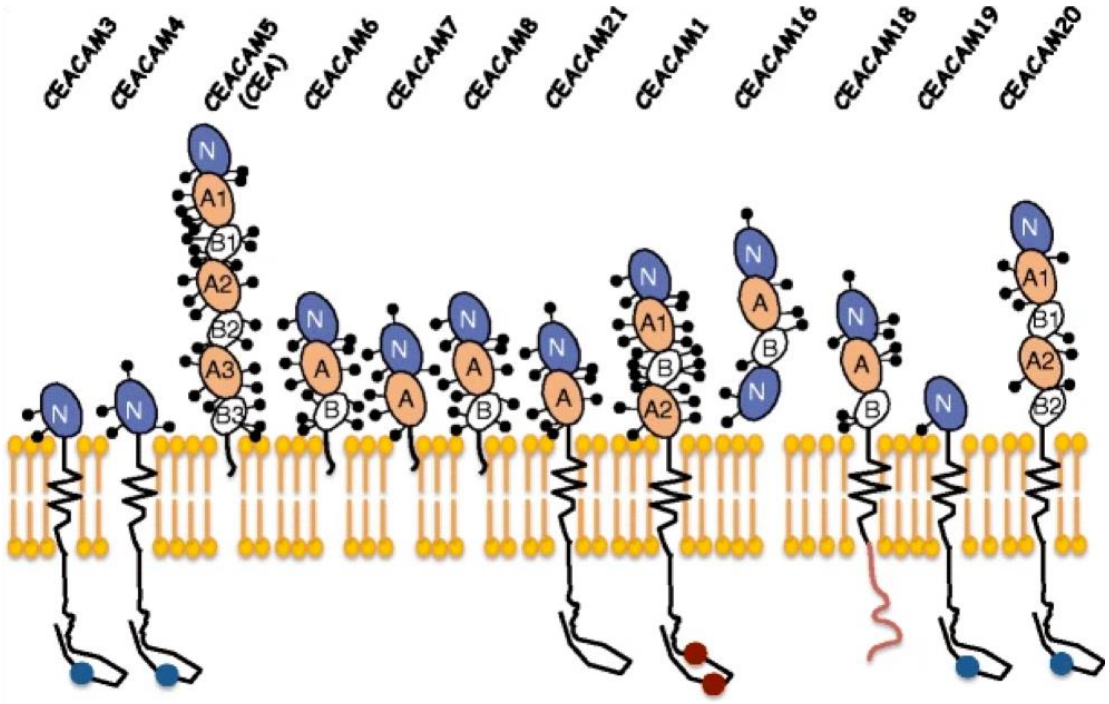


Figure 2.1: Human CEACAM family members. *Reprinted from [8] with permission.* Twelve members of the human CEACAM family generally have one variable (V)-like Ig domain (N-domain, blue) and possess a different number of constant C2-like Ig domains, denoted A (orange) or B (white). Six CEACAM family members (CEACAM1, CEACAM3, CEACAM4, CEACAM19, CEACAM20, and CEACAM21) are anchored to the cellular membrane via transmembrane domains. CEACAM5, CEACAM6, CEACAM7, and CEACAM8 are associated with the membrane through a GPI anchor. CEACAM16 is a secreted version with two N-domains and no membrane anchorage. The CEACAM1 cytoplasmic domain has ITIM motifs (red circles), whereas CEACAM3, CEACAM4, CEACAM19, and CEACAM20 carry ITAM motifs (blue circles). CEACAM extracellular domains are highly glycosylated (black stick and balls).

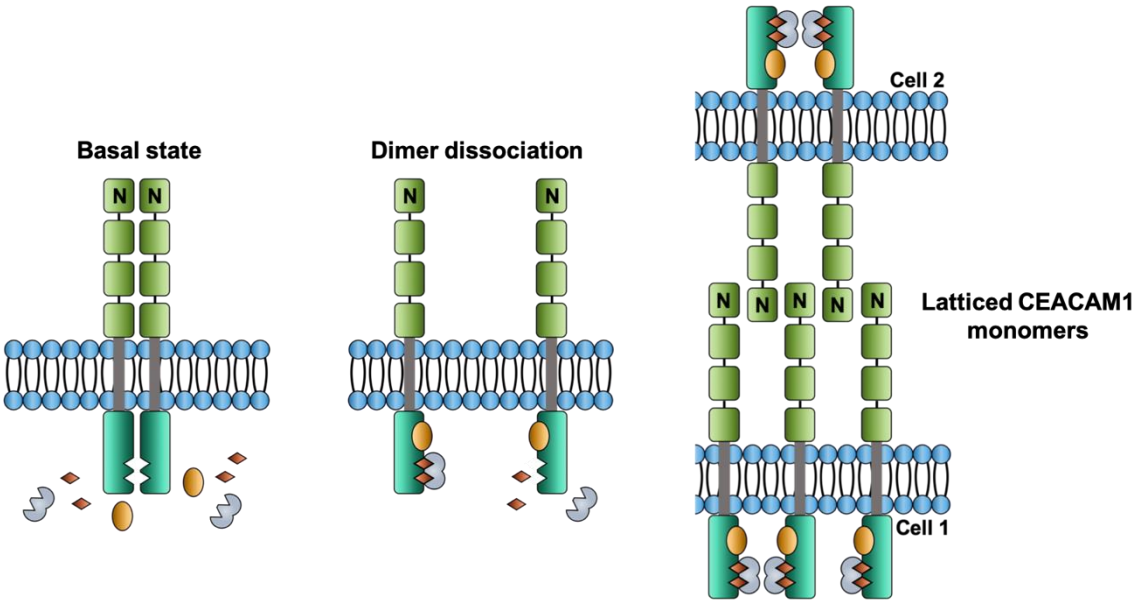


Figure 2.2: Calmodulin-mediated signaling controls human CEACAM1 oligomerization and cell adhesion properties. *Adapted from* [7] *with permission*. Ca^{2+}-activated calmodulin (yellow oval) binds specifically to the cytoplasmic domain of human CEACAM1, triggering dimer dissociation on the apical cell surface (*cis* dimers). Src tyrosine kinases (gray lobes) associate with CEACAM1 monomers, phosphorylating the cytoplasmic domain. Cell-cell adhesion occurs between CEACAM1 interactions in *trans* that are formed by latticed monomers.

Table 2.1: Recombinant human CEACAM N-domain (CCM) structures

	Oligomer	PDB ID	Dimerization K_D (μM)
CCM1	Dimer	4WHD [25], 2GK2 [17], 5DZL [20], 4QXW [20]	0.45 (+0.37/-0.23) [19], 2.8 ± 0.2 [5]
CCM3	Dimer	6AW1 [5]	N/A [5]‡
CCM5	Dimer	2QSQ [11]	1.3 ± 0.6 [19]
CCM6	Dimer	4Y8A [19], 4WHC	60 ± 15 [19]
CCM7	Dimer	4Y89 [18]	0.10 (+0.02/-0.06) [18]
CCM8	Monomer	4Y88 [19]	650 (+350/-300) [19]
CCM6–CCM8	Heterodimer	4YIQ [19], 4WTZ	2.0 ± 0.8 [19], 4.5 ± 0.5* [19]
HopQ–CCM1	Heterodimer	6AW2 [5]	0.208 ± 0.004 [5]
HopQ–CCM3	Heterodimer	6AVZ [5]	0.435 ± 0.009 [5]
Dr Adhesin–CCM5	Heterodimer	2VER [11]	13.1 ± 2.5 [4]

* Measured with glycosylated constructs expressed from HEK293 cells

‡ N/A indicates monomeric by analytical ultracentrifugation (AUC) analysis

Table 2.2: RMSD of aligned CCM monomer structures

	4WHD	2GK2	5DZL	4QXW	6AW1	2QSQ	4Y8A	4WHC	4Y89	4Y88	4YIQ	4WTZ	6AW2	6AVZ	2VER
4WHD															
2GK2	0.60														
5DZL	0.48	0.61													
4QXW	0.27	0.64	0.47												
6AW1	0.58	0.70	0.62	0.66											
2QSQ	0.52	0.61	0.46	0.45	0.69										
4Y8A	0.54	0.61	0.55	0.55	0.51	0.58									
4WHC	0.49	0.51	0.49	0.56	0.47	0.48	0.34								
4Y89	0.74	0.84	0.73	0.84	0.87	0.74	0.86	0.67							
4Y88	0.57	0.82	0.67	0.60	0.89	0.59	0.78	0.61	0.61						
4YIQ	0.50	0.69	0.58	0.61	0.63	0.66	0.48	0.46	0.73	0.39					
4WTZ	0.47	0.55	0.52	0.55	0.47	0.49	0.32	0.23	0.71	0.64	0.39				
6AW2	0.41	0.67	0.52	0.46	0.73	0.49	0.70	0.49	0.75	0.64	0.63	0.53			
6AVZ	0.58	0.80	0.59	0.64	0.55	0.65	0.65	0.48	0.91	0.70	0.59	0.45	0.40		
2VER	0.76	0.77	0.63	0.76	0.81	0.55	0.74	0.63	1.1	0.73	0.72	0.66	0.69	0.71	

Table 2.3: RMSD of aligned CCM dimer structures

	4WHD	2GK2*	5DZL	4QXW	6AW1	2QSQ	4Y8A	4WHC*	4Y89	4Y88	4YIQ	4WTZ
4WHD												
2GK2*	19.0 1.1											
5DZL	0.76	18.8 1.1										
4QXW	0.32	19.1 1.1	0.72									
6AW1	0.82	18.9 1.3	0.85	0.96								
2QSQ	0.97	18.9 1.3	0.78	0.88	1.4							
4Y8A	0.89	18.7 0.77	0.75	0.96	0.94	1.1						
4WHC*	19.0 1.9	1.3 18.7	19.0 1.9	19.3 1.9	19.0 1.7	19.2 2.4	19.0 1.6					
4Y89	0.77	19.1 1.5	1.1	0.96	1.1	1.3	1.2	19.3 2.3				
4Y88												
4YIQ	1.1	18.7 1.5	1.1	1.2	0.75	1.4	0.99	18.9 1.7	1.3			
4WTZ	0.93	18.7 1.4	0.99	1.1	0.70	1.4	0.89	19.0 1.8	1.1		0.68	

*For 2GK2 and 4WHC, the first RMSD was calculated from alignments between the biological assemblies reported in the PDB. The second number listed for an RMSD corresponds to alternative 2GK2 and 4WHC dimer structures that were identified by generating symmetry mates in the asymmetric unit.

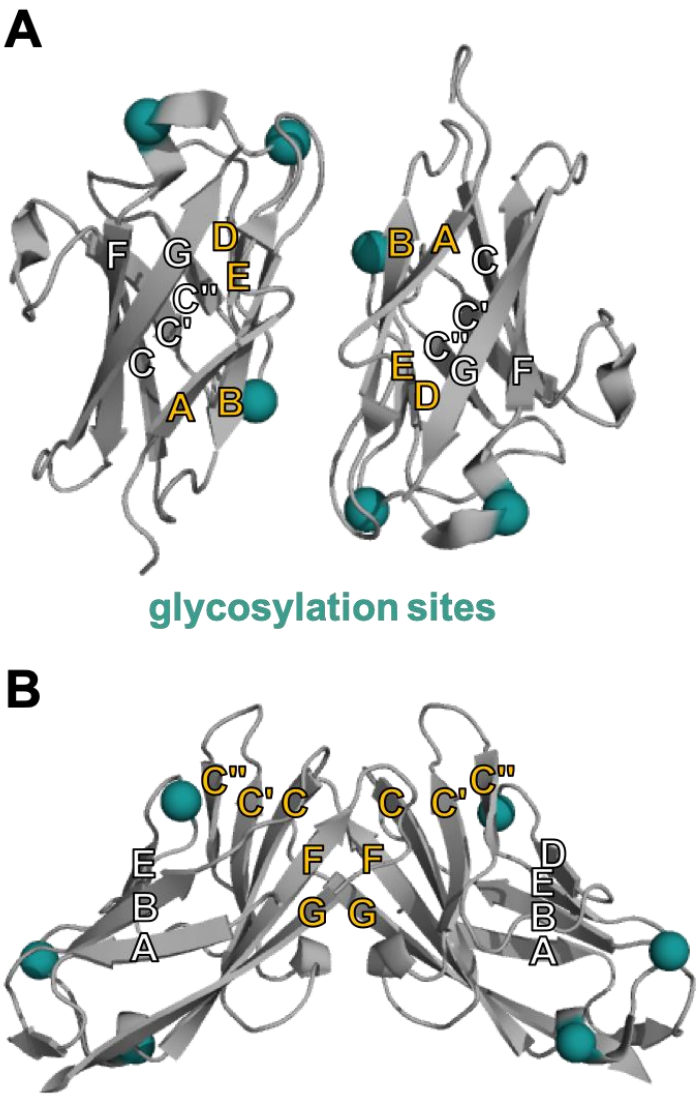


Figure 2.3: Crystal structures of CCM1 dimers. Two unique CCM1 dimer structures are reported in the literature. β-strands are labeled according to the scheme described by [17]. Published CCM1 dimer interfaces (yellow letters) are formed by strands (A) ABED (PDB ID: 2GK2 [17]) and (B) GFCC′C″ (PDB ID: 4QXW [20]). C_α atoms of N-linked glycosylation sites (teal) are rendered as spheres.

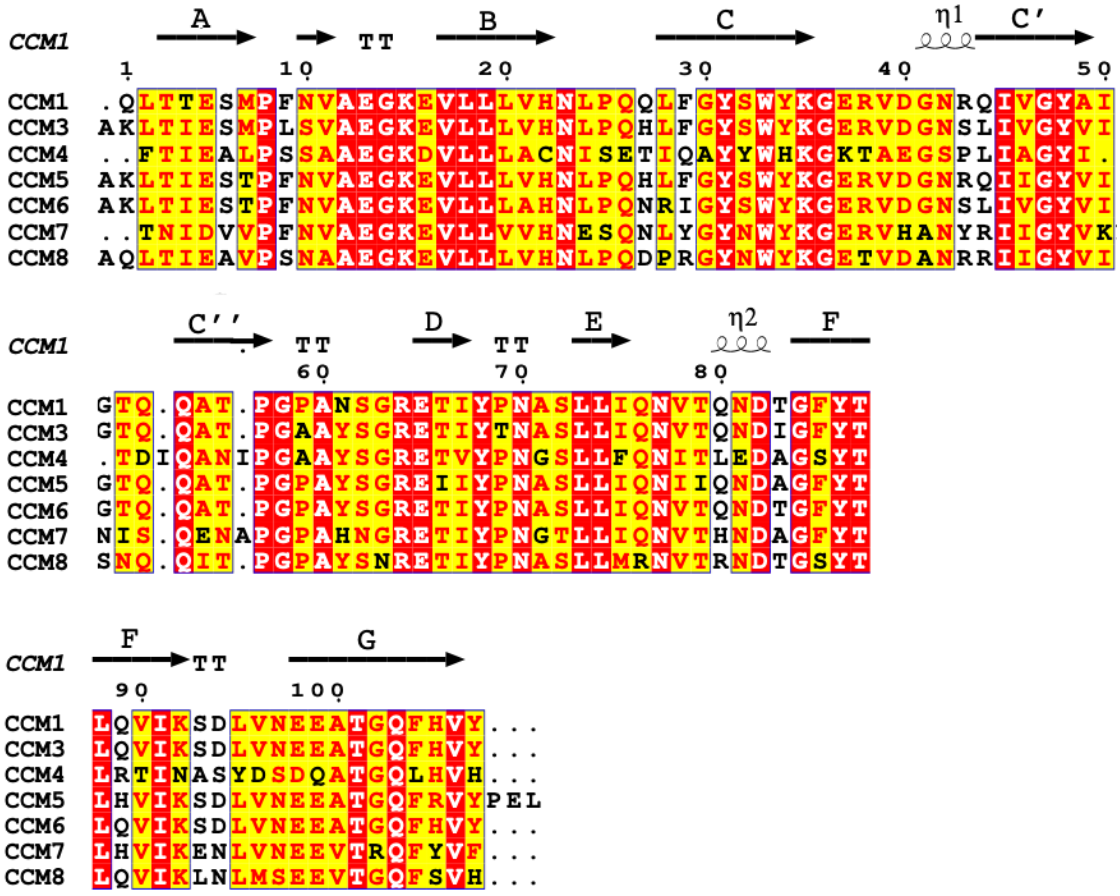


Figure 2.4: Human CCM multiple sequence alignment. N-domain sequences of human CEACAMs (CCM) -1 and 3-8 were aligned using Expresso [26]. The resulting multiple sequence alignment was uploaded into the ESPript server [27], which renders alignments based on secondary structure of the query protein (human CCM1, PDB ID: 4QXW). Arrows represent β-strand sequences, which are labeled according to the scheme reported by [17]. Strict β-turns are represented by the characters TT. Loops appear for helical structures, represented by η (310-helices). Strict identity of residues in the alignment are

rendered as red boxes with white characters, while similarities within a group are depicted as red characters. Finally, similarities across groups are filled in with yellow.

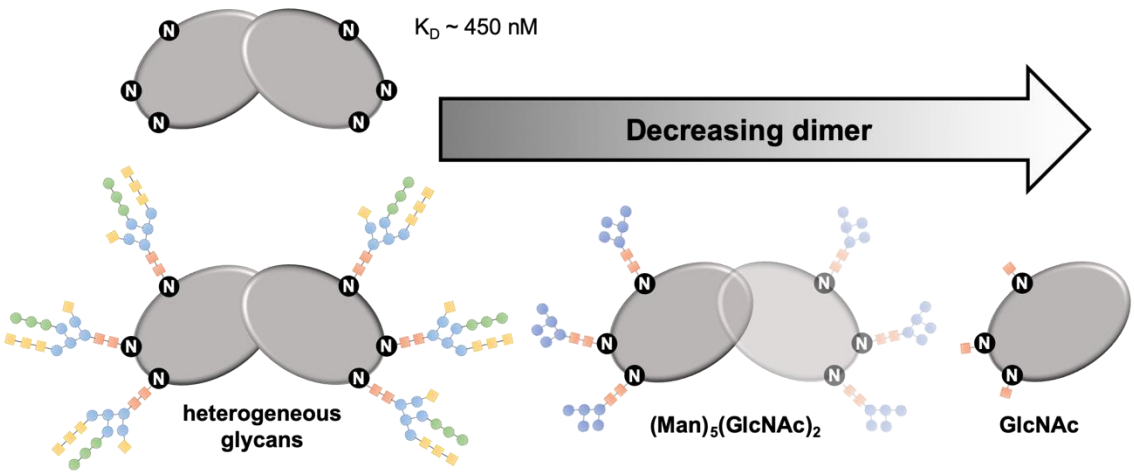


Figure 2.5: Proposed glycan-dependent CCM1 dimer formation from Zhuo *et al*. CCM1 forms a homodimer with high affinity in the absence of glycan modifications [19]. Asparagine sites that undergo N-linked glycosylation are depicted according to approximate location (assumes GFCC′C″ dimers). Zhuo *et al.* [6] reported that CCM1 with single GlcNAc modifications (gCCM1) inhibits dimer formation, while larger $(Man)_5(GlcNAc)_2$ and heterogeneously-expressed glycans limit dimers to a lesser extent. CCM1 with large, heterogeneous glycan modifications (cgCCM1) are predominantly dimers.

2.3 Materials and methods

2.3.1 Mutagenesis, expression and purification of non-glycosylated human CEACAM N-terminal domains (CCM1, 3, 8)

Cysteine mutations were introduced using PIPE mutagenesis [28]. All mutations were sequenced with primers in both directions (GENEWIZ). The expression and purification protocol of the N-terminal domain of CEACAMs (CCMs) has been published previously [16] and was adapted from Fedarovich *et al* [17]. A pGEX-2T vector containing the N-terminal D1 domain of the human *ceacam1* gene (amino acids 35-141 of the full length protein) was generously provided by Rob Nicholas (University of North Carolina at Chapel Hill). Equivalent constructs were provided for the N-terminal domains of the human *ceacam3* and *ceacam8* genes. Plasmids containing *ceacam* genes were transformed into MC1061 *E. coli* cells, which were grown in LB media to an OD_{600} 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 with a microfluidizer (Microfluidics model 110L). 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. CCM1 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 CCM1 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. CCM1 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.

2.3.2 Mutagenesis, expression, and purification of the human CEACAM1 N-terminal domain with single GlcNAc modifications (gCCM1)

Cysteine mutations were introduced using PIPE mutagenesis [28]. All mutations were sequenced with primers in both directions (GENEWIZ). An expression and purification protocol for CCM1 modified by GlcNAc (gCCM1) has been adapted from Zhuo *et al.* [6]. HEK293S GnTI$^-$ cells (which express primarily $(Man)_5$ – $(GlcNAc)_2$ glycans) were generously provided by Kelley Moremen (University of Georgia) and were maintained using FreeStyle™ 293 expression medium (Thermo Fisher Scientific) in a humidified CO_2 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 8xHis tag, an AviTag, green fluorescent protein (GFP), a TEV cleavage sequence, and the human *ceacam1* gene encoding the N-terminal domain of CEACAM1 (CCM1, residues 34-141, UniProt P13688). A 250 mL suspension culture of HEK293S (GnT1$^-$) cells in a 9:1 ratio of FreeStyle^T 293 media and EX-CELL media (Sigma) was transfected with the CCM1-pGEn2 plasmid using polyethyleneimine (Polysciences, Inc., Warrington, PA) as described previously [6]. After incubating for 24 hours, 250 mL of a 9:1 ratio of FreeStyle^T 293 EX-CELL medias and 2.2 mM valproic acid (Sigma) were added to the suspension culture. Glycosylated CCM1 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 CCM1 proteins were purified from the supernatant via Co^{2+} 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^{2+} IMAC), removing the GFP tag from CCM1 and truncating the glycans to single GlcNAc residues on CCM1 (gCCM1). Excess GFP was removed using Co^{2+} IMAC, where the flow-through containing gCCM1 proteins was collected. gCCM1 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.

2.3.3 Mutagenesis, expression, and purification of the human CEACAM1 N-terminal domain with complex glycan modifications (cgCCM1)

An expression and purification protocol for CCM1 proteins with complex glycan modifications (cgCCM1) has been adapted from Zhuo *et al.* [6] using a similar approach described above for preparation of gCCM1, with the following modifications. HEK293F cells (which express complex glycans), rather than HEK293S GnTI$^-$ cells, were used and generously provided by Kelley Moremen (University of Georgia). PIPE mutagenesis for introducing cysteine mutations, transfection of HEK293F cells with the CCM1-pGEn2 plasmid, protein expression, and IMAC purification protocols for gCCM1 production were followed. The cgCCM1 IMAC 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 cgCCM1. Further polishing steps were conducted as described for gCCM1.

2.3.4 CCM spin labeling

Fractions containing pure protein (assessed via SDS-PAGE) were concentrated, and DTT was removed with a PD-10 column. Proteins were eluted using buffer [20 mM sodium phosphate, 150 mM NaCl, and 10% glycerol] directly into five molar excess S-(2, 2, 5, 5-tetramethyl-2,5-dihydro-1H-pyrrol-3-yl)methyl methanesulfonothioate (MTSL; Toronto Research Chemicals Inc., Toronto, Canada, stored as 100 mM stock in acetonitrile). The reaction was incubated overnight in the dark at 4°C, after which excess spin label was removed with a second PD10 column. The eluted proteins were concentrated to approximately 100-200 µM. The cysteine modified with the nitroxide label is referred to as R1.

2.3.5 Continuous-wave Electron Paramagnetic Resonance spectroscopy

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. For CW-EPR measurements of samples containing ficoll, a 50% w/v solution of Ficoll® PM 70 (Sigma) in 20 mM sodium phosphate, 150 mM NaCl, 10% glycerol buffer was prepared. Protein samples were then mixed with the 50% w/v ficoll solution at a 1:1 ratio such that the final concentration was 25% w/v ficoll. Five µL of each sample (100 – 300 µM) 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).

2.3.6 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 μM 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 π/2, two 32 ns π observed pulses (at an observed frequency υ_1), and a π pump pulse (at a frequency υ_2) optimized at approximately 32 ns [29]. A pump frequency (υ_2) is set at the maximum of the nitroxide spectrum, and the observed frequency (υ_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 μs. Dipolar evolution data were processed using DEERAnalysis2016 software using Tikhonov regularization to generate distance distributions [30]. 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 (RMSD) of the best fit.

2.3.7 DEER distribution simulations

CCM1 crystal structures representing GFCC′C′′ (PDB ID: 4QXW [20]) and ABED (PDB ID: 2GK2 [17]) dimer interfaces were selected for structure refinement, with water

and ligand molecules removed. To generate MTSL-labeled CCM1 dimer structures at sites used in DEER experiments, the open-source package Multiscale Modeling of Macromolecules (MMM) [31] was used, with implementation of the spin label rotamer library approach described by Polyhach *et al* [32]. Restricted conformational space prevented MTSL labeling at site 7 for the ABED dimer structure. Sites were additionally not *in silico* labeled or used in structure refinement if clashing between neighboring R1 ensembles was observed. The m, m (-60, -60) R1 rotamer at each labeled site was selected in PyMol to give the final MTSL-labeled GFCC′C′′and ABED CCM1 parent structures.

Following *in silico* MTSL labeling with GFCC′C′′ and ABED dimer structures, DEER distributions were simulated with MMM for all experimental CCM1 DEER sites. Initial simulated DEER distributions were modeled using the predicted R1 rotamer distributions for each dimer pair. Rotamer pairs from the R1 ensembles were then used to determine the best fit of simulated to experimental data. Expected β-carbon (C_β) distances were additionally calculated with MMM.

2.3.8 Solution CCM1 structure refinement

Two separate rounds of refinement were performed for each CCM1 parent structure: (1) with spin-labeled CCM1 and (2) with unmodified CCM1. For spin-labeled structures, distance restraints were generated based on experimental DEER results and placed between the ring nitrogen atom (NS1) of the R1 side chain for identical sites of each CCM1 protomer. Distance restraints for unmodified CCM1 structures were placed between C_β atoms. With the exception of sites 12 and 16, all distance restraint uncertainties were

set to 8 Å. Distance restraints for sites 12 and 16 were set to an average distance of 65 Å with an uncertainty of 15 Å to reflect the maximum distance that would produce a DEER signal with the instrumentation and pulse sequence used in this study. Though Xplor-NIH does not currently support energy potential terms that are specific for DEER measurements, square NOE potentials serve as appropriate treatments for the purpose of this study.

Structure calculations were carried out with Xplor-NIH [33] using the following simulated annealing protocol, adapted from Sarver *et al* [34]. Initial high temperature dynamics was run by heating the protein to 3000 K for 5000 time steps. During the high temperature run, covalent energy terms for bonds, angles, and improper dihedral angles were implemented. The bath temperature was then lowered from 3000 K to 100 K in 15 K decrements using the covalent energy terms from the dynamics step in addition to the NOE potential. Potentials for non-crystallographic symmetry (NCS) and distance symmetry restraints were added to the cooling protocol to keep the CCM1 subunits identical and maintain C_2 symmetry. During both dynamics and annealing steps, the backbone atoms of the first CCM1 subunit were fixed, while the backbone atoms of the second CCM1 subunit were left free to allow for dimer rearrangement based on the DEER distance restraints. A total of 200 refined structures were generated from each CCM1 parent structure. CCM1 structural ensembles were produced from the 20 lowest energy structures, which were classified by NOE energy. Statistics for CCM1 ensemble restraint violations were generated by Xplor-NIH, and root mean square deviation (RMSD) calculations were performed with MOLMOL [35].

2.3.9 CCM1 dimer contact predictions

To identify residues that may be important in stabilizing the solution CCM1 dimer structure and thus evolutionarily selected for, the monomeric CCM1 amino acid sequence was submitted to the GREMLIN server for sequence coevolution analysis [36, 37]. A statistical model for predicted CCM1 residue contacts based on conservation and coevolution patterns was subsequently produced from a multiple sequence alignment consisting of 5,964 sequences within the CCM protein family. Residue contacts predicted from the GREMLIN server were mapped onto the CCM1 dimer structures to determine which contacts may be formed specifically across unique interfaces but not within the same protomer. Electrostatic-dependent interfacial contacts that may play a role in defining the CCM1 GFCC′C″ dimer orientation were additionally identified from structure mapping. These contacts are referenced against CCM1 GFCC′C″ dimer electrostatic potential maps at pH 7 that were generated using the APBS PyMol plugin.

2.4 Results and discussion

2.4.1 CCM1 oligomeric states persist with glycan modifications

To investigate the oligomeric state of CCM1 in solution, CW EPR spectra were recorded for 10 spin-labeled sites on CCM1 to observe if EPR line shapes are consistent with CCM1 oligomers and/or show evidence of interface contacts between protomers [38, 39]. EPR line shapes are dependent on the rotational motion of the nitroxide, which are influenced by (i) internal rotations of the nitroxide side chain, (ii) backbone motions, and (iii) domain or overall protein correlation time for spin-labeled proteins [39] (Figure 2.6).

CW-EPR is well-suited for distinguishing between CCM1 monomeric and dimeric species, as spin-labeled sites are expected to have different contributions from the overall tumbling of the protein as well as have unique EPR line shapes according to CCM1 crystal structures (Figure 2.7).

At microwave frequencies used in this study (X-band), the nuclear hyperfine splitting determines the rotational range of ~ 0.1 to 40 ns motions [40, 41]. The nitroxide correlation time (τ_n) is related to the individual dynamic modes by the following equation:

$$\frac{1}{\tau_n} = \frac{1}{\tau_s} + \frac{1}{\tau_b} + \frac{1}{\tau_p}$$

Equation 2.1

where τ_s is the correlation time of the side chain rotations, τ_b is the correlation time resulting from backbone dynamics, and τ_p is the overall protein correlation time [42]. The relationship assumes that dynamics are not correlated. Overall, protein correlation times greater than ~20 ns do not contribute to the EPR line shape [43]. Both the CCM1 monomer (11.8 kDa) and dimer (23.6 kDa) are predicted to have correlation times [44] less than 20 ns and, therefore, contribute to the EPR line shapes. For proteins with correlation times less than ~20 ns, a viscous agent such as ficoll or sucrose can be used to increase the correlation time of the protein so as not to contribute to the EPR line shape [43, 45]. The addition of a viscous agent on the line shape typically manifests as spectral component broadening [43] (Figure 2.8), of which the magnitude will be increased for CCM1 monomers compared to dimers.

A comparison of the CW EPR line shapes for the ten spin-labeled sites of CCM1 with and without ficoll (Figure 2.9) shows apparent spectral line broadening upon the addition of ficoll, which indicates that the overall protein correlation time is contributing to the nitroxide dynamics. For all of the sites except 94R1, EPR line shapes alone cannot distinguish monomeric from dimeric CCM1. However, the observed line shape for 94R1 indicates that the nitroxide is at a tertiary contact site with restricted mobility, which supports that CCM1 is an oligomer (Figure 2.10A). If CCM1 was a monomer or ABED dimer, 94R1 would be a solvent-exposed site based on the crystal structures, thus producing a mobile nitroxide side chain and EPR line shape similar to the other residues (Figure 2.7A and B). The hyperfine splitting present in the 94R1 CW-EPR spectrum alone ($2A_{zz}'$, Figure 2.10A) can additionally be used to quantify the nitroxide correlation time (τ_n) [42] with the relationship in Equation 2.1. CCM1 EPR line shapes without ficoll contain all contributions, while those with ficoll report solely on τ_s and τ_b. Therefore, CCM1 τ_p can be determined from calculating τ_n based on the hyperfine splitting ($2A_{zz}'$) quantified from 94R1 EPR spectra recorded with and without ficoll. Nitroxide correlation times (τ_n) can be estimated from experimental $2A_{zz}'$ values using the relationship described by Freed [42]:

$$\tau_{\mathrm{n}} = a \cdot (1 - S)^b$$

Equation 2.2

where constants a = 5.4 x 10^{-10} and b = -1.36 are given for a protein tumbling according to the Brownian diffusion model and a Lorentzian line width of 3 G. A ratio (S) of the measured splitting ($A_{zz}e$) to the maximum splitting for R1 (A_{zz} = 37 G) is given by:

$$S = \frac{\left(\frac{2A_{zz}'}{2}\right)}{A_{zz}} = \frac{A_{zz}e}{A_{zz}}$$

Equation 2.3

From the experimental $2A_{zz}'$ values measured from CCM1 94R1 EPR spectra, τ_n without ficoll is 3.3 ns and with ficoll is 4.5 ns. The CCM1 94R1 nitroxide correlation time with ficoll ($\tau_s^{-1} + \tau_b^{-1} = [4.5]^{-1}$ ns) can be used to solve for τ_p of CCM1according to Equation 2.1, giving an experimentally-determined CCM1 tumbling correlation time of 12.4 ns. The estimated correlation time range [44] for the CCM1 monomer is 5.5 – 7.1 ns and 9.8 – 14.0 ns for the dimer based on the monomeric molecular weight of approximately 12 kDa. Therefore, the protein correlation time determined with CW EPR experiments is consistent with a CCM1 dimer. Higher-order CCM1 oligomers (tetramers) are not apparent from CW-EPR experiments because the overall tetramer correlation time is > 20 ns, which indicates that the line shape is not expected to change with ficoll addition.

To compare with CCM1 EPR spectra, a subset of residues was chosen to spin label in CCM1 with single GlcNAc (gCCM1) and complex glycan (cgCCM1) modifications (Figure 2.11). The EPR line shapes for CCM1 and gCCM1 are very similar and upon addition of ficoll decrease in overall mobility (Figure 2.9B and Figure 2.10B-C) with the same magnitude, indicating that CCM1 and gCCM1 are identical oligomeric states – specifically, they are both dimers. Generating CCM1 94C mutations for glycosylated constructs was not successful at the mutagenesis stage, which prevented precise quantification of protein tumbling correlation times to confidently distinguish monomer

from dimer based on gCCM1 and complex-glycosylated CCM1 (cgCCM1) EPR line shapes. However, similarity of the line shapes of gCCM1 and cgCCM1 with and without ficoll indicates that different glycosylation patterns do not alter the oligomeric state of the protein.

2.4.2 CCM1 forms GFCC′C′′-interface dimers in solution

To investigate the structure of CCM1 dimers, which is evident in both glycosylated and non-glycosylated constructs, DEER-derived distances were determined for each of the labeled residues. DEER spectroscopy, like CW-EPR, is well-suited for distinguishing between CCM1 monomeric and dimeric species because monomers are expected to have no DEER signal, and different CCM1 dimers will have unique DEER distributions (Figure 2.12). The background-corrected dipolar evolution functions (DEFs) [46] and the corresponding distance distributions derived from the DEFs [30] for each of the spin-labeled CCM1 sites are shown in Figure 2.13 and Figure 2.14. All ten spin-labeled CCM1 proteins produced DEER signals, agreeing with CW-EPR results and previously established CCM1 dimer studies [6, 19]. High populations of CCM1 dimers are recapitulated with size-exclusion chromatography (Figure 2.15A), which is evident from comparing the relative elution profile of CCM1 to CCM3 and CCM8 (monomeric by analytical ultracentrifugation; [5, 19]). To demonstrate that DEER can appropriately distinguish between monomeric and dimeric CCMs using the methods in this study, CCM3 was spin-labeled and produces no DEER signal as expected for a monomer (Figure 2.15B-C).

Reliability of distance distributions in the four-pulse DEER experiment depends on the maximum dipolar evolution time, t_{max}, which is observed in the decay time of dipolar oscillations in the DEF [47]. An upper limit of 5.0 nm can reliably be achieved with t_{max} = 2 μs, scaling as $t_{max}^{1/3}$. All DEER data were obtained with $t_{max} \geq 3$ μs, corresponding to an upper limit of ~ 5.7 nm for an accurate mean distance. With the exception of A12R1 and E16R1 (Figure 2.14), all mean distances are ≤ 5.2 nm and fall within the reliable limit for data collected out to ~ 3 μs. Although non-zero DEF signals were observed for A12R1 and E16R1, these distances were greater than 6.0 nm and require $t_{max} > 3.5$ μs for generating a reliable distance distribution. Most distance distributions are unimodal with widths ranging from ~ 0.5 – 2.0 nm. Many of the narrower distributions correspond to more immobilized spin labels as assessed by CW line shapes (e.g. 94R1, 56R1, 20R1, and 9R1), and the broader distributions (e.g. 5R1) correspond to residues with more mobile line shapes (Figure 2.9A) indicating the distributions are due to conformations sampled by side chain rotamers.

Two approaches were used to investigate if CCM1 ABED (PDB ID: 2GK2) or GFCC′C′-interface (PDB ID: 4QXW) dimers are supported by the DEER-derived distance distributions. First, using the software package Multiscale Modeling of Macromolecules (MMM) [31], predicted distance distributions were calculated by modeling MTSL at each spin-labeled site for both dimer structures and then compared to the experimental distance distributions. All ten DEER distance distributions were recapitulated in MMM with the GFCC′C′-interface CCM1 dimer (Figure 2.16A and Figure 2.17) whereas only 3 experimental DEER distributions (T56R1, T83R1, and Q103R1) matched those calculated

with the ABED dimer (Figure 2.16B and Figure 2.18). At position 7R1 in the ABED dimer, *in silico* modeling with MTSL did not produce any possible rotameric states from the MMM spin label library due to restricted conformational space (Figure 2.18). Unimodal CCM1 distance distributions support that one predominant dimer structure is observed in solution, specifically a GFCC′C″-interface dimer, based on similarities between experimental DEER data and data simulated from the crystal structure. If a mixture of both GFCC′C″ and ABED-interface CCM1 dimers existed in solution (at sufficiently high concentrations), two distance populations might be expected for E5R1, M7R1, F9R1, L20R1, and D94R1 as these positions deviate in simulated DEER distributions for each dimer structure. Additionally, A12R1 and E16R1 DEFs would possess evidence of short distance contributions, which is not observed (Figure 2.14).

The second approach used to assess which CCM1 dimer structure is supported by DEER distributions was structure refinement with DEER-derived distance restraints using the software package Xplor-NIH [33]. Each DEER-refined structure was then compared to the parent crystal structure. For Xplor-NIH refinement, the lowest-energy MTSL rotamers for all R1 labels were modeled onto CCM1 GFCC′C″ and ABED dimer structures, yielding models with 9 MTSL labels each. Position 9R1 was removed from the GFCC′C″ dimer structure to prevent clashing with 7R1. Removal of 9R1 was not required for the ABED structure as 7R1 could not be modeled with MMM (Figure 2.18). Ensemble statistics improved when using distance restraints between modeled R1 nitroxide moieties compared to C_β atoms (Table 2.4). Since DEER restraints are derived from R1 nitroxide moieties, extending up to ~8 Å from the backbone, utilizing CCM1 modeled with R1 labels was

therefore preferable to unmodified crystal structures. Structure refinement of the CCM1 GFCC′C″ dimer with the DEER restraints resulted in an average structure that closely resembles the parent structure (Figure 2.19A). The average backbone RMSD between DEER-refined and parent GFCC′C″ structures was 0.4 ± 0.1 Å with no distance restraint violations (Table 2.4). In contrast, refinement of the CCM1 ABED dimer using the same DEER restraints and simulated annealing protocol produced an average structure that deviates significantly from the protomer orientations in the parent structure (Figure 2.19B). The backbone RMSD of the refined ABED dimer to its parent was 13.0 ± 0.3 Å with seven DEER restraint violations. A maximum distance restraint violation of 17.6 Å was observed for CCM1 ABED dimer refinement and corresponded to the T83R1 distance restraint (Table 2.4).

2.4.3 CCM1 homodimer conformers are consistent between glycoforms

To determine the glycosylated gCCM1 dimer conformation in solution, DEER distributions were obtained for E5R1 and Q103R1 gCCM1 and compared with non-glycosylated CCM1 measurements (Figure 2.20). Neither gCCM1 spin-labeled mutants change the mean DEER distribution with respect to non-glycosylated CCM1, with E5R1 and Q103R1 gCCM1 producing mean distances of 5.0 nm and 3.4 nm, respectively. Furthermore, both glycosylated states of CCM1 (gCCM1 and cgCCM1) possess the same Q103R1 DEER distributions and CW-EPR spectra (Figure 2.21). Since glycosylated CCM1 has the same DEER distributions as CCM1, this suggests that the GFCC′C″-interface dimer is the dominant species in solution with or without glycan modifications.

Previous reports of glycosylation modulating CCM1 dimers noted that the phenomenon was difficult to rationalize, as the glycosylated residues N70, N77, and N81 are located far from the GFCC′C″ interface [6]. Aside from CCM1, only CCM6-CCM8 heterodimerization studies have been performed with glycosylated constructs *in vitro*. The CCM6-CCM8 heterodimer (PDB ID: 4YIQ) and CCM1 homodimer (4QXW) rmsd is 1.2 Å (Table 2.3), and the K_D of each complex is ~2 μM and 450 nM, respectively [19] (Table 2.1). Glycosylated CCM6-CCM8 was reported to decrease the heterodimer affinity ~ two-fold (4.5 μM, Table 2.1) [19]. A similar reduction in CCM1 homodimer affinity (CCM1 K_D = 2.8 μM (Table 2.1) to K_D ~ 6-7 μM with glycosylation) would result in ~7% glycosylated CCM1 monomer at concentrations used for NMR and EPR experiments (~100 μM). Seven percent CCM1 monomer under conditions for this study is not expected to noticeably change the observed DEER DEF echo intensities and EPR line shapes.

2.4.4 GFCC′C″ interface contacts are predicted from co-evolution patterns

Due to the prevalence of CCM1 GFCC′C″ interface dimers observed both *in vitro* and *in vivo*, the evolutionary basis for its formation across the CCM protein family was investigated by co-evolution based prediction of residue-residue contacts with the GREMLIN server [37, 48]. Co-evolution observations of CCM proteins have not been described to date and may be important for identifying GFCC′C″ interface contacts that are relevant to dimer formation and evolutionarily selected for, supplementing experimental CCM1 mutagenesis studies. Using the monomeric CCM1 protein sequence, contact predictions were generated from conservation and co-evolution patterns based on

a multiple-sequence alignment (MSA) containing 5,964 homologous, non-redundant (no pair of sequences > 90% identical) protein sequences. Contact predictions are most accurate when the number of sequences in the MSA is at least five times the query sequence length [37]. Given the CCM1 query sequence length of 106 amino acids, the sequence/length parameter was well above the minimum value required for accurate contact predictions.

The evolutionarily-coupled residues (ECs) within the CCM1 sequence predicted to be most probable by the GREMLIN server are provided in (Table 2.5). Each EC pair was mapped onto both the CCM1 GFCC′C″ and ABED dimer structures. To identify EC pairs at the dimer interface, a 5 Å distance separation cutoff between residues in opposite protomers was applied using terminal side chain atoms. Based on the initial filter, Figure 2.22A shows the following proposed GFCC′C″ interface residues that are evolutionarily coupled: Y34/Q89 (prob. = 0.999), Y34/E99 (prob. = 0.551), and G41/N97 (prob. = 0.393). EC pairs identified across the CCM1 ABED interface (Figure 2.22B) are L18/N70 (prob. = 0.912) and L18/Y68 (prob. = 0.848). To distinguish between intraprotomer ECs and those specific to GFCC′C″ and ABED interfaces, each EC pair was additionally mapped onto the same protomer in CCM1 crystal structures (Figure 2.22). Using an intraprotomer distance separation cutoff of 5 Å between terminal side chain atoms, only Y34/E99 and G41/N97 EC pairs specific to the CCM1 GFCC′C″ interface could be identified as residues that are coupled and at a dimer interface (rather than a tertiary contact within the CCM1 monomer fold).

Basing meaningful interpretations of CCM1 co-evolved dimer interface residues on EC scores alone is difficult. However, extensive mutagenesis studies of CCM proteins both *in vivo* and *in vitro* have been performed to date, which crucially allows assessment of CCM1 ECs using both bioinformatics data and published experimental observations. For example, neither Y34 nor Q89 play roles in CEACAM1 homophilic adhesion [22] and dimer formation [5] even though the pair's EC probability score is considered significant (0.999). However, Y34 in CEACAM5, but not the analogous residue to Q89 (H89) [11] is suggested to be important for homophilic adhesion [21]. Mutagenesis studies of Y34 and Q89 combined with mapping the EC pair onto the CCM1 structure thus supports that this pair has likely not co-evolved to form the GFCC′C″ dimer interface. However, both Y34 and Q89 are adhesiotopes crucial for binding Opa proteins, Dr adhesins, and HopQ proteins expressed by *Neisseria* bacteria [3], pathogenic *Escherichia coli* [11], and *Helicobacter pylori* [5], respectively.

The proposed CCM1 dimer interface pair Y34/E99 is highly conserved among the CCM protein family (Figure 2.4), except for CCM4 (H34 and Q99). E99 point mutations have only been performed with recombinantly-expressed CCM5, in which the E99A mutation converts CCM5 from a dimer to a monomer in solution but does not impact Dr-adhesin binding [11]. It is likely that CCM1 E99 plays a similar role in dimer formation, as both CCM1 and CCM5 are highly related in terms of residues important for CCM recognition, homotypic and heterotypic interactions, as well as homodimer affinities [11]. Finally, the EC pair G41/N97 is the most convincing CCM dimer interface pair based on structure mapping, albeit with a low EC probability score (0.393). Recently, mutation of

CCM1 N97A was shown to convert the protein from dimer to monomer [5]. G41 and N97 additionally form hydrophilic contacts at the CCM5 dimer interface [11]. While N97A appears to disrupt the CCM1 homodimer in solution, the mutation has little effect on binding HopQ [5]. CCM G41 on the other hand plays an important role in binding select Opa proteins [3] and HopQ [5], though G41 mutations have not been explored in CCM homodimerization. Combined, data from evolutionary coupling analysis and previous CCM literature support that Y34/E99 and G41/N97 are residue pairs that have likely co-evolved to facilitate CCM1 homodimer interactions. We additionally propose that E99 and N97 are most crucial for maintaining the CCM1 and possibly CCM5 dimer interface, strengthening experimental observations for CCM1 and CCM5 specificity [11].

Aside from information about the CCM1 dimer interface, EC analysis of residue pairs with high probability scores (> 0.7) may provide crucial insight into residues that play general roles in CCM dimer orientation and interactions with proteins expressed on pathogen surfaces. The residue pair R64/D82 ranked third in EC scores (probability = 1.0) and was previously found to form a critical salt bridge required for maintaining the CCM1 protein fold that indirectly impacts its general adhesion function [22]. A key feature of the CCM1 dimer is that the protomers are "wedged" apart, owing to a positively-charged patch on the protein surface and network of salt bridges underneath the GFCC′C″ face (Figure 2.23). These residues unsurprisingly appear in the EC analysis, such as E37/D40 (prob. = 0.998), R38/R43 (prob. = 0.775), and E37/R43 (prob. = 0.621). Residue pairs that serve as adhesiotopes for binding *Neisseria* Opa proteins, *E. coli* Dr adhesins, and *H. pylori* HopQ proteins similarly have high EC scores, including I91/V96 (prob. = 0.996), S32/Q89 (prob.

= 0.988), S32/I91 (prob. = 0.984), S32/Q44 (prob. = 0.941), and Y34/Q44 (prob. = 0.709) [3, 5, 11]. Interestingly, glycosylation sites N70, N77, and N81 do not constitute many EC pairs, with L20/N70 (prob. = 0.968) and L18/N70 (prob. = 0.912) separated by more than 5 Å. Neither L18 nor L20 play roles in CEACAM1 adhesion [22]. A pitfall of utilizing EC analysis for residues relevant to CCM1 dimerization is that identical residues between pairs cannot be identified, specifically F29, R38, V39, Q89, and I91. Among these residues, V39 is crucial for CEACAM1 and CEACAM5 homophilic interactions and dimerization [11, 22] as well as important for heterophilic interactions [3, 11]. F29 mutations additionally reduce CEACAM dimerization [11] and, along with I91, is implicated in binding Dr adhesins [11] and Opa proteins [3].

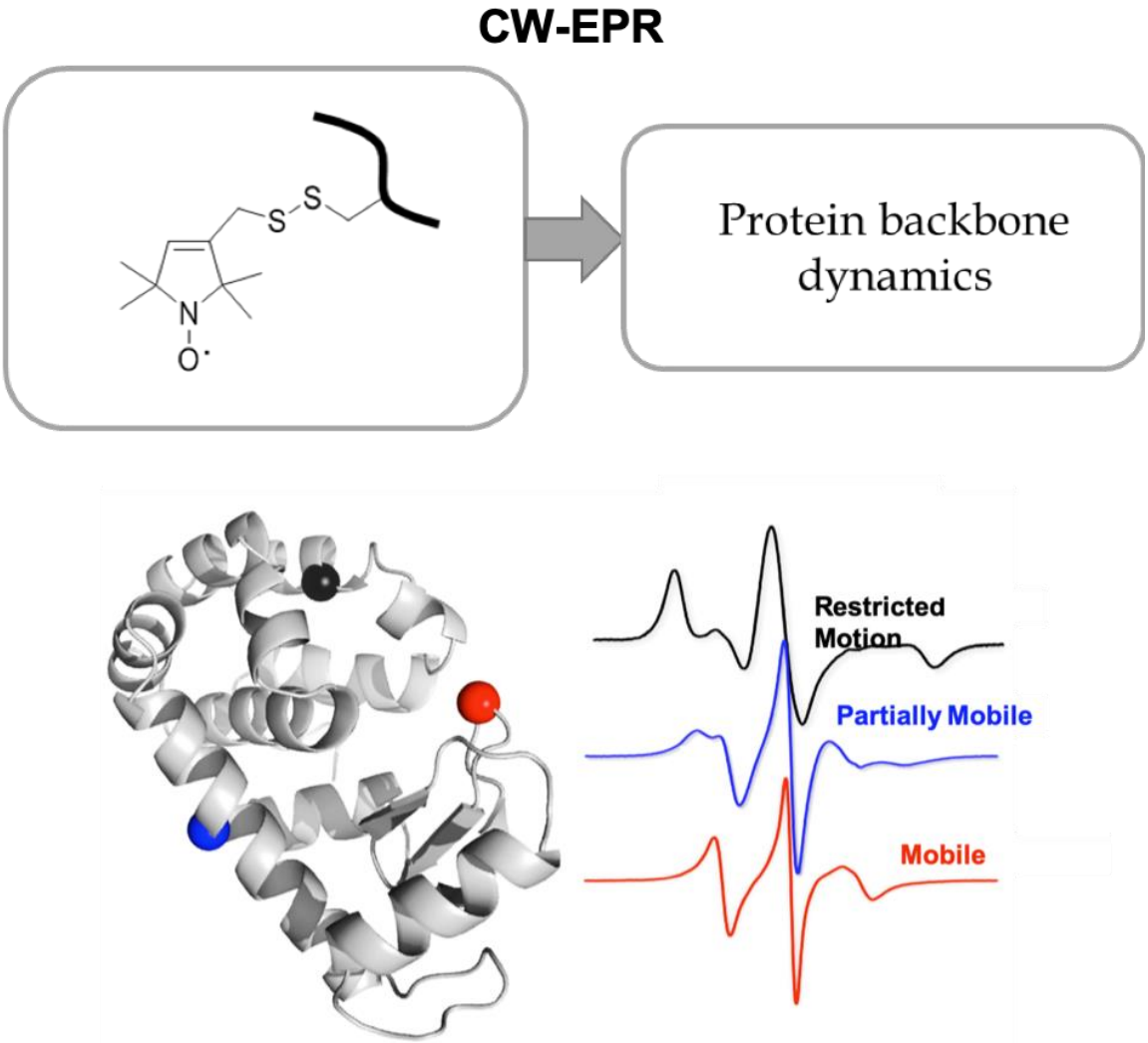


Figure 2.6: Continuous-wave EPR spectra report on protein backbone dynamics. *Adapted from [45] with permission.* General approach used for site-directed spin labeling described in this work. Cysteine point mutations within the protein sequence are used to conjugate thiol-reactive MTSL label (R1), which possesses a nitroxide moiety. The label's paramagnetic center may be exploited by continuous-wave EPR (CW-EPR) spectroscopy methods to study protein backbone dynamics. Characteristic CW-EPR spectra for MTSL-

labeled T4 lysozyme at buried (black), solvent-exposed helix (blue), and loop (red) sites are shown.

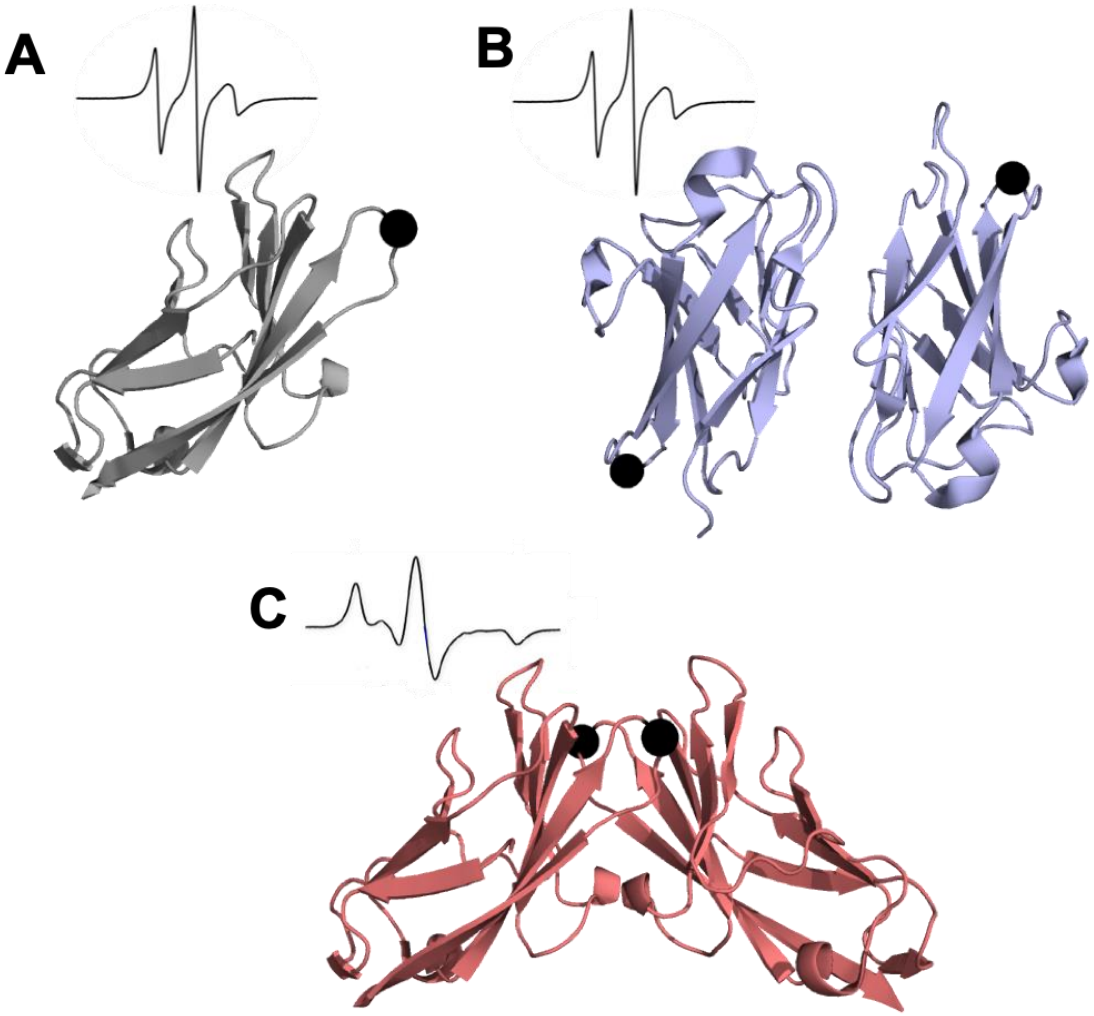


Figure 2.7: Expected CW-EPR spectra for CCM1 monomer and dimers using a selected site. Spin-labeled CCM1 sites may be used to distinguish monomeric and dimeric species based on CW-EPR line shapes according to the examples for T4 lysozyme shown in Figure 2.6. (A) The selected CCM1 spin-labeled site is located on a loop, which is expected to produce a CW-EPR spectrum reflecting a mobile nitroxide side chain if CCM1 is a monomer. (B) If ABED-interface CCM1 dimers are present in solution, the same spin-labeled site is expected to remain as a solvent-exposed loop and show a similar mobile line

shape as monomeric CCM1. (C) If GFCC′C′′-interface CCM1 dimers form, the site shown is expected to bury the label within the dimer interface and produce an EPR line shape reflecting a nitroxide side chain with restricted motion.

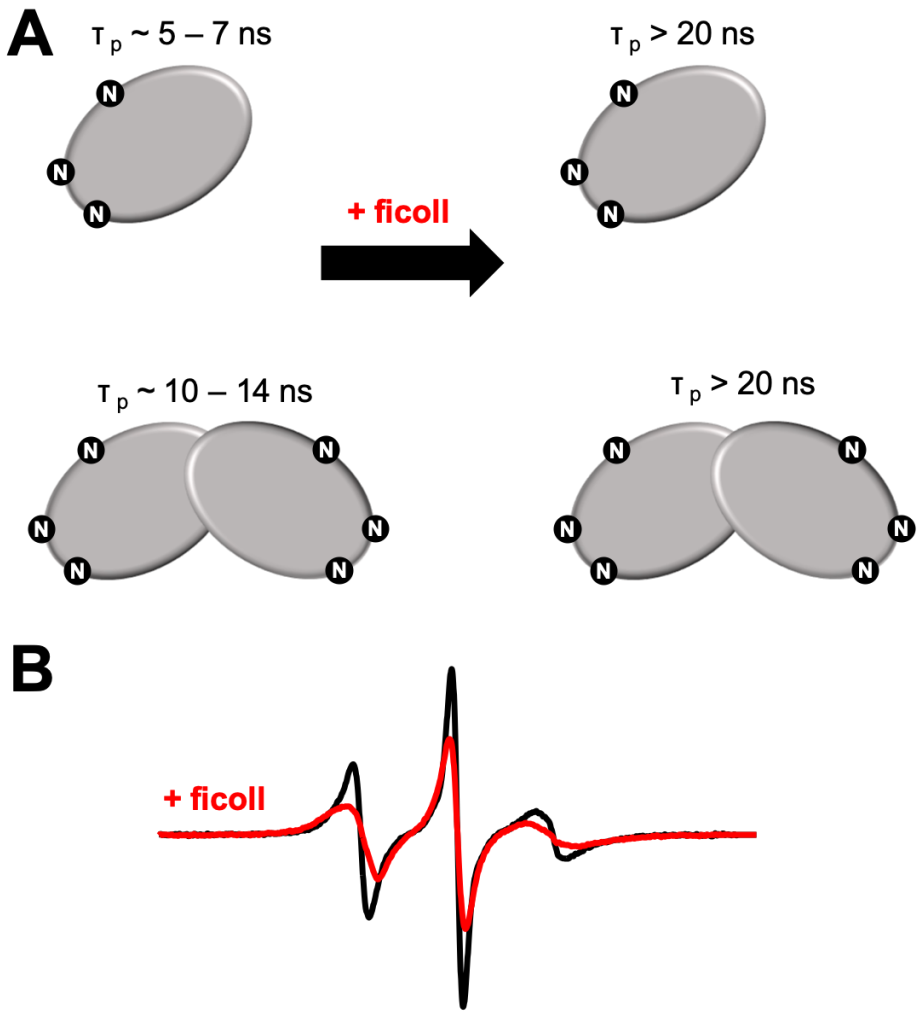


Figure 2.8: Expected CCM1 CW-EPR spectra upon ficoll addition. (A) Both the CCM1 monomer and dimer are predicted to have correlation times [44] less than 20 ns and contribute to EPR line shapes [43]. Three GlcNAc modifications per subunit are expected to increase the CCM1 correlation time by ~ 0.5 ns to 1.0 ns for monomer and dimer, respectively. Addition of ficoll is expected to increase the correlation time for monomeric and dimeric CCM1 to > 20 ns, which will broaden the EPR spectral components (B).

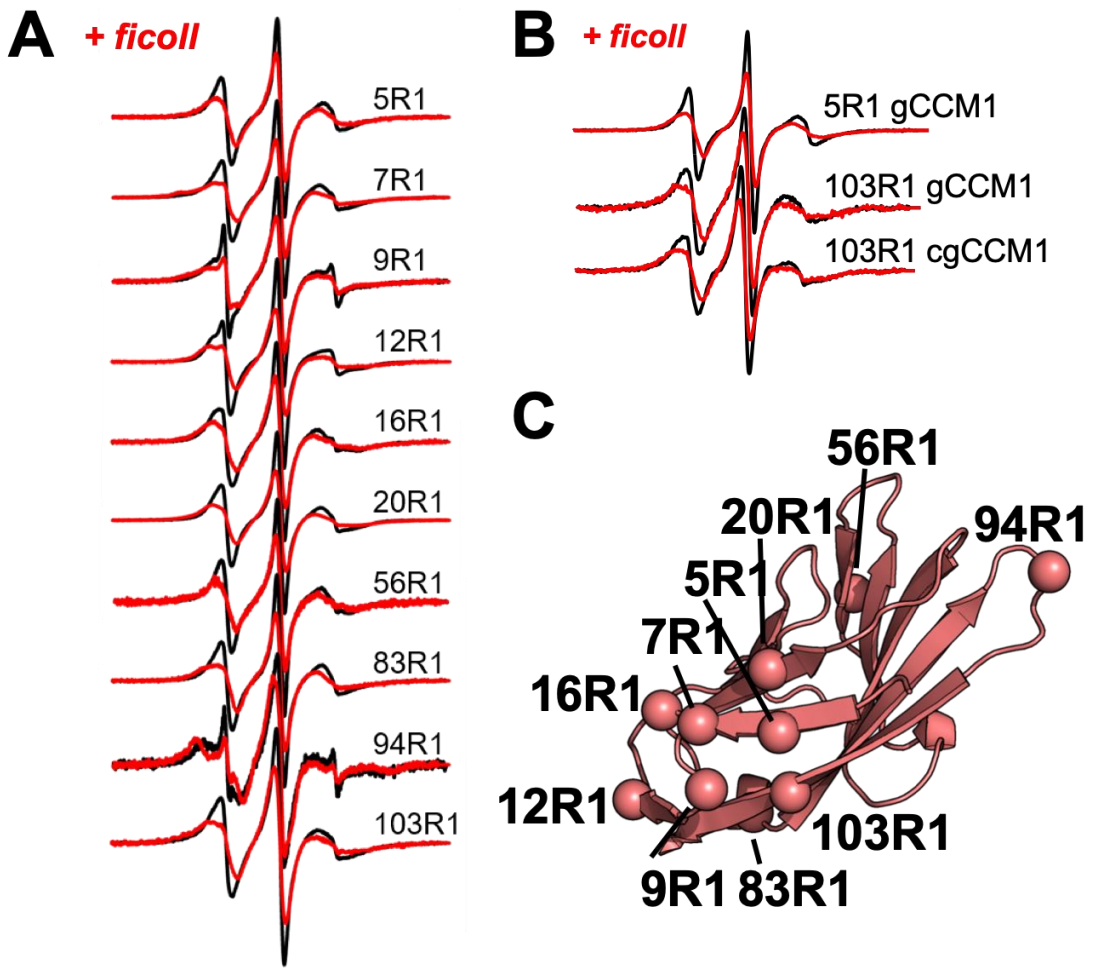


Figure 2.9: CCM1 with various glycan modifications is predominantly dimer in solution. Continuous-wave (CW) electron paramagnetic resonance (EPR) spectra with (red) and without (black) ficoll of all spin-labeled CCM1 samples generated in this study: non-glycosylated CCM1 (A) and CCM1 with either single GlcNAc (gCCM1) or complex glycan (cgCCM1) modifications (B). With the exception of CCM1 94R1, all spectra exhibit line broadening in the presence of ficoll, which suggests that the main population of both non-glycosylated and glycosylated forms of CCM1 corresponds to the dimeric state. (C) Spin-labeled sites of CCM1 investigated in this study are represented by spheres.

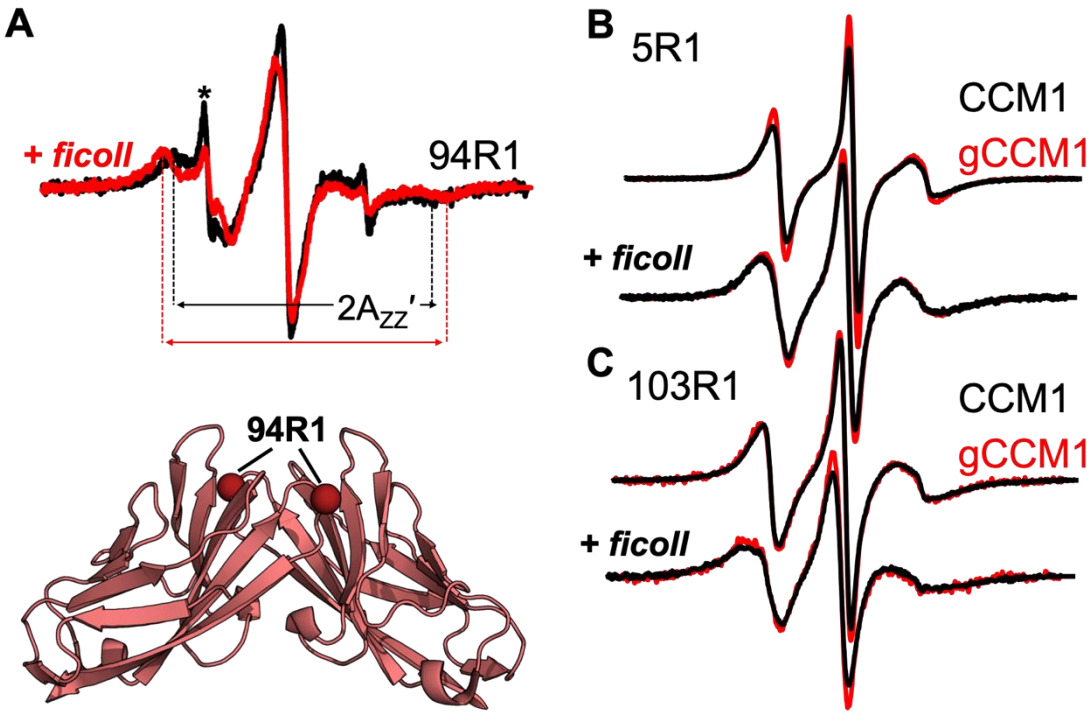


Figure 2.10: CCM1 and gCCM1 have the same solution oligomeric state. (A) The EPR lineshape of 94R1 indicates that the nitroxide is highly restricted and at a buried site. Spectra were recorded both with (red) and without (black) ficoll. The effective hyperfine splitting, 2Azz′, of restricted nitroxides can be used to estimate the nitroxide correlation time. Asterisk denotes component reflecting the presence of free spin label in the sample. The buried CCM1 94R1 line shape is consistent with GFCC′C′′-interface dimers rather than ABED-interface dimers or monomeric CCM1. EPR spectra of spin-labeled CCM1 (black) and gCCM1 (red) for 5R1 (B) and 103R1 (C) with and without ficoll show similar levels of line-broadening between samples, suggesting CCM1 and gCCM1 have similar populations of dimer.

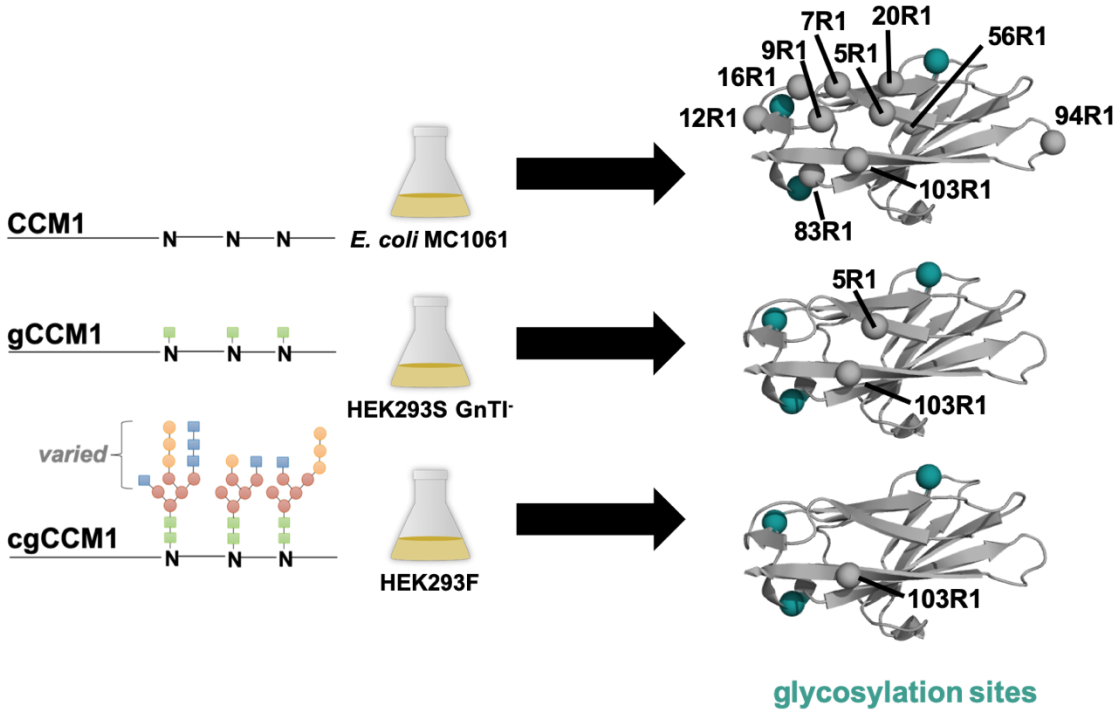


Figure 2.11: Schematic of spin-labeled CCM1 constructs. Overview of non-glycosylated and glycosylated CCM1 proteins generated in this study. Ten non-glycosylated CCM1 variants were spin-labeled as well as several spin-labeled variants of CCM1 with GlcNAc (gCCM1) and complex glycan modifications (cgCCM1). R1 (gray spheres) is the designation of the thiol-reactive MTSL spin label (Figure 2.6) covalently attached to the protein via cysteine point mutations. Asparagine residues specific to N-linked CCM1 glycosylation are shown as teal spheres on the monomer (PDB ID: 4QXW [20]).

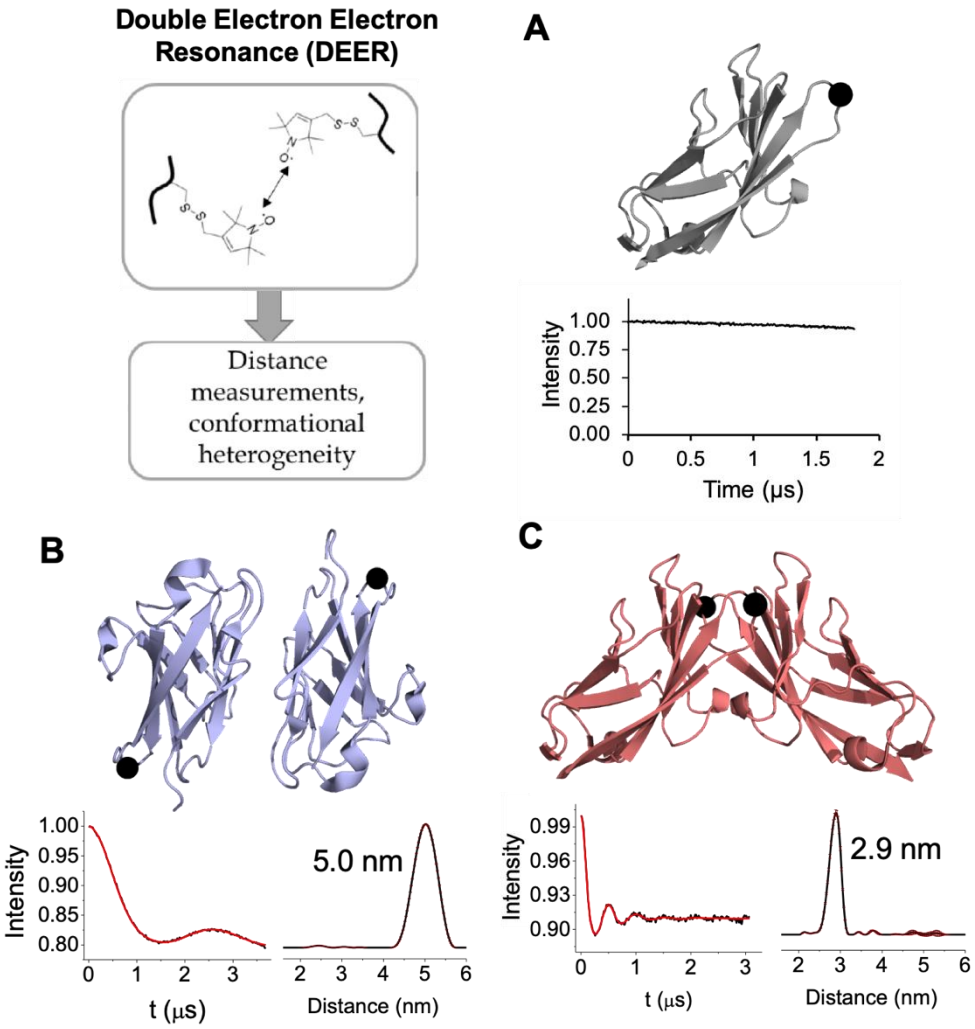


Figure 2.12: Expected DEER distributions for CCM1 monomer and dimers using a selected site. Double electron-electron resonance (DEER) spectroscopy is used to determine distance measurements between two paramagnetic spin labels (R1 in this study). (A) Monomeric CCM1 will produce no DEER signal. (B) A DEER signal (left, black trace) will be observed for R1 distance separations between ~2.5 – 6.0, which is fit with Tikhonov regularization (left, red trace) to generate distance distributions (right, black trace) [30]. Error within the distance distribution (right) is plotted as maroon bars. An expected DEER

distance distribution between two R1 labels for CCM1 ABED dimers is given. (C) If CCM1 GFCC′C′′ dimers are present, the same R1 labels will produce an average distance < 3.0 nm less than the observed distribution for ABED dimers.

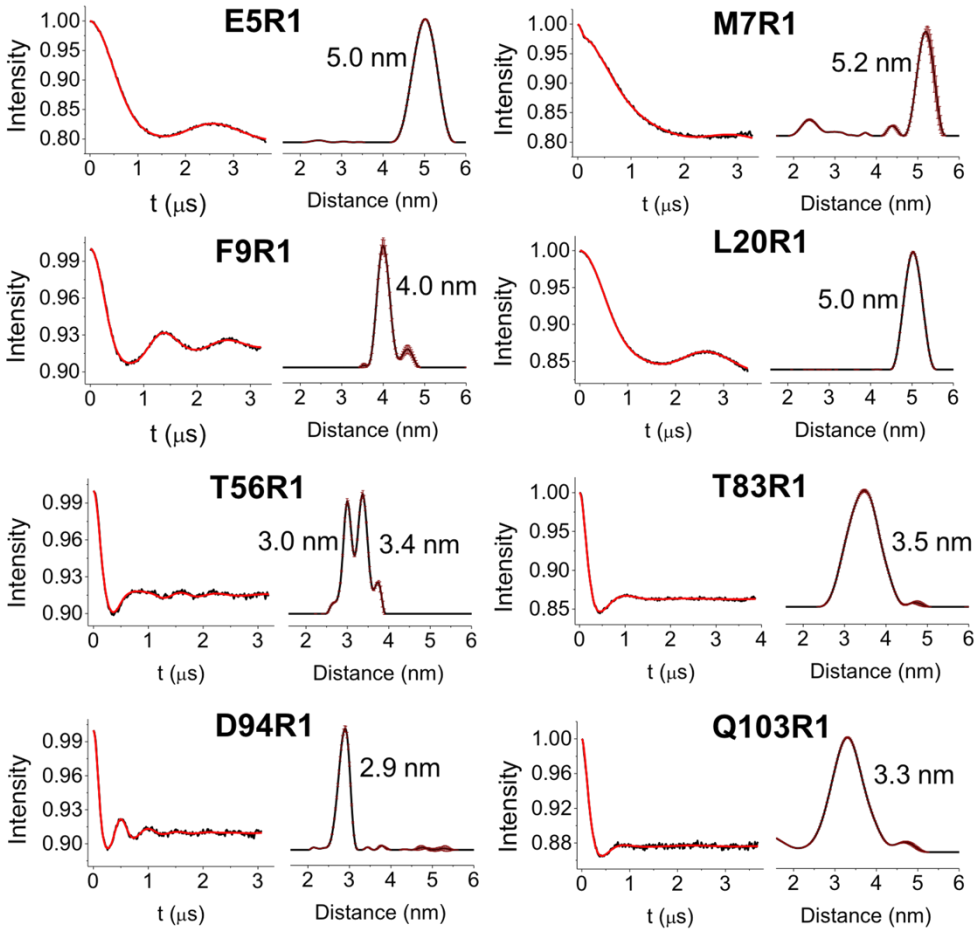


Figure 2.13: CCM1 DEER distance distributions. Background-corrected dipolar evolution functions (DEF; left, black trace) fit with Tikhonov regularization (left, red trace) for the indicated spin-labeled CCM1 mutants. The resulting normalized DEER distance distributions (right, black trace) were generated from the fits to the DEF for pairs of R1 residues in the CCM1 dimer. Error within the distance distribution is plotted as maroon bars (right). Average distances are noted in nm.

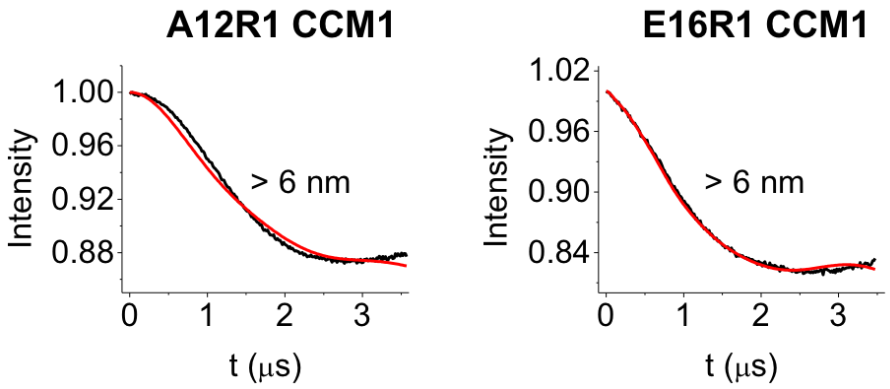


Figure 2.14: CCM1 A12R1 and E16R1 DEER dipolar evolution functions. Background-corrected dipolar evolution functions (DEF) for spin-labeled CCM1 mutants that contain a distance contribution greater than 6 nm. The maximum dipolar evolution time is not sufficient to generate reliable DEER distributions from the fit DEFs, which are therefore excluded for this sample set.

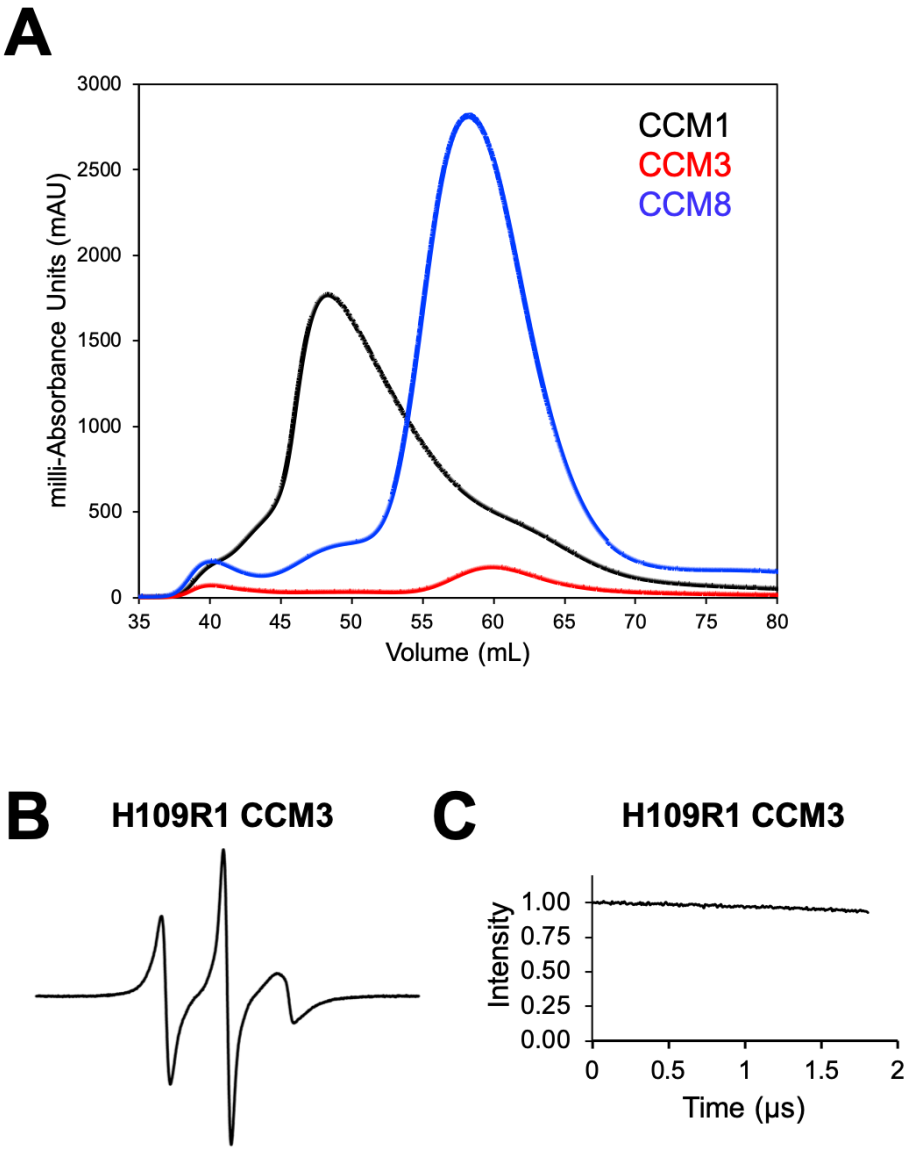


Figure 2.15: Size-exclusion chromatography and DEER results from additional CCM family members. (A) A_{280} traces of CCM1 (black), CCM3 (red), and CCM8 (blue) elution by size-exclusion chromatography. The major elution peak of CCM1 relative to CCM3 and CCM8 (monomers) suggests that CCM1 predominantly exists as a dimer (elution at ~50 mL) with a minor population of monomer (elution at ~60 mL). CCM3 H109R1 produces

a CW-EPR signal (B) but no DEER signal (C), which is expected for monomeric CCM proteins.

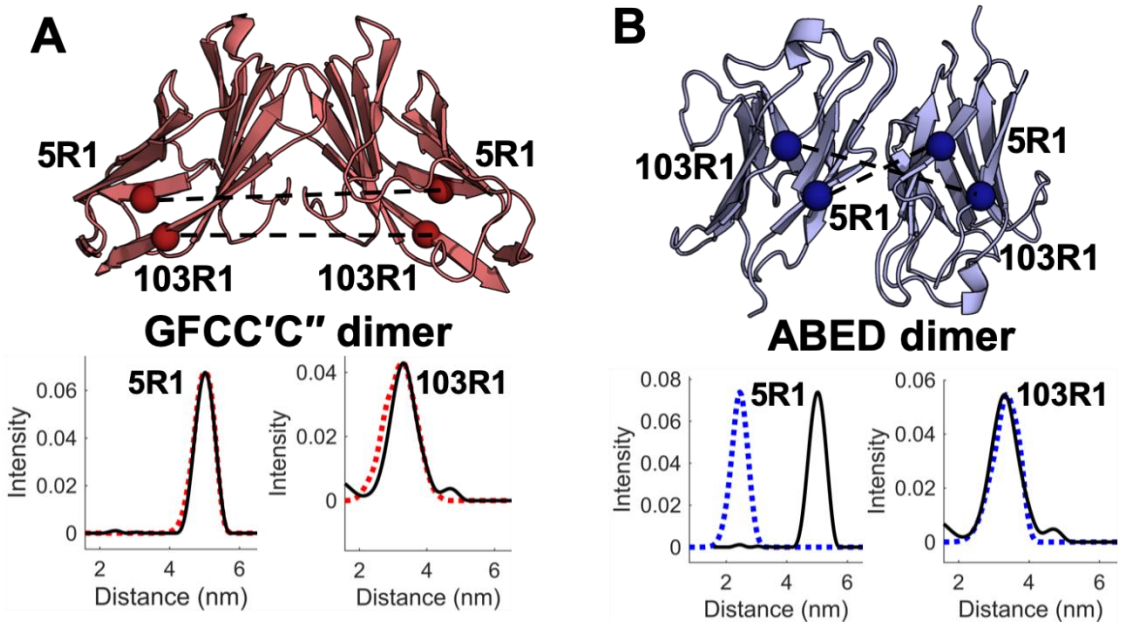


Figure 2.16: CCM1 DEER distance distributions demonstrate that GFCC′C″-interface dimers are preferred in solution. Simulated and experimental DEER distances for a subset of CCM1 R1 labels (5R1 and 103R1) based on GFCC′C″ (A, red) and ABED (B, blue) dimer structures. (Top) dashed lines shown on the crystal structures indicate separations between labeled sites (spheres). (Bottom) simulated (dashed trace) and experimental (solid trace) DEER distance distributions for 5R1 and 103R1 based on GFCC′C″ and ABED dimer structures. The chosen R1 subset represents a distance which is expected to be different (5R1) and the same (103R1) between conformers.

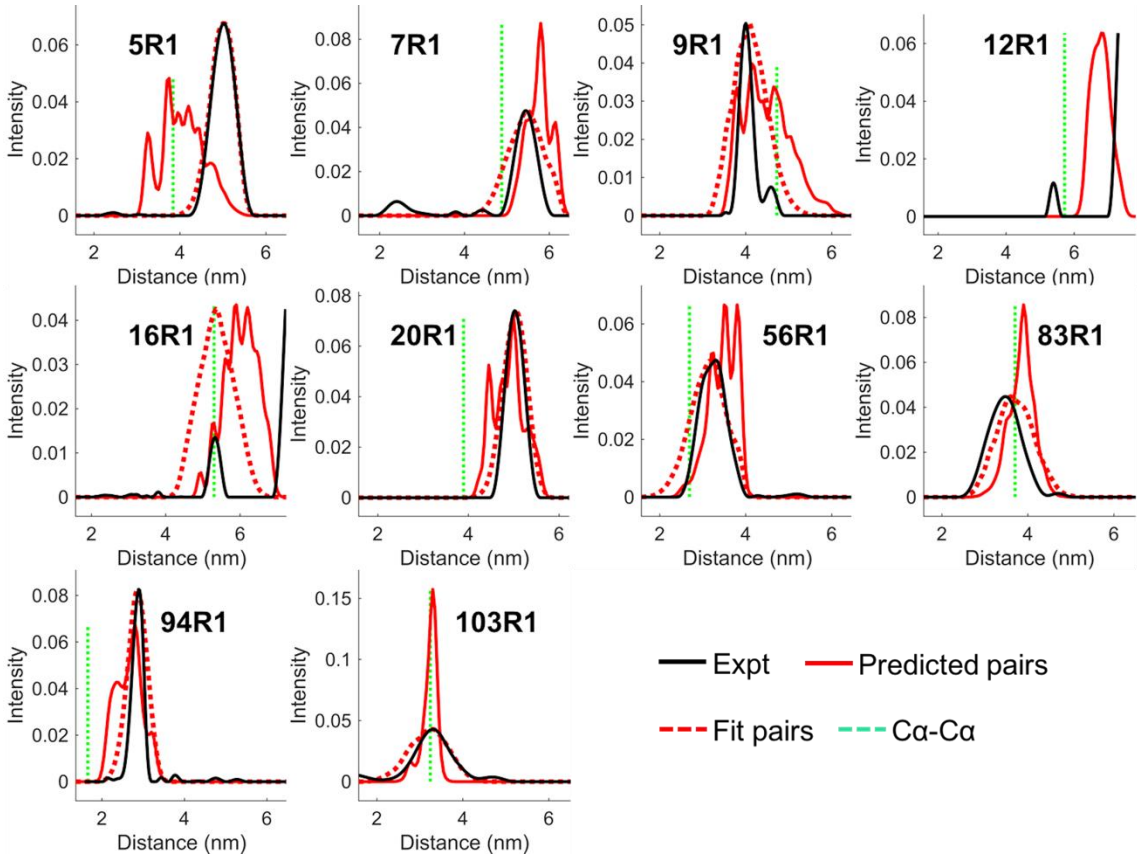


Figure 2.17: Comparison of simulated and experimental DEER distributions for CCM1 GFCC′C″ dimers. Spin-labeled CCM1 was modeled with the GFCC′C″ dimer structure, followed by generation of calculated DEER distributions using MMM [49]. For each modeled CCM1 label, distributions were calculated for all R1 side chain rotamers (solid red trace) and selected rotamers (dashed red trace) based on fits to the experimental distance distributions (black trace). The average predicted distance between alpha carbon atoms of modeled R1 labels (dashed green line) are additionally included from MMM calculations.

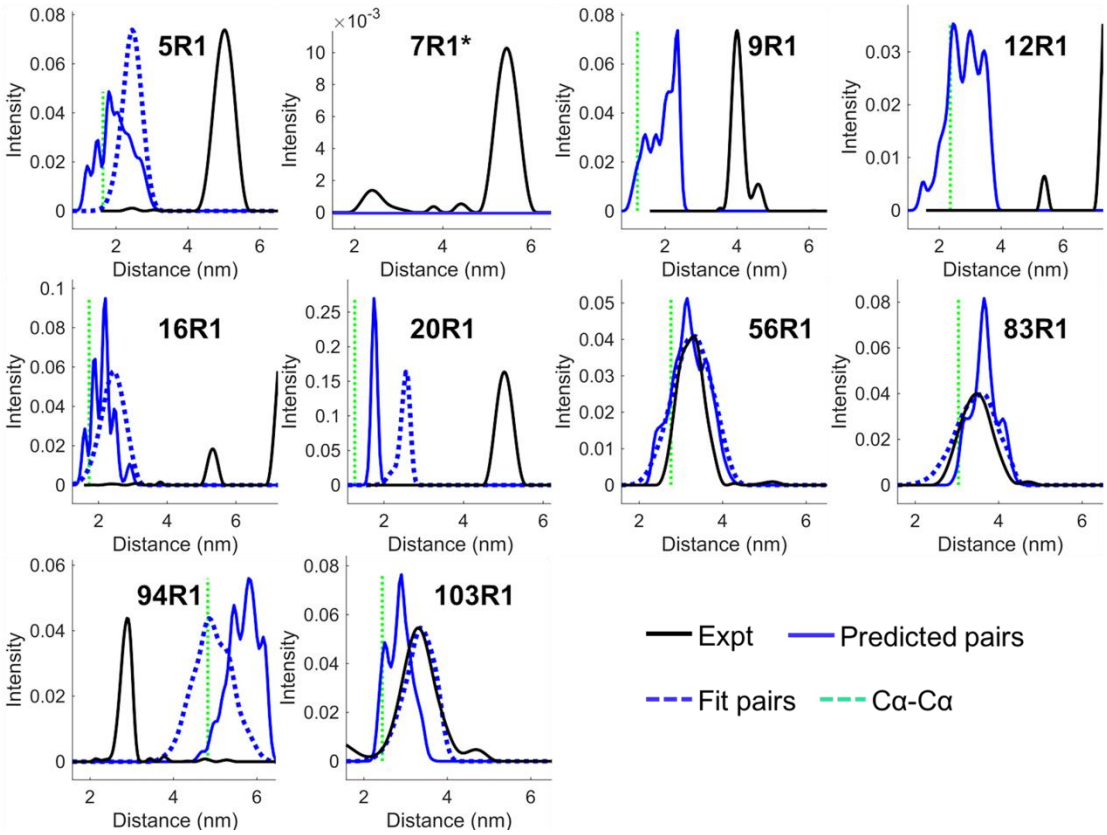


Figure 2.18: Comparison of simulated and experimental DEER distributions for CCM1 ABED dimers. Spin-labeled CCM1 was modeled with the ABED dimer structure, followed by generation of calculated DEER distributions using MMM [49]. For each modeled CCM1 label, distributions were calculated for all R1 side chain rotamers (solid blue trace) and selected rotamers (dashed blue trace) based on fits to the experimental distance distributions (black trace). The average predicted distance between alpha carbon atoms of modeled R1 labels (dashed green line) are additionally included from MMM calculations. Asterisk denotes a position that did not generate any predicted R1 rotamers in the ABED structure due to restricted conformational space at the labeling site.

Table 2.4: Xplor-NIH restraints and structural statistics for CCM1 ensembles*

Parent Structure	GFCC′C″	GFCC′C″	ABED	ABED
Distance Restraint	C_β–C_β	NS1–NS1†	C_β–C_β	NS1–NS1†
# Distance Restraints	10	9‡	10	8‡
# Restraint Violations	2	0	9	6
Maximum Restraint Violation (Å)	0.7	-	19.1	17.6
Mean Ensemble Backbone RMSD (Å)	0.2 ± 0.1	0.2 ± 0.1	0.8 ± 0.7	0.4 ± 0.2
Mean Ensemble Heavy Atom RMSD (Å)	0.9 ± 0.2	0.6 ± 0.2	1.3 ± 0.6	0.6 ± 0.2
Average Backbone RMSD (Å) of Parent to CCM1 ensemble	0.7 ± 0.1	0.4 ± 0.1	19.4 ± 0.1	13.0 ± 0.3

*Statistics calculated for the 20 lowest-energy structures
†NS1 represents the nitrogen atom of the R1 nitroxide moiety
‡Restraint(s) removed for sites that were not *in silico* labeled with R1 due to clashing

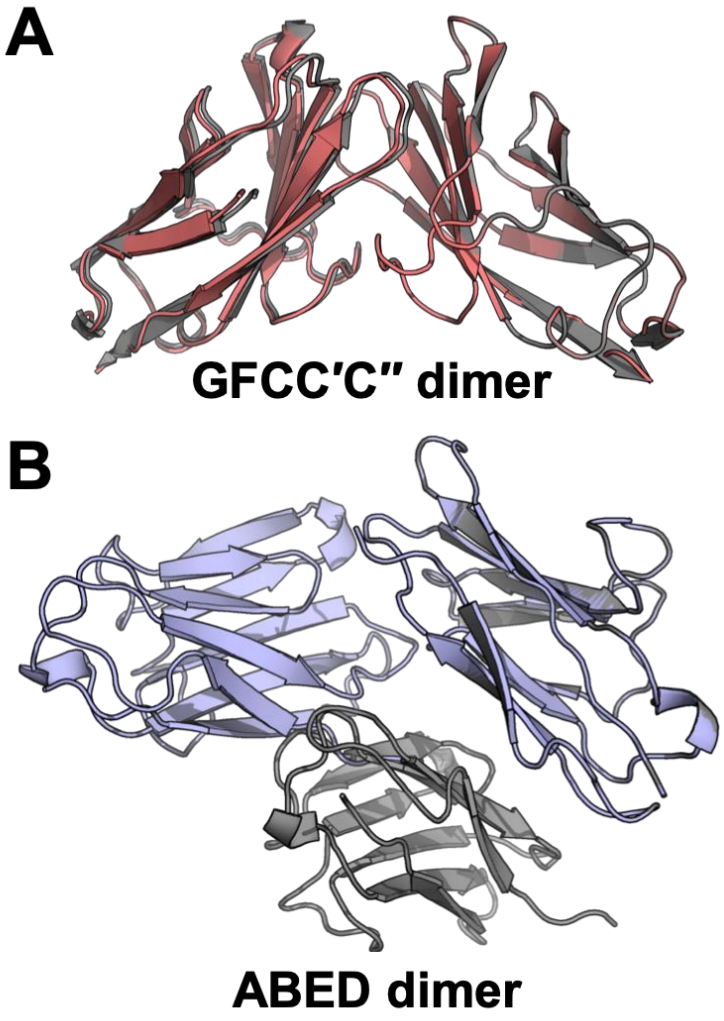


Figure 2.19: CCM1 dimer refinement from DEER distances recapitulates the GFCC′C″-interface structure. Overlay of parent crystal structures (gray) and DEER-refined structures for CCM1 GFCC′C″ (A, red) and ABED (B, blue) dimers.

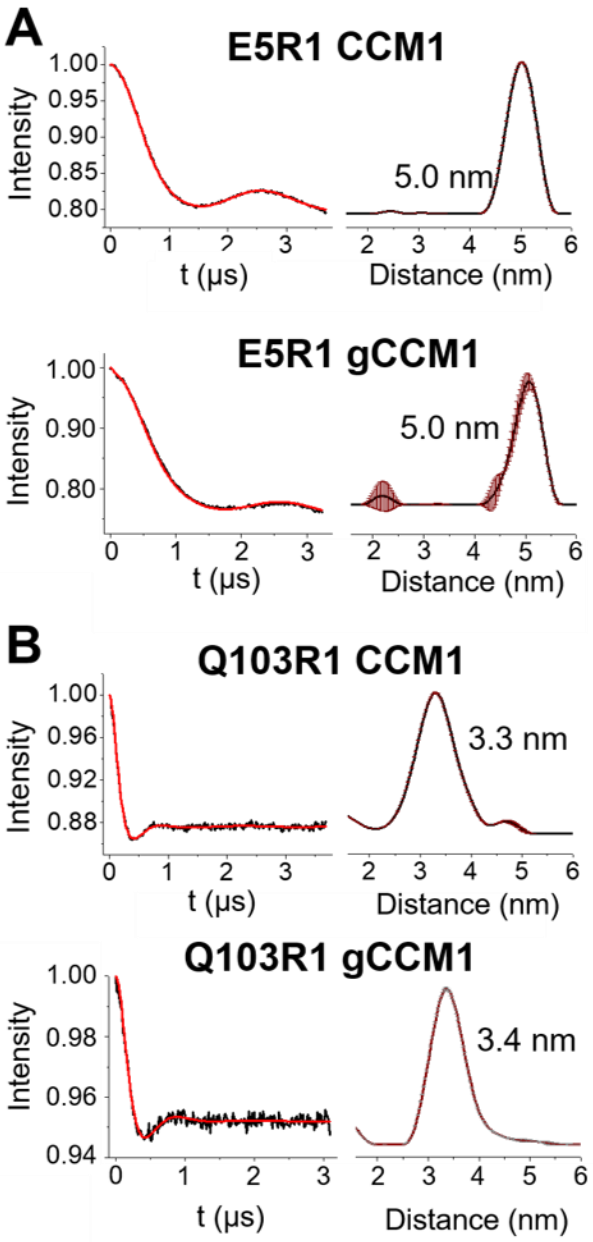


Figure 2.20: Glycosylated and non-glycosylated CCM1 share a similar dimer interface. Background-corrected DEFs (left, black trace) and the resulting normalized DEER distance distributions (right, black trace) from the fits to the DEF (left, red trace)

for 5R1 (A) and 103R1 (B) labeled CCM1 and gCCM1. Error within the distance distribution is plotted as maroon bars (right). Average distances are noted in nm.

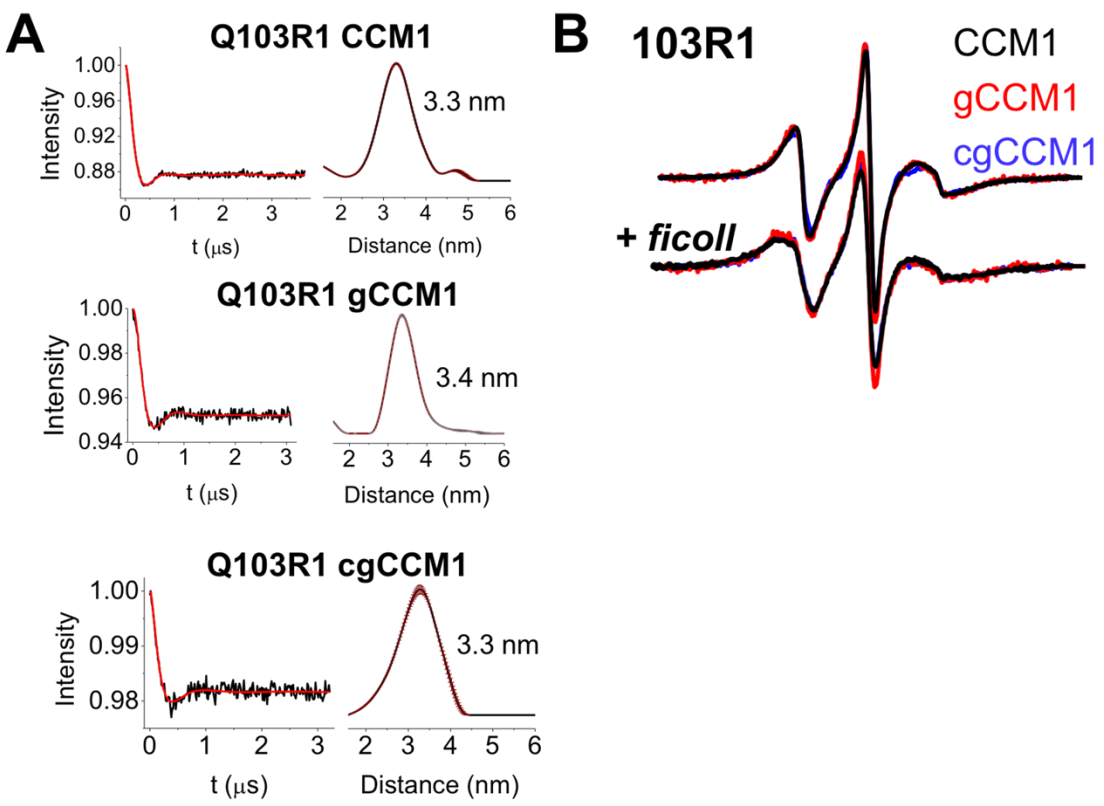


Figure 2.21: Q103R1 DEER distance distributions and CW-EPR spectra are consistent between CCM1 glycoforms. (A) Background-corrected DEFs (left, black trace) and the resulting normalized DEER distance distributions (right, black trace) from the fits to the DEF (left, red trace) for Q103R1 CCM1, gCCM1, and cgCCM1. Error within the distance distribution is plotted as maroon bars (right). Average distances are noted in nm. (B) CW-EPR spectra for CCM1 Q103R1 glycoforms with and without ficoll addition.

Table 2.5: CCM1 evolutionary couplings from GREMLIN

i_id	j_id	Raw Score	Scaled Score	Probability
11_V	78_V	0.9964	3.782	1
21_V	88_L	0.9515	3.611	1
64_R	82_D	0.745	2.828	1
18_L	74_L	0.6411	2.434	1
17_V	104_F	0.6388	2.425	1
11_V	17_V	0.5917	2.246	1
35_K	84_G	0.5229	1.985	1
85_F	103_Q	0.5209	1.977	1
66_T	76_Q	0.5197	1.973	1
46_V	65_E	0.5191	1.97	1
25_P	29_F	0.4992	1.895	1
46_V	53_Q	0.4981	1.891	1
12_A	15_K	0.4876	1.851	1
16_E	74_L	0.4774	1.812	1
68_Y	74_L	0.4703	1.785	1
19_L	104_F	0.4324	1.641	0.999
18_L	72_S	0.4289	1.628	0.999
13_E	80_Q	0.4268	1.62	0.999
44_Q	47_G	0.4259	1.617	0.999
34_Y	89_Q	0.4177	1.586	0.999
61_N	65_E	0.4139	1.571	0.998
37_E	40_D	0.4038	1.533	0.998
85_F	101_T	0.394	1.495	0.997
91_I	96_V	0.382	1.45	0.996
66_T	74_L	0.3752	1.424	0.996
82_D	86_Y	0.3695	1.402	0.995
20_L	72_S	0.3643	1.383	0.994
92_K	95_L	0.3625	1.376	0.994
67_I	73_L	0.362	1.374	0.994
23_N	31_Y	0.3551	1.348	0.992
16_E	76_Q	0.3542	1.345	0.992
32_S	89_Q	0.3385	1.285	0.988
19_L	102_G	0.3371	1.279	0.988

87_T	101_T	0.3274	1.243	0.985
6_S	102_G	0.3269	1.241	0.984
32_S	91_I	0.3252	1.234	0.984
87_T	99_E	0.3244	1.231	0.983
11_V	15_K	0.3244	1.231	0.983
14_G	77_N	0.3195	1.213	0.981
50_I	54_Q	0.316	1.199	0.979
10_N	105_H	0.3127	1.187	0.978
65_E	75_I	0.3099	1.176	0.976
64_R	67_I	0.3021	1.147	0.971
20_L	70_N	0.2989	1.134	0.968
23_N	71_A	0.2967	1.126	0.966
78_V	83_T	0.2892	1.098	0.96
8_P	105_H	0.2853	1.083	0.955
22_H	70_N	0.2853	1.083	0.955
23_N	29_F	0.2849	1.081	0.955
28_L	93_S	0.2818	1.069	0.951
36_G	87_T	0.279	1.059	0.948
32_S	44_Q	0.2742	1.041	0.941
31_Y	71_A	0.2702	1.026	0.935
2_L	5_E	0.2695	1.023	0.934
53_Q	57_P	0.2658	1.009	0.928
36_G	39_V	0.2631	0.999	0.923
21_V	33_W	0.2618	0.994	0.921
11_V	14_G	0.2614	0.992	0.92
18_L	70_N	0.2577	0.978	0.912
9_F	19_L	0.2512	0.954	0.898
16_E	77_N	0.2477	0.94	0.889
75_I	82_D	0.2426	0.921	0.875
83_T	106_V	0.2379	0.903	0.861
83_T	103_Q	0.2359	0.895	0.855
17_V	90_V	0.2347	0.891	0.851
23_N	90_V	0.2342	0.889	0.849
18_L	68_Y	0.2337	0.887	0.848
34_Y	87_T	0.2306	0.875	0.837
59_P	86_Y	0.2283	0.866	0.828
78_V	104_F	0.2264	0.859	0.821

48_Y	55_A	0.2177	0.826	0.785
38_R	43_R	0.2153	0.817	0.775
32_S	47_G	0.2151	0.817	0.775
45_I	48_Y	0.2135	0.81	0.766
46_V	51_G	0.2134	0.81	0.766
33_W	73_L	0.212	0.805	0.76
4_T	15_K	0.2119	0.804	0.759
52_T	56_T	0.2105	0.799	0.752
26_Q	93_S	0.2092	0.794	0.746
12_A	24_L	0.2078	0.789	0.739
51_G	54_Q	0.2041	0.775	0.721
16_E	78_V	0.203	0.771	0.715
6_S	9_F	0.2021	0.767	0.709
34_Y	44_Q	0.202	0.767	0.709
47_G	53_Q	0.2018	0.766	0.708
49_A	54_Q	0.198	0.752	0.688
68_Y	72_S	0.1972	0.749	0.683
29_F	90_V	0.1955	0.742	0.673
93_S	96_V	0.1951	0.741	0.671
44_Q	48_Y	0.195	0.74	0.67
8_P	103_Q	0.1926	0.731	0.656
27_Q	92_K	0.1925	0.731	0.656
83_T	105_H	0.1923	0.73	0.654
100_A	103_Q	0.1922	0.729	0.653
78_V	106_V	0.1896	0.72	0.639
52_T	58_G	0.1886	0.716	0.632
37_E	43_R	0.1867	0.709	0.621
21_V	35_K	0.1852	0.703	0.611
43_R	58_G	0.185	0.702	0.61
48_Y	52_T	0.1839	0.698	0.603
14_G	78_V	0.1832	0.695	0.598
89_Q	98_E	0.1832	0.695	0.598
64_R	76_Q	0.183	0.695	0.598
47_G	52_T	0.1809	0.687	0.585
50_I	53_Q	0.1807	0.686	0.583
71_A	88_L	0.1787	0.678	0.57
62_S	66_T	0.1787	0.678	0.57

13_E	83_T	0.1779	0.675	0.565
96_V	99_E	0.1765	0.67	0.556
34_Y	99_E	0.1757	0.667	0.551
9_F	17_V	0.1755	0.666	0.549
77_N	80_Q	0.1752	0.665	0.548
26_Q	79_T	0.1746	0.663	0.544
2_L	90_V	0.1742	0.661	0.541
49_A	53_Q	0.1736	0.659	0.538
2_L	11_V	0.1727	0.655	0.531
24_L	66_T	0.172	0.653	0.527
53_Q	69_P	0.1717	0.652	0.526
38_R	58_G	0.171	0.649	0.52
42_N	52_T	0.1698	0.645	0.514
23_N	32_S	0.1692	0.642	0.508
46_V	49_A	0.1683	0.639	0.503
51_G	56_T	0.1682	0.638	0.502
31_Y	90_V	0.1654	0.628	0.484
47_G	50_I	0.1651	0.627	0.483
75_I	84_G	0.1642	0.623	0.476
11_V	104_F	0.1632	0.62	0.471
62_S	65_E	0.1628	0.618	0.467
36_G	101_T	0.162	0.615	0.462
35_K	82_D	0.1615	0.613	0.459
4_T	22_H	0.1613	0.612	0.457
48_Y	51_G	0.1613	0.612	0.457
26_Q	94_D	0.1603	0.609	0.452
50_I	56_T	0.1598	0.607	0.449
9_F	20_L	0.1576	0.598	0.433
47_G	55_A	0.157	0.596	0.43
45_I	50_I	0.157	0.596	0.43
72_S	75_I	0.1558	0.591	0.422
19_L	33_W	0.1556	0.591	0.422
4_T	9_F	0.1555	0.59	0.42
42_N	56_T	0.1548	0.588	0.417
61_N	64_R	0.1547	0.587	0.415
12_A	78_V	0.1538	0.584	0.41
36_G	85_F	0.1537	0.584	0.41

46_V	52_T	0.1533	0.582	0.407
71_A	90_V	0.1529	0.58	0.403
27_Q	90_V	0.1527	0.58	0.403
17_V	78_V	0.152	0.577	0.398
43_R	47_G	0.1515	0.575	0.395
44_Q	49_A	0.1514	0.575	0.395
102_G	107_Y	0.1514	0.574	0.393
41_G	97_N	0.1513	0.574	0.393
15_K	100_A	0.1509	0.573	0.392
27_Q	44_Q	0.1507	0.572	0.39
27_Q	51_G	0.1502	0.57	0.387
102_G	105_H	0.1492	0.566	0.38
66_T	69_P	0.1492	0.566	0.38
8_P	31_Y	0.1487	0.564	0.377
89_Q	99_E	0.1487	0.564	0.377

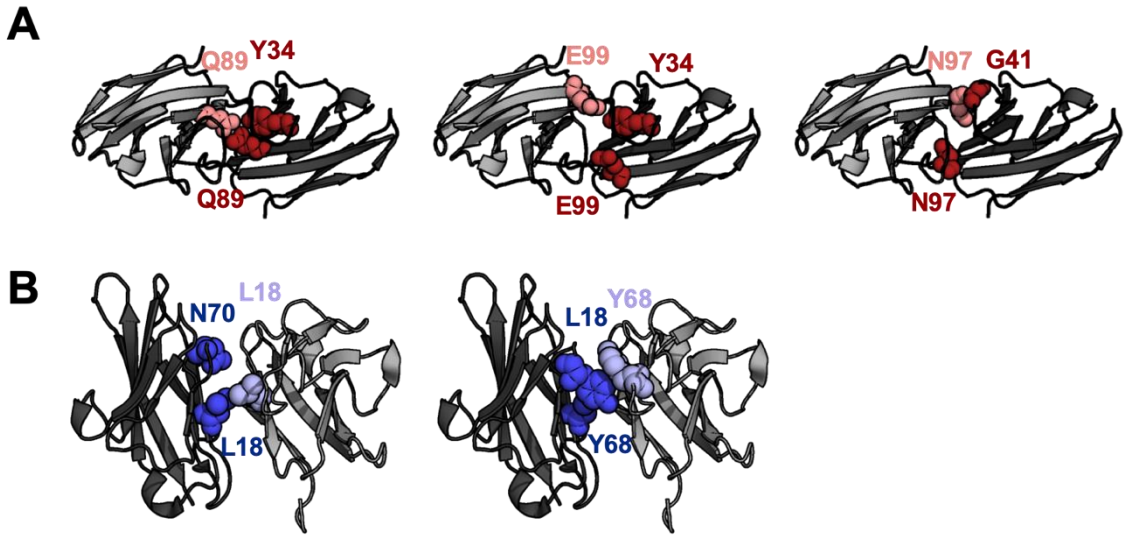


Figure 2.22: CCM1 evolutionarily-coupled residues mapped onto GFCC′C″ and ABED dimers. Evolutionarily-coupled (EC) residue pairs (spheres) that are within 5 Å across the GFCC′C″ (A, red) and ABED (B, blue) CCM1 dimer interfaces. Terminal side chain atoms were used for distance cutoffs. Three EC pairs were identified across the GFCC′C″ interface (Q89/Y34, E99/Y34, and N97/G41), while two EC pairs were identified across the ABED interface (L18/N70 and L18/Y68). EC pairs are arranged in order of decreasing score (Table 2.5). Protomers are represented by dark and light shades, respectively. EC pairs Q89/Y34 (GFCC′C″) and L18/Y68 (ABED) are in contact both within the monomer and across dimer interfaces and therefore cannot be used to support one structure over another. EC pair L18/N70 is sufficiently spaced within the CCM1 monomer, but L18 contacts itself across the ABED interface.

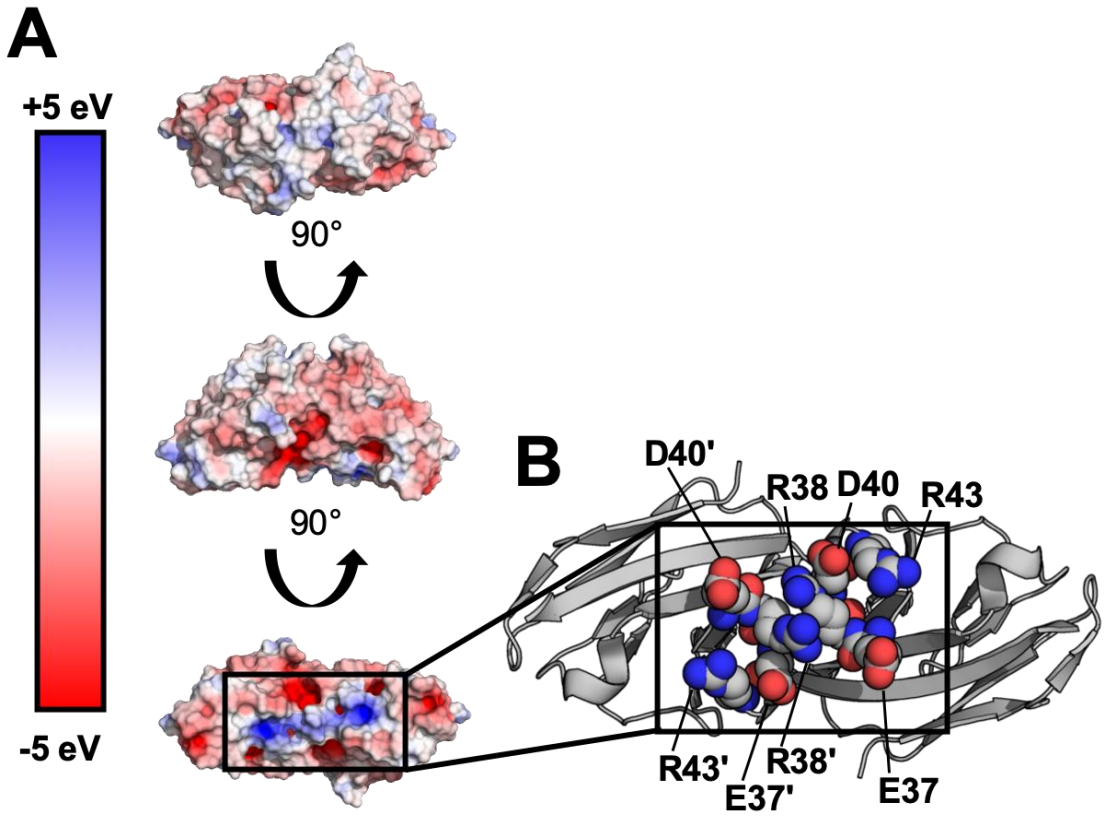


Figure 2.23: CCM1 GFCC′C″ dimer electrostatic surface potential. (A) The electrostatic surface potential of the CCM1 GFCC′C″ structure shows a positively-charged patch on the bottom surface of the protein that likely wedges the two protomers apart. A surface potential gradient of -5.0 to +5.0 eV was used to generate the maps. (B) CCM1 EC pairs E37/D40, R43/R48, and E37/R43 are rendered as spheres on the GFCC′C″ dimer structure with oxygen (red) and nitrogen (blue) atoms highlighted. The network of charged EC pairs on the bottom of the dimer interface is additional evidence of conserved sequence and structure among CCM family members that form dimers.

2.5 Concluding remarks

Although this study suggests that glycan modifications do not perturb the ability of CCM1 to form GFCC′C″ interface dimers in solution, mammalian glycosylation is highly regulated and important for immune recognition, glycan-dependent protein binding, and cell signaling events [50]. The prevalence of CCM1 GFCC′C″ dimers importantly leaves glycans exposed and may promote specific types of CEACAM1 association *in vivo*. CEACAM1 has been reported to form dimers *in vivo* both in *cis* (apical cell surface) and *trans* (between cells), with GFCC′C″ interface mutations abrogating CEACAM1 *trans*-homophilic binding [7]. Thus, solution-based studies of glycosylated CCM1 *in vitro* support observations of *trans* binding between HeLa cells expressing CEACAM1. However, electron tomography structures of soluble rat CEACAM1 containing all four extracellular Ig domains reported a minor population of dimers distinct from the major GFCC′C″ interface species [24] that was not captured by DEER spectroscopy. The minor CEACAM1 dimer was identified based on in-parallel, close binding between three or all four Ig domains. Interestingly, the authors report enhanced formation of this alternative CEACAM1 dimer upon membrane attachment in which CEACAM1 clusters were greater on free liposome surfaces (*cis*-type). As the CEACAM1 N-terminal domain alone was assessed in this study, additional CEACAM1 domains and/or membrane tethering may be necessary to observe the minor dimer population (possibly ABED dimers) reported by Klaile *et al.* [24].

While *cis* and *trans* CEACAM1 binding *in vivo* is controlled via calmodulin-dependent signaling [7] (Figure 2.2), it is possible that CEACAM1 glycan-glycan

interactions may promote clustering on cells or provide an additional regulatory mechanism for interactions in *cis*. For example, glycosylation of CEACAM5 modulates the strength of homotypic CEACAM5 interactions but not adhesion functionality nor specificity [51]. Due to reported similarities between CEACAM1 and CEACAM5 specificity, glycosylation may play a similar role in the strength of CEACAM1 homotypic interactions. DEER spectroscopy of CEACAM1 glycoforms and evolutionary coupling analysis strongly supports a preference for GFCC′C″ interface dimers that would expose glycans for homotypic CEACAM1 interactions either in *cis* or *trans* (Figure 2.24).

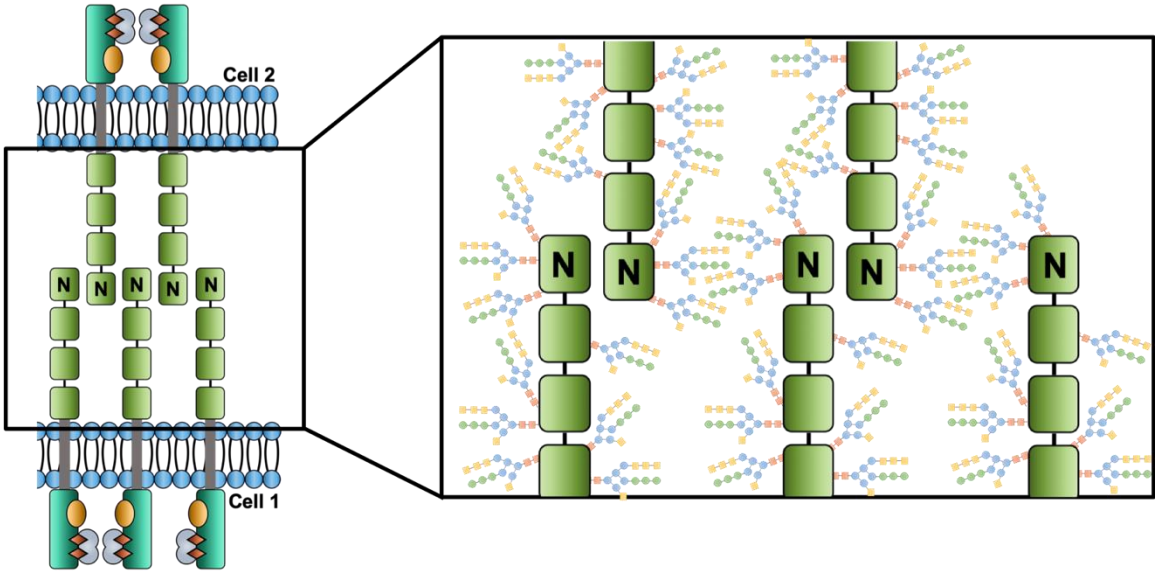


Figure 2.24: Proposed glycan-glycan interactions between CEACAM1 dimers *in vivo*. The preference of CEACAM1 N-domains to form dimers at the GFCC′C″ interface leaves glycans exposed in the lattice monomer adhesion model (Figure 2.2) proposed by [7]. Therefore, CEACAM1 N-domain glycosylation may strengthen homotypic interactions in the *trans* binding configuration.

2.6 Acknowledgments

I would like to thank Prof. David Cafiso and Dr. David Nyenhuis for helpful discussions and guidance with incorporating DEER restraints into Xplor-NIH. This work was funded by the National Institutes of Health under Grant (2R01GM087828-07 to Prof. Linda Columbus), resource Grant (P41-GM103309 to Prof. Kelley Moremen and Prof. James Prestegard), Training Grant (5T32GM008715-17 to Nicole Swope), and by the National Science Foundation Graduate Research Fellowship under Grant (DDGE-1315231 to Nicole Swope).

2.6.1 Author contributions

Nicole Swope, Dr. Marissa Kieber, and Prof. Linda Columbus coordinated and designed this study. Nicole Swope designed, conducted, and analyzed the experiments shown in Figure 2.4, Figure 2.16, Figure 2.17, Figure 2.18, Figure 2.19, Figure 2.22, and Figure 2.23 and Table 2.2, Table 2.3, Table 2.4, and Table 2.5. Dr. Marissa Kieber designed and conducted the experiments shown in Figure 2.9, Figure 2.10, Figure 2.13, Figure 2.14, Figure 2.20, and Figure 2.21, and both Marissa Kieber and Nicole Swope analyzed this experiment set. Meagan Belcher Dufrisne designed, conducted, and analyzed the experiments shown in Figure 2.15. Ji Han and Jason Li aided in sample preparation for experiments shown in Figure 2.9 and Figure 2.10. Jeong-Yeh Yang, Prof. Kelley Moremen, and Prof. James Prestegard established expression and purification protocols for glycosylated CCM1 constructs. The manuscript was drafted and edited by Nicole Swope and Prof. Linda Columbus.

2.7 References

1. Kuespert, K., S. Pils, and C.R. Hauck, *CEACAMs: their role in physiology and pathophysiology.* Current Opinion in Cell Biology, 2006. **18**(5): p. 565-571.
2. Dankner, M., et al., *CEACAM1 as a multi-purpose target for cancer immunotherapy.* OncoImmunology, 2017. **6**(7): p. e1328336.
3. 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.
4. Korotkova, N., et al., *A Subfamily of Dr Adhesins of Escherichia coli Bind Independently to Decay-accelerating Factor and the N-domain of Carcinoembryonic Antigen.* Journal of Biological Chemistry, 2006. **281**(39): p. 29120-29130.
5. Bonsor, D.A., et al., *The Helicobacter pylori adhesin protein HopQ exploits the dimer interface of human CEACAMs to facilitate translocation of the oncoprotein CagA.* The EMBO Journal, 2018. **37**(13).
6. Zhuo, Y., et al., *Glycosylation Alters Dimerization Properties of a Cell-surface Signaling Protein, Carcinoembryonic Antigen-related Cell Adhesion Molecule 1 (CEACAM1).* J Biol Chem, 2016. **291**(38): p. 20085-95.
7. Patel, P.C., et al., *Inside-out signaling promotes dynamic changes in the CEACAM1 oligomeric state to control its cell adhesion properties.* Journal of Biological Chemistry, 2013.
8. 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.
9. Kammerer, R. and W. Zimmermann, *Coevolution of activating and inhibitory receptors within mammalian carcinoembryonic antigen families.* BMC Biology, 2010. **8**(1): p. 12.
10. 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.
11. Korotkova, N., et al., *Binding of Dr adhesins of Escherichia coli to carcinoembryonic antigen triggers receptor dissociation.* Molecular Microbiology, 2008. **67**(2): p. 420-434.
12. Moonens, K., et al., *Helicobacter pylori adhesin HopQ disrupts trans dimerization in human CEACAMs.* The EMBO Journal, 2018. **37**(13): p. e98665.
13. 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 : CCS, 2014. **12**: p. 27-27.
14. Öbrink, B., *CEA adhesion molecules: multifunctional proteins with signal-regulatory properties.* Current Opinion in Cell Biology, 1997. **9**(5): p. 616-626.

15. Virji, M., et al., *The N-domain of the human CD66a adhesion molecule is a target for Opa proteins of Neisseria meningitidis and Neisseria gonorrhoeae.* Molecular Microbiology, 1996. **22**(5): p. 929-939.
16. 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.
17. 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 Crystallogr D Biol Crystallogr, 2006. **62**(Pt 9): p. 971-9.
18. Bonsor, D.A., D. Beckett, and E.J. Sundberg, *Structure of the N-terminal dimerization domain of CEACAM7.* Acta Crystallogr F Struct Biol Commun, 2015. **71**(Pt 9): p. 1169-75.
19. Bonsor, D.A., et al., *Diverse oligomeric states of CEACAM IgV domains.* Proc Natl Acad Sci U S A, 2015. **112**(44): p. 13561-6.
20. Huang, Y.-H., et al., *CEACAM1 regulates TIM-3-mediated tolerance and exhaustion.* Nature, 2014. **517**: p. 386.
21. Taheri, M., et al., *Self Recognition in the Ig Superfamily: IDENTIFICATION OF PRECISE SUBDOMAINS IN CARCINOEMBRYONIC ANTIGEN REQUIRED FOR INTERCELLULAR ADHESION.* Journal of Biological Chemistry, 2000. **275**(35): p. 26935-26943.
22. 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. 1469-79.
23. Markel, G., et al., *The Critical Role of Residues ⁴³R and ⁴⁴Q of Carcinoembryonic Antigen Cell Adhesion Molecules-1 in the Protection from Killing by Human NK Cells.* The Journal of Immunology, 2004. **173**(6): p. 3732.
24. Klaile, E., et al., *The CEACAM1 N-terminal Ig domain mediates cis- and trans-binding and is essential for allosteric rearrangements of CEACAM1 microclusters.* J Cell Biol, 2009. **187**(4): p. 553-67.
25. Kirouac, K.N. and G.G. Prive, *Human CEACAM1 N-domain homodimer.* To be published, 2015.
26. Armougom, F., et al., *Expresso: automatic incorporation of structural information in multiple sequence alignments using 3D-Coffee.* Nucleic Acids Research, 2006. **34**(suppl_2): p. W604-W608.
27. Robert, X. and P. Gouet, *Deciphering key features in protein structures with the new ENDscript server.* Nucleic Acids Research, 2014. **42**(W1): p. W320-W324.
28. Klock, H.E. and S.A. Lesley, *The Polymerase Incomplete Primer Extension (PIPE) Method Applied to High-Throughput Cloning and Site-Directed Mutagenesis*, in *High Throughput Protein Expression and Purification: Methods and Protocols*, S.A. Doyle, Editor. 2009, Humana Press: Totowa, NJ. p. 91-103.
29. 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.

30. Jeschke, G., et al., *DeerAnalysis2006 - a comprehensive software package for analyzing pulsed ELDOR data.* Applied Magnetic Resonance, 2006. **30**(3-4): p. 473-498.
31. Danielli, J.F. and H. Davson, *A contribution to the theory of permeability of thin films.* Journal of Cellular and Comparative Physiology, 1935. **5**(4): p. 495-508.
32. Polyhach, Y., E. Bordignon, and G. Jeschke, *Rotamer libraries of spin labelled cysteines for protein studies.* Physical Chemistry Chemical Physics, 2011. **13**(6): p. 2356-2366.
33. Schwieters, C.D., et al., *The Xplor-NIH NMR molecular structure determination package.* Journal of Magnetic Resonance, 2003. **160**(1): p. 65-73.
34. Sarver, J.L., et al., *A Dynamic Protein–Protein Coupling between the TonB-Dependent Transporter FhuA and TonB.* Biochemistry, 2018. **57**(6): p. 1045-1053.
35. Koradi, R., M. Billeter, and K. Wüthrich, *MOLMOL: A program for display and analysis of macromolecular structures.* Journal of Molecular Graphics, 1996. **14**(1): p. 51-55.
36. Balakrishnan, S., et al., *Learning generative models for protein fold families.* Proteins: Structure, Function, and Bioinformatics, 2011. **79**(4): p. 1061-1078.
37. Kamisetty, H., S. Ovchinnikov, and D. Baker, *Assessing the utility of coevolution-based residue–residue contact predictions in a sequence- and structure-rich era.* Proceedings of the National Academy of Sciences, 2013. **110**(39): p. 15674.
38. Cornish, V.W., et al., *Site-specific incorporation of biophysical probes into proteins.* Proceedings of the National Academy of Sciences, 1994. **91**(8): p. 2910.
39. 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.
40. Columbus, L. and W.L. Hubbell, *A new spin on protein dynamics.* Trends in Biochemical Sciences, 2002. **27**(6): p. 288-295.
41. Nesmelov, Y.E. and D.D. Thomas, *Protein structural dynamics revealed by site-directed spin labeling and multifrequency EPR.* Biophysical reviews, 2010. **2**(2): p. 91-99.
42. Freed, J.H., *Theory of Slow Tumbling ESR Spectra of Nitroxides*, in *Spin Labeling Theory and Applications*, L.J. Berliner, Editor. 1976, Academic Press: New York. p. 53-132.
43. Lopez, C.J., et al., *Osmolyte perturbation reveals conformational equilibria in spin-labeled proteins.* Protein Sci, 2009. **18**(8): p. 1637-52.
44. Cavanagh, J., et al., *CHAPTER 1 - CLASSICAL NMR SPECTROSCOPY*, in *Protein NMR Spectroscopy (Second Edition)*, J. Cavanagh, et al., Editors. 2007, Academic Press: Burlington. p. 1-28.
45. 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. 7692-704.
46. Jeschke, G. and Y. Polyhach, *Distance measurements on spin-labelled biomacromolecules by pulsed electron paramagnetic resonance.* Phys Chem Chem Phys, 2007. **9**(16): p. 1895-910.

47. Jeschke, G., *DEER Distance Measurements on Proteins.* Annual Review of Physical Chemistry, 2012. **63**(1): p. 419-446.
48. Anishchenko, I., et al., *Origins of coevolution between residues distant in protein 3D structures.* Proceedings of the National Academy of Sciences, 2017. **114**(34): p. 9122.
49. Jeschke, G., *MMM: A toolbox for integrative structure modeling.* Protein Sci, 2018. **27**(1): p. 76-85.
50. Marth, J.D. and P.K. Grewal, *Mammalian glycosylation in immunity.* Nature reviews. Immunology, 2008. **8**(11): p. 874-887.
51. Charbonneau, J. and C.P. Stanners, *Role of carbohydrate structures in CEA-mediated intercellular adhesion.* Cell Adhesion and Communication, 1999. **7**(3): p. 233-244.

Chapter 3: Physical Properties of Phosphatidylcholine Lipid-Detergent Bicelles

3.1 Overview

A brief introduction to applications of detergent micelles and lipid-detergent bicelles in membrane protein investigations was presented in Chapter 1. Principles of micelle and bicelle formation were additionally outlined and are critical for understanding physical properties of both assemblies. I am broadly interested in determining how isotropically-tumbling lipid/detergent mixtures stabilize the fold of membrane proteins, which requires establishing bicelle properties with and without protein present. Here I will discuss my work on bicelle morphology without an embedded membrane protein, which is a continuation of previous studies in the laboratory by Tracy Caldwell. The possible dependence between bicelle components and the lipid segregation point for a given mixture was explored with small-angle X-ray scattering (SAXS) using four phosphatidylcholine (PC) lipids and nine unique detergents. This work indicates that bicelles with PC lipids undergo a composition-independent phase transition when the ratio of lipid to detergent is greater than or equal to ~ 1:2. Results suggest that the bicelle phase transition from a mixed lipid/detergent assembly to a lipid-segregated core is a cooperative event requiring a minimum number of lipids, regardless of the rim-forming agent. The detergent component does, however, influence the maximum bicelle thickness (short dimension) observed when the same PC lipid is used to form the bicelle. Detergent-dependent changes in bicelle

morphology are explored at the end of this chapter (section 3.6.6 and 3.7), in order to investigate effects on an embedded membrane protein.

3.2 Introduction

3.2.1 Morphology of lipid-detergent mixtures

Bicelles are typically formed with long-chain lipid and short-chain detergent. The lipid-detergent assembly combines the properties of both bilayers and micelles in the "ideal bicelle" model, for which lipid and detergent molecules are phase-separated into a disc-like structure with a lipid bilayer core encircled by a detergent rim (Figure 3.1A) [1, 2]. Due to the bilayer environment, ideal bicelles are thought to provide a more native environment for embedded membrane proteins compared to detergent micelles [3]. However, the ideal bicelle model holds for a limited set of conditions and depends on lipid and detergent structure, total concentration of amphiphiles, ratio of amphiphile components, and temperature [4, 5]. The molar ratio of amphiphile molecules (q, q-value) is defined by

$$q = \frac{[\text{lipid}]}{[\text{detergent}]}$$

Equation 3.1

and represents the parameter most often associated with bicellar phase transitions and morphology [3]. The dependence of bicelle morphology on q has been described for the most common bicelle mixture using the long-chain lipid DMPC (14:0 PC) and short-chain detergent DH_6PC (6:0 PC) (Figure 3.1B) [6]. When $q < 2.5$, small, isotropically-tumbling aggregates form, and $2.5 < q < 7.5$ describes bicelle conditions for extended bilayer lamellae that become perforated with detergent as q increases [7] (Figure 3.2). Bicelles

with $2.5 < q < 7.5$ additionally undergo magnetic alignment, and phase-separated lipid and detergent occurs in this regime for DMPC/DH_6PC [2]. Above $q = 7.5$, multilamellar vesicles form [8].

The existence of spatially-separated lipid and detergent in bicelles when $q < 0.5$ is widely-debated [5, 9-14]. Bicelles formed with $q < 0.5$, specifically $q \sim 0.3$, are routinely used for solution NMR spectroscopy studies of membrane proteins due to favorable tumbling properties [15]. As a bilayer-like environment is desirable for membrane proteins embedded in bicelles, it is important to establish spatial segregation of lipid and detergent under conditions for solution NMR. To date, isotropically-tumbling ($q < 2.5$) bicelle morphologies have been predominantly characterized with dynamic light scattering (DLS), ^{31}P NMR, ^{1}H NMR, and electron microscopy [5, 10, 11, 14, 15]. The hydrodynamic radius (R_H) of bicelle aggregates may be measured by DLS, but R_H calculations are often biased to a specific geometric model that may not reflect the actual shape of the assembly [12, 15]. Relative chemical shifts of ^{31}P resonances for DMPC and DH_6PC headgroups are considered reliable indicators of spatial segregation within a bicelle. Bicelles with $q > 0.5$ produce unique ^{31}P resonances for DMPC and DH_6PC, supporting that bicelles above $q = 0.5$ possess phase-separated lipid and detergent [10]. Terminal ^{1}H methyl resonances of DMPC and DH_6PC are similarly unique and used to infer lipid/detergent segregation [5]. However, proximity of DH_6PC methyl protons to the electron-rich headgroup relative to DMPC methyl protons may induce the observed difference in chemical shift.

Combined, DLS, ^{31}P NMR, and ^{1}H NMR derive bicelle dimensions and extent of lipid segregation indirectly, while electron microscopy provides direct visualization of

bicelle shape but contain 1.0 nm or greater measurement uncertainty [10]. Various interpretations of DMPC/DH_6PC bicelle morphology from these experiments from different publications include (1) that all bicelles below q = 1.0 follow the ideal bicelle model [11], (2) only q > 0.5 follows the ideal bicelle model [10], and (3) spatially-segregated lipid and detergent does not occur for q < 1.0 [16, 17]. The morphology of isotropically-tumbling bicelle compositions other than DMPC/DH_6PC are similarly unclear from the literature [11, 12, 15, 18]. Conflicting results from DLS, NMR, and electron microscopy therefore demonstrate that establishing the presence of ideal bicelles requires direct measurements of the assembly size and shape as well as accurately determining lipid and detergent localization.

3.2.2 Low-q DMPC/DH_6PC bicelles are mixed micelles

The Columbus laboratory recently combined small-angle scattering, fluorescence anisotropy, and molecular dynamics to directly determine the size, shape, and lipid segregation present in DMPC/DH_6PC bicelles for 0.1 < q < 1.0 [9]. The major finding from this study was that when q < 0.5, DMPC/DH_6PC follows mixing behavior similar to mixed micelles [19, 20] rather than phase-separated lipid and detergent. Both the model-independent aggregate thickness from SAXS and mixture T_m from fluorescence anisotropy supported mixed lipid/detergent for 0.1 < q < 0.5 with evidence of internal lipid organization when q > 0.5. To assess bicelle morphology and lipid/detergent mixing in each regime (q = 0.3 and 0.7), contrast-variation experiments with small-angle neutron scattering (SANS) were performed. Fitting SANS profiles to a core-shell bicelle model (Figure 3.3) provided the dimensions for assembly radius, length, face thickness, and rim

thickness in addition to the percentage of DMPC and DH_6PC in the core and rim. Parameters derived from SANS showed that q = 0.3 DMPC/DH_6PC mixtures are mixed, recapitulating results from SAXS and fluorescence anisotropy, while q = 0.7 DMPC/DH_6PC suggest internal lipid organization with some detergent present in the core. Molecular dynamics simulations performed with DMPC/DH_6PC assemblies reflecting q = 0.3 and q = 0.7 validated dimensions and lipid/detergent mixing derived from SANS. As the title of the study suggests, for DMPC/DH_6PC, low-q (< 0.5) bicelles are therefore mixed micelles.

3.2.3 Evidence of low-q lipid segregation in lipid-detergent mixtures

Aside from DMPC, PC lipids may also be combined with bile salt and fos-choline detergents to form fast-tumbling bicelles [11, 12, 15]. The bile salt detergent 3-[(3-cholamidopropyl)-dimethylammonio]-1-propane sulfonate (CHAPS) and 6-cyclohexyl-1-hexylphosphocholine (Cyclofos-6) (Figure 3.8) represent two rim-forming agents for PC bicelles that are proposed to produce ideal bicelle behavior at low q-values (< 0.5) and are suitable environments for transmembrane proteins [12, 15]. For CHAPS bicelles, Mineev *et al.* [12] derived hydrodynamic radii (R_H) from translational diffusion coefficients measured by NMR spectroscopy. The authors concluded that the dependence of R_H on the bicelle q-value demonstrated ideal bicelle behavior for all low-q (< 1.0) bicelles in the study, which included combinations of PC lipids with DH_6PC, CHAPS, and CHAPSO detergents. However, R_H calculations relied on the *a priori* assumption that all aggregates formed cylindrical/discoidal morphologies. The height of the disc was also biased to the expected bilayer height of the PC lipid in the assembly.

Translational diffusion coefficients were also used to infer bicelle size for DMPC/Cyclofos-6 mixtures based on the measured R_H from NMR spectroscopy. When combined with DMPC, the detergent Cyclofos-6 is suggested to form bicelles under highly dilute conditions ($C_L < 0.5\%$) when q = 0.3 [15]. The authors of the latter study presented a reasonable hypothesis that Cyclofos-6 possesses a bulky cyclohexyl group at the end of the alkyl tail (Figure 3.8) that may decrease partitioning with DMPC and allow bicelle formation under dilute conditions. As the study was designed to qualitatively compare bicelle assembly rather than quantify dimensions of the bicelles, the authors state that future experiments are needed to assess DMPC/Cyclofos-6 morphology. CHAPS may similarly partition less with PC lipids as it also contains a bulky hydrophobic face (Figure 3.8), but direct measurements of bicelle dimensions are required to support this hypothesis.

3.2.4 X-ray scattering of lipid-detergent aggregates

Small-angle X-ray scattering (SAXS) may be used to study the size and shape of biological macromolecules in solution [21-23]. A general illustration of the SAXS experiment setup and data reduction are shown in Figure 3.4. The SAXS scattering profile (Figure 3.4B) of a macromolecule is given by the scattering intensity I(Q) as a function of the magnitude of the scattering vector Q:

$$Q = \frac{4\pi \sin(\theta)}{\lambda}$$

Equation 3.2

where 2θ is the angle between the incident beam and scattered intensity, and λ is the wavelength of incident radiation [24]. The scattering profile originates from all molecules in solution with random orientations and positions (Figure 3.4A). Therefore, the observed scattering pattern is equivalent to the macromolecule spherically-averaged over all orientations and related to real-space interatomic distances (*d*) by Equation 3.3 as a result of constructive interference between X-rays [24].

$$Q = \frac{2\pi}{d}$$

Equation 3.3

The net macromolecular scattering profile is given by subtracting the scattering of matched buffer conditions from the original sample. Amphipathic assemblies such as micelles and bicelles often produce unique scattering profiles with a characteristic second peak in the scattering intensity at intermediate Q values ($0.15 < Q < 0.3$ Å^{-1}) (Figure 3.5A) [9, 13, 20, 25, 26]. A first maximum at Q = 0 for the forward scattering followed by a second maximum is generally observed for core-shell structures [27, 28].

The second maximum present in micelle and bicelle scattering profiles is due to large differences in scattering contrast between detergent/lipid headgroups and aliphatic cores (Figure 3.5B) [26]. Lipfert *et al.* [26] discovered that the length scale of the second maximum corresponded to the average interatomic distance between detergent head groups across the short dimension of a micelle (*L*) (Figure 3.5C). Detergent and lipid headgroups are generally electron-dense relative to the low electron density in the hydrophobic core of a micelle or bicelle. Therefore, the scattering interference between detergent/lipid head

groups across the core determines the position of the second maximum (Q_{max}). The dominant headgroup to headgroup distance, *L*, may be determined from Q_{max} using the relationship in Equation 3.3 (Figure 3.5A). Crucially, the *L* parameter is specific to the short dimension of elongated core-shell structures because more monomers are separated by approximately the short dimension than by other distances across the core [26].

3.2.5 Evidence that the *L* parameter from SAXS measures bicelle phase transitions

For DMPC/DH_6PC bicelles at q-values ranging from 0.1 to 1.0, the *L* parameter derived from SAXS profiles (Figure 3.6A) corresponds well to the extent of lipid segregation determined from SANS contrast variation experiments and MD simulations [9]. *L* increases linearly as a function of the DMPC/DH_6PC q-value for $0.1 < q < 0.5$ and may be extrapolated to give the experimental *L* parameter for pure DH_6PC ($q = 0.0$, $L = 22$ Å) (Figure 3.6B). The linear increase in *L* for $0.1 < q < 0.5$ is indicative of mixing behavior that has been reported for mixed detergent micelles [19, 20] rather than the expected *L* parameter for a DMPC core with extended acyl chains according to the ideal bicelle model. Once the DMPC/DH_6PC q-value is approximately 0.5, the bicelle undergoes a phase transition from a mixed lipid/detergent aggregate to a structure with a lipid-segregated core (bicelle). Internal lipid organization across the short dimension of the bicelle is evident from the L parameters for $0.5 < q < 1.0$ (~42 Å), which plateau at a reasonable approximation of L for a pure DMPC bilayer (Figure 3.6B).

SANS contrast variation experiments and MD simulations report that there is still some DH_6PC present in the bicelle core when $q = 0.7$, so SAXS cannot be used to infer the

degree of lipid segregation within bicelle aggregates. However, the work by Caldwell *et al.* [9] highlights that the SAXS-derived headgroup to headgroup distance, *L*, across the short dimension of a bicelle is well-suited to study phase transitions between lipid/detergent mixing and internal lipid organization. As *L* for a bicelle may be calculated directly from the second maximum without model-dependent fitting of the scattering profile, SAXS may be used to screen a large number of bicelle compositions to approximate the minimum ratio of lipid to detergent required for a bicelle phase transition in the low-q regime.

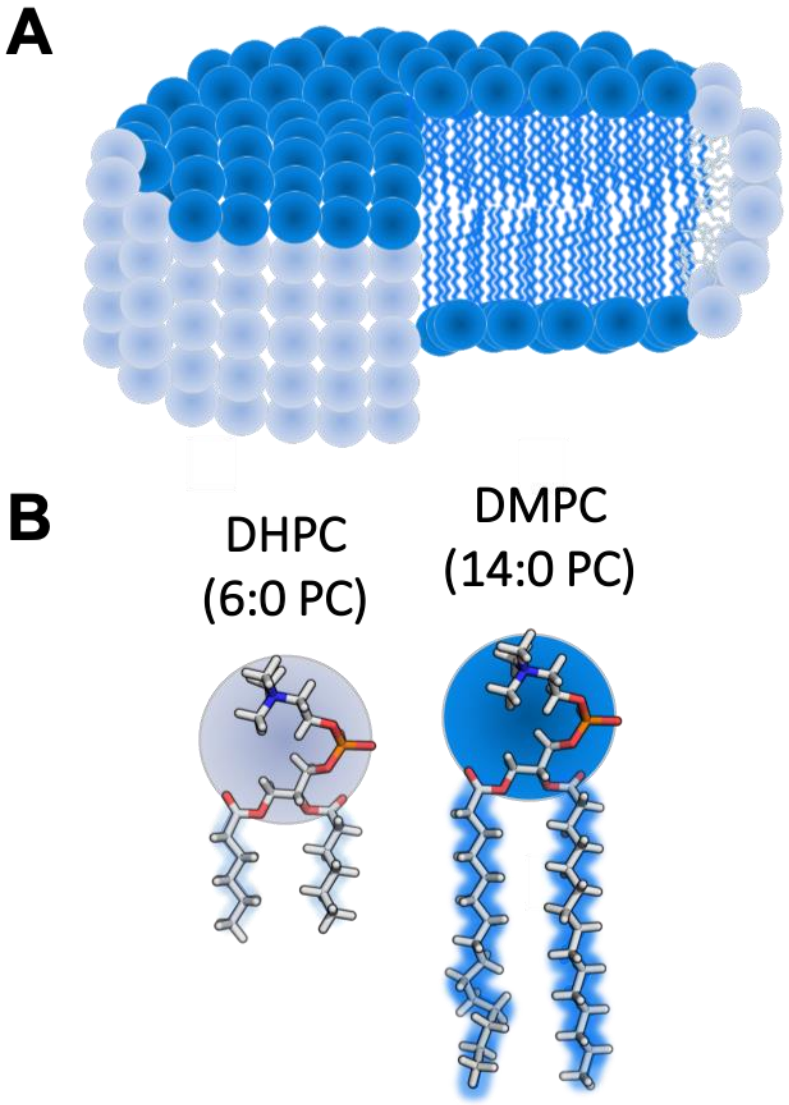


Figure 3.1: The ideal bicelle model. *Adapted from [9] with permission.* (A) A core of lipid (dark blue) encircled by a rim of detergent (light blue) represents the ideal bicelle morphology. (B) Three-dimensional structures of short-chain DH_6PC detergent (light blue shadow) and long-chain DMPC lipid (dark blue shadow) that form bicelles. Structures are rendered as sticks and include carbon-hydrogen bonds (gray) as well as oxygen (red), nitrogen (blue), and phosphorous (orange) atoms.

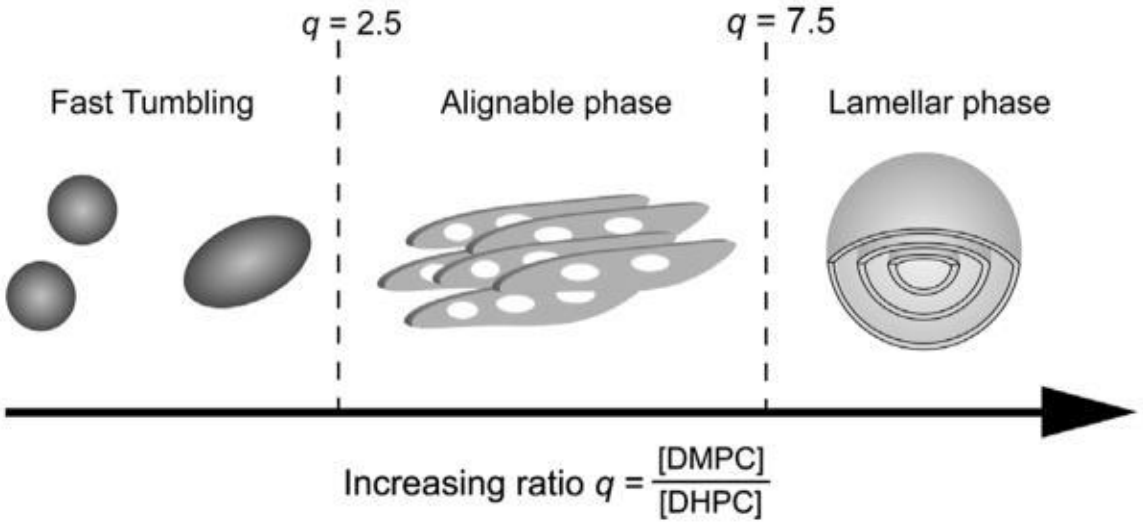


Figure 3.2: Bicelle morphology depends on the ratio of lipid to detergent (q). *Reprinted from [3] with permission.* Fast-tumbling (isotropic) bicelles exist when $q < 2.5$ (left panel). For $2.5 \leq q \leq 7.5$, perforated lamellae structures are present that undergo alignment within a magnetic field (middle panel). Above $q = 7.5$, multilamellar vesicles form (right panel).

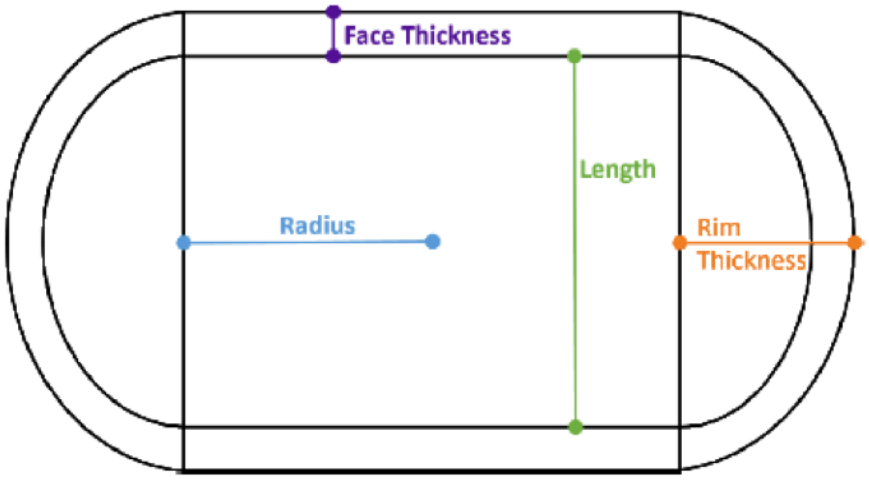


Figure 3.3: Dimensions in the core-shell bicelle model. *Reprinted from [9] with permission.* The core-shell bicelle model was used to fit experimental SANS scattering profiles in the study by Caldwell *et al* [9]. The thickness of the PC headgroup is represented by the face of the core-shell structure. SANS fitting also provides the length of the hydrophobic core region. The total radius of the bicelle across the major axis is determined by the sum of the core radius and rim thickness parameters.

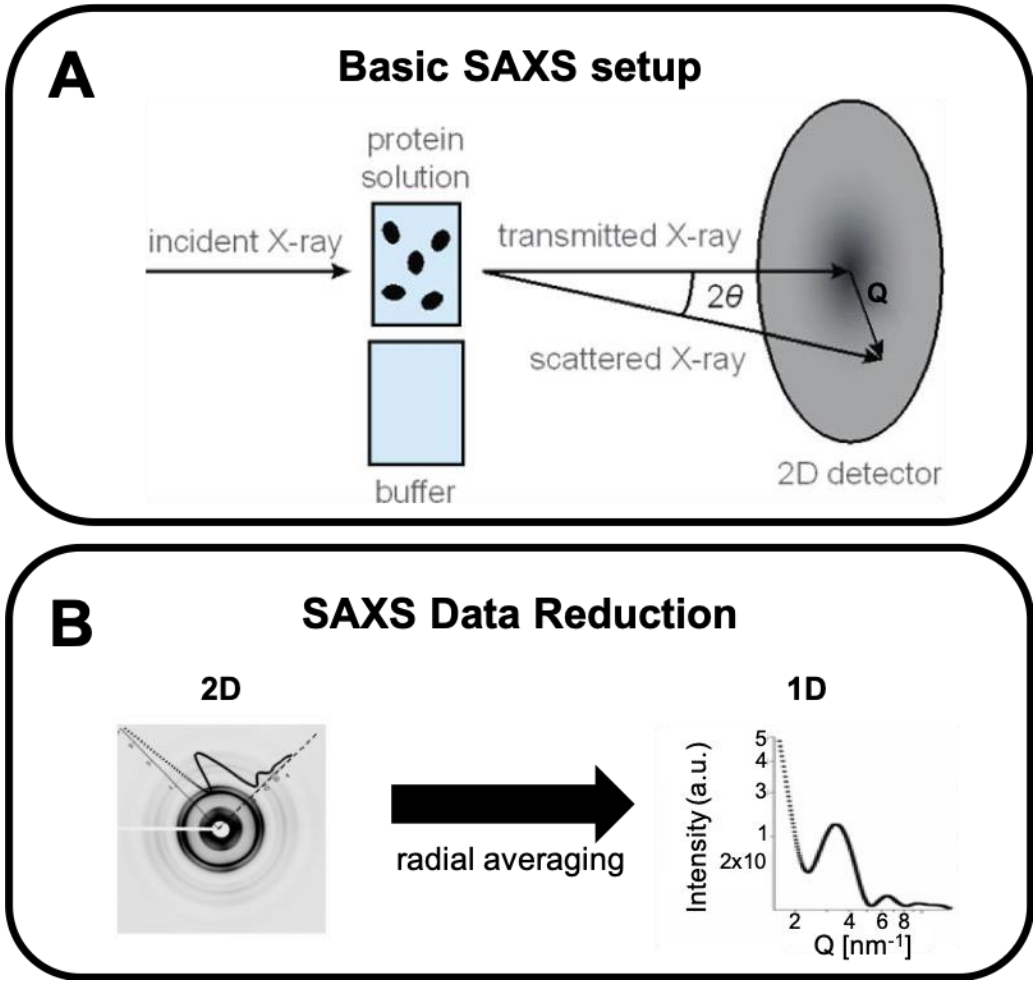


Figure 3.4: Basic SAXS setup and data reduction. (A) *Reprinted from [29] with permission*. A collimated, monochromatic X-ray beam incident on the macromolecule solution generates scattered X-rays, which are imaged by a 2D detector. The shadow in the image is the result of the transmitted beam blocked by a beamstop. The scattering vector, Q, describes the change in direction of the scattered X-rays. (B) *Adapted from the Safinya laboratory website at UC Santa Barbara [30]*. 2D images from SAXS are integrated by radial averaging to yield scattering intensity (also known as I(Q)) as a 1D function of the scattering vector Q.

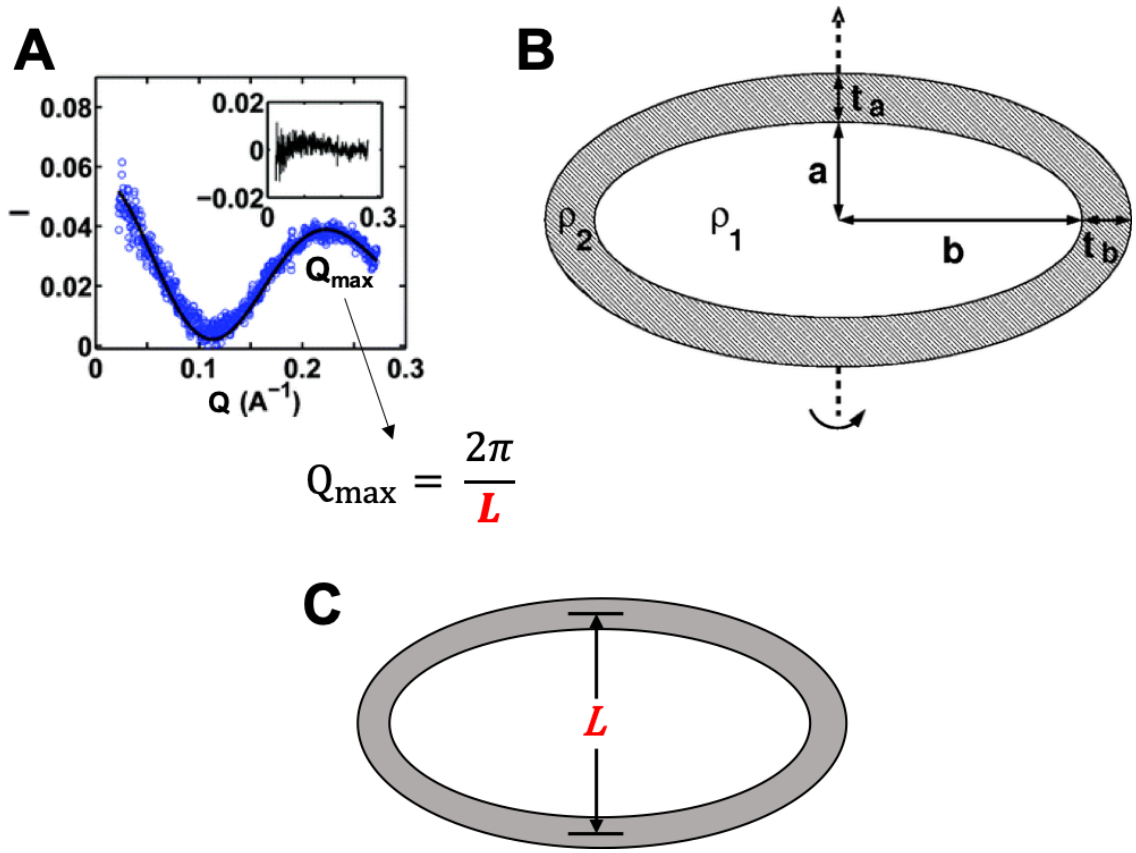


Figure 3.5: Scattering of two-shell ellipsoid structures. (A) Scattering intensity (I, blue dots) as a function of the scattering vector Q for 25 mM decylphosphocholine detergent. The second maximum (Q_{max}) in the scattering is characteristic of core-shell ellipsoid structures. The scattering fits to a two-component ellipsoid model (B), and residuals of the fit are shown in the inset. (B) Schematic of the two-component ellipsoid model used to fit detergent micelle scattering. The dashed line represents the rotation symmetry axis. The radii of the core minor axis (a) and major axis (b) are shown along with the shell minor axis (t_a) and major axis (t_b) thicknesses. The electron density of the hydrophobic core (ρ_1) is typically much lower than the head group region (ρ_2), and the contrast difference gives rise to the scattering second maximum in (A). (C) The position of Q_{max} approximates the

headgroup to headgroup distance, *L*, across the minor axis of the micelle. The *L* parameter is biased to the minor axis because more monomers are approximately separated by the short dimension relative to other distances in the ellipsoid. *Panels A and B adapted from [26] with permission.*

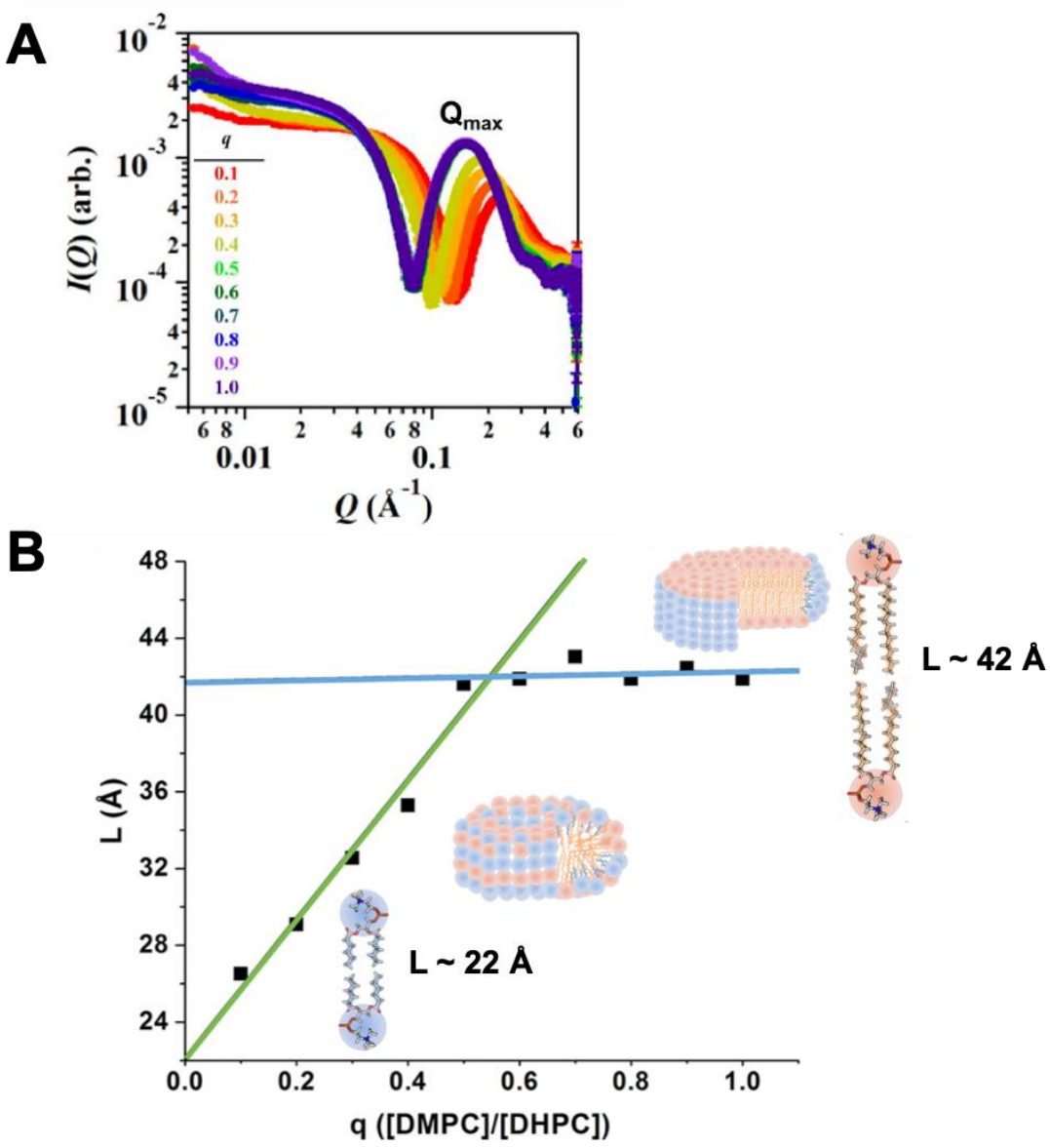


Figure 3.6: DMPC/DHPC (6:0 PC) bicelle scattering and headgroup to headgroup distance *L* as a function of the q-value. *Adapted from [9] with permission.* (A) Scattering of DMPC/DH_6PC bicelles for $0.1 \leq q \leq 1.0$. The position of the second maximum in the scattering (Q_{max}) is used to determine the headgroup to headgroup distance *L* (Figure 3.5C) as a function of the lipid to detergent q-value (B). When $q < 0.5$, L increases linearly with the q-value and supports mixing behavior between lipid and detergent. The linear region

below q = 0.5 may be extrapolated to the experimental *L* parameter for pure DH_6PC, which is ~ 22 Å (inset). When $q \geq 0.5$, a phase transition occurs from mixed lipid and detergent to a structure reflecting the approximate thickness of a pure DMPC bilayer (~42 Å, inset).

3.3 Study approach

As bicelle size for a number of lipid/detergent mixtures in the low-q regime have relied on model-dependent calculations from diffusion coefficients [12, 15], the model-independent approach to directly quantify bicelle size from SAXS measurements is better-suited to clarify the phase transition point for a given bicelle composition. It is possible that bicelle phase transitions in the low-q regime may depend on lipid and detergent physicochemical properties, which is suggested for CHAPS and Cyclofos-6-containing bicelles (section 3.2.3). Since membrane protein investigations using low-q bicelles routinely employ mixtures other than DMPC/DH_6PC [3, 6], it is imperative to establish the minimum ratio of lipid to detergent required to induce internal lipid organization for a given mixture if a predominantly lipid core is desirable for the protein of interest.

This study utilized synchrotron SAXS data to determine the dominant headgroup to headgroup distance across the short dimension (L) for 16 binary lipid-detergent mixtures ($0.0 < q\text{-value} < 2.0$). Bicelles containing PC lipids (DLPC, DMPC, DPPC, and POPC, Figure 3.7) are the focus of this study, which represent the majority of bicelles reported to date. Detergents used as rim-forming agents in bicelles for membrane protein applications were selected, including CHAPS, Cyclofos-6, FC12, DH_6PC, and DH_7PC (Figure 3.8). Additional detergents that have not been used in bicelles for membrane proteins, such as FC10, SDS, LDAO, and β-OG (Figure 3.8), are explored to assess bicelle formation and low-q phase transitions with PC lipids. Temperature-dependent changes in L are furthermore reported for each composition.

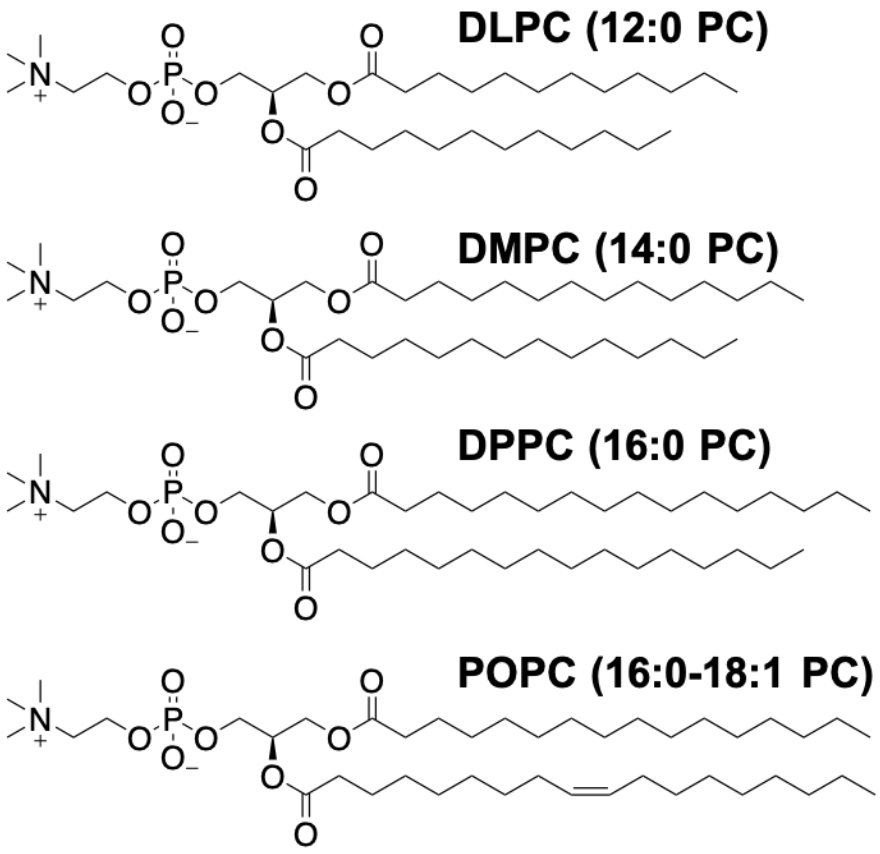


Figure 3.7: PC lipids selected for bicelle studies. Chemical structures of the PC lipids used in this work: 1,2-dilauroyl-*sn*-glycero-3-phosphocholine (12:0 PC, DLPC), 1,2-dimyristoyl-*sn*-glycero-3-phosphocholine (14:0 PC, DMPC), 1,2-dipalmitoyl-*sn*-glycero-3-phosphocholine (16:0 PC, DPPC), and 1-palmitoyl-2-oleoyl-glycero-3-phosphocholine (16:0-18:1 PC, POPC)

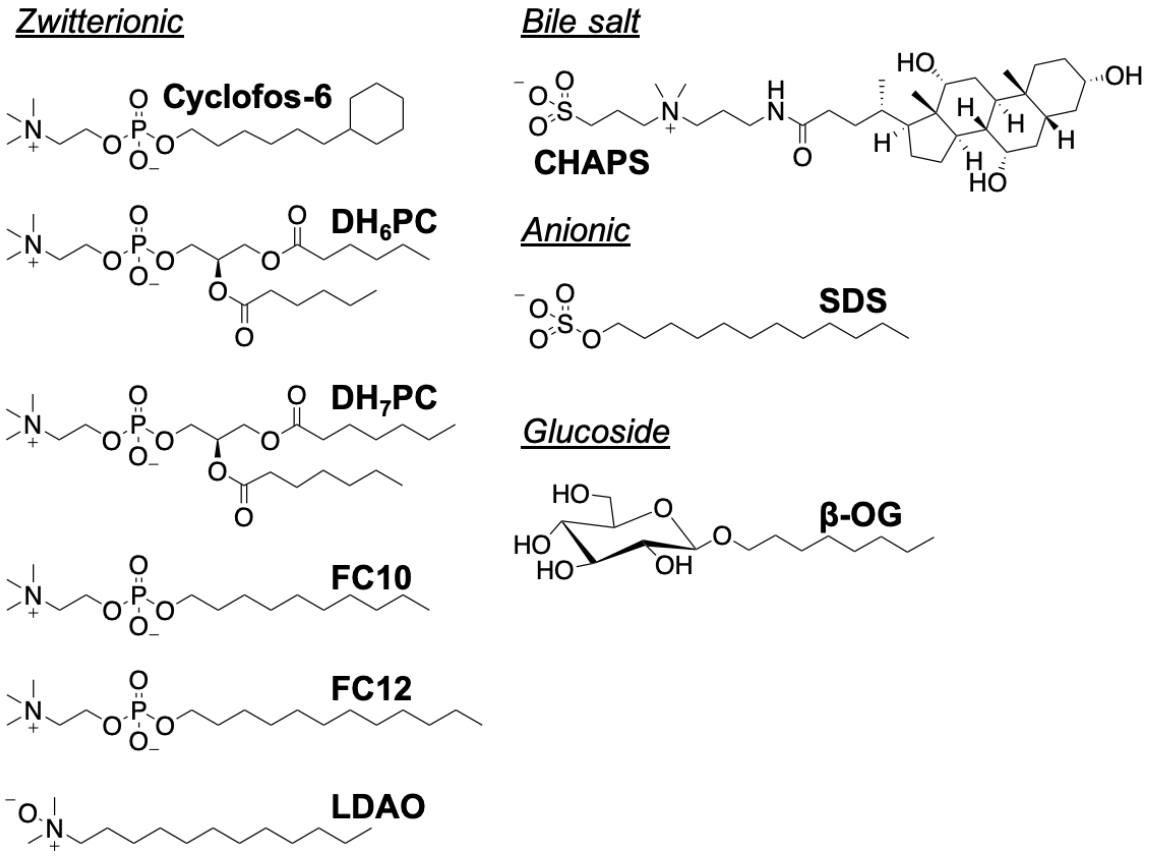


Figure 3.8: Detergents selected for bicelle studies. Chemical structures of the detergents used in this work. Detergents are grouped by detergent nomenclature based on headgroup (zwitterionic, anionic, glucoside) and/or structure. Zwitterionic: 6-cyclohexyl-1-hexylphosphocholine (Cyclofos-6); 1,2-dihexanoyl-*sn*-glycero-3-phosphocholine (6:0 PC, DH_6PC); 1,2-diheptanoyl-*sn*-glycero-3-phosphocholine (7:0 PC, DH_7PC); n-decylphosphocholine (FC10); n-dodecylphosphocholine (FC12); and n-dodecyl-*N*,*N*-dimethylamine-*N*-oxide (LDAO). Bile salt: 3-[(3-cholamidopropyl)-dimethylammonio]-1-propane sulfonate (CHAPS). CHAPS is both a bile salt and zwitterionic. Anionic: sodium dodecyl sulfate (SDS). Glucoside: n-octyl-β-D-glucopyranoside (β-OG).

3.4 Rationale and background for detergents used to form bicelles

3.4.1 CHAPS

CHAPS is a zwitterionic detergent that is a sulfobetaine derivative of cholic acid (Figure 3.8) [31]. CHAPS is generally classified as a "mild" (non-denaturing) detergent and may be used to solubilize membrane proteins in folded, functional states [32]. Both CHAPS and its hydroxylated form, CHAPSO, are considered among the top ten widely-used detergents in membrane protein structural biology studies [33]. Each have high CMC values (~8-10 mM) and form small micelles (~6-7 kDa) relative to other common detergents in the maltoside and fos-choline families [33]. Compared to traditional detergents with a polar headgroup and nonpolar acyl chain, CHAPS possesses a steroid structure (cholic acid) with a convex, hydrophobic back (β-plane) and a concave, hydrophilic face (α-plane) with three hydroxyl groups. The sulfobetaine chain is also hydrophilic and analogous to a traditional detergent headgroup. To form micelle aggregates, CHAPS detergent molecules therefore pack with the cholic acid α-plane and sulfobetaine chain exposed to the solvent, while the hydrophobic β-plane faces toward the micelle interior [34].

Bicelles containing DMPC and CHAPSO were first reported by the Prestegard laboratory in 1990 [1] at q-values sufficient to induce magnetic alignment. Isotropically-tumbling DMPC/CHAPS bicelles have since been used in studies with bacteriorhodopsin [35], opsin [36], and the membrane domain of the rat TrkA receptor [12]. Authors examining the TrkA receptor in bicelle compositions proposed that CHAPS is preferable

for bicelle applications with membrane proteins if detergent deuteration is not required (not commercially-available for CHAPS). Due to the enhanced stability and folding properties of TrkA in DMPC/CHAPS bicelles and inferred lipid segregation from NMR diffusion coefficients, the authors proposed that CHAPS is better than DH_6PC for maintaining the ideal bicelle model in the low-q regime.

3.4.2 Cyclofos-6

Cyclofos-6 is a zwitterionic detergent that contains a phosphocholine headgroup and an acyl chain tail terminated with a cyclohexyl group (Figure 3.8). To date, Cyclofos detergent micelle morphologies have not been characterized, but CMC values are available from manufacturer analytical studies (Anatrace). Cyclofos detergents are not widely used to study membrane proteins, but a notable example comes from the structure and function of bacterial outer membrane protein PhoPQ-activated gene P (PagP) solubilized in Cyclofos-7 [37]. The first report of Cyclofos detergents in bicelle systems was reported by Lu *et al.* [15]. When combined with DMPC at q = 0.3, Cyclofos-6 was considered most promising for bicelle formation and is able to form lipid-detergent assemblies under highly dilute conditions ($C_L < 0.5\%$). DMPC/Cyclofos-6 mixtures (q = 0.3) have additionally been employed to study the transmembrane and C-terminal domain of the human integrin β1 subunit [15]. Similar to CHAPS, Lu *et al.* [15] proposed that Cyclofos-6 improves bicelle stability relative to DH_6PC and is less likely to partition into the lipid core based on their results.

3.4.3 DH_6PC and DH_7PC

DH_6PC is the detergent component most often used in bicelle compositions, along with DMPC lipid [3, 6]. Like DMPC, DH_6PC shares the same phosphatidylcholine headgroup and two acyl chains (Figure 3.8). However, DH_6PC occupies a cylindrical volume due to the surface area of the short hydrocarbon tails (6:0 PC) relative to the headgroup. DH_6PC, along with the longer-chain DH_7PC (7:0 PC, Figure 3.8), is therefore considered a detergent that forms micellar aggregates [3]. The CMC of DH_6PC is high (~14 – 15 mM) relative to DH_7PC (1.4 mM), and both have been used to solubilize membrane proteins [38]. Isotropic bicelles formed with DH_7PC are less common than DH_6PC for membrane protein studies, but a recent example is lipoprotein signal peptidase A (LspA) in DMPC/DH_7PC from DMPC nanodiscs [39]. Although the lipid/detergent mixing properties in low-q isotropic bicelles has been established for DMPC/DH_6PC, previous literature suggests that DH_6PC may partition more with shorter-chain lipids such as DLPC (12:0 PC) [40]. Following the same logic, DH_7PC may partition more with DMPC due to the longer chain length relative to DH_6PC. If miscibility between lipid and detergent is increased for a bicelle relative to DMPC/DH_6PC, the headgroup to headgroup thickness from SAXS is expected to plateau at q-values above 0.5.

3.4.4 Fos-cholines

Fos-choline detergents possess a phosphocholine headgroup and single acyl chain. The acyl chain length ranges from eight (FC8) to sixteen carbons (FC16). Two detergents from the fos-choline family, FC10 and FC12 (Figure 3.8), have been used to solubilize a

large number of membrane proteins, particularly in the NMR spectroscopy field [33]. The CMCs of FC10 and FC12 are 11 mM and 1.5 mM, respectively. Isotropic DMPC/FC12 bicelles have been used to study amyloid-β peptide(42) barrel formation [41], outer membrane protein X [42], and human ether-a-go-go related gene (hERG) voltage-gated potassium channels [43]. There are no reports of DMPC/FC10 bicelles in membrane protein preparations to date.

3.4.5 SDS, LDAO, and β-OG

The last set of detergents used in this study are among the top ten found in membrane protein crystallization (LDAO, β-OG) and NMR studies with bicelles (SDS) [33] (Figure 3.8). These specific detergents were also selected to examine bicelle formation based on headgroup properties that deviate from phosphocholine-like groups, including anionic (SDS) and glucoside (β-OG) detergents. For example, it is possible that the large glucoside headgroup of β-OG would prevent detergent partitioning with bicelle lipids. DMPC bicelles formed with LDAO and β-OG are not reported in the literature, and SDS is only present in ternary bicelle mixtures [44].

3.5 Materials and methods

3.5.1 Bicelle preparation

The detergents 1,2-dihexanoyl-*sn*-glycero-3-phosphocholine (6:0 PC, DH_6PC) and 1,2-diheptanoyl-*sn*-glycero-3-phosphocholine (7:0 PC, DH_7PC) and lipids 1,2-dilauroyl-*sn*-glycero-3-phosphocholine (12:0 PC, DLPC), 1,2-dimyristoyl-*sn*-glycero-3-phosphocholine (14:0 PC, DMPC), 1,2-dipalmitoyl-*sn*-glycero-3-phosphocholine (16:0 PC, DPPC), and 1-palmitoyl-2-oleoyl-glycero-3-phosphocholine (16:0-18:1 PC, POPC) were purchased from Avanti Polar Lipids in powder form. Detergents 3-[(3-cholamidopropyl)-dimethylammonio]-1-propane sulfonate (CHAPS, sol-grade), 6-cyclohexyl-1-hexylphosphocholine (Cyclofos-6, anagrade), n-decylphosphocholine (FC10, sol-grade), n-dodecylphosphocholine (FC12, sol-grade), sodium dodecyl sulfate (SDS), n-dodecyl-*N*,*N*-dimethylamine-*N*-oxide (LDAO, sol-grade), and n-octyl-β-D-glucopyranoside (β-OG, sol-grade) were purchased from Anatrace in powder form. Chemical structures for all lipids and detergents are provided in Figure 3.7 and Figure 3.8

Prior to forming bicelle mixtures, powder lipids and detergents were resuspended in 20 mM sodium phosphate buffer, 150 mM NaCl, pH = 6.2 and mixed by nutation at room temperature for 30 minutes. The phosphate buffer and pH chosen reflect conditions commonly used for solution NMR spectroscopy experiments [9]. Lipids and detergents were combined to give a total amphiphile concentration (C_L) of 6% in 0.5 mL working volume. Each mixture was prepared with effective bicelle q-values of 0.0, 0.1, 0.2, 0.3, 0.4,

0.5, 0.6, 0.7, 0.8, 0.9, 1.0, 1.5, and 2.0, accounting for the concentration of free detergent in solution according to:

$$q_{eff} = \frac{[\text{lipid}]}{[\text{detergent}] - CBC_{\text{detergent}}}$$

Equation 3.4

where q_{eff} is the effective q-value and $CBC_{detergent}$ represents the critical bicelle concentration (CBC) of a given detergent [36]. For example, the CBC of DH_6PC is ~ 7-9 mM [9, 12], which is roughly half of its critical micelle concentration (CMC). All q-values listed for SAXS experiments reflect the q_{eff} value rather than the q-value given by Equation 3.1. A C_L of 6% was chosen because previous studies have reported that more dilute conditions produce larger bicelle aggregates [9], and C_L = 6% has been successful in forming a wide number of lipid-detergent bicelle mixtures [15]. Under dilute C_L conditions (< 6%), the effective detergent concentration in the bicelle aggregate is greatly reduced by the concentration of free detergent in solution, which increases the q_{eff} and likely accounts for the larger aggregates observed in earlier studies. When C_L = 6%, the CBC term in Equation 3.4 becomes negligible. As experimental CBC values are reported for only DH_6PC, the higher amphiphile concentration allowed a reasonable approximation to be used for any detergents with unknown CBC values (half of the detergent CMC) without introducing significant error in the effective q-value.

After forming the final lipid/detergent mixture with the desired q-value and % C_L, samples were vortexed for 60 sec. Samples were then subjected to freeze/thaw cycles of -80°C and 37°C for 30 minutes each, with vortexing for 60 sec before each freeze step.

Cycles of freezing, thawing, and vortexing were performed until the solution was optically clear. At the point of optical clarity, three more freeze/thaw cycles were performed. For lipid/detergent mixtures that did not become clear after several freeze/thaw cycles, the suspension was subjected to pulse sonication (10% amplitude, 10 sec on, 10 sec off) for ~15 minutes (Qsonica). Several more freeze/thaw cycles were performed by flash-freezing in liquid nitrogen followed by a thawing step at 37°C and repeated sonication. If the mixture failed to become a transparent solution after both preparation methods, the mixture was considered unviable for bicelle formation and SAXS experiments. Final lipid/detergent solutions were filtered using 0.5 mL centrifugal filters with a 0.2 μm nylon membrane (VWR). Blank buffer was filtered using the same method to provide optimal background subtraction for scattering measurements. Samples were stored at -80°C, thawed in a room temperature water bath, and vortexed immediately prior to SAXS experiments.

3.5.2 SAXS data collection

Synchotron SAXS data were measured with the 12-ID-C beamline at the Advanced Photon Source of the Argonne National Laboratory. Samples were loaded into quartz capillary tubes (2.0 mm O.D.) with 0.01 mm wall thickness (Charles Supper) for data collection. The incident photon energy was 18 keV (wavelength = 0.69 Å), and the sample-to-detector distance was adjusted to provide a scattering wave-vector range Q of $0.001 < Q < 1.140$ Å^{-1}. For initial experiments with DLPC/DH_6PC, DPPC/DH_6PC, POPC/DH_6PC, and DMPC/FC10 ($0.0 \leq q \leq 1.0$), the incident photon energy was 12 keV (wavelength = 1 Å), and the sample-to-detector distance was adjusted to provide a scattering wave-vector

range Q of $0.006 < Q < 1.079$ $Å^{-1}$. A mosaic X-ray charge-coupled device detector was used to acquire images with typical exposure times of 1.0 sec. A temperature-controlled sample holder with 13 capillary slots was used to record data from 10 to 60°C in 5°C increments with 3-minute equilibration periods between measurements. A final measurement was recorded at 10°C to assess reversible phenomena. For initial experiments with DLPC/DH_6PC, DPPC/DH_6PC, POPC/DH_6PC, and DMPC/FC10 ($0.0 \leq q \leq 1.0$), data was recorded for 25°C, 45°C, and 60°C. SAXS measurements with filtered buffer were collected in parallel for background subtraction. 2D scattering data were radially-averaged upon acquisition to give the measured scattering intensity I(Q) as a function of Q ($Å^{-1}$). A total of ~2100 SAXS measurements were acquired over the course of 30 hours.

3.5.3 SAXS data processing

SAXS profiles [I(Q) vs. Q ($Å^{-1}$)] from matched buffer conditions were subtracted from sample scattering profiles using SasView [45]. Programs to combine and plot background-subtracted SAXS profiles were written in R. Scattering intensities, I(Q), were normalized to the second maximum for each profile. Q-values of second maxima were used to calculate the dominant headgroup to headgroup distance across the short dimension, *L*, of lipid/detergent assemblies according to Equation 3.3. Source code for combining SAXS data, generating scattering plots, and calculating *L* parameters is provided in Appendix I.

3.5.4 Calculation of L_{max} constraint for bicelles with PC lipids

For the ideal bicelle model, the maximum thickness (L_{max}) across the short dimension is constrained to the length of fully-extended lipid acyl chains. Exceeding this

distance results in unoccupied volume at the bicelle center [46]. The maximum extension of the lipid alkyl chain length (l_c (Å), Equation 3.5) can be determined from Tanford's formula [46] based on the number of aliphatic carbons (n_c).

$$l_c = 1.5 + 1.265n_c$$

Equation 3.5

The constraint L_{max} (Å) for a given bicelle is then calculated using Equation 3.6, where t is the shell thickness corresponding to the lipid headgroup.

$$L_{max} = 2(1.5 + 1.265n_c) + t$$

Equation 3.6

For DMPC/DH_6PC bicelles, the average shell thickness is ~ 8 Å (range = 7-12 Å) when q = 0.3 and 0.7, which is consistent with the PC headgroup for the two components [9]. As all lipids in this study possess PC headgroups, L_{max} constraints for DLPC, DMPC, DPPC, and POPC bicelles were calculated based on Equation 3.6. Due to the range of reported DMPC/ DH_6PC shell thicknesses for q = 0.3 and 0.7, Table 3.1 summarizes L_{max} constraints based on the average PC shell thickness (t = 8 Å) and range of experimental shell thicknesses (t = 7 – 12 Å).

Table 3.1: PC lipid T_m and theoretical L_{max} constraints

Lipid	T_m (°C)	Average L_{max} (Å)	Range L_{max} (Å)
DLPC	-2	39	38-43
DMPC	24	44	43-49
DPPC	41	49	48-53
POPC	-2	49	48-53

3.6 Results and discussion

3.6.1 CHAPS bicelles

To test the hypothesis that CHAPS may induce lipid organization within a bicelle at low-q ratios (< 0.5), binary mixtures of CHAPS with DLPC, DMPC, and POPC lipids were formed with $0.0 \leq q \leq 2.0$. If lipids are localized in the planar region of CHAPS bicelles when $q < 0.5$, the headgroup to headgroup thickness L across the short dimension from SAXS measurements should reflect the predicted L_{max} for the PC lipids. All compositions formed isotropic bicelles up to q = 2.0 except for POPC/CHAPS, which became gel-like at q = 2.0. The SAXS profiles of DLPC/CHAPS, DMPC/CHAPS, and POPC/CHAPS at 25°C are shown in Figure 3.9 for $0.1 \leq q \leq 1.0$. Compared to amphiphiles with acyl chains, CHAPS micelles (q = 0.0) do not produce a second maximum in the scattering profile due to insufficient contrast difference between the electron densities of the hydrophobic interior and hydrophilic head group [26]. This feature of CHAPS prevents model-independent structure calculations for pure CHAPS micelles. However, a pronounced second maximum is expected in the scattering profile for a segregated lipid core, making SAXS particularly advantageous for interpreting the morphology of CHAPS bicelles. For all CHAPS bicelles, a second maximum appears at q = 0.2 with increasing intensity as the ratio of lipid to detergent increases (Figure 3.9). The second maximum intensity becomes constant when $q \sim 0.7$ for DLPC/CHAPS, $q \sim 0.5$ for DMPC/CHAPS, and $q \sim 0.4$ for POPC/CHAPS. The position of the second maximum, Q_{max}, simultaneously shifts to smaller values with larger bicelle q, reflecting an increase across the short

dimension of the bicelle core. Minor changes are observed in the SAXS profiles when q > 1.0 (Figure 3.10).

Experimental *L* parameters were calculated from Q_{max} for each scattering profile according to Equation 3.3. The observed L_{max} values for DLPC/CHAPS, DMPC/CHAPS, and POPC/CHAPS were approximately 33 Å, 37 Å, and 43 Å, respectively. All observed L_{max} parameters are well below the theoretical L_{max} for DLPC, DMPC, and POPC lipids in a planar bilayer phase (Table 3.1). Furthermore, the observed L_{max} remains constant when q ~ 0.8 for each composition (Figure 3.11). Above q = 1.0, experimental L_{max} values increase by ~ 2 Å, which improves agreement between the measured *L* for high q-values and the lower range of expected L_{max} parameters. While it is possible that CHAPS induces thinning in the lipid core, the low observed L_{max} values are likely due to intermittent CHAPS molecules in the core that decrease the effective contrast difference. As the concentration of CHAPS is lowered in the bicelles (Figure 3.10), the experimental L_{max} approaches the theoretical value, supporting that the core contrast becomes closer to a pure PC lipid core.

Small temperature-dependent differences in *L* parameters were observed for q = 0.3 and q = 1.0 CHAPS bicelles (Figure 3.12, Figure 3.13, Figure 3.14, panel D). However, the second maximum intensity for q = 0.3 CHAPS bicelles noticeably increased with temperature (Figure 3.12, Figure 3.13, Figure 3.14, panel B). Temperature-dependent second maximum intensities increased with lipid chain length, suggesting that each aggregate possesses the properties of mixed lipid and detergent for q = 0.3 CHAPS bicelles. Any changes to bicelle shape and size with temperature were reversible, as scattering

profiles were similar following a temperature cycle (10°C → 60°C → 10°C). Combined, the SAXS profiles and experimental *L* parameters for PC lipids mixed with CHAPS suggest that CHAPS does not induce lipid organization when $q < 0.8$.

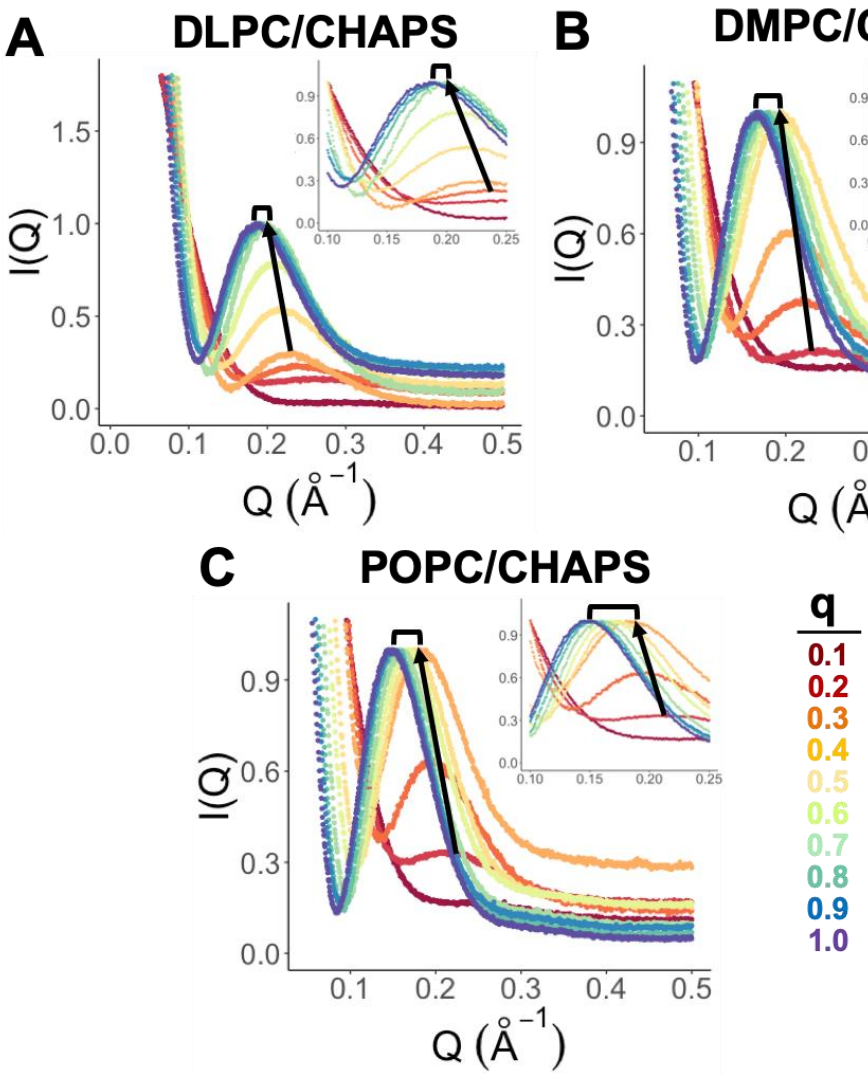


Figure 3.9: Scattering of CHAPS bicelles with q-values ranging from 0.1 to 1.0. SAXS profiles of scattering intensity, I(Q), as a function of the magnitude of the scattering vector Q (Å^{-1}) for DLPC/CHAPS (A), DMPC/CHAPS (B), and POPC/CHAPS (C) lipid-detergent mixtures at 25°C. Scattering profiles were plotted as rainbow gradients for bicelle q-values between 0.1 (red) and 1.0 (violet). Zoomed scattering in the second maximum region is inset for each plot. Arrows indicate the shift in the second maximum position and intensity, while braces indicate the shift in scattering Q (Å^{-1}) with increasing bicelle q-values.

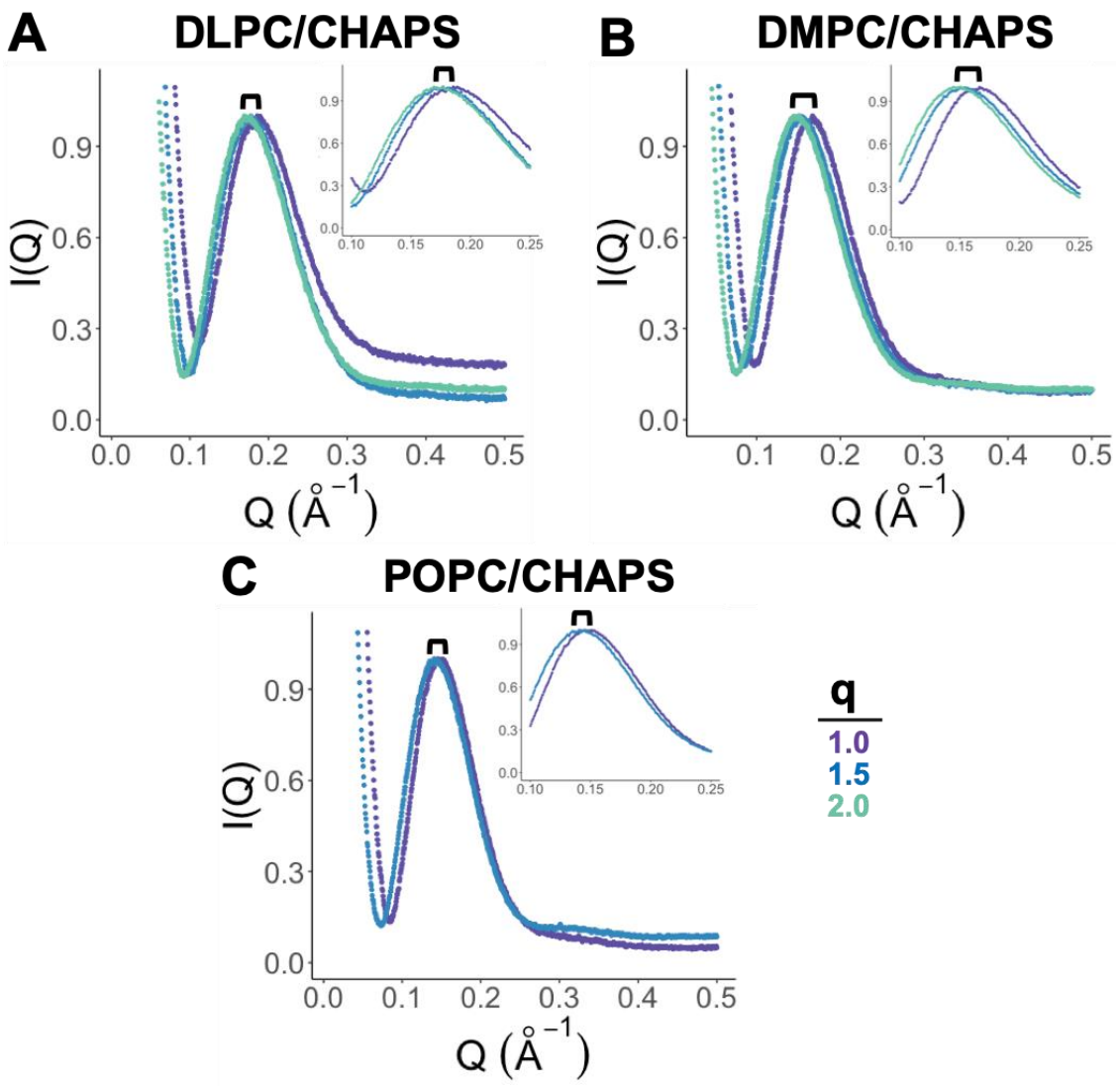


Figure 3.10: Scattering of CHAPS bicelles with q-values > 1.0. SAXS profiles of scattering intensity, I(Q), as a function of the magnitude of the scattering vector Q (Å^{-1}) for DLPC/CHAPS (A), DMPC/CHAPS (B), and POPC/CHAPS (C) lipid-detergent mixtures above q-value = 1.0 at 25°C. Zoomed scattering in the second maximum region is inset for each plot. Braces indicate the shift in scattering Q (Å^{-1}) with increasing bicelle q-values. POPC/CHAPS did not form isotropically-tumbling bicelles above q = 1.5.

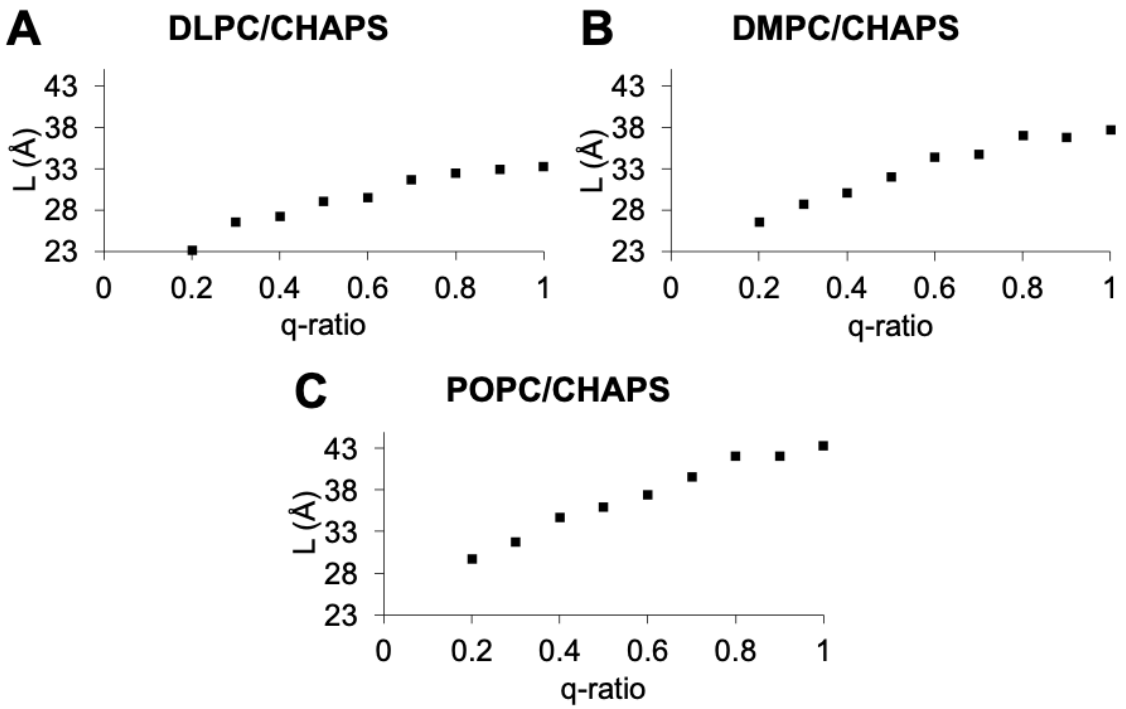
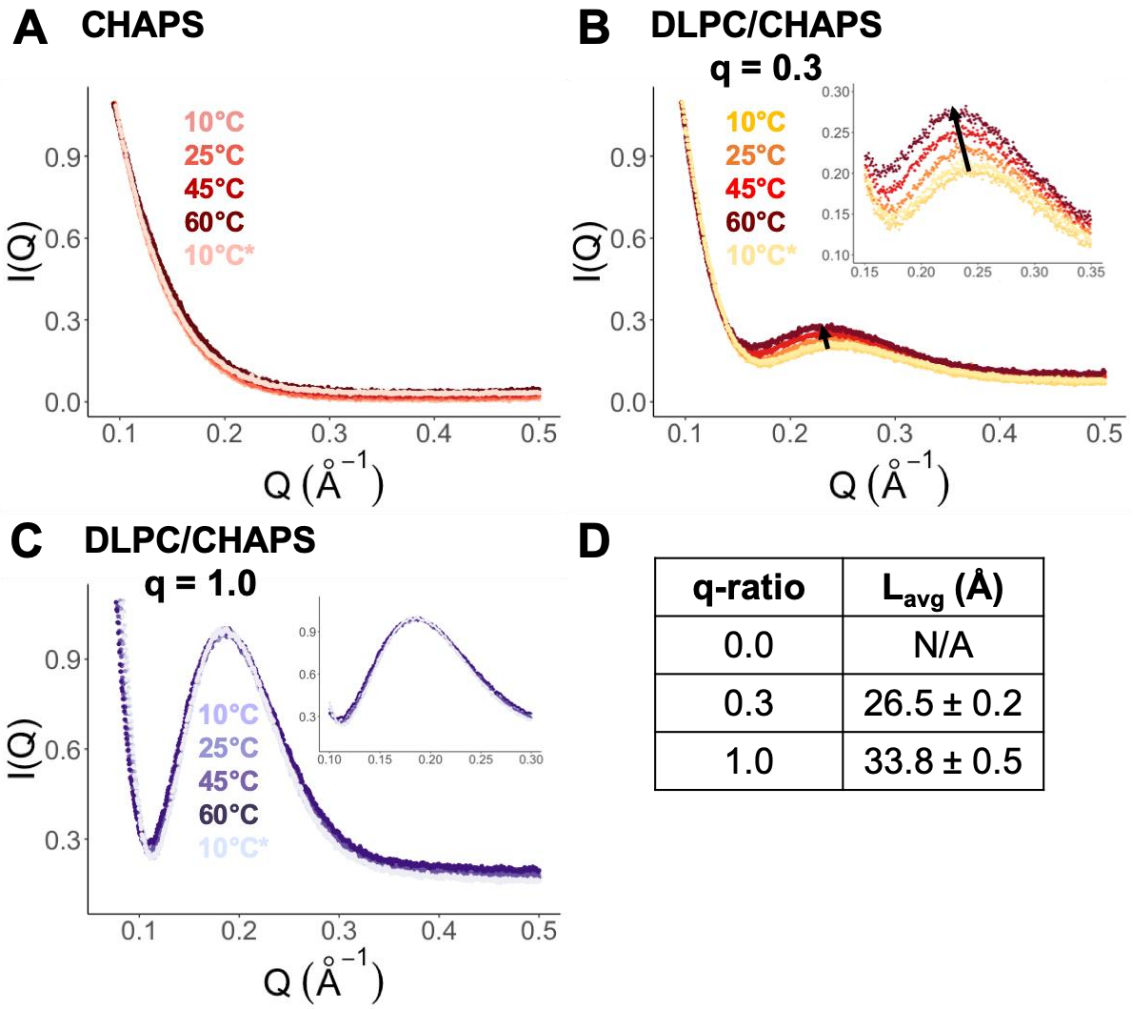


Figure 3.11: Headgroup to headgroup thickness (L) of CHAPS bicelles according to bicelle q-values. The headgroup to headgroup thickness (L) across the short dimension of the bicelle/mixture core at 25°C was determined from the second maximum position, Q_{max} ($Å^{-1}$) using Equation 3.3. Experimental L parameters were determined for (A) DLPC/CHAPS, (B), DMPC/CHAPS, and (C), POPC/CHAPS bicelles for $0.2 \leq q \leq 1.0$. The headgroup to headgroup thickness of CHAPS mixtures below q = 0.2 could not be determined from SAXS profiles because a second maximum was not observed in the scattering.



q-ratio	L_{avg} (Å)
0.0	N/A
0.3	26.5 ± 0.2
1.0	33.8 ± 0.5

Figure 3.12: Temperature-dependent DLPC/CHAPS scattering and averaged parameters. Temperature-dependent SAXS profiles for DLPC/CHAPS when q = 0.0 (pure CHAPS, A), q = 0.3 (B), and q = 1.0 (C). Scattering was recorded for each mixture during a temperature cycle from 10°C to 60°C, with a final temperature of 10°C (denoted by asterisks). Zoomed scattering in the second maximum region is inset for each plot. Arrows indicate the shift in the second maximum position and intensity. (D) Average headgroup to headgroup thickness, L, for bicelle q-values in A-C during the temperature cycle.

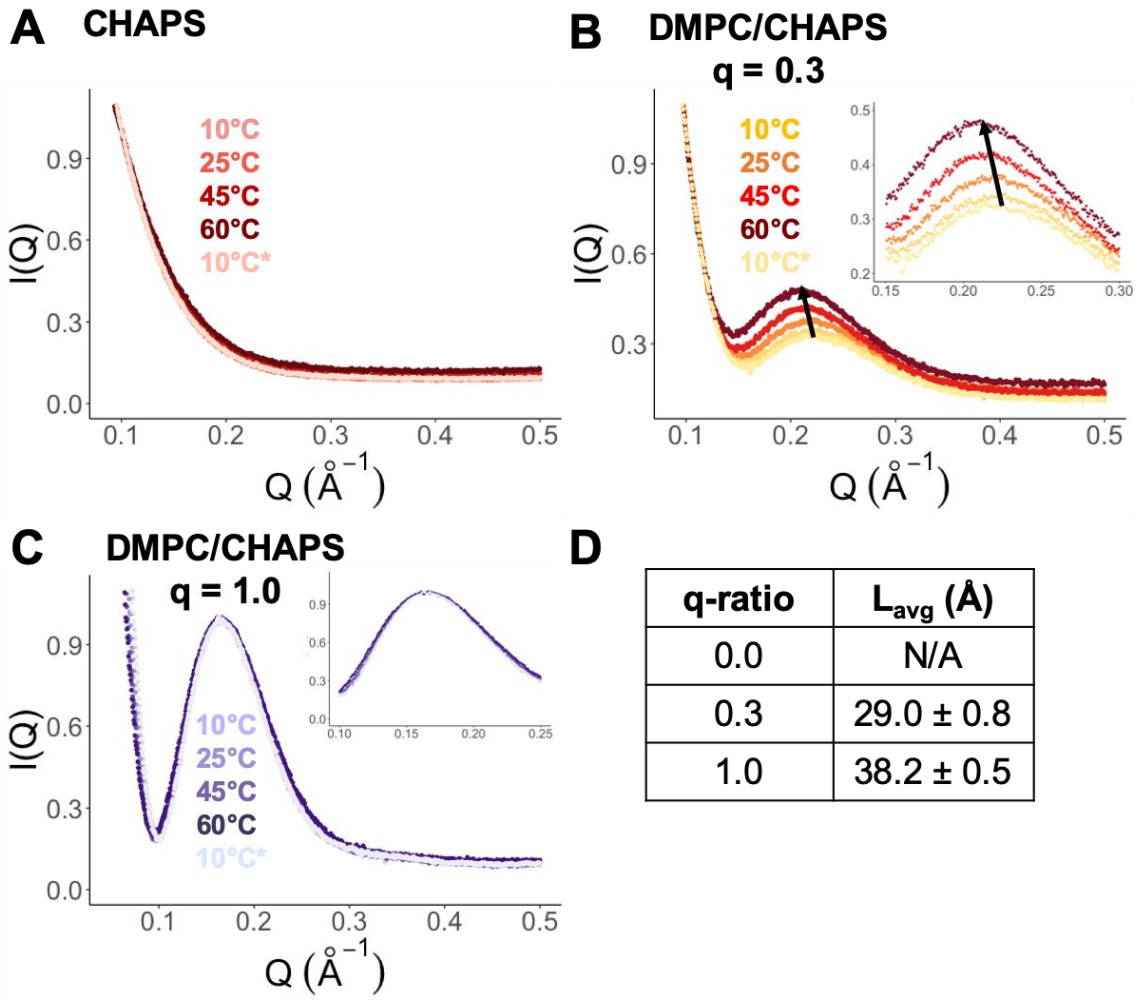
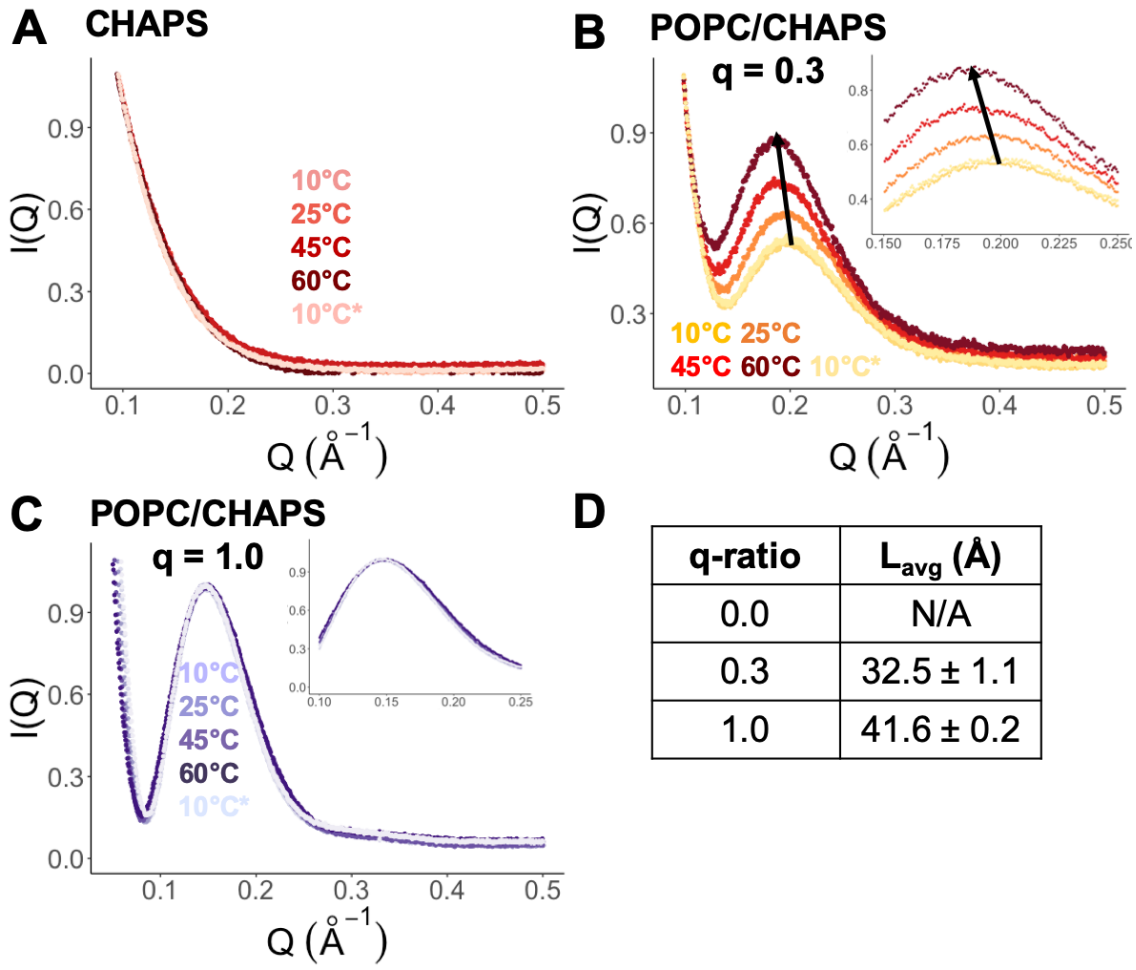


Figure 3.13: Temperature-dependent DMPC/CHAPS scattering and averaged parameters. Temperature-dependent SAXS profiles for DMPC/CHAPS when q = 0.0 (pure CHAPS, A), q = 0.3 (B), and q = 1.0 (C). Scattering was recorded for each mixture during a temperature cycle from 10°C to 60°C, with a final temperature of 10°C (denoted by asterisks). Zoomed scattering in the second maximum region is inset for each plot. Arrows indicate the shift in the second maximum position and intensity. (D) Average headgroup to headgroup thickness, L, for bicelle q-values in A-C during the temperature cycle.



q-ratio	L_{avg} (Å)
0.0	N/A
0.3	32.5 ± 1.1
1.0	41.6 ± 0.2

Figure 3.14: Temperature-dependent POPC/CHAPS scattering and averaged parameters. Temperature-dependent SAXS profiles for POPC/CHAPS when q = 0.0 (pure CHAPS, A), q = 0.3 (B), and q = 1.0 (C). Scattering was recorded for each mixture during a temperature cycle from 10°C to 60°C, with a final temperature of 10°C (denoted by asterisks). Zoomed scattering in the second maximum region is inset for each plot. Arrows indicate the shift in the second maximum position and intensity. (D) Average headgroup to headgroup thickness, L, for bicelle q-values in A-C during the temperature cycle.

3.6.2 Cyclofos-6 bicelles

To compare with SAXS measurements of CHAPS bicelles, Cyclofos-6 bicelles were formed with DLPC, DMPC, and POPC lipids for $0.0 \leq q \leq 2.0$. DLPC/Cyclofos-6 mixtures above q = 1.0 formed gels and were not included in SAXS experiments. Unlike CHAPS, all Cyclofos-6 bicelles display a prominent second maximum in the scattering profile characteristic of amphipathic core-shell structures (Figure 3.15). For Cyclofos-6 bicelle q-values with $0.1 \leq q \leq 1.0$, the scattering profiles stop changing in the forward scattering and second maximum regions when q ~ 0.6 at 25°C. Like CHAPS bicelles, DMPC/Cyclofos-6 and POPC/Cyclofos-6 bicelles above q = 1.0 displayed minor changes with SAXS measurements (Figure 3.16).

Experimental *L* parameters were calculated from Q_{max} for each scattering profile according to Equation 3.3. The headgroup to headgroup distance across the core of Cyclofos-6 micelles is ~ 28-29 Å based on SAXS measurements (Figure 3.17), which is most consistent with FC10 micelles [19, 20, 25, 26]. As FC10 and Cyclofos-6 share the same headgroup and a similar acyl chain length, these SAXS results agree well with previous literature and theoretical calculations from **Equation 3.5** and **Equation 3.6**. The observed *L* parameters remain within 1 Å of the maximum *L* reached when q ~ 0.4 for DLPC/Cyclofos-6, q ~ 0.7 for DMPC/Cyclofos-6, and q ~ 0.5 for POPC/Cyclofos-6 (Figure 3.17). The experimental L_{max} values for DLPC/Cyclofos-6, DMPC/Cyclofos-6, and POPC/Cyclofos-6 were approximately 40 Å, 48 Å, and 49 Å, respectively. L_{max} for DLPC/Cyclofos-6 and POPC/Cyclofos-6 bicelles agrees well with the average theoretical L_{max} values for DLPC (39 Å) and POPC (49 Å) bilayers. However, DMPC/Cyclofos-6

displayed an L_{max} approximately 4 Å larger than the expected L_{max} for DMPC, albeit within the range of theoretical values (Table 3.1). Above q = 1.0, experimental L_{max} values increase by < 1 Å.

Small temperature-dependent differences in scattering profiles and *L* parameters were observed for most Cyclofos-6 bicelle conditions (Figure 3.18, Figure 3.19, Figure 3.20). However, POPC/Cyclofos-6 q = 1.0 bicelles showed a sensitivity to increasing temperature above 25°C (Figure 3.20, panel C). Two sharp peaks appear at 45°C in the second maximum region for scattering profiles of POPC/Cyclofos-6 q = 1.0 bicelles, which indicates formation of multilamellar vesicles [47]. Bicelle fusion to form multilamellar vesicles at high temperatures appears to be a property of POPC/Cyclofos-6 bicelles, as evidence of multilamellar vesicles was not observed for DLPC/Cyclofos-6 and DMPC/Cyclofos-6 bicelles. Notably, the temperature-dependent transition of POPC/Cyclofos-6 bicelles to vesicles at q = 1.0 is irreversible based on the scattering profiles before and after the temperature cycle (Figure 3.20, panel C). Combined, the SAXS profiles and experimental *L* parameters for PC lipids mixed with Cyclofos-6 suggest that Cyclofos-6 does not induce lipid organization when q < 0.5.

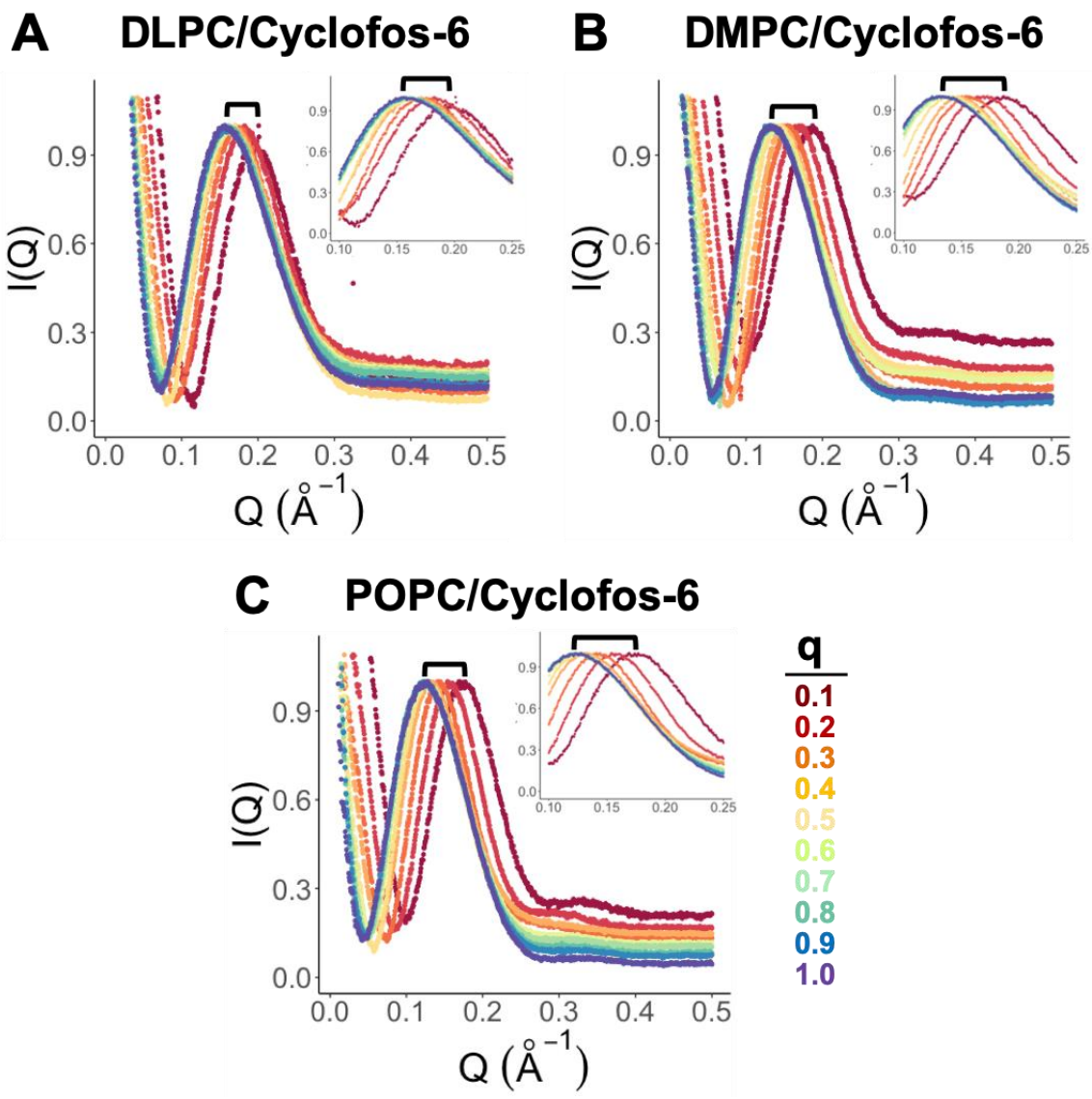


Figure 3.15: Scattering of Cyclofos-6 bicelles with q-values ranging 0.1 to 1.0. SAXS profiles of scattering intensity, I(Q), as a function of the magnitude of the scattering vector Q (Å^{-1}) for DLPC/Cyclofos-6 (A), DMPC/Cyclofos-6 (B), and POPC/Cyclofos-6 (C) lipid-detergent mixtures at 25°C. Scattering profiles were plotted as rainbow gradients for bicelle q-values between 0.1 (red) and 1.0 (violet). Zoomed scattering in the second maximum

region is inset for each plot. Braces indicate the shift in scattering Q ($Å^{-1}$) with increasing bicelle q-values.

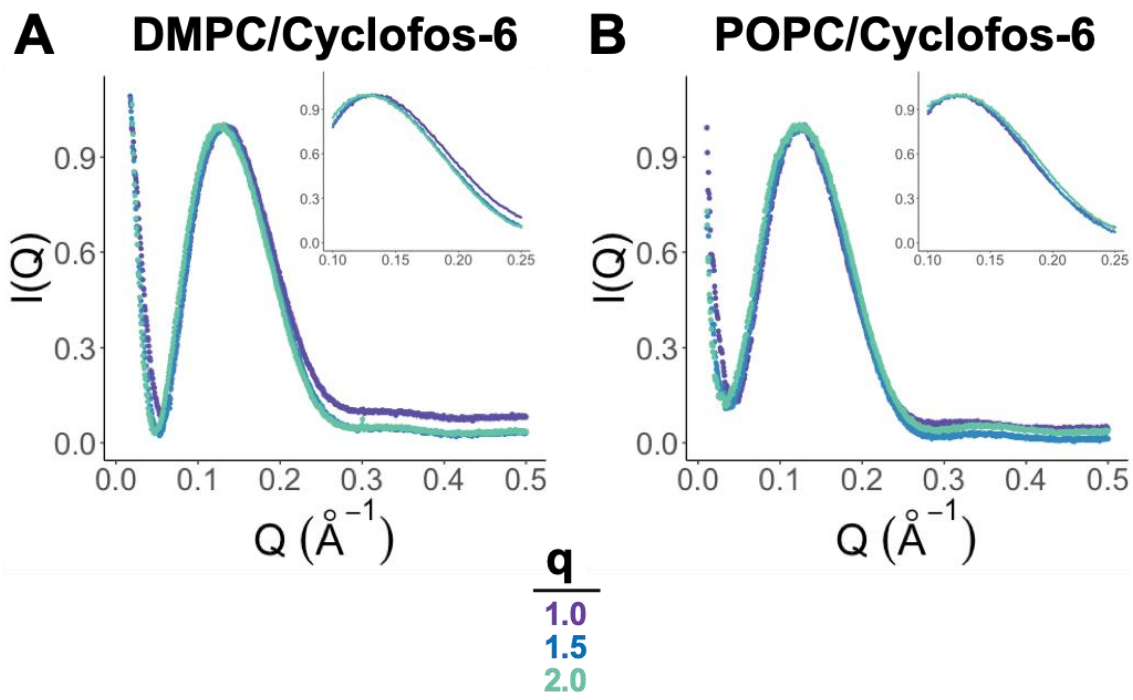


Figure 3.16: Scattering of DMPC/Cyclofos-6 and POPC/Cyclofos-6 bicelles with q-values > 1.0. SAXS profiles of scattering intensity, I(Q), as a function of the magnitude of the scattering vector Q (Å^{-1}) for DMPC/Cyclofos-6 (A) and POPC/Cyclofos-6 (B) above q-value = 1.0 at 25°C. Zoomed scattering in the second maximum region is inset for each plot. DLPC/Cyclofos-6 did not form isotropically-tumbling bicelles above q = 1.0.

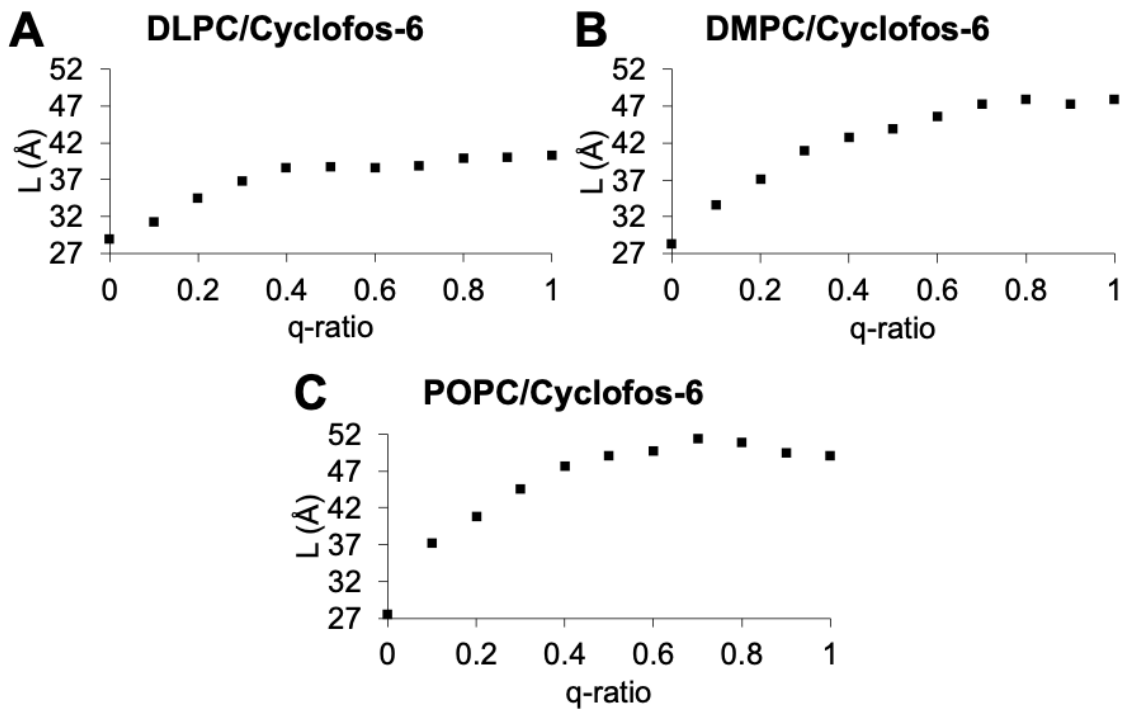
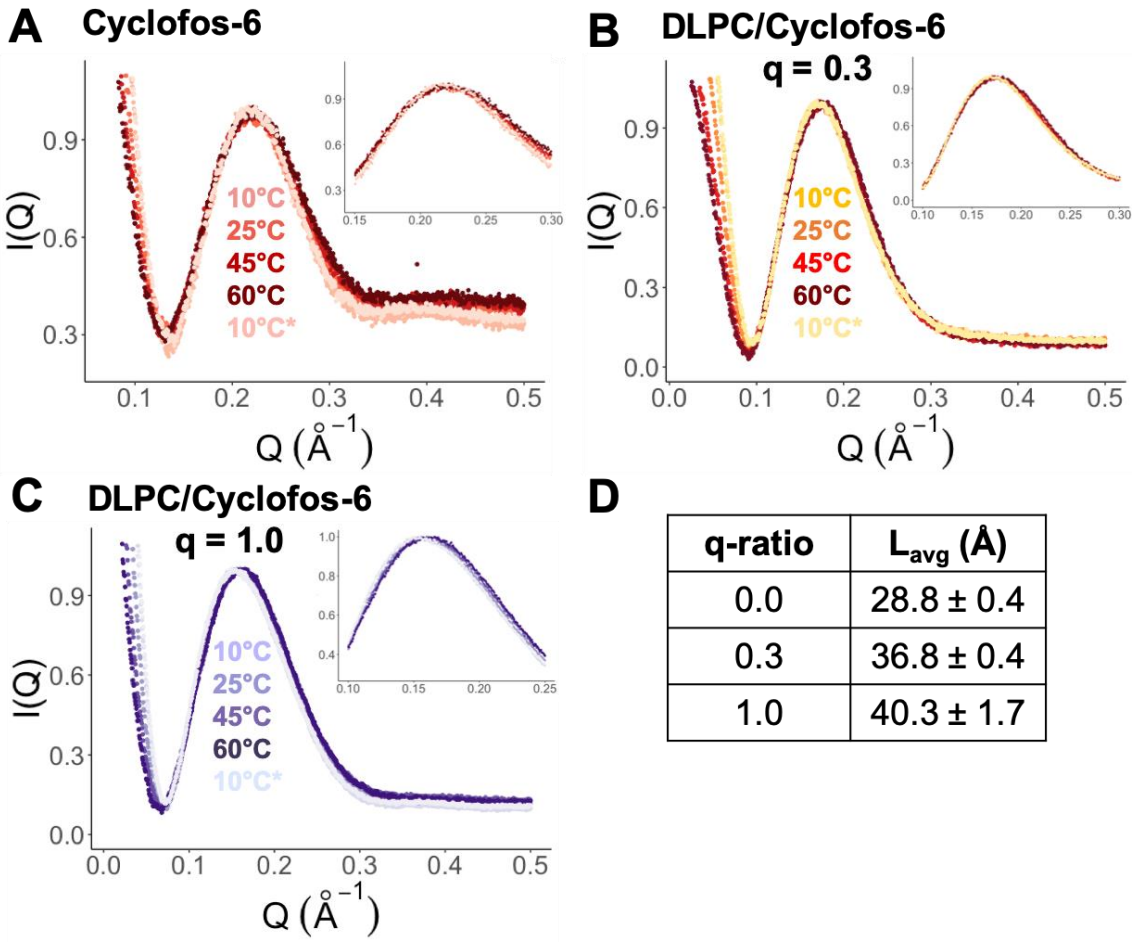
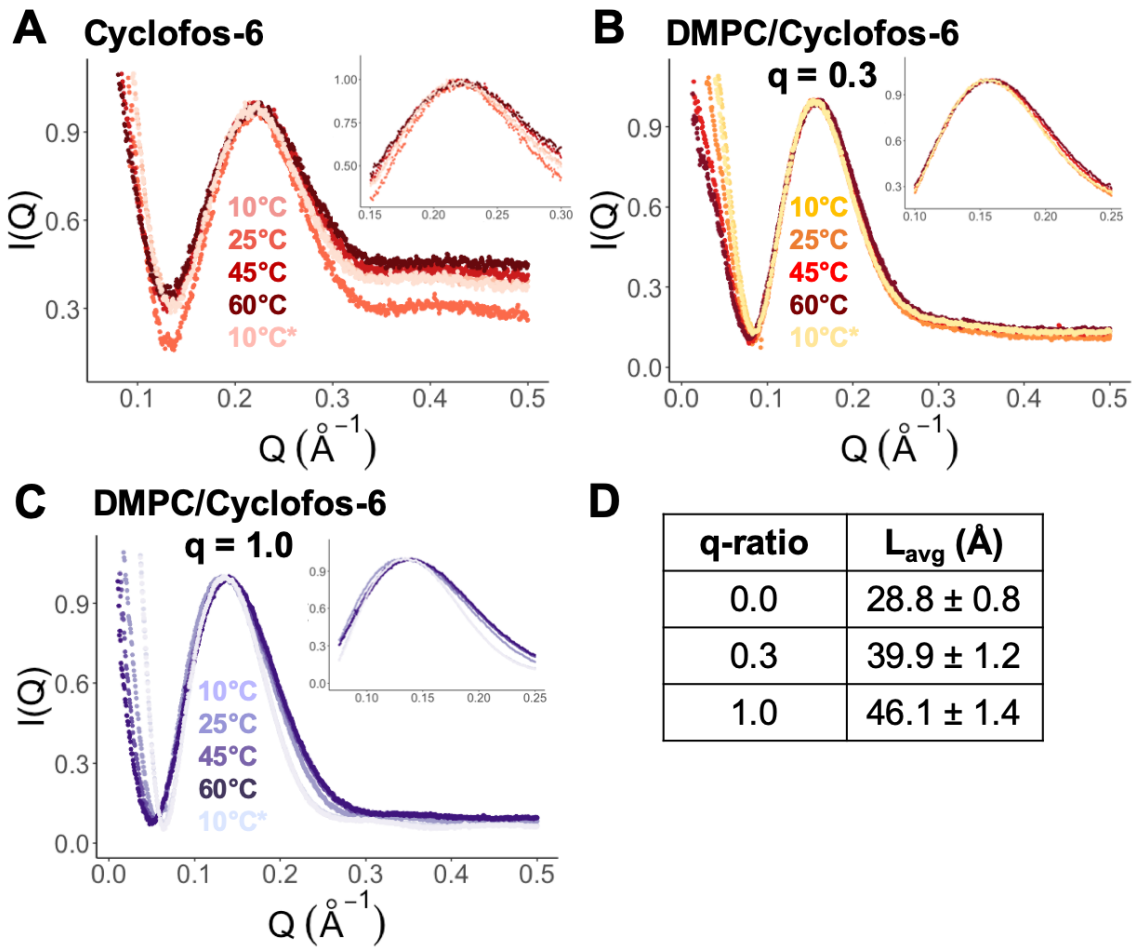


Figure 3.17: Headgroup to headgroup thickness (L) of Cyclofos-6 bicelles according to bicelle q-values. Experimental L parameters were determined for (A) DLPC/Cyclofos-6, (B), DMPC/Cyclofos-6, and (C), POPC/Cyclofos-6 bicelles for $0.0 \leq q \leq 1.0$.



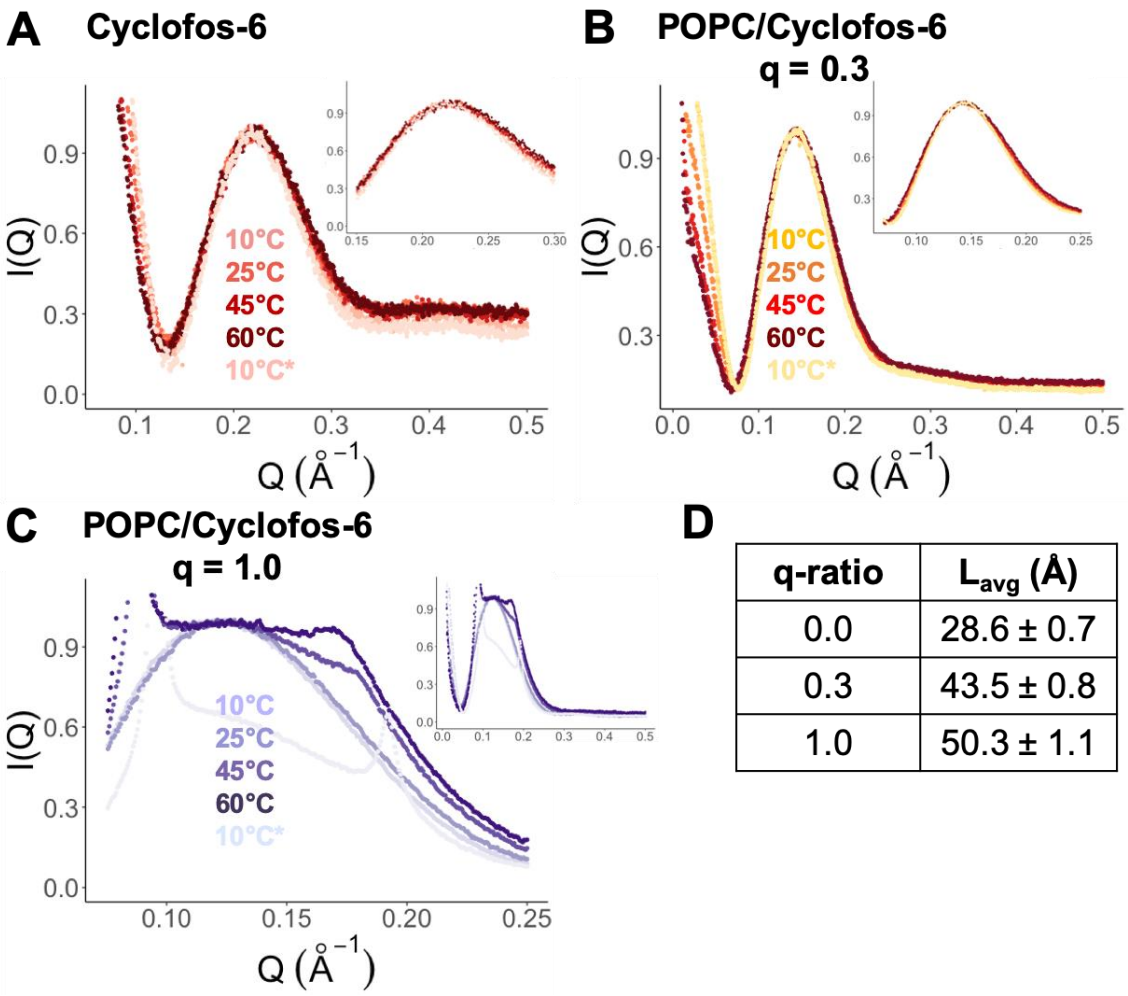
q-ratio	L_{avg} (Å)
0.0	28.8 ± 0.4
0.3	36.8 ± 0.4
1.0	40.3 ± 1.7

Figure 3.18: Temperature-dependent DLPC/Cyclofos-6 scattering and averaged parameters. Temperature-dependent SAXS profiles for DLPC/Cyclofos-6 when q = 0.0 (pure Cyclofos-6, A), q = 0.3 (B), and q = 1.0 (C). Scattering was recorded for each mixture during a temperature cycle from 10°C to 60°C, with a final temperature of 10°C (denoted by asterisks). Zoomed scattering in the second maximum region is inset for each plot. (D) Average headgroup to headgroup thickness, L, for bicelle q-values in A-C during the temperature cycle.



q-ratio	L_{avg} (Å)
0.0	28.8 ± 0.8
0.3	39.9 ± 1.2
1.0	46.1 ± 1.4

Figure 3.19: Temperature-dependent DMPC/Cyclofos-6 scattering and averaged parameters. Temperature-dependent SAXS profiles for DMPC/Cyclofos-6 when q = 0.0 (pure Cyclofos-6, A), q = 0.3 (B), and q = 1.0 (C). Scattering was recorded for each mixture during a temperature cycle from 10°C to 60°C, with a final temperature of 10°C (denoted by asterisks). Zoomed scattering in the second maximum region is inset for each plot. (D) Average headgroup to headgroup thickness, L, for bicelle q-values in A-C during the temperature cycle.



D

q-ratio	L_{avg} (Å)
0.0	28.6 ± 0.7
0.3	43.5 ± 0.8
1.0	50.3 ± 1.1

Figure 3.20: Temperature-dependent POPC/Cyclofos-6 scattering and averaged parameters. Temperature-dependent SAXS profiles for POPC/Cyclofos-6 when q = 0.0 (pure Cyclofos-6, A), q = 0.3 (B), and q = 1.0 (C). Scattering was recorded for each mixture during a temperature cycle from 10°C to 60°C, with a final temperature of 10°C (denoted by asterisks). Zoomed scattering in the second maximum region is inset for each plot. (D) Average headgroup to headgroup thickness, L, for bicelle q-values in A-C during the temperature cycle.

3.6.3 DH_6PC and DH_7PC bicelles

DH_6PC was combined with DLPC, DPPC, and POPC lipids to form bicelles with $0.0 \leq q \leq 2.0$. To demonstrate that the SAXS method in this study agrees with Caldwell *et al.* [9], DMPC/DH_6PC bicelles above q = 1.0 were formed. Use of DPPC lipid in this study was limited to DH_6PC because the DPPC melting temperature (41°C) is not compatible with most membrane proteins [18]. Similarly, DH_7PC was combined with DMPC and POPC to reflect bicelle compositions used for membrane proteins [39]. Compositions with DH_6PC and DH_7PC formed isotropic bicelles up to q = 2.0 except for POPC/DH_6PC (up to q = 1.0) and POPC/DH_7PC (up to q = 0.9). POPC/DH_6PC above q = 1.0 was not stable at room temperature, and POPC/DH_7PC above q = 0.9 failed to form soluble aggregates.

Trends in the scattering for DH_6PC bicelles with q = 0.1 to 1.0 at 25°C (Figure 3.21 and Figure 3.22) is similar to Cyclofos-6 bicelles. The scattering profiles stop changing in the forward scattering and second maximum regions when q ~ 0.6. Scattering for DH_6PC bicelles above q = 1.0 changes minimally (Figure 3.23), except for POPC/DH_6PC. Like POPC/Cyclofos-6 bicelles, two sharp peaks indicative of multilamellar vesicle formation appears in the second maximum region for POPC/DH_6PC above q = 1.0. A freeze/thaw cycle would clarify high-q POPC/DH_6PC mixtures for approximately 5 minutes before the solution showed evidence of insoluble lipid aggregates forming. Experimental *L* parameters for DH_6PC bicelles were calculated from the scattering Q_{max} according to Equation 3.3. The headgroup to headgroup distance across the core of DH_6PC micelles (22-23 Å) is consistent with previous literature [19, 20, 25]. The observed *L* parameters remain within 1 Å of the maximum *L* reached when q ~ 0.7 for DLPC/DH_6PC and q ~ 1.0

for DPPC/DH_6PC and POPC/DH_6PC (Figure 3.24). The experimental L_{max} values for DLPC/DH_6PC, DPPC/DH_6PC, and POPC/DH_6PC were approximately 37 Å, 50 Å, and 49 Å, respectively. L_{max} for DPPC/DH_6PC and POPC/ DH_6PC bicelles agree well with the average theoretical L_{max} values for DPPC and POPC (49 Å) bilayers. The L_{max} for DLPC/DH_6PC is approximately 1 Å below the range of theoretical L_{max} for DLPC bilayers (Table 3.1).

Small temperature-dependent differences in scattering profiles and *L* parameters were observed for most DH_6PC bicelle conditions (Figure 3.25, Figure 3.26, Figure 3.27). The second maximum intensity of pure DH_6PC increased above 25°C (Figure 3.25, Figure 3.26, Figure 3.27, panel A) without changing the Q_{max} value and resulting *L* parameter (Figure 3.25, Figure 3.26, Figure 3.27, panel D). Temperature-dependent increases in the second maximum were not observed when DH_6PC was mixed with lipid. Aside from changes in the pure DH_6PC scattering, DPPC/DH_6PC exhibited a steeper slope in the forward scattering that indicates transition to a larger structure at 45°C. Larger DPPC/DH_6PC aggregates with increased temperature likely reflects the DPPC transition from the gel to liquid-crystalline phase (T_m = 41°C).

Unlike the scattering trends for DH_6PC, DH_7PC bicelle scattering continues to change up to q = 0.9 for DMPC/DH_7PC (Figure 3.28) at 25°C. Changes in POPC/DH_7PC scattering are similarly observed until isotropic aggregates no longer form above q = 0.9. A minor peak begins to appear in the second maximum region when q = 0.9 for POPC/DH_7PC that is evidence of lipid aggregates beginning to form. Scattering for DMPC/DH_7PC bicelles above q = 1.0 shifts the second maximum at q = 1.5 and remains

constant for q = 2.0 (Figure 3.29). Experimental *L* parameters for DH_7PC bicelles were calculated from the scattering Q_{max} according to Equation 3.3. The headgroup to headgroup distance across the core of DH_7PC micelles is ~25 Å (Figure 3.30) and agrees with the addition of one carbon to the acyl chain relative to DH_6PC. The observed *L* parameters for DH_7PC bicelles do not exhibit an apparent plateau for $0.0 \leq q \leq 1.0$ (Figure 3.30). For DMPC/DH_7PC, L_{max} is reached when q = 1.5 (47 Å), and *L* appears to plateau for POPC/DH_7PC when q = 0.9 (47 Å). L_{max} for DMPC/DH_7PC is in line with the expected range for DMPC, and POPC/ DH_7PC L_{max} is approximately 1 Å below the POPC range. No apparent temperature-dependent changes in DH_7PC bicelle scattering or *L* parameters were present (Figure 3.31 and Figure 3.32).

As expected, neither DH_6PC nor DH_7PC induces PC lipid organization when $q < 0.5$ in any of the lipid mixtures. When combined with DLPC (12:0), DH_6PC did shift the apparent phase transition from a lipid/detergent mixture to q ~ 0.7. DMPC/DH_7PC mixtures similarly did not reach a constant headgroup to headgroup thickness until a higher q-value (~1.5). Taken together, results from DLPC/DH_6PC and DMPC/DH_7PC would indicate that a closer match between the lipid/detergent chain length causes increased detergent partitioning with lipid. However, DPPC (16:0)/DH_6PC, POPC (16:0-18:1)/DH_6PC, and POPC/DH_7PC are further-separated in the relative length of lipid and detergent, and all three mixtures did not reach L_{max} below q = 0.5. While results from DPPC/DH_6PC, POPC/DH_6PC, and POPC/DH_7PC does not preclude increased mixing for DLPC/DH_6PC and DMPC/DH_7PC, it is clear that bicelle morphology cannot simply be predicted from the relative number of carbons in each component.

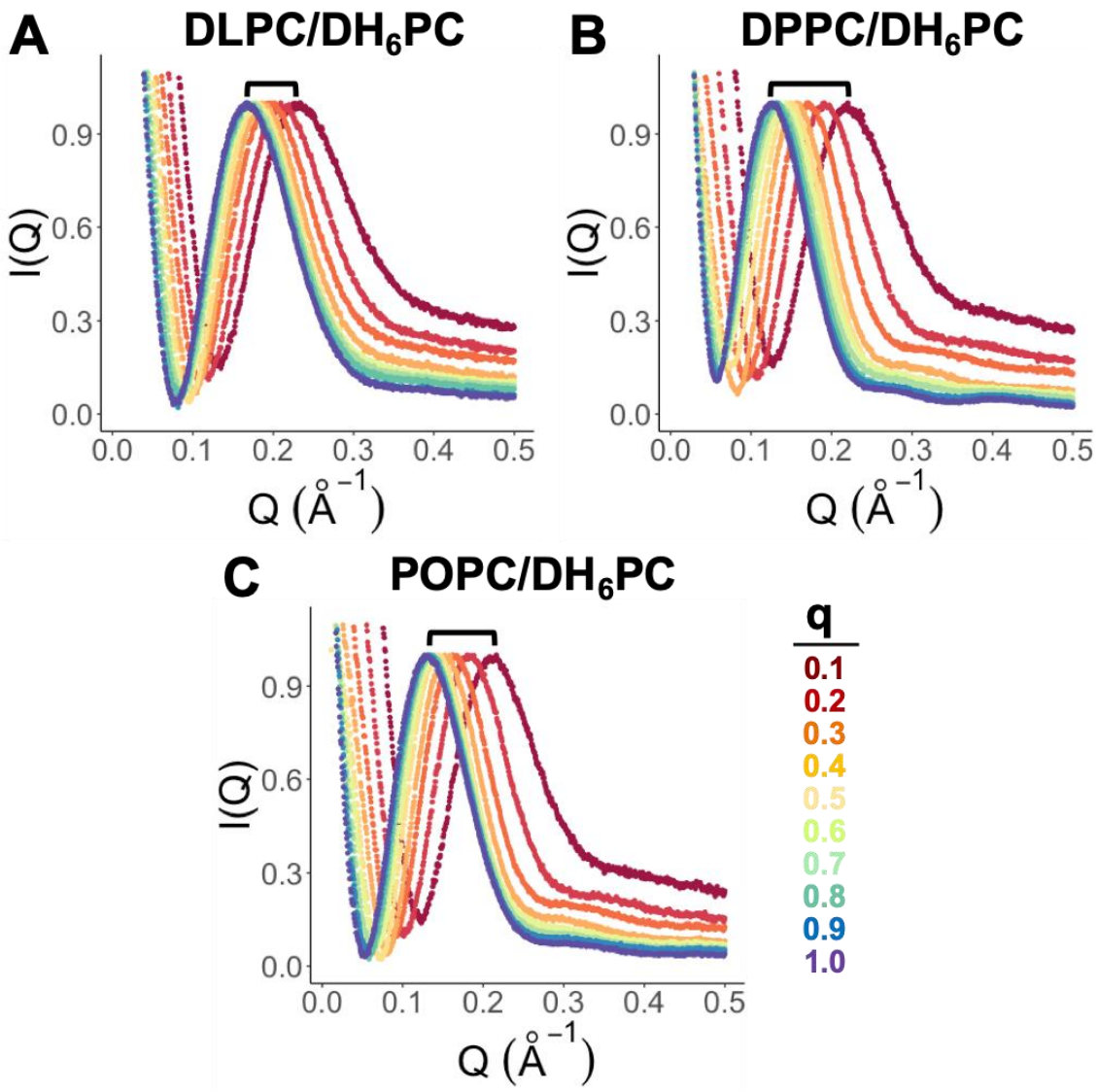


Figure 3.21: Scattering of DH_6PC bicelles with q-values ranging 0.1 to 1.0. SAXS profiles of scattering intensity, I(Q), as a function of the magnitude of the scattering vector Q ($Å^{-1}$) for DLPC/DH_6PC (A), DPPC/DH_6PC (B), and POPC/DH_6PC (C) lipid-detergent mixtures at 25°C. Scattering profiles were plotted as rainbow gradients for bicelle q-values between 0.1 (red) and 1.0 (violet). Braces indicate the shift in scattering Q ($Å^{-1}$) with increasing bicelle q-values.

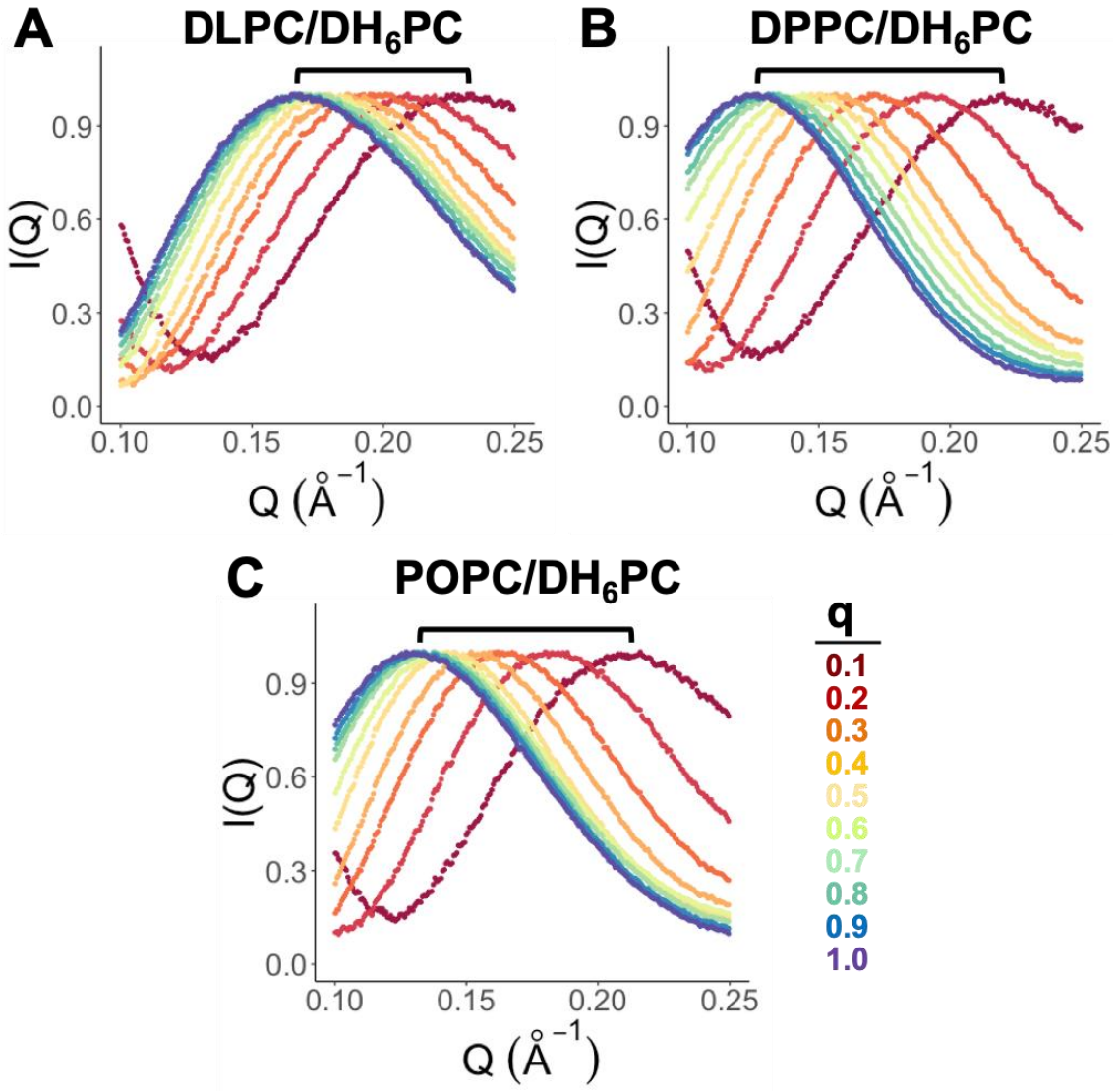


Figure 3.22: Scattering at second maximum of DH_6PC bicelles with q-values ranging 0.1 to 1.0. Zoomed scattering in the second maximum region for DLPC/DH_6PC (A), DPPC/DH_6PC (B), and POPC/DH_6PC (C) lipid-detergent mixtures at 25°C. Scattering profiles were plotted as rainbow gradients for bicelle q-values between 0.1 (red) and 1.0 (violet). Braces indicate the shift in scattering Q ($Å^{-1}$) with increasing bicelle q-values.

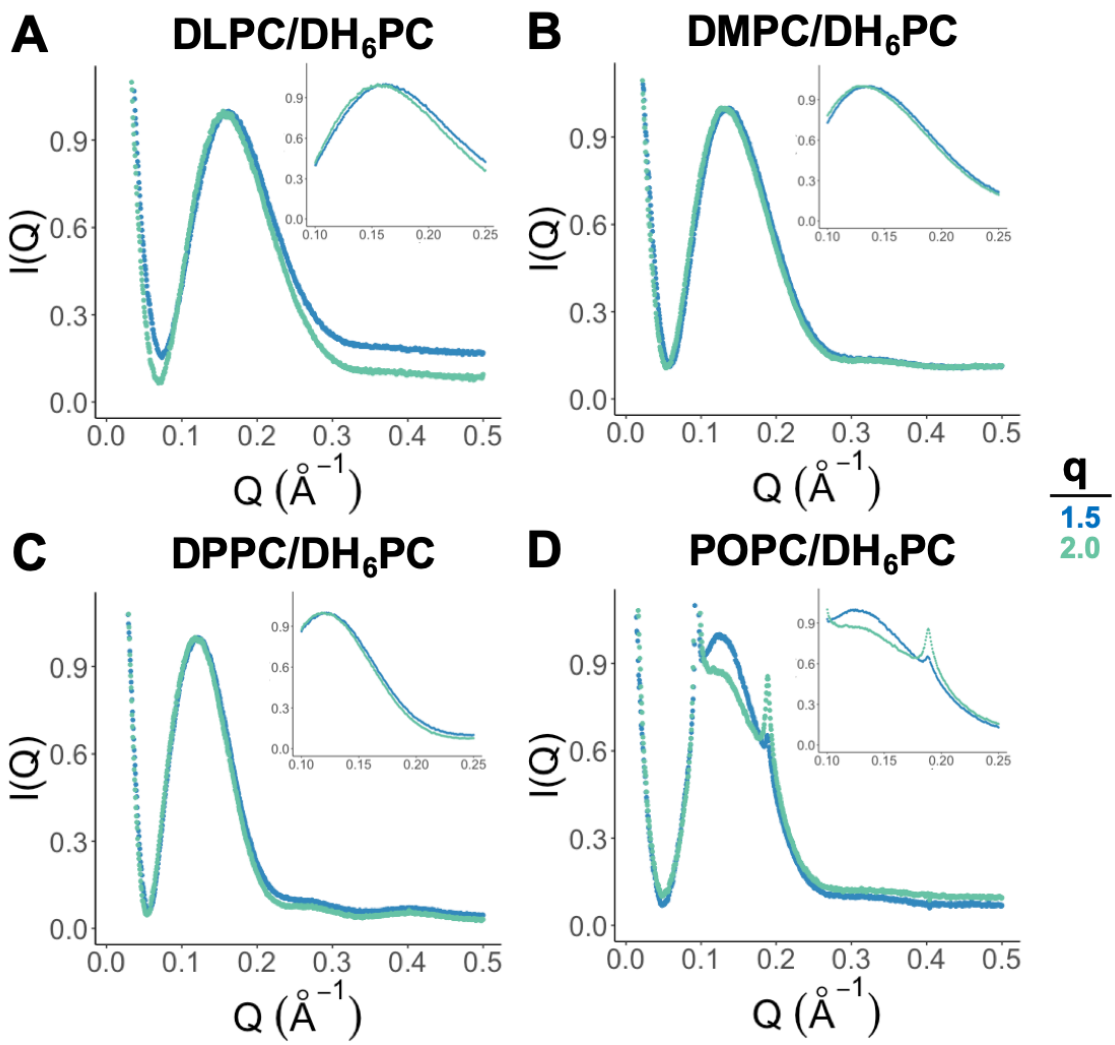


Figure 3.23: Scattering of DH_6PC bicelles with q-values > 1.0. SAXS profiles of scattering intensity, I(Q), as a function of the magnitude of the scattering vector Q ($Å^{-1}$) for DLPC/DH_6PC (A), DMPC/DH_6PC (B), DPPC/DH_6PC (C), and POPC/ DH_6PC (D) bicelles above q-value = 1.0 at 25°C. Zoomed scattering in the second maximum region is inset for each plot. The two sharp peaks in the second maximum region for POPC/DH_6PC q = 1.5 and 2.0 are evidence of multilamellar vesicle formation.

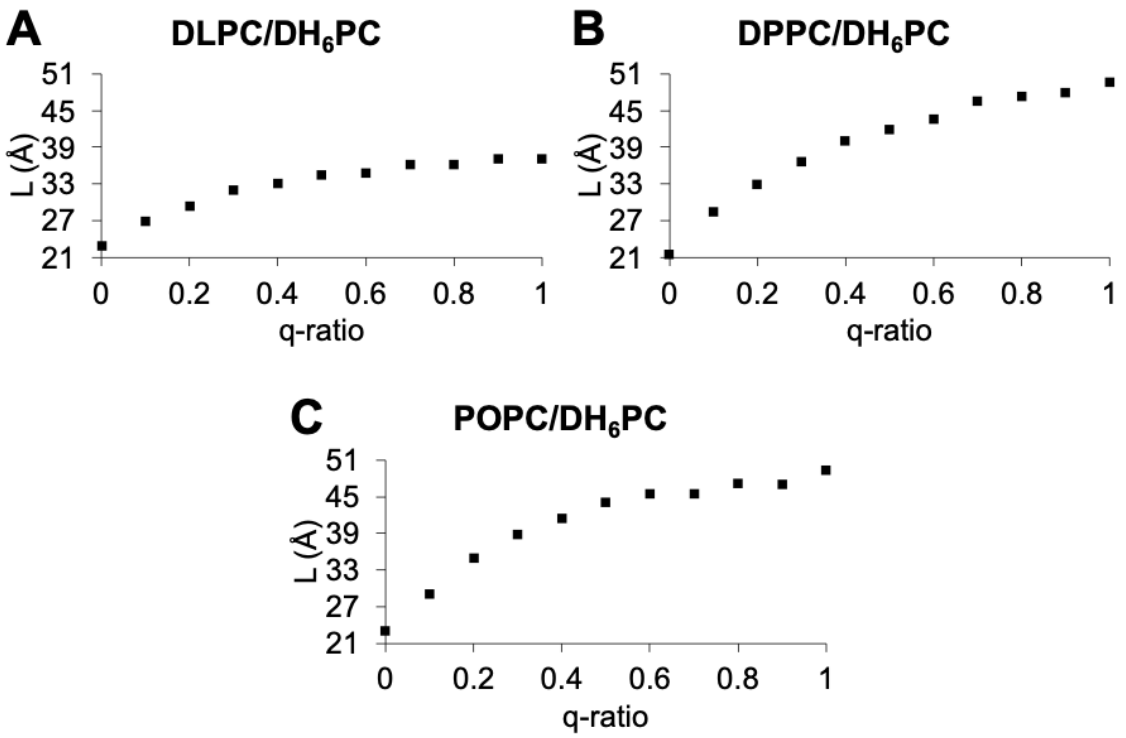
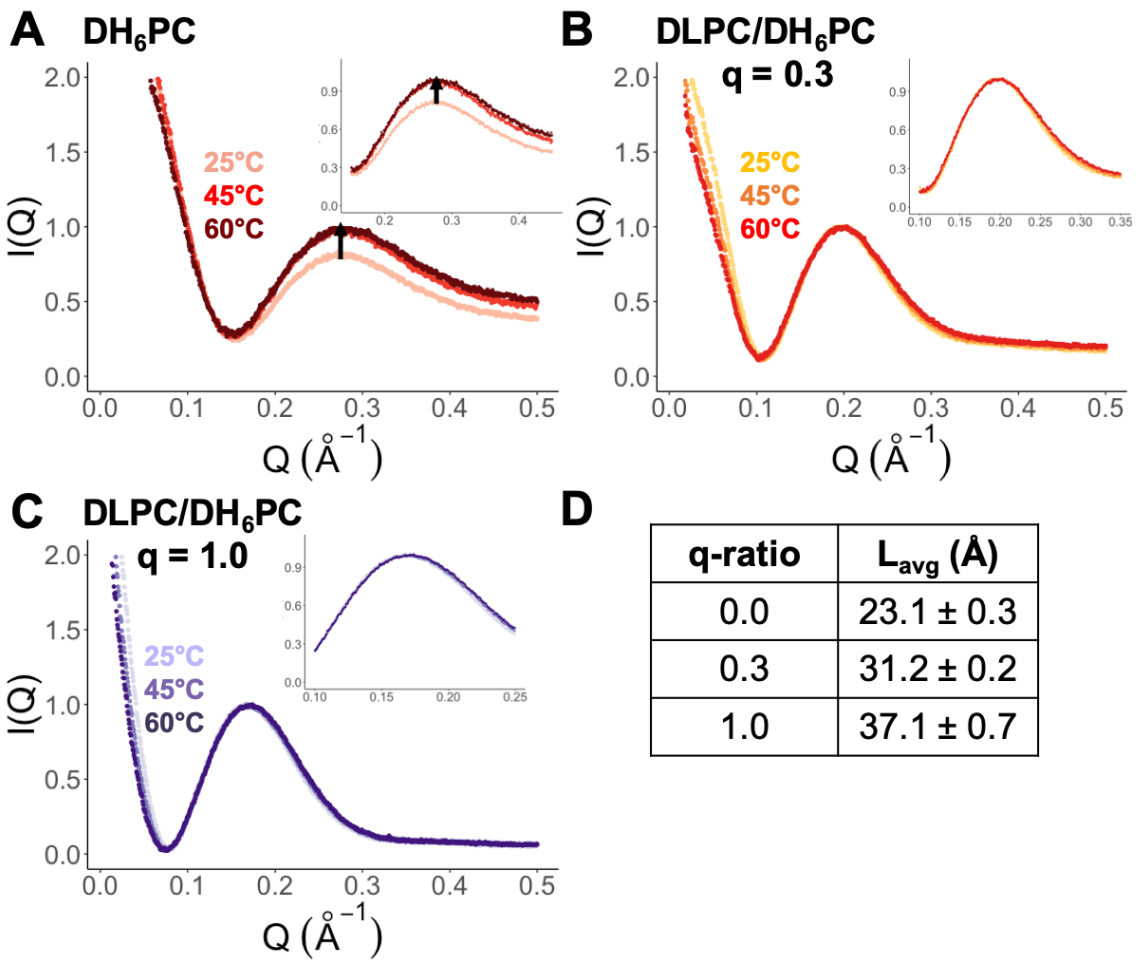
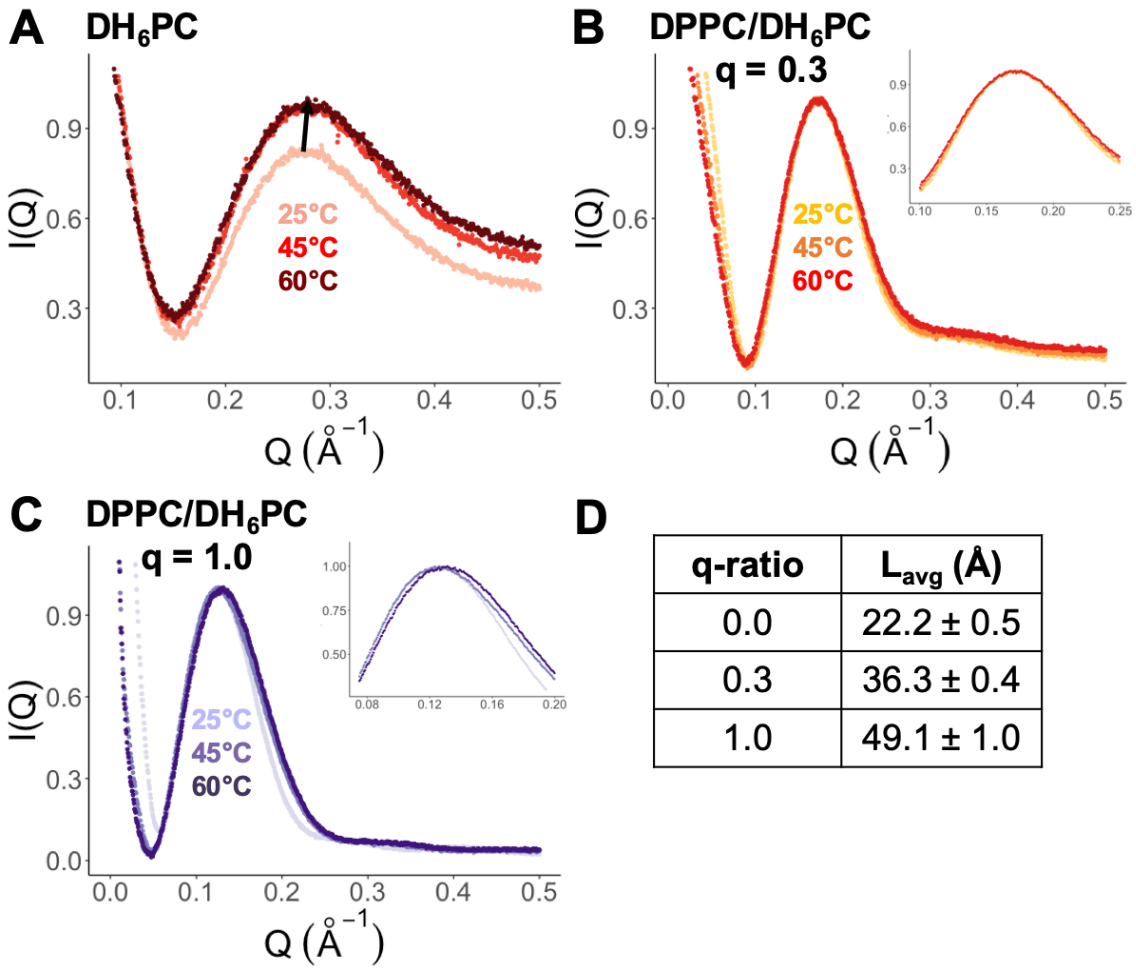


Figure 3.24: Headgroup to headgroup thickness (L) of DH_6PC bicelles according to bicelle q-values. Experimental L parameters were determined for (A) DLPC/DH_6PC, (B), DPPC/DH_6PC, and (C), POPC/DH_6PC bicelles for $0.0 \leq q \leq 1.0$.



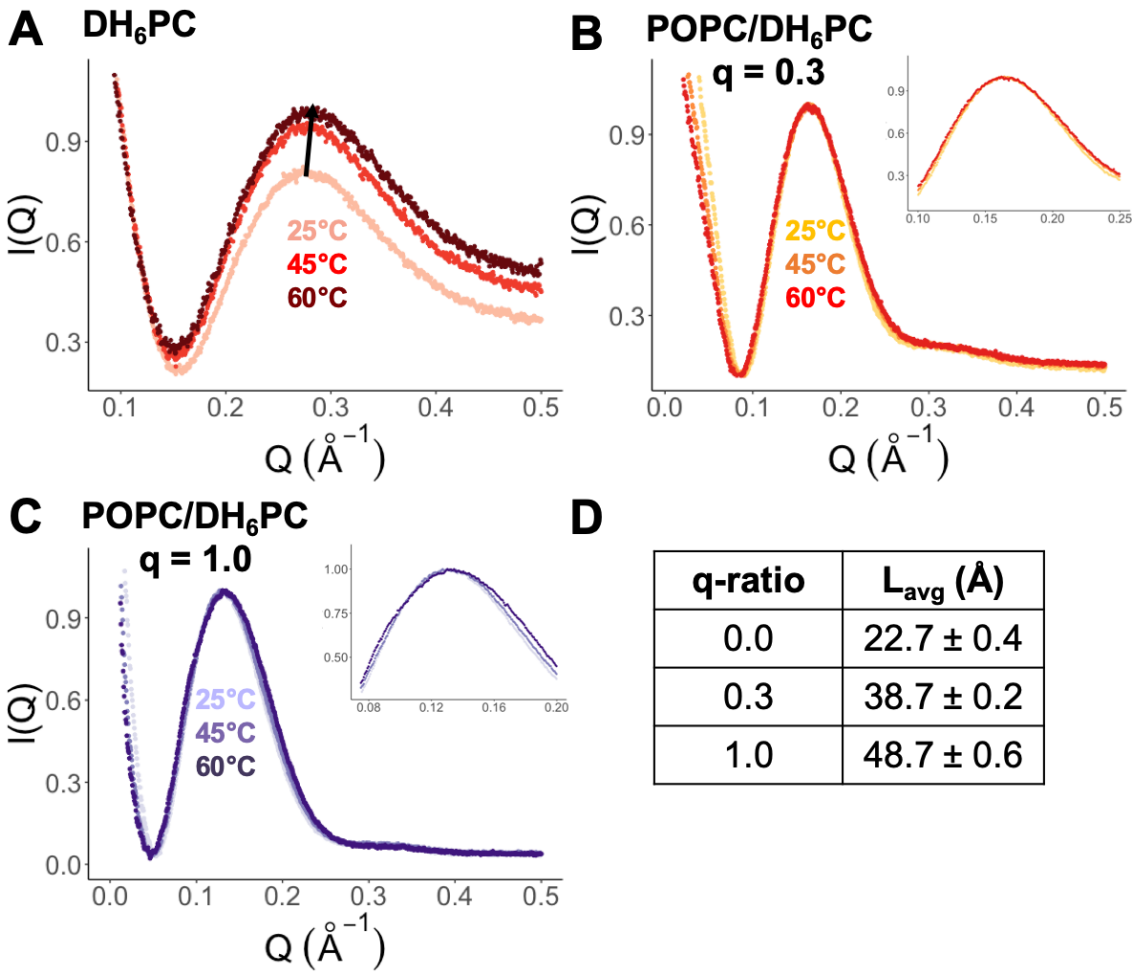
q-ratio	L_{avg} (Å)
0.0	23.1 ± 0.3
0.3	31.2 ± 0.2
1.0	37.1 ± 0.7

Figure 3.25: Temperature-dependent DLPC/DH$_6$PC scattering and averaged parameters. Temperature-dependent SAXS profiles for DLPC/DH$_6$PC when q = 0.0 (pure DH$_6$PC, A), q = 0.3 (B), and q = 1.0 (C). Scattering was recorded for each mixture from 25°C to 60°C. Zoomed scattering in the second maximum region is inset for each plot. Arrows indicate the shift in the second maximum position and intensity. (D) Average headgroup to headgroup thickness, L, for bicelle q-values in A-C during the temperature cycle.



q-ratio	L_{avg} (Å)
0.0	22.2 ± 0.5
0.3	36.3 ± 0.4
1.0	49.1 ± 1.0

Figure 3.26: Temperature-dependent DPPC/DH$_6$PC scattering and averaged parameters. Temperature-dependent SAXS profiles for DPPC/DH$_6$PC when q = 0.0 (pure DH$_6$PC, A), q = 0.3 (B), and q = 1.0 (C). Scattering was recorded for each mixture from 25°C to 60°C. Zoomed scattering in the second maximum region is inset for each plot. Arrows indicate the shift in the second maximum position and intensity. (D) Average headgroup to headgroup thickness, L, for bicelle q-values in A-C during the temperature cycle.



q-ratio	L_{avg} (Å)
0.0	22.7 ± 0.4
0.3	38.7 ± 0.2
1.0	48.7 ± 0.6

Figure 3.27: Temperature-dependent POPC/DH$_6$PC scattering and averaged parameters. Temperature-dependent SAXS profiles for POPC/DH$_6$PC when q = 0.0 (pure DH$_6$PC, A), q = 0.3 (B), and q = 1.0 (C). Scattering was recorded for each mixture from 25°C to 60°C. Zoomed scattering in the second maximum region is inset for each plot. Arrows indicate the shift in the second maximum position and intensity. (D) Average headgroup to headgroup thickness, L, for bicelle q-values in A-C during the temperature cycle.

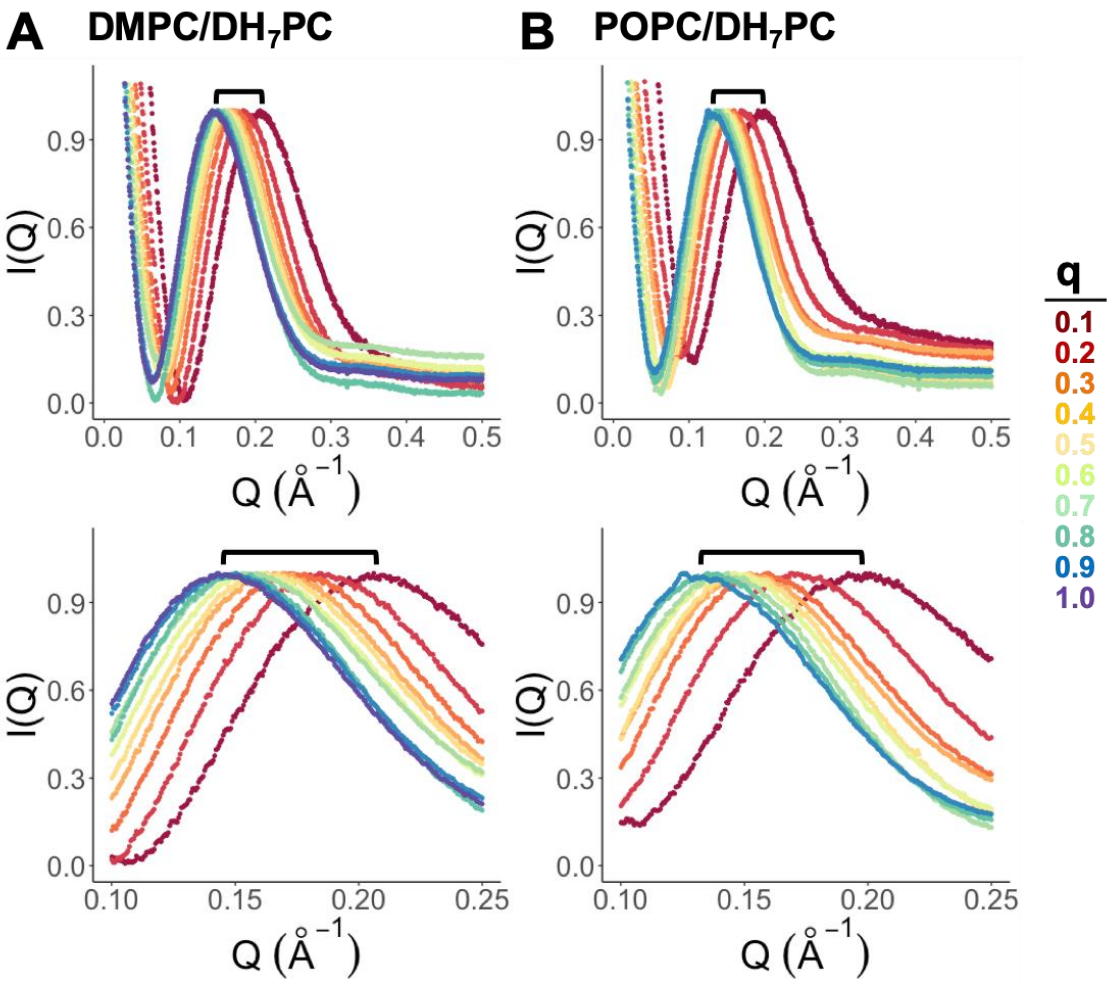


Figure 3.28: Scattering profiles of DH_7PC bicelles. SAXS profiles of scattering intensity, I(Q), as a function of the magnitude of the scattering vector Q ($Å^{-1}$) for DMPC/DH_7PC (A) and POPC/DH_7PC (B) lipid-detergent mixtures at 25°C. Scattering profiles were plotted as rainbow gradients for bicelle q-values between 0.1 (red) and 1.0 (violet) for DMPC/DH_7PC, and between 0.1 and 0.9 (blue) for POPC/DH_7PC. POPC/DH_7PC above q = 0.9 did not form isotropically-tumbling bicelles. Zoomed scattering in the second maximum region is below each plot. Braces indicate the shift in scattering Q ($Å^{-1}$) with increasing bicelle q-values.

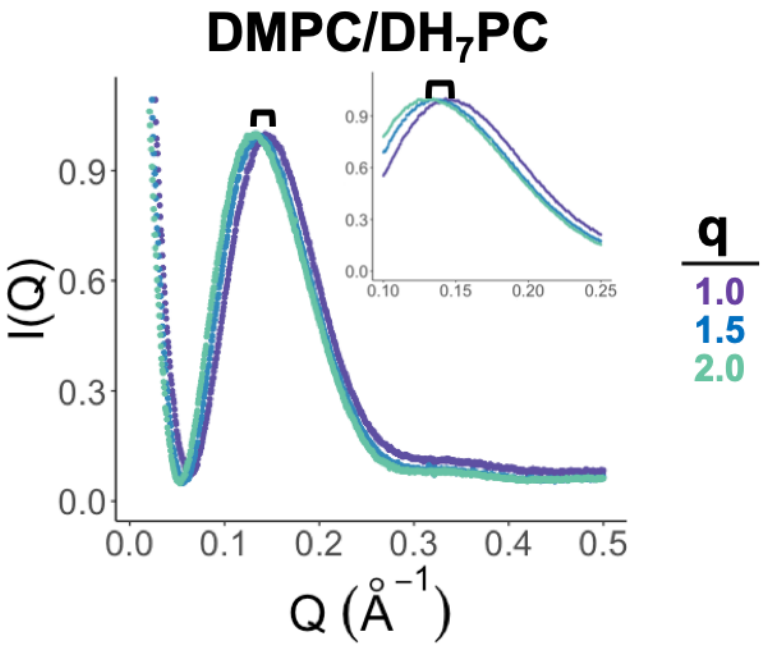


Figure 3.29: Scattering of DMPC/DH$_7$PC bicelles with q-values > 1.0. SAXS profiles of scattering intensity, I(Q), as a function of the magnitude of the scattering vector Q (Å^{-1}) for DMPC/DH$_7$PC bicelles above q-value = 1.0 at 25°C. Zoomed scattering in the second maximum region is inset.

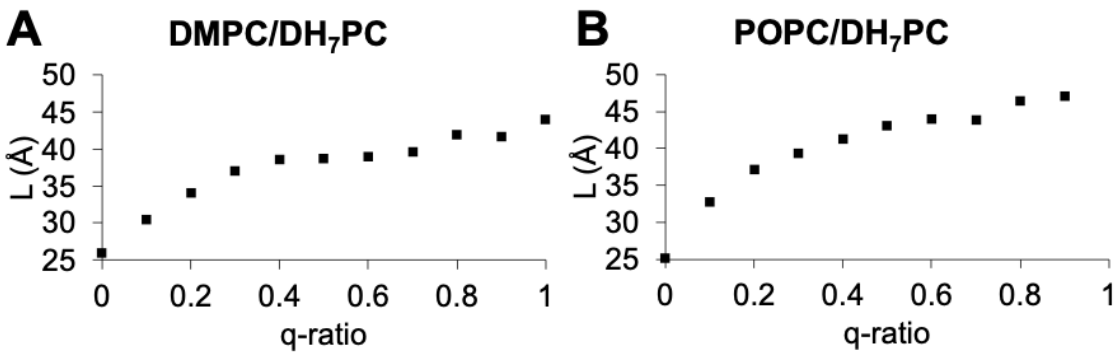
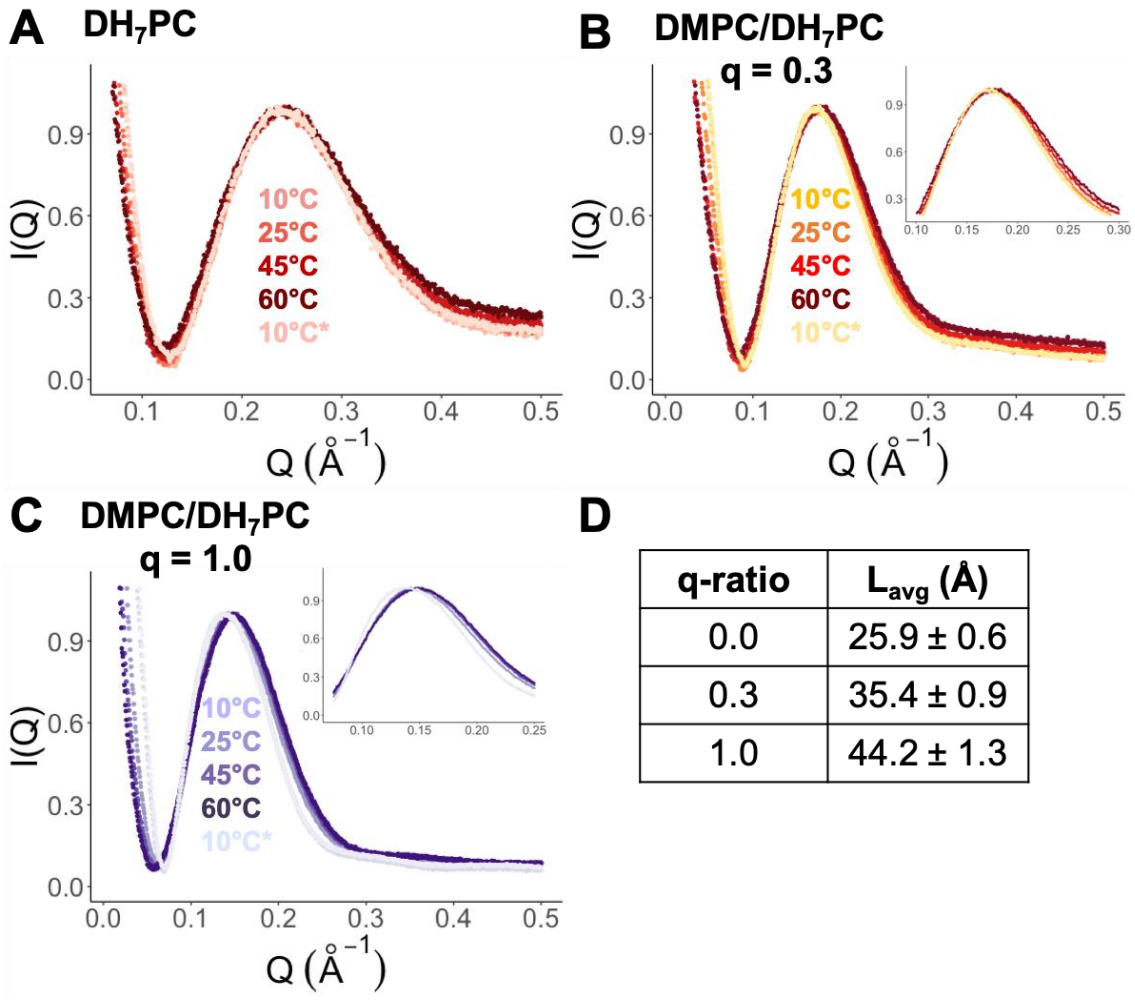
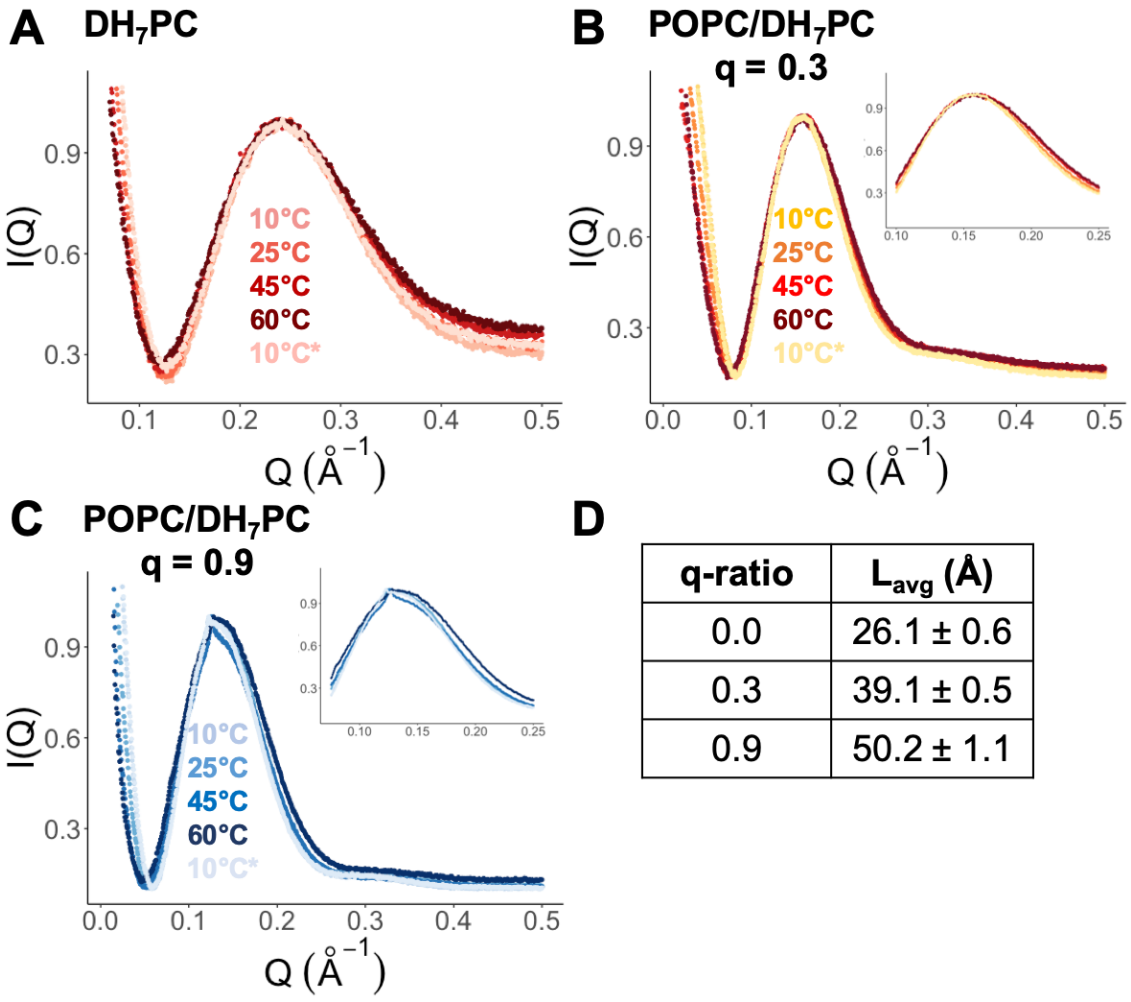


Figure 3.30: Headgroup to headgroup thickness (L) of DH_7PC bicelles according to bicelle q-values. Experimental L parameters were determined for (A) DMPC/DH_7PC bicelles for $0.0 \leq q \leq 1.0$ and (B) POPC/DH_7PC bicelles for $0.0 \leq q \leq 0.9$. POPC/DH_7PC did not form isotropic bicelles above $q = 0.9$.



q-ratio	L_{avg} (Å)
0.0	25.9 ± 0.6
0.3	35.4 ± 0.9
1.0	44.2 ± 1.3

Figure 3.31: Temperature-dependent DMPC/DH$_7$PC scattering and averaged parameters. Temperature-dependent SAXS profiles for DMPC/DH$_7$PC when q = 0.0 (pure DH$_7$PC, A), q = 0.3 (B), and q = 1.0 (C). Scattering was recorded for each mixture during a temperature cycle from 10°C to 60°C, with a final temperature of 10°C (denoted by asterisks). Zoomed scattering in the second maximum region is inset for each plot. (D) Average headgroup to headgroup thickness, L, for bicelle q-values in A-C during the temperature cycle.



q-ratio	L_{avg} (Å)
0.0	26.1 ± 0.6
0.3	39.1 ± 0.5
0.9	50.2 ± 1.1

Figure 3.32: Temperature-dependent POPC/DH$_7$PC scattering and averaged parameters. Temperature-dependent SAXS profiles for POPC/DH$_7$PC when q = 0.0 (pure DH$_7$PC, A), q = 0.3 (B), and q = 0.9 (C). Scattering was recorded for each mixture during a temperature cycle from 10°C to 60°C, with a final temperature of 10°C (denoted by asterisks). Zoomed scattering in the second maximum region is inset for each plot. (D) Average headgroup to headgroup thickness, L, for bicelle q-values in A-C during the temperature cycle.

3.6.4 Fos-choline bicelles

Due to the prevalence of DMPC/FC12 bicelles in recent literature, we sought to characterize fos-choline bicelles in this work. FC10 bicelles were included to examine miscibility between a shorter-chain fos-choline detergent and DMPC relative to FC12. Both DMPC/FC10 and DMPC/FC12 bicelles were formed with $0.0 \leq q \leq 2.0$ and produced clear solutions within the tested q-range. Changes in the forward scattering were observed for both DMPC/FC10 and DMPC/FC12 bicelles with increasing q-values. For fos-choline bicelles at 25°C, the scattering profiles stop changing in the second maximum region when q ~ 0.6 (Figure 3.33). However, the scattering Q for the DMPC/FC12 second maximum appears constant at q ~ 0.4. Scattering for fos-choline bicelles above q = 1.0 changes minimally in the second maximum region (Figure 3.33). Experiments varying the amphiphile concentration of fos-choline bicelles are needed to clarify if the forward scattering changes are due to interparticle interference. *L* parameters for fos-choline micelles were calculated from the scattering Q_{max} according to Equation 3.3. The headgroup to headgroup distance across the core of FC10 micelles was ~ 28 Å, and ~ 34 Å for FC12 micelles (Figure 3.34). Both values are consistent with previous literature [19, 20, 25]. The observed *L* parameters remain within 1 Å of the maximum *L* reached when q ~ 0.6 for DMPC/FC10 (L_{max} ~ 48 Å) and q ~ 0.5 for DMPC/FC12 (L_{max} ~ 48 Å) (Figure 3.34). L_{max} for both fos-choline bicelles are approximately 4 Å thicker than the average theoretical L_{max} for DMPC but fall within the expected range (Table 3.1).

Minor temperature-dependent differences in *L* parameters were observed for fos-choline bicelle conditions (Figure 3.35 and Figure 3.36). Similar to increasing the q-value

at 25°C, the forward scattering slope (low scattering Q values) increased for all compositions with higher temperatures. One explanation is that forward scattering changes are due to interparticle interference as temperature increases; however, to confirm, scattering measurements with different fos-choline bicelle concentrations are needed. For DMPC/FC12, the temperature effects were reversible (Figure 3.36). Similar to DH_6PC and DH_7PC bicelles, lipid miscibility predictions for fos-choline detergents cannot be predicted from hydrocarbon chain length based on the similar transition from mixed lipid/detergent at q ~ 0.5-0.6 for FC10 and FC12.

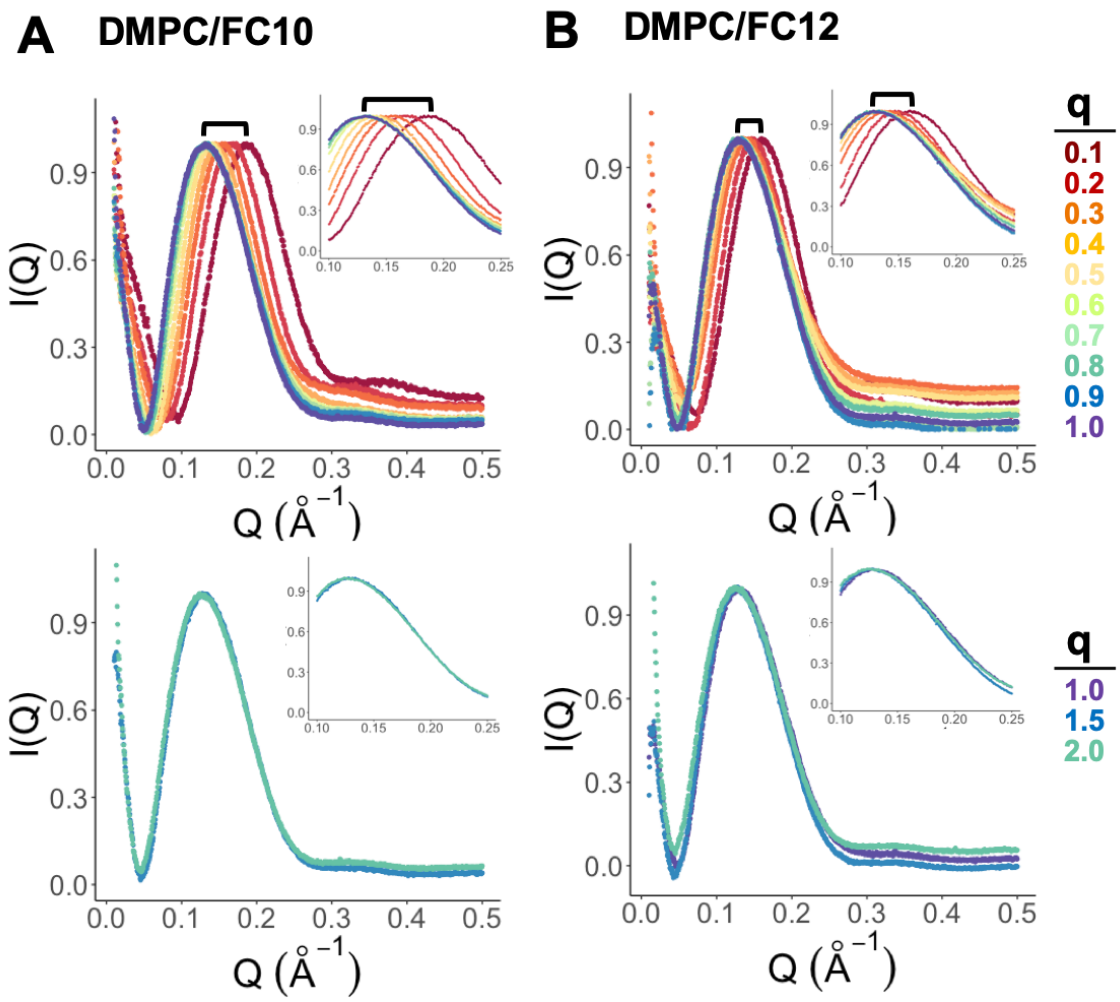


Figure 3.33: Scattering profiles of FC10 and FC12 bicelles. SAXS profiles of scattering intensity, I(Q), as a function of the magnitude of the scattering vector Q (Å^{-1}) for DMPC/FC10 (A) and DMPC/FC12 (B) lipid-detergent mixtures at 25°C. Scattering profiles were plotted as rainbow gradients for bicelle q-values between 0.1 (red) and 1.0 (violet) (top plots). Bottom plots represent bicelle scattering above q-value = 1.0 at 25°C. DMPC/FC10 q = 1.0 bicelle scattering could not be combined with data from higher q-values because DMPC/FC10 q ≤ 1.0 measurements were recorded under different

conditions. Zoomed scattering in the second maximum region is below each plot. Braces indicate the shift in scattering Q ($Å^{-1}$) with increasing bicelle q-values.

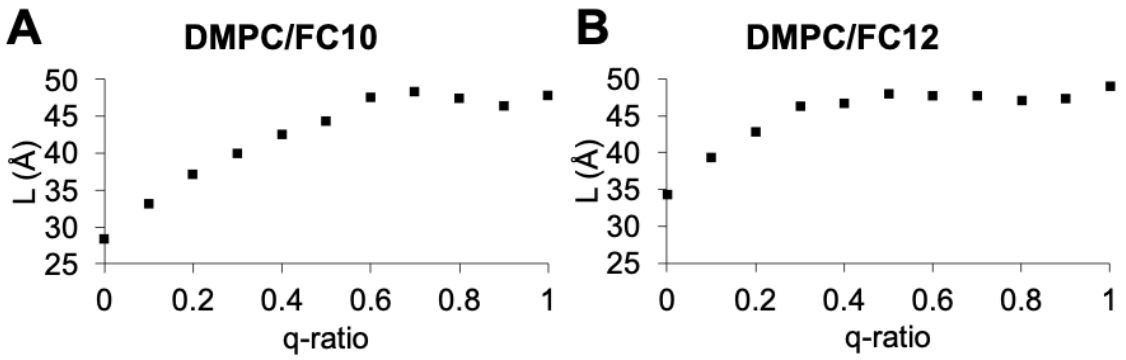
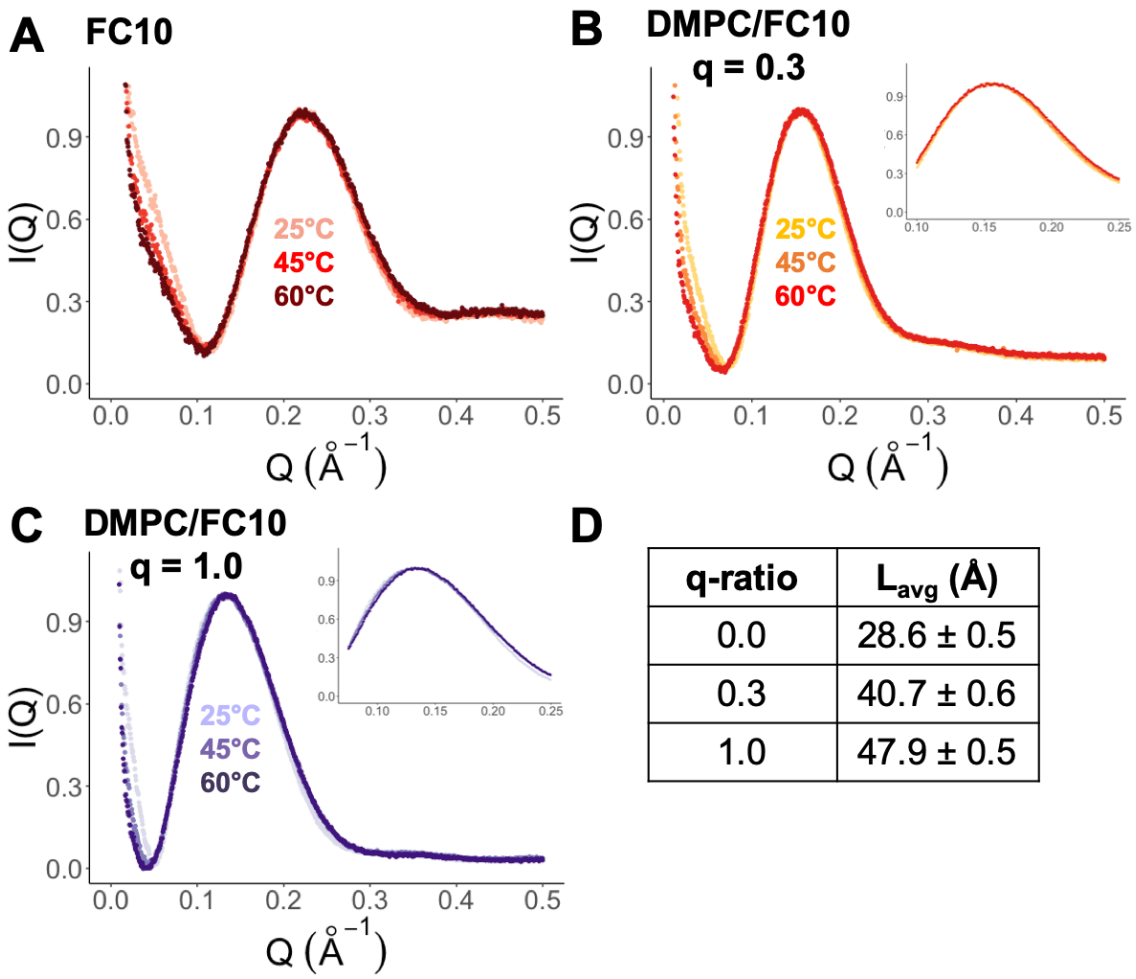
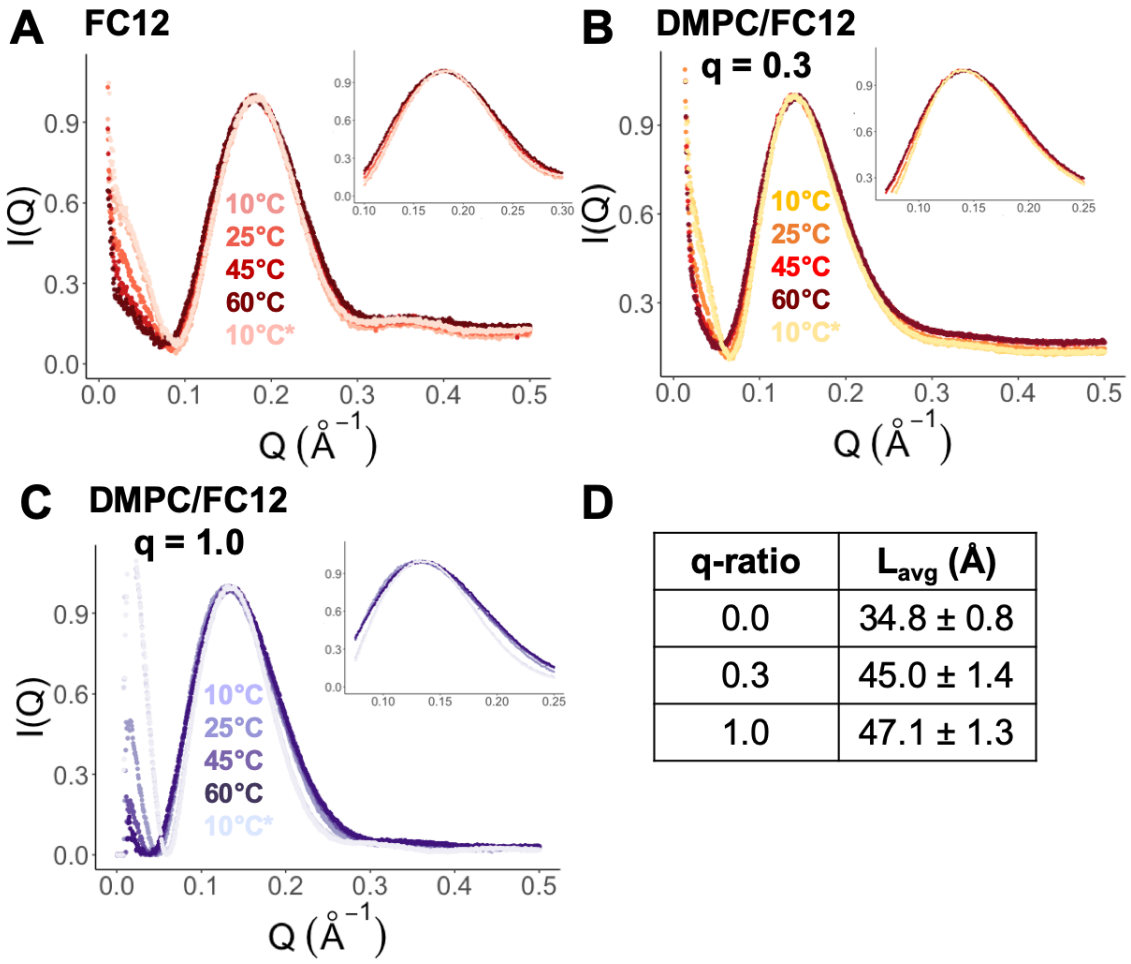


Figure 3.34: Headgroup to headgroup thickness (L) of FC10 and FC12 bicelles according to bicelle q-values. Experimental L parameters were determined for (A) DMPC/FC10 and (B) DMPC/FC12 bicelles for $0.0 \leq q \leq 1.0$.



q-ratio	L_{avg} (Å)
0.0	28.6 ± 0.5
0.3	40.7 ± 0.6
1.0	47.9 ± 0.5

Figure 3.35: Temperature-dependent DMPC/FC10 scattering and averaged parameters. Temperature-dependent SAXS profiles for DMPC/FC10 when q = 0.0 (pure FC10, A), q = 0.3 (B), and q = 1.0 (C). Scattering was recorded for each mixture from 25°C to 60°C. Zoomed scattering in the second maximum region is inset for each plot. (D) Average headgroup to headgroup thickness, L, for bicelle q-values in A-C during the temperature cycle.



q-ratio	L_{avg} (Å)
0.0	34.8 ± 0.8
0.3	45.0 ± 1.4
1.0	47.1 ± 1.3

Figure 3.36: Temperature-dependent DMPC/FC12 scattering and averaged parameters. Temperature-dependent SAXS profiles for DMPC/FC12 when q = 0.0 (pure FC12, A), q = 0.3 (B), and q = 1.0 (C). Scattering was recorded for each mixture during a temperature cycle from 10°C to 60°C, with a final temperature of 10°C (denoted by asterisks). Zoomed scattering in the second maximum region is inset for each plot. (D) Average headgroup to headgroup thickness, L, for bicelle q-values in A-C during the temperature cycle.

3.6.5 SDS, LDAO, and β-OG bicelles

Bicelle preparation with $0.0 \leq q \leq 2.0$ was attempted for DMPC/SDS, DMPC/LDAO, and DMPC/β-OG mixtures. DMPC/SDS produced transparent solutions for $q \leq 0.3$ when the sample temperature was at least 25°C. DMPC/LDAO mixtures for $q \leq 0.4$ formed clear solutions with freeze/thaw cycles but phase-separated after several minutes at room temperature. DMPC/β-OG was able to form a soluble mixture with the highest q-value (0.5) across this specific detergent set.

Figure 3.37 shows the scattering profiles for DMPC/SDS, DMPC/LDAO, and DMPC/β-OG mixtures and the pure detergents at 25°C. The second maximum shifts between SDS and DMPC/SDS q = 0.1, and the forward scattering starts to reflect core-shell ellipsoid structures when q = 0.2. Assessing changes in the *L* parameter for DMPC/SDS mixtures demonstrates that *L* quickly reaches ~44 Å (L_{max} for DMPC) when q = 0.2 and remains constant for q = 0.3 (Figure 3.38). However, DMPC/SDS required specific temperature conditions and failed to form solutions above q = 0.3, indicating that SDS likely does not form a stable rim around segregated DMPC. Scattering for LDAO and DMPC mixtures lacked a prominent second maximum but was not absent as for CHAPS. Although the chemical structure of LDAO appears similar to the fos-choline detergents, the missing phosphoryl group decreases the electron density in the headgroup region. LDAO micelles therefore do not possess the required contrast difference between the shell and hydrophobic core necessary to produce a characteristic second maximum in the scattering. The calculated *L* parameters from the minor second maximum are much higher than what is reasonably expected for LDAO structures at low-q ratios but decrease with

more DMPC (Figure 3.38). Finally, DMPC/β-OG displayed scattering profiles (Figure 3.37) and *L* parameters (Figure 3.38) similar to other mixtures of lipid and detergent reported in this study up to q = 0.5. As DMPC/β-OG failed to form solutions above q = 0.5, results combined from SAXS suggests that lipid ordering occurs above q = 0.5, and β-OG is unable to form a stable rim around segregated DMPC. Key conclusions from DMPC combinations with SDS, LDAO, and β-OG are that (1) bicelle and/or soluble aggregate formation is not universal to all lipids and detergents, and (2) detergent-rich conditions (q ≤ 0.3) produced soluble mixtures across all of the samples in this study.

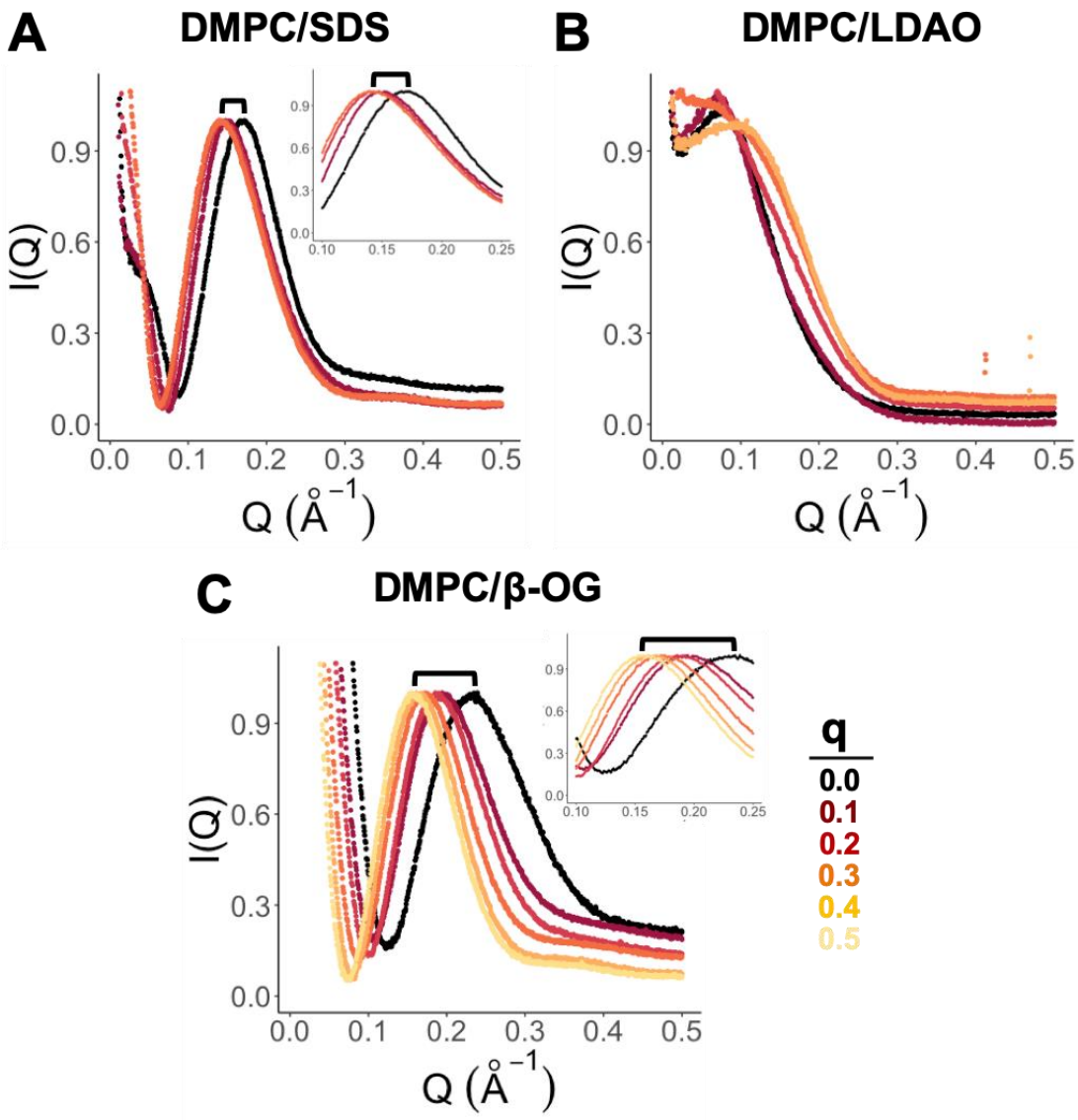


Figure 3.37: Scattering profiles of DMPC and SDS, LDAO, β-OG mixtures. SAXS profiles of scattering intensity, I(Q), as a function of the magnitude of the scattering vector Q (Å^{-1}) for DMPC/SDS (A), DMPC/LDAO (B), and DMPC/β-OG (C) lipid-detergent mixtures at 25°C. Scattering profiles were plotted for bicelle q-values according to the color schemes in previous rainbow gradient plots, with the addition of q = 0.0 (pure

detergent) scattering (black). Zoomed scattering in the second maximum region is inset. Braces indicate the shift in scattering Q ($Å^{-1}$) with increasing bicelle q-values.

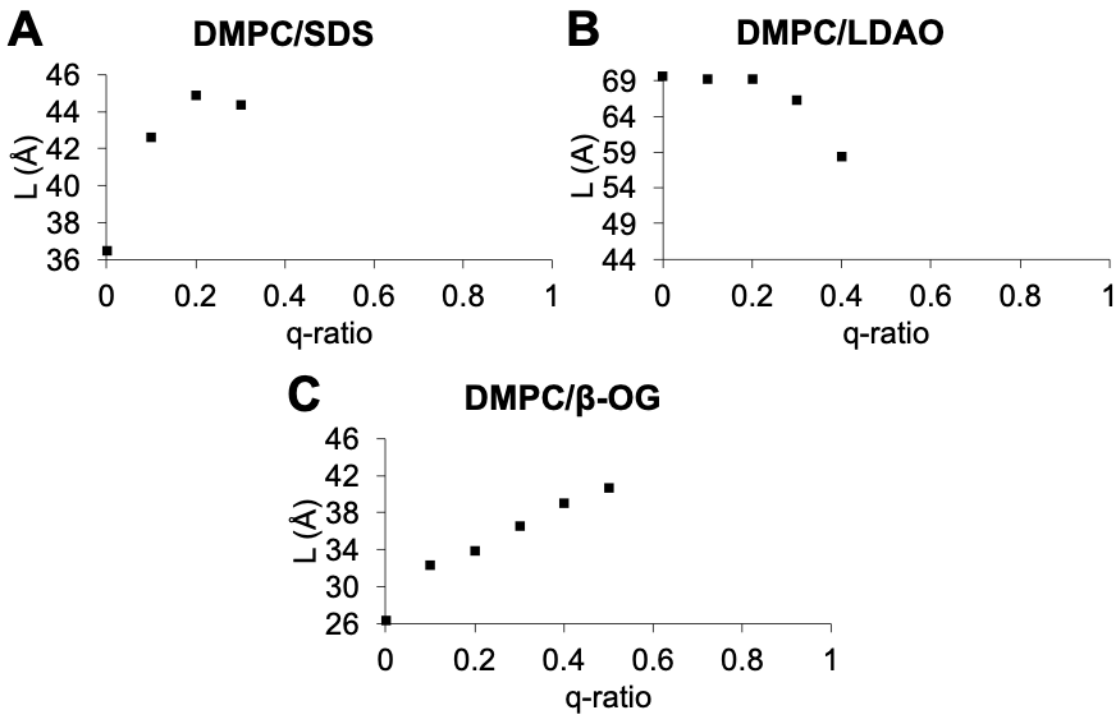


Figure 3.38: Headgroup to headgroup thickness (L) of DMPC mixed with SDS, LDAO, and β-OG according to bicelle q-values. Experimental L parameters were determined for (A) DMPC/SDS mixtures for $0.0 \leq q \leq 0.3$, (B) DMPC/LDAO mixtures for $0.0 \leq q \leq 0.4$, and (C) DMPC/β-OG mixtures for $0.0 \leq q \leq 0.4$. Mixtures precipitated above the q-values used to determine L parameters from SAXS measurements.

3.6.6 Detergent influences maximum length of the bicelle short dimension

To determine if the headgroup to headgroup thickness *L* was consistent between bicelles formed with the same PC lipid, the experimental L_{max} was calculated for DLPC, DMPC, DPPC, and POPC bicelles (Table 3.2). Average bicelle thicknesses for DLPC (39 ± 2 Å), DMPC (44 ± 4 Å), and POPC (49 ± 3 Å) match the theoretical L_{max} values but display detergent-dependent thicknesses. As DPPC was only combined with DH_6PC, the bicelle L_{max} represents the average value between q = 1.5 and q = 2.0. While DMPC/DH_6PC (L_{max} ~ 42 Å) and DMPC/DH_7PC (L_{max} ~ 44 Å) bicelles agree with the expected thickness of a DMPC bilayer, DMPC/Cyclofos-6, DMPC/FC10, and DMPC/FC12 all produce thicker bicelles across the core (L_{max} ~ 48 Å). Bicelles larger than the expected thicknesses of DLPC and POPC were not observed for either DLPC/Cyclofos-6 (L_{max} ~ 40 Å) or POPC/Cyclofos-6 (L_{max} ~ 49 Å), suggesting that the phenomenon is specific to the DMPC/Cyclofos-6 combination. Future experiments are required to demonstrate if the same results hold true for DLPC and POPC mixed with fos-choline detergents. Hypotheses regarding detergent-dependent bicelle thicknesses are presented in the following section.

Table 3.2: Comparisons for theoretical L_{max} constraints and experimental bicelle L_{max}

Lipid	Average L_{max} (Å)	Range L_{max} (Å)	Bicelle L_{max} (Å)
DLPC	39	38-43	39 ± 2
DMPC	44	43-49	44 ± 4
DPPC	49	48-53	50*
POPC	49	48-53	49 ± 3

*Average for DPPC/DH_6PC q = 1.5 and 2.0

3.7 Concluding remarks

All bicelles formed with PC lipids in this study demonstrate mixing behavior between lipid and detergent for $q < 0.5$ based on SAXS measurements. The ratio of lipid to detergent that supports a transition from a mixture to a bicelle was at least 1:2 ($q = 0.5$), suggesting that the phase transition is a cooperative event and requires a minimum number of lipids to induce lipid organization. However, the detergent component gives rise to unique bicelle dimensions above the phase transition point, even when the same PC lipid is used. This finding indicates that detergent packing around the bicelle rim may influence lipid packing and rearrangement within the core to produce unique thickness properties (Figure 3.39). Further study is required to determine the effect of the rim-forming agent on lipid properties within the bicelle core, which could profoundly impact the structure, fold, and/or function of membrane proteins reconstituted in bicelles.

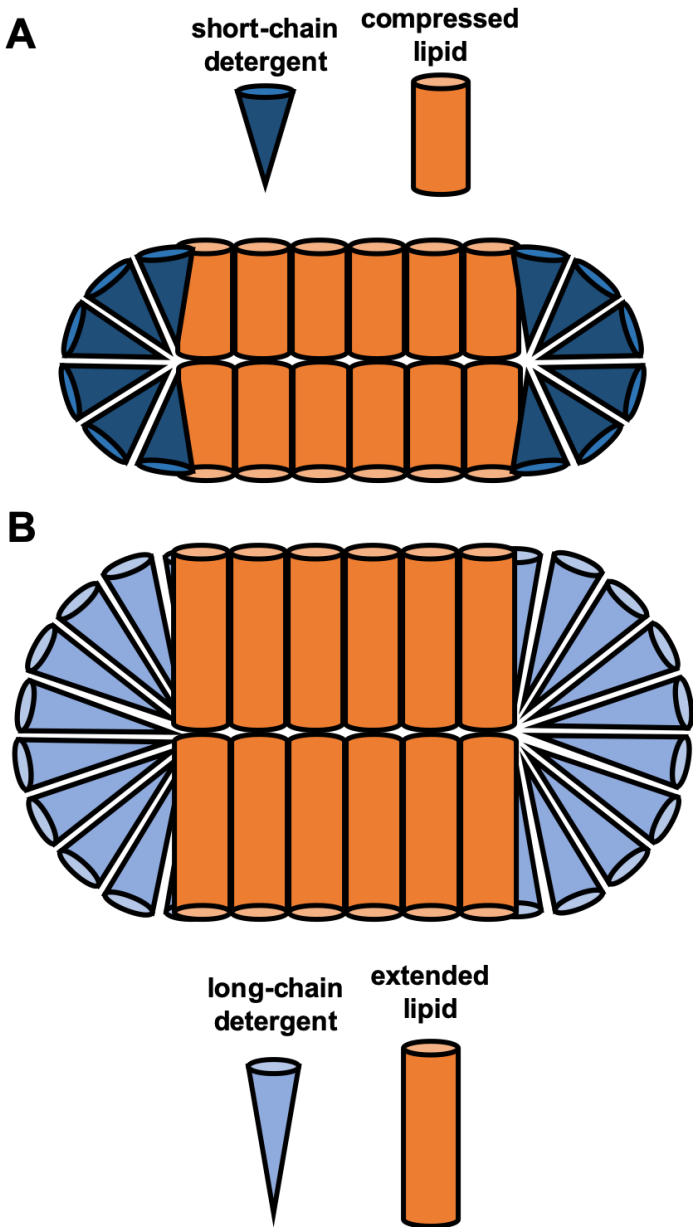


Figure 3.39: Potential hypothesis for detergent-dependent changes in bicelle thickness. Bicelle thickness may be correlated with detergent packing around the rim of the aggregate. (A) Short-chain detergent (such as DH_6PC) may compress lipid in the tail and/or headgroup region to accommodate the rim morphology. (B) Long-chain detergent

(such as FC12) may result in fully-extended lipid acyl chains, and further extension in the headgroup region is possible.

3.8 Acknowledgments

I would like to thank 12-ID-C,D beamline staff at Argonne National Laboratory for help with SAXS data collection. I would especially like to thank Soenke Seifert for helping collect preliminary data, allocating me four shifts of experiment time, and on-site support during experiments. Bicelle characterization for sixteen lipid/detergent mixtures with SAXS would not have been possible without his guidance and collaboration. This work was supported by the NSF (MCB 1817735 to Prof. Linda Columbus; DDGE 1315231 to Nicole Swope) and the NIH (T32 GM008715 to Nicole Swope).

3.9 References

1. Sanders, C.R. and J.H. Prestegard, *Magnetically orientable phospholipid bilayers containing small amounts of a bile salt analogue, CHAPSO.* Biophysical Journal, 1990. **58**(2): p. 447-460.
2. Sanders, C.R. and J.P. Schwonek, *Characterization of magnetically orientable bilayers in mixtures of dihexanoylphosphatidylcholine and dimyristoylphosphatidylcholine by solid-state NMR.* Biochemistry, 1992. **31**(37): p. 8898-8905.
3. Dürr, U.H.N., R. Soong, and A. Ramamoorthy, *When detergent meets bilayer: birth and coming of age of lipid bicelles.* Progress in nuclear magnetic resonance spectroscopy, 2013. **69**: p. 1-22.
4. Arnold, A., et al., *Cation Modulation of Bicelle Size and Magnetic Alignment as Revealed by Solid-State NMR and Electron Microscopy.* Biophysical Journal, 2002. **83**(5): p. 2667-2680.
5. Luchette, P.A., et al., *Morphology of fast-tumbling bicelles: a small angle neutron scattering and NMR study.* Biochimica et Biophysica Acta (BBA) - Biomembranes, 2001. **1513**(2): p. 83-94.
6. Dürr, U.H.N., M. Gildenberg, and A. Ramamoorthy, *The magic of bicelles lights up membrane protein structure.* Chemical reviews, 2012. **112**(11): p. 6054-6074.
7. Marcotte, I. and M. Auger, *Bicelles as model membranes for solid- and solution-state NMR studies of membrane peptides and proteins.* Concepts in Magnetic Resonance Part A, 2005. **24A**(1): p. 17-37.
8. Yamada, N.L., M. Hishida, and N. Torikai, *Nanopore formation on unilamellar vesicles of long- and short-chain lipids.* Physical Review E, 2009. **79**(3): p. 032902.
9. Caldwell, T.A., et al., *Low-q Bicelles Are Mixed Micelles.* The Journal of Physical Chemistry Letters, 2018. **9**(15): p. 4469-4473.

10. Glover, K.J., et al., *Structural Evaluation of Phospholipid Bicelles for Solution-State Studies of Membrane-Associated Biomolecules.* Biophysical Journal, 2001. **81**(4): p. 2163-2171.
11. Kot, E.F., et al., *Phase Transitions in Small Isotropic Bicelles.* Langmuir, 2018. **34**(11): p. 3426-3437.
12. Mineev, K.S., et al., *Characterization of Small Isotropic Bicelles with Various Compositions.* Langmuir, 2016. **32**(26): p. 6624-6637.
13. Mortensen, H.G., et al., *Structure of Phospholipid Mixed Micelles (Bicelles) Studied by Small-Angle X-ray Scattering.* Langmuir, 2018. **34**(48): p. 14597-14607.
14. Triba, M.N., D.E. Warschawski, and P.F. Devaux, *Reinvestigation by Phosphorus NMR of Lipid Distribution in Bicelles.* Biophysical Journal, 2005. **88**(3): p. 1887-1901.
15. Lu, Z., et al., *Bicelles at low concentrations.* Molecular pharmaceutics, 2012. **9**(4): p. 752-761.
16. Soong, R. and P.M. Macdonald, *Water Diffusion in Bicelles and the Mixed Bicelle Model.* Langmuir, 2009. **25**(1): p. 380-390.
17. Beaugrand, M., et al., *Lipid Concentration and Molar Ratio Boundaries for the Use of Isotropic Bicelles.* Langmuir, 2014. **30**(21): p. 6162-6170.
18. Kot, E.F., A.S. Arseniev, and K.S. Mineev, *Behavior of Most Widely Spread Lipids in Isotropic Bicelles.* Langmuir, 2018. **34**(28): p. 8302-8313.
19. Columbus, L., et al., *Mixing and Matching Detergents for Membrane Protein NMR Structure Determination.* Journal of the American Chemical Society, 2009. **131**(21): p. 7320-7326.
20. Oliver, R.C., et al., *Tuning Micelle Dimensions and Properties with Binary Surfactant Mixtures.* Langmuir, 2014. **30**(44): p. 13353-13361.
21. Lipfert, J. and S. Doniach, *Small-Angle X-Ray Scattering from RNA, Proteins, and Protein Complexes.* Annual Review of Biophysics and Biomolecular Structure, 2007. **36**(1): p. 307-327.
22. Koch, M.H.J., P. Vachette, and D.I. Svergun, *Small-angle scattering: a view on the properties, structures and structural changes of biological macromolecules in solution.* Quarterly Reviews of Biophysics, 2003. **36**(2): p. 147-227.
23. Jacques, D.A. and J. Trewhella, *Small-angle scattering for structural biology—Expanding the frontier while avoiding the pitfalls.* Protein Science, 2010. **19**(4): p. 642-657.
24. Boldon, L., F. Laliberte, and L. Liu, *Review of the fundamental theories behind small angle X-ray scattering, molecular dynamics simulations, and relevant integrated application.* Nano Reviews, 2015. **6**(1): p. 25661.
25. Lipfert, J., et al., *Analysis of small-angle X-ray scattering data of protein-detergent complexes by singular value decomposition.* Journal of Applied Crystallography, 2007. **40**(s1): p. s235-s239.
26. Lipfert, J., et al., *Size and Shape of Detergent Micelles Determined by Small-Angle X-ray Scattering.* The Journal of Physical Chemistry B, 2007. **111**(43): p. 12427-12438.

27. Schurtenberger, P., et al., *Cross-Section Structure of Cylindrical and Polymer-like Micelles from Small-Angle Scattering Data. 2. Experimental Results.* Langmuir, 1996. **12**(10): p. 2433-2440.
28. Bagger-Jörgensen, H., et al., *A Nonionic Microemulsion with Adsorbing Polyelectrolyte*. Langmuir, 1997. **13**(22): p. 5820-5829.
29. Skou, S., R.E. Gillilan, and N. Ando, *Synchrotron-based small-angle X-ray scattering of proteins in solution.* Nature Protocols, 2014. **9**(7): p. 1727-1739.
30. Safinya, C.R. *Safinya Group Research: Small Angle X-Ray Scattering and Diffraction*. Available from: http://www.mrl.ucsb.edu/~safinyaweb/XRD.htm.
31. Hjelmeland, L.M., *A nondenaturing zwitterionic detergent for membrane biochemistry: design and synthesis.* Proceedings of the National Academy of Sciences, 1980. **77**(11): p. 6368.
32. Kalipatnapu, S. and A. Chattopadhyay, *Membrane Protein Solubilization: Recent Advances and Challenges in Solubilization of Serotonin1A Receptors.* IUBMB Life, 2005. **57**(7): p. 505-512.
33. Stetsenko, A. and A. Guskov, *An Overview of the Top Ten Detergents Used for Membrane Protein Crystallization.* Crystals, 2017. **7**(7).
34. Herrera, F.E., A.S. Garay, and D.E. Rodrigues, *Structural Properties of CHAPS Micelles, Studied by Molecular Dynamics Simulations.* The Journal of Physical Chemistry B, 2014. **118**(14): p. 3912-3921.
35. Booth, P.J., A. Farooq, and S.L. Flitsch, *Retinal Binding during Folding and Assembly of the Membrane Protein Bacteriorhodopsin.* Biochemistry, 1996. **35**(18): p. 5902-5909.
36. McKibbin, C., et al., *Opsin Stability and Folding: Modulation by Phospholipid Bicelles.* Journal of Molecular Biology, 2007. **374**(5): p. 1319-1332.
37. Hwang, P.M., R.E. Bishop, and L.E. Kay, *The integral membrane enzyme PagP alternates between two dynamically distinct states.* Proceedings of the National Academy of Sciences of the United States of America, 2004. **101**(26): p. 9618.
38. Hein, C., et al., *Hydrophobic supplements in cell-free systems: Designing artificial environments for membrane proteins.* Engineering in Life Sciences, 2014. **14**(4): p. 365-379.
39. Laguerre, A., et al., *From Nanodiscs to Isotropic Bicelles: A Procedure for Solution Nuclear Magnetic Resonance Studies of Detergent-Sensitive Integral Membrane Proteins.* Structure (London, England : 1993), 2016. **24**(10): p. 1830-1841.
40. Lind, J., J. Nordin, and L. Mäler, *Lipid dynamics in fast-tumbling bicelles with varying bilayer thickness: Effect of model transmembrane peptides.* Biochimica et Biophysica Acta (BBA) - Biomembranes, 2008. **1778**(11): p. 2526-2534.
41. Serra-Batiste, M., et al., *Aβ42 assembles into specific β-barrel pore-forming oligomers in membrane-mimicking environments.* Proceedings of the National Academy of Sciences, 2016. **113**(39): p. 10866.
42. Chaturvedi, D. and R. Mahalakshmi, *Position—Specific contribution of interface tryptophans on membrane protein energetics.* Biochimica et Biophysica Acta (BBA) - Biomembranes, 2018. **1860**(2): p. 451-457.

43. Beaugrand, M., et al., *Comparative study of the structure and interaction of the pore helices of the hERG and Kv1.5 potassium channels in model membranes.* European Biophysics Journal, 2017. **46**(6): p. 549-559.
44. Gruenhagen, T.C., J.J. Ziarek, and J.P. Schlebach, *Bicelle size modulates the rate of bacteriorhodopsin folding.* Protein Science, 2018. **27**(6): p. 1109-1112.
45. Doucet, M., et al. *SasView version 4.2.2*. Available from: http://www.sasview.org.
46. Tanford, C., *The hydrophobic effect: formation of micelles and biological membranes*. 2nd edition ed. 1980: Wiley.
47. Wang, L., et al., *Formation of a Mimetic Biomembrane from the Hydrophobic Protein Zein and Phospholipids: Structure and Application.* The Journal of Physical Chemistry C, 2017. **121**(36): p. 19999-20006.

Chapter 4: α-helical membrane protein folding in lipid-detergent bicelles

4.1 Overview

Principles of membrane protein folding as well as the use of bicelles to solubilize membrane proteins were discussed in Chapter 1. Chapter 3 described the morphology of isotropically-tumbling bicelles formed with phosphatidylcholine (PC) lipids. Bicelles using PC lipids generally undergo a phase transition from a mixed lipid/detergent assembly to a structure with internal lipid organization. The study outlined in Chapter 3 also suggests that bicelle composition determines the thickness across the core of the assembly in both mixed-lipid and segregated-lipid bicelle phases. An important extension of this work is examining a bicelle system containing a membrane protein because each facet of bicelle morphology (lipid mixing and size/shape) may impact the embedded protein fold. Here I will discuss ongoing work to establish the fold of a two transmembrane α-helical protein, TM0026, in micelles and bicelles using nuclear magnetic resonance (NMR) spectroscopy and electron paramagnetic resonance (EPR) spectroscopy. The fold of TM0026 is sensitive to the thickness of detergent micelles, but lipid-detergent bicelles varying in size and mixing regimes maintain both the protein fold and tertiary helical contacts. TM0026 tertiary contacts were additionally assessed in four different bicelle compositions above the bicelle phase transitions determined in Chapter 3. TM0026 helical contacts change in each environment, indicating that both components of a bicelle influence properties within the core which surrounds a given membrane protein. Possible properties of lipid-detergent

bicelles that stabilize the fold of membrane proteins are explored along with future work to evaluate TM0026 lipid/detergent contacts and protein-bicelle complexes.

4.2 Introduction

Isotropic bicelles with lipid to detergent ratios (q-values) less than 0.5 represent the majority of bicelle conditions used to study membrane proteins by solution NMR spectroscopy [1]. Membrane proteins prepared in bicelles with $q < 0.5$ have empirically produced high-quality spectra for solution-based experiments relative to larger bicelle q-values, which is attributed to their favorable tumbling properties [2]. The most common bicelle q-value found in solution NMR studies is $q \sim 0.3$ because, until recently, the field generally considered $q = 0.3$ bicelles to be the smallest assemblies that would maintain an ideal bicelle morphology [3]. Work by Caldwell *et al.* [4] and Chapter 3 of this dissertation demonstrated that bicelles formed with PC lipids generally form mixed lipid/detergent assemblies below $q = 0.5$. Therefore, results from empty bicelles suggest that membrane proteins prepared in bicelles with $q = 0.3$ likely possess an environment resembling mixed micelles rather than a bilayer-like core. However, it is possible that an embedded protein perturbs lipid/detergent assemblies compared to mixtures without protein.

Little is known about the morphology of protein-bicelle complexes. The environment surrounding a membrane protein in isotropic bicelles is typically inferred from lipid contacts reported by NMR spectroscopy [5, 6]. For example, NMR studies with the β-barrel outer membrane protein X (OmpX) in DMPC/DH_6PC bicelles ($q = 0.5$) reported contacts between backbone amide protons and DMPC lipid by nuclear Overhauser effect (NOE) experiments [5]. From NOE experiments, the authors concluded that OmpX was embedded in a predominantly lipid core. As the phase transition of DMPC/DH_6PC occurs approximately when $q = 0.5$, results from OmpX in this specific study agrees with

empty DMPC/DH_6PC characterization by Caldwell *et al* [4]. Similarly, Piai *et al.* recently explored the optimal DMPC/DH_6PC q-value for solution NMR spectroscopy using the α-helical transmembrane domain of the human Fas receptor [7]. The authors reported that the environment around the Fas receptor was perturbed $q < 0.6$, assessed by NMR chemical shifts [7]. In some cases, however, perdeuterated (^{2}H-labeled) detergent is used in bicelle preparations for membrane proteins, which renders the detergent "invisible" for experiments that probe contacts between backbone amide protons and detergent protons [6].

Interestingly, some proteins require preformed lipid interactions to be stable in bicelles. For example, the small multidrug resistance transporter EmrE must be reconstituted in liposomes before detergent addition to form bicelles ($q = 0.3$), as opposed to direct purification in bicelle solutions [8]. Similarly, lipoprotein signal peptidase II (LspA) from *E. coli* is sensitive to detergent and requires initial assembly in lipid nanodiscs with cell-free expression before detergent addition to form bicelles ($q = 0.3$) [9]. The authors from both studies propose that a lipid boundary is maintained around EmrE and LspA with these bicelle preparation methods, and the lipid boundary is not disrupted with detergent addition [8, 9].

The ongoing work described in this chapter is designed to assess the properties of bicelles that stabilize alpha-helical membrane protein folds using a model membrane protein, TM0026. Specifically, the fold of TM0026 is probed under mixed lipid/detergent conditions and at q-values above the bicelle phase transition using NMR and EPR spectroscopy. The following section (4.3) describes the rationale of using TM0026 to

establish basic principles of α-helical membrane protein folding in lipid/detergent systems. Results from TM0026 prepared in DLPC/DH_6PC, DMPC/DH_6PC, DPPC/DH_6PC, and DMPC/FC10 bicelles are presented in section 4.5. Interpretations of results thus far and future work to test hypotheses are explored in the final section of this chapter (4.5.1).

4.3 TM0026 background and approach

4.3.1 TM0026: a model, polytopic α-helical transmembrane protein

To establish basic principles of α-helical membrane protein folding in lipid/detergent systems, the model, polytopic protein TM0026 from *Thermotoga maritima* [10-16] was selected for this study. TM0026 contains only two transmembrane helices (Figure 4.1), providing a simple system to interpret the effect of the surrounding protein environment on the overall protein fold and tertiary helical contacts. The function of TM0026 is not known, but the gene is located in the *Thermotoga maritima* β-glucoside catabolism regulon BglR. The TM0026 gene is located downstream from TM0024, a β-glucanase, and TM0025, a β-glucosidase, while TM0026 is located upstream from genes expressing components of the β-glucosidase ABC transport system (TM0027 – TM0031) [17]. The final gene in the BglR regulon, TM0032, is the cellobiose-responsive regulator of β-glucoside utilization in *Thermotoga maritima*. Expression of all proteins in the *Thermotoga maritima* BglR regulon is upregulated in the presence of β-linked glucans, indicating that TM0026 likely plays a role in sugar glucose transport [17]. The TM0026 secondary structure [12, 13], low-resolution tertiary structure [16], backbone NMR

assignments [14], and backbone dynamics [15] have been established in detergent micelles by the Columbus laboratory.

4.3.2 Micelle hydrophobic thickness affects the structure of TM0026

Columbus *et al.* demonstrated that the fold of TM0026 is sensitive to matching between the hydrophobic surface area of the protein and detergent micelle thickness [12]. From studying protein-detergent complexes by SAXS [18], the authors discovered that detergents compatible with TM0026 had a headgroup to headgroup thickness (*L*) of approximately 34 Å (Figure 4.2; refer to Chapter 3 for a detailed discussion of SAXS and the *L* parameter of amphipathic core-shell structures). Evidence from NMR and CW EPR spectra suggested that detergents such as *n*-decyl-β-D-maltopyranoside (DM; *L* ~ 34 Å) and *n*-dodecylphosphocholine (FC12; *L* ~ 34 Å) avoided TM0026 conformational exchange processes due to hydrophobic matching. TM0026 conformational heterogeneity was present in *n*-decylphosphocholine (FC10; *L* ~ 28 Å) and *n*-dodecyl-β-D-maltopyranoside DDM (DDM; *L* ~ 39 Å) (Figure 4.2). By mixing FC10 and DDM at molar ratios designed to produce mixed micelles with *L* ~ 34 Å, Columbus *et al.* demonstrated that mixed FC10/DDM micelles resulted in a homogenous protein conformer nearly identical to the conformers observed in DM and FC12 [12]. The same phenomenon was recapitulated in TM0026 by mixing DHPC (*L* ~ 22 Å) and LPPG (*L* ~ 46 Å) (pure detergents incompatible with TM0026) to form mixed micelles with *L* ~ 34 Å. Based on the folding requirements of TM0026 in micelles, an interesting question is whether the same behavior would be observed in lipid-containing environments such as bicelles.

4.3.3 Assessing TM0026 fold and tertiary helical contacts

To evaluate the fold of TM0026 in lipid-detergent bicelles, NMR and CW EPR spectroscopy were employed in this study. Two-dimensional ^{15}N, ^{1}H-HSQC spectra report on the overall fold of the protein. A homogenous TM0026 conformer (a single fold) produces NMR HSQC spectra with the number of resonances approximately reflecting the number of amino acids in the protein sequence, and the resonances are dispersed throughout the spectrum and of similar intensity (Figure 4.3A). TM0026 that is partially unfolded or undergoing conformational exchange in the micro- to millisecond timescale produces resonances that are broadened beyond detection in the HSQC spectrum, and the remaining resonances are typically clustered together (Figure 4.3B). Therefore, a homogenous TM0026 conformer in bicelles can be identified with HSQC experiments as well as partially unfolded TM0026 and/or TM0026 undergoing conformational exchange.

CW EPR was selected as a complementary method to NMR experiments to evaluate a TM0026 site (A13R1) previously identified as a tertiary helical contact by site-directed spin labeling (Figure 4.4). Crucially, A13R1 spin labeling does not perturb the fold of TM0026 [15]. An explanation of protein backbone and side chain dynamics assessed by CW EPR spectra is provided in Chapter 2. Columbus *et al.* demonstrated that TM0026 A13R1 in DM and FC12 detergents (L ~ 34 Å) produce a multicomponent CW EPR spectrum reflecting a spin label with limited mobility (Figure 4.4A) [12]. Multicomponent spectra are typically observed from tertiary contact sites within proteins [19]. When TM0026 A13R1 is solubilized in FC10 (L ~ 28 Å), for example, the relative populations of CW EPR spectral components, α and β, decrease and increase, respectively

(Figure 4.4B). A decrease in the population of the α component, concurrent with an increase in the β component, reflects a spin label with increased mobility. The more mobile A13R1 label is interpreted as a partially unfolded state of the protein (Figure 4.4B). Multicomponent CW EPR spectra for A13R1 are also observed when the protein is undergoing conformational exchange [12], but the α and β components are broadened compared to the spectrum shown in Figure 4.4A.

4.3.4 Study approach

To test if TM0026 folding in lipid-detergent bicelles is unique from detergent micelles, TM0026 was predominantly prepared in DMPC/DH_6PC bicelles. DMPC/DH_6PC was chosen because *L* parameters as well as internal lipid/detergent organization for empty bicelles are known for $0.1 \leq q \leq 1.0$ [4], and TM0026 is soluble in DMPC/DH_6PC over the same q-range. DMPC/DH_6PC mixtures span a range of thickness properties (26 – 42 Å) and undergo a phase transition from mixed to segregated lipid and detergent within the range $0.1 \leq q \leq 1.0$ (Figure 4.5) [4]. If the sole requirement to produce a single conformer of TM0026 is matching the hydrophobic surface area of the protein, only DMPC/DH_6PC q-values with L ~ 34 Å (q ~ 0.3 – 0.4) are expected to produce TM0026 NMR and EPR spectra similar to Figure 4.3A and Figure 4.4A. It is possible, however, that a bilayer-like lipid environment ($q \geq 0.5$) stabilizes a homogenous fold of TM0026 even though the expected DMPC bilayer thickness (L ~ 42 Å) is well above the TM0026 “preference” in detergent environments (Figure 4.5).

An expected pitfall of this study prior to beginning experiments was that TM0026 in DMPC/DH_6PC would produce poor-quality HSQC spectra for $q \geq 0.5$, which is reported for several transmembrane proteins similar in size to TM0026 [20]. Authors of previous studies attribute poor NMR spectra to the large size of proteins in bicelles with higher q-values [1]. However, it is impossible to interpret whether resonance broadening in HSQC spectra is due to protein unfolding, conformational exchange, or protein complex size without complementary EPR experiments and biophysical techniques [11, 12]. In the event of challenges with NMR spectroscopy, additional spin-labeled TM0026 sites (V15R1, F34R1, and V52R1; Figure 4.6A) were prepared for CW EPR experiments. V15R1 and V52R1 represent sites that face detergent according to the low-resolution structure of TM0026 [16] as well as CW EPR line shapes in FC10/DDM [15] (Figure 4.6, panels B and D). Two components are present in V15R1 and V52R1 CW EPR spectra, which are similar to a detergent-facing spin-labeled site in the leucine transporter, LeuT, of *A. aeolicus* [21]. Electron density of a specific detergent-facing R1 label in the LeuT mutant crystal structure showed a spin label rotamer against the surface of the protein [21]. Collectively, the R1-labeled LeuT crystal structure and LeuT CW EPR spectra suggest that line shape components for detergent-facing residues come from spin label rotamers facing detergent and lying against the protein surface. TM0026 F34R1 is located near the expected micelle surface, and the CW EPR spectrum in FC10/DDM suggests a spin label with limited mobility (Figure 4.6C) [15]. Restricted F34R1 motion is interpreted as spin label interactions with micelle headgroups. Combined, V15R1, V52R1, and F34R1 provide

information about the TM0026 environment and protein backbone/side chain dynamics that supplements tertiary contact evaluation using A13R1.

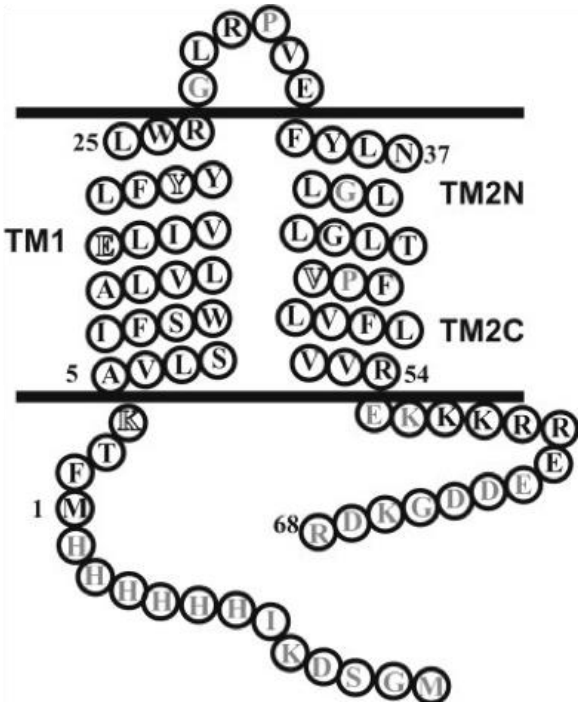


Figure 4.1: Topology of TM0026. *Reprinted from [15] with permission*. Transmembrane helices are labeled TM1 (residues 5-26) and TM2 (residues 34-54). TM2 is divided into two regions, TM2N and TM2C, because backbone dynamics are uncoupled in each region. Letter color corresponds to results from spin label incorporation into the protein. Gray residues were not spin labeled, and labels disrupting the TM0026 fold are indicated as white with an outline. Spin labeling black residues does not disrupt the fold of TM0026.

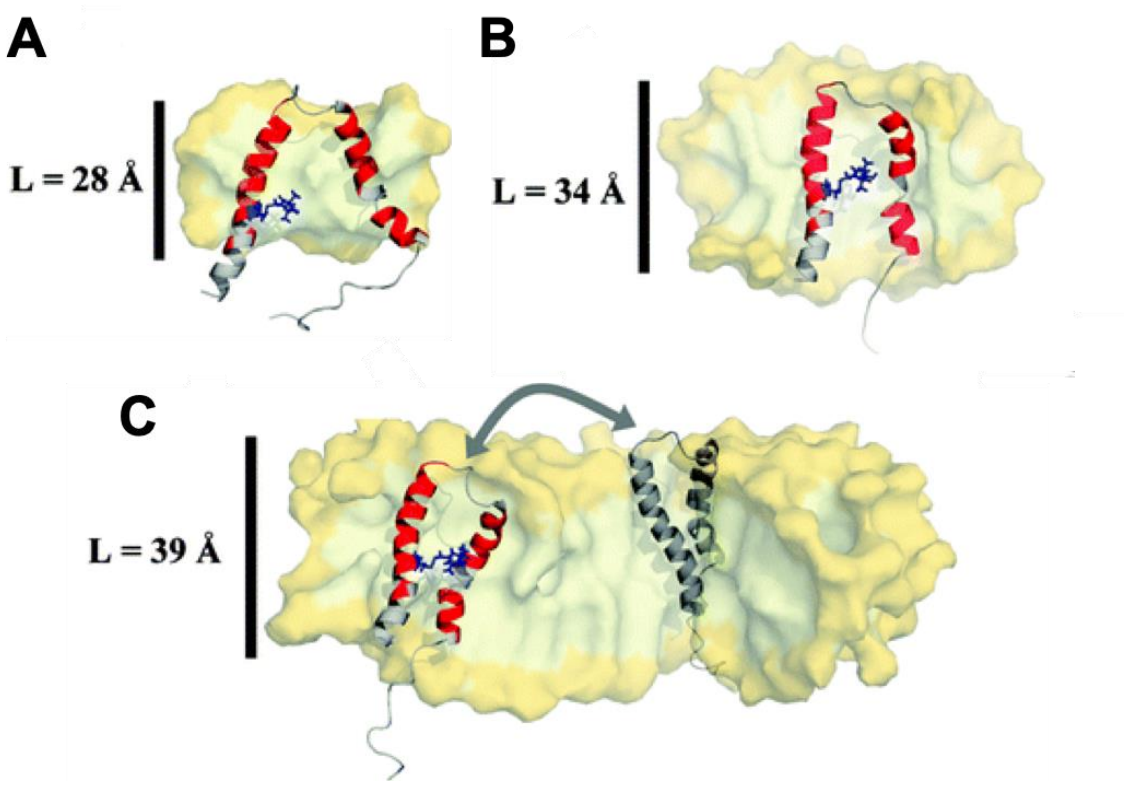


Figure 4.2: Proposed model of TM0026 folding influenced by detergent micelle thickness. *Adapted from [12] with permission*. (A) TM0026 in FC10 (L ~ 28 Å) results in partially-unfolded protein due to insufficient hydrophobic surface area to cover the transmembrane helices. (B) TM0026 in pure micelles or mixed micelles with L ~ 34 Å matches the hydrophobic surface area of the protein and stabilizes one conformer. (C) TM0026 in DDM (L ~ 39 Å) provides a sufficient hydrophobic for the transmembrane helices, but the protein undergoes conformational exchange within the micelle.

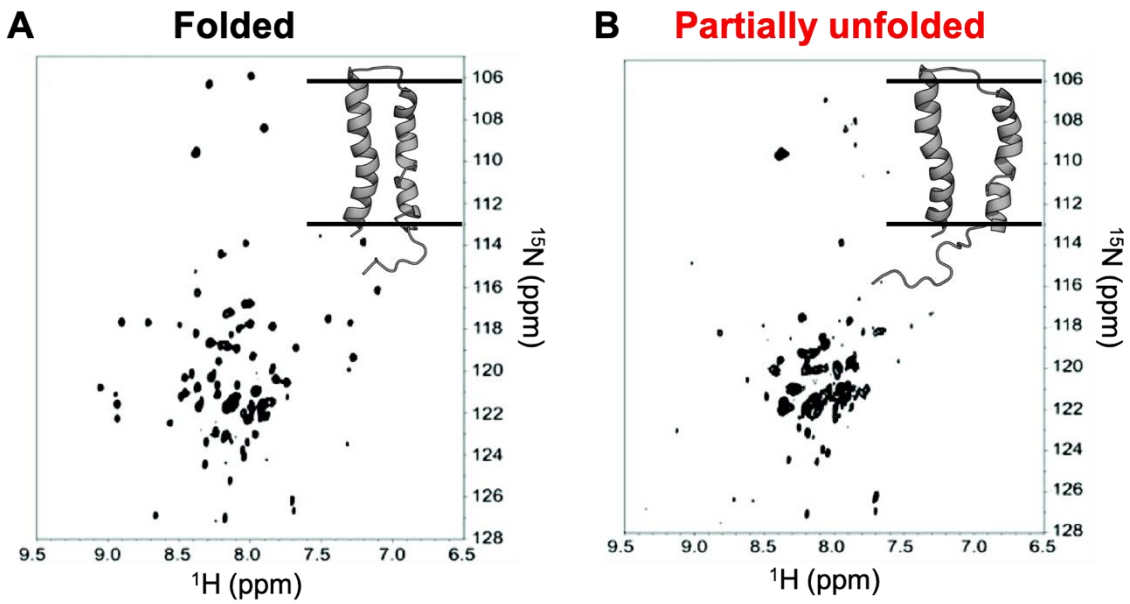


Figure 4.3: Evaluation of TM0026 fold from ^{15}N, ^{1}H-HSQC spectra. *Adapted from [12] with permission*. ^{15}N, ^{1}H-HSQC NMR spectra report on the overall fold of TM0026. (A) TM0026 in DM micelles (L ~ 34 Å) stabilizes one conformer of the protein and produces a ^{15}N, ^{1}H-NMR spectrum with well-dispersed peaks that reflect the number of amino acids in the protein sequence. (B) TM0026 in FC10 (L ~ 28 Å) results in partially-unfolded protein and produces a ^{15}N, ^{1}H-NMR spectrum with a limited number of poorly-dispersed peaks. The low-resolution TM0026 structures from data reported in [16] are inset and used in figures throughout this chapter.

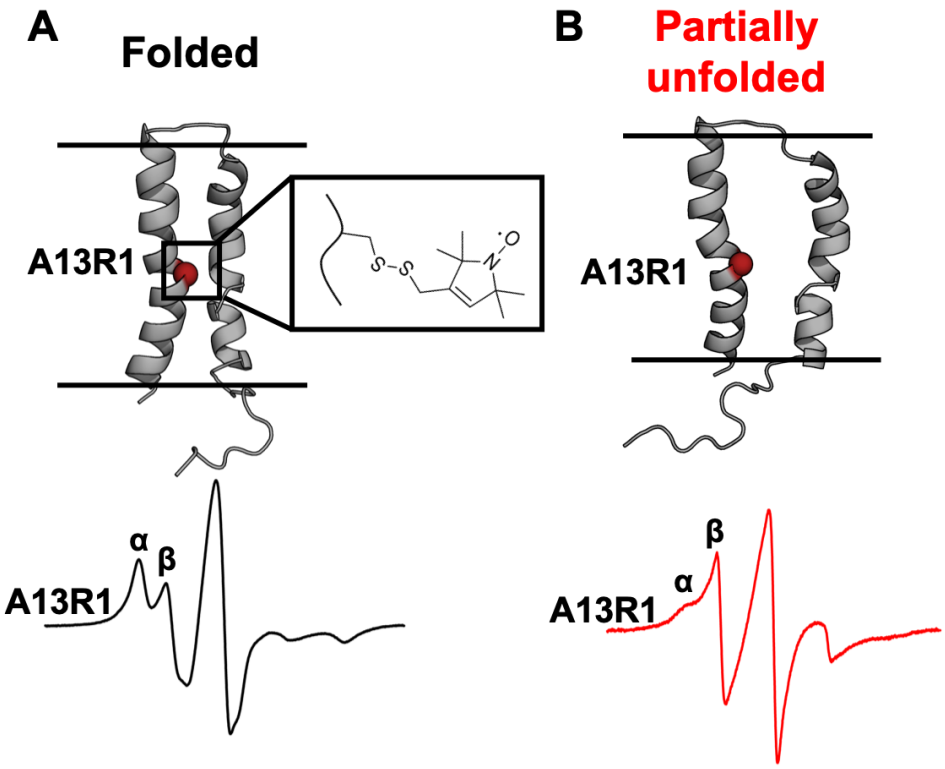


Figure 4.4: Evaluation of TM0026 tertiary helical contact site with CW EPR spectra. TM0026 A13R1 is a reporter of tertiary helical contacts between TM1 and TM2. Multiple components in CW EPR spectra are labeled α and β. (A) TM0026 A13R1 (red sphere) in DM results in folded protein and a CW EPR line shape reflecting a spin label with limited mobility due to forming a tertiary contact with the opposing helix. The chemical structure of the R1 label is inset. (B) Partially-unfolded TM0026 A13R1 in FC10 produces a CW EPR line shape reflecting a spin label with enhanced mobility relative to panel A, indicated by a decrease in the α component decreases and an increase in the β component.

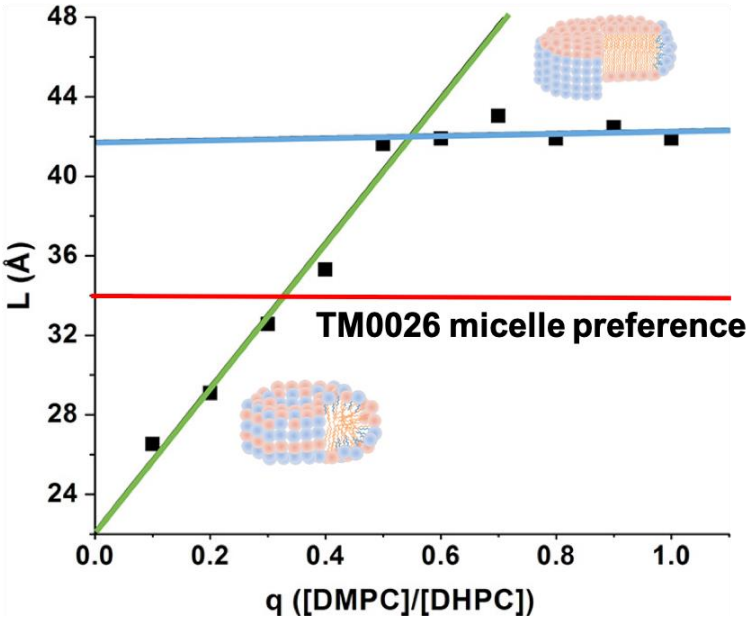


Figure 4.5: Expected DMPC/DH$_6$PC headgroup to headgroup thickness compatible with the fold of TM0026. *Adapted from [4] with permission*. Plot of headgroup to headgroup thickness, L, of DMPC/DH$_6$PC bicelles by q-value [4]. The green line represents the linear increase of L that corresponds to mixed lipid/detergent, and the blue line corresponds to the approximate L parameter of a pure DMPC bilayer (Chapter 3). Ideal bicelles and mixed lipid/detergent assemblies are inset above and below the phase transition point for DMPC/DH$_6$PC. The red line corresponds to the micelle thickness that stabilizes one conformer of TM0026 (L ~ 34 Å).

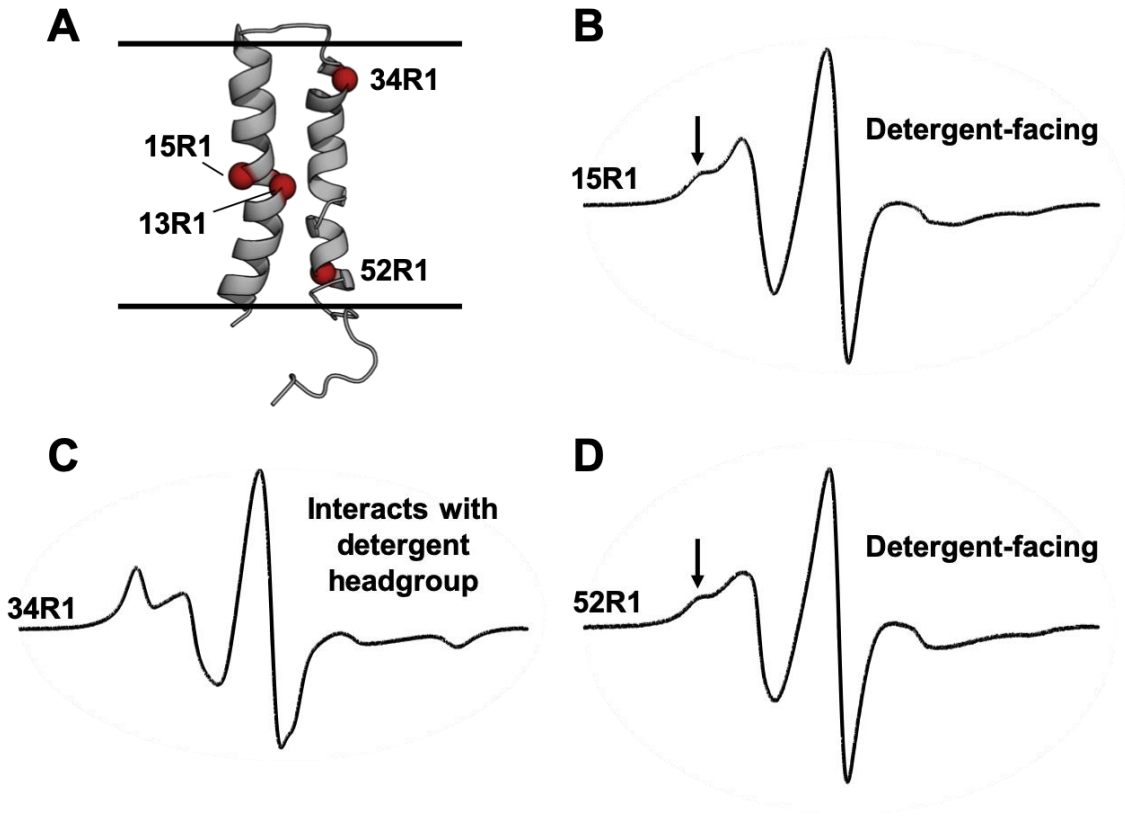


Figure 4.6: TM0026 R1 sites in this study and CW EPR spectra in detergent micelles. (A) Spin labeled sites mapped onto the low-resolution structure of TM0026, rendered as red spheres. CW EPR spectra in panels B-D were recorded for TM0026 in FC10/DDM micelles (χ_{DDM} ~ 0.5). (B) V15R1 is a detergent-facing residue in TM1. An additional component in the spectrum is noted by an arrow and likely corresponds to a spin label rotamer that interacts with the helical protein surface according to studies by Kroncke *et al* [21]. (C) The EPR line shape of F34R1 reflects a spin label with limited motion. Due to the residue location near the estimated interface of the micelle environment, F34R1 is proposed to interact with detergent head groups. (D) V52R1 is a detergent-facing residue in TM2. Similar to V15R1, the CW EPR spectrum shows an additional component (arrow).

4.4 Materials and methods

4.4.1 Bicelle preparation

The detergent 1,2-dihexanoyl-*sn*-glycero-3-phosphocholine (6:0 PC, DH_6PC) and lipids 1,2-dilauroyl-*sn*-glycero-3-phosphocholine (12:0 PC, DLPC), 1,2-dimyristoyl-*sn*-glycero-3-phosphocholine (14:0 PC, DMPC), and 1,2-dipalmitoyl-*sn*-glycero-3-phosphocholine (16:0 PC, DPPC) were purchased from Avanti Polar Lipids in powder form. Detergents n-decylphosphocholine (FC10, sol-grade) and n-dodecylphosphocholine (FC12, sol-grade) were purchased from Anatrace in powder form. Chemical structures for lipids and detergents are provided in Chapter 3 (Figure 3.7 and Figure 3.8).

Prior to forming bicelle mixtures, powder lipids and detergents were resuspended in the appropriate purification buffer and mixed by nutation at room temperature for 30 minutes. Lipids and detergents were combined to give a total amphiphile concentration (C_L) of 2% in a 10 mL working volume. A 2% bicelle solution represents the most dilute condition that gives lipid/detergent assemblies across all mixtures in this study [3], which was required to minimize the final C_L in proteobicelle samples. Each mixture was prepared with effective bicelle q-values of 0.3 or 1.0, accounting for the concentration of free detergent in solution according to Equation 3.4 (Chapter 3). Free detergent concentrations are 7 mM [4, 22] for DH_6PC and 6 mM for FC10. After forming the final lipid/detergent mixture with the desired q-value and % C_L, samples were vortexed for 60 sec. Samples were then subjected to freeze/thaw cycles of -80°C and 37°C for 30 minutes each, with vortexing for 60 sec before each freeze step. Cycles of freezing, thawing, and vortexing

were performed until the solution was optically clear. At the point of optical clarity, three more freeze/thaw cycles were performed. Samples were stored at -80°C.

4.4.1 Expression and purification of ^{15}N, ^{1}H-labeled TM0026

Previously published protocols were followed [13-15]. Briefly, the TM0026 gene with an N-terminal tag (MGSDKIHHHHHH) was cloned into pET25b [13]. Plasmid containing the TM0026 gene was transformed into BL21(DE3)-RIL cells (Stratagene) for expression in M9 minimal media containing $^{15}NH_4Cl$ (99%; Cambridge Isotope Laboratories) and 50 μg/mL ampicillin. Cells were grown to an OD_{600} of 0.8, and protein expression was induced with 1 mM IPTG for 4 hours at 37°C. Cells were harvested and resuspended in lysis buffer containing 50 mM Tris (pH 8.0) and 150 mM NaCl, and 1 mM tris(2-carboxyethyl)phosphine (TCEP). Cell lysis was performed with either a microfluidizer (Microfluidics model 110L) or Nano DeBEE homogenizer (BEE International). Following cell lysis, cell debris was removed by centrifugation at 15,000 *g* for 60 min. To solubilize cell membranes, the supernatant was supplemented with 10 mM *n*-decyl-β-D-maltopyranoside (DM; Anatrace) for 3 hours at room temperature.

After incubation with DM, the suspension was passed through a Co^{2+} immobilized metal affinity chromatography (IMAC) column equilibrated with 10 column volumes (CV) of lysis buffer with 5 mM DM. For TM0026 exchange into bicelles, the column was washed with 20 CV of 2% bicelle solution in 20 mM phosphate buffer (pH 7.8), 150 mM NaCl, and 15 mM imidazole. TM0026 was eluted in 10 CV of 2% bicelle solution in 20 mM phosphate buffer (pH 7.0), 150 mM NaCl, and 600 mM imidazole. TM0026

purification in DH_6PC was performed in the same wash and elution buffer conditions with 20 mM DH_6PC rather than bicelles. Fractions containing pure protein were assessed via SDS-PAGE. Eluted ^{15}N, ^{1}H-labeled TM0026 was concentrated and dialyzed with a 3.5 kDa molecular weight cutoff (MWCO) dialysis cassette (Millipore) against 4 L of 20 mM phosphate buffer (pH 6.2), 150 mM NaCl for one hour, three times to remove imidazole. Imidazole removal was confirmed with ^{1}H NMR. Final protein concentrations were approximately 400-600 μM. ^{1}H NMR spectra confirmed that protein in exchange into bicelles and detergent removal was successful.

4.4.2 Mutagenesis, expression, and purification of TM0026 cysteine mutants

Previously published protocols were followed [13-15]. Cysteine mutations were introduced using PIPE mutagenesis [23]. All mutations were confirmed by sequencing the entire gene (GENEWIZ). Plasmid containing the TM0026 gene was transformed into BL21(DE3)-RIL cells (Stratagene) for expression in Luria-Bertani (LB) containing 50 μg/mL ampicillin. Protein expression, cell lysis, and membrane solubilization followed the same protocol described for isotopically-labeled TM0026. After incubation with DM to solubilize membranes, the suspension was passed through a Co^{2+} immobilized metal affinity chromatography (IMAC) column equilibrated with 10 column volumes (CV) of lysis buffer with 5 mM DM. The column was washed with 20 CV of 20 mM phosphate buffer (pH 7.8), 150 mM NaCl, 15 mM imidazole, 1 mM TCEP, and 5 mM *n*-dodecylphosphocholine (FC12). TM0026 was eluted in 10 CV of 20 mM phosphate buffer (pH 7.0), 150 mM NaCl, 600 mM imidazole, 1 mM TCEP, and 5 mM FC12. Fractions containing pure protein were assessed via SDS-PAGE.

4.4.3 TM0026 spin labeling and micelle/bicelle exchange

TM0026 cysteine mutants were spin labeled as previously described [13]. Protein eluted from Co^{2+} IMAC purification was concentrated, and TCEP and imidazole were removed with a PD10 column (GE Healthcare) equilibrated with 20 mM phosphate buffer (pH 7.0), 150 mM NaCl, and 5 mM FC12. Proteins were eluted using the PD10 equilibration buffer directly into (1-oxyl-2,2,5,5-tetramethyl-3-pyrroline-3-methyl)-methanethiosfonate (MTSSL, R1; Toronto Research Chemicals Inc) at a 1:15 molar ratio of protein to spin label (stored as 100 mM stock in acetonitrile). TM0026 was incubated with the spin label overnight in the dark at 4°C. After the labeling reaction overnight, the sample was passed through a second Co^{2+} immobilized metal affinity chromatography (IMAC) column equilibrated with 10 column volumes (CV) of 20 mM phosphate buffer (pH 7.0), 150 mM NaCl, and 5 mM FC12. To remove excess spin label, the column was washed with 40 CV of the IMAC equilibration buffer. 20 mM phosphate buffer (pH 7.8), 150 mM NaCl, 15 mM imidazole, 1 mM TCEP, and 5 mM *n*-dodecylphosphocholine (FC12). Exchange into the appropriate micelle and/or bicelle condition was performed by washing with 10 CV of detergent and/or 2% bicelle solution in 20 mM phosphate buffer (pH 7.0), 150 mM NaCl. TM0026 was eluted with 10 CV of detergent and/or 2% bicelle solution in 20 mM phosphate buffer (pH 7.0), 150 mM NaCl, and 300 mM imidazole. For exchange into micelles, detergent concentrations were 5 mM above the critical micelle concentration (CMC). Fractions containing pure protein were assessed via SDS-PAGE. Eluted protein was concentrated and dialyzed with a 3.5 kDa MWCO dialysis cassette (Millipore) against 4 L of 20 mM phosphate buffer (pH 6.2), 150 mM NaCl for one hour,

three times to remove imidazole. ^{1}H NMR spectra confirmed that protein in exchange into bicelles and detergent removal was successful. Final protein concentrations were approximately 200 µM.

4.4.4 Nuclear Magnetic Resonance spectroscopy

Isotopically ^{15}N, ^{1}H-labeled TM0026 for NMR experiments included addition of 10% D_2O for lock. Chemical shifts were obtained from the published TM0026 assignment (BMRB 18494) [14]. NMR experiments were performed on a Bruker AVANCE spectrometer operating at a proton frequency of 800 MHz and equipped with a 5 mM TXI cryoprobe. ^{15}N-^{1}H transverse relaxation optimized spectroscopy (TROSY)-based experiments were used to record two-dimensional heteronuclear single quantum coherence (HSQC) spectra. All spectra were recorded at 40°C. Data were processed with NMRPipe [24] and analyzed with NMRFAM-SPARKY [25].

4.4.5 Continuous-wave Electron Paramagnetic Resonance spectroscopy

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 µL of each sample were used for measurements 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).

4.4.6 Bicelle q-titrations with TM0026

Starting q-values of TM0026 bicelle samples were confirmed by the integral ratio of lipid and detergent methyl resonances from ^{1}H NMR spectra. NMR and/or CW EPR experiments were performed on the first sample prior to beginning the titration. “Forward” q-titrations (q = 0.3 $\rightarrow$ 1.0) were performed by adding lipid (Figure 4.7A), and “reverse” q-titrations (q = 1.0 $\rightarrow$ 0.3) were done with detergent addition (Figure 4.7B). Lipid and detergent stocks were prepared at 600 mM in 20 mM phosphate buffer (pH 6.2), 150 mM NaCl. High stock concentrations were used to minimize dilution effects during titration experiments. Lipid stocks were subjected to pulse sonication (10% amplitude, 10 sec on, 10 sec off) for ~15 minutes (Qsonica) immediately before addition to TM0026 bicelle samples. For each target q-value, lipid or detergent stock was added to TM0026 samples in 20 μL increments, with mixing between each addition, until the desired effective q-value was reached. To facilitate incorporation of new lipid or detergent into the bicelle assemblies, the titrated sample was then subjected to three cycles of cooling (4°C) and warming (room temperature) for 30 minutes each, with nutation during each step to mix. After cooling and warming cycles, NMR and/or CW EPR spectra were recorded for the titrated sample. The process of lipid or detergent addition to TM0026 samples, cooling/warming, and recording spectra was repeated for each desired q-value in the titration experiment until the final q-value was achieved.

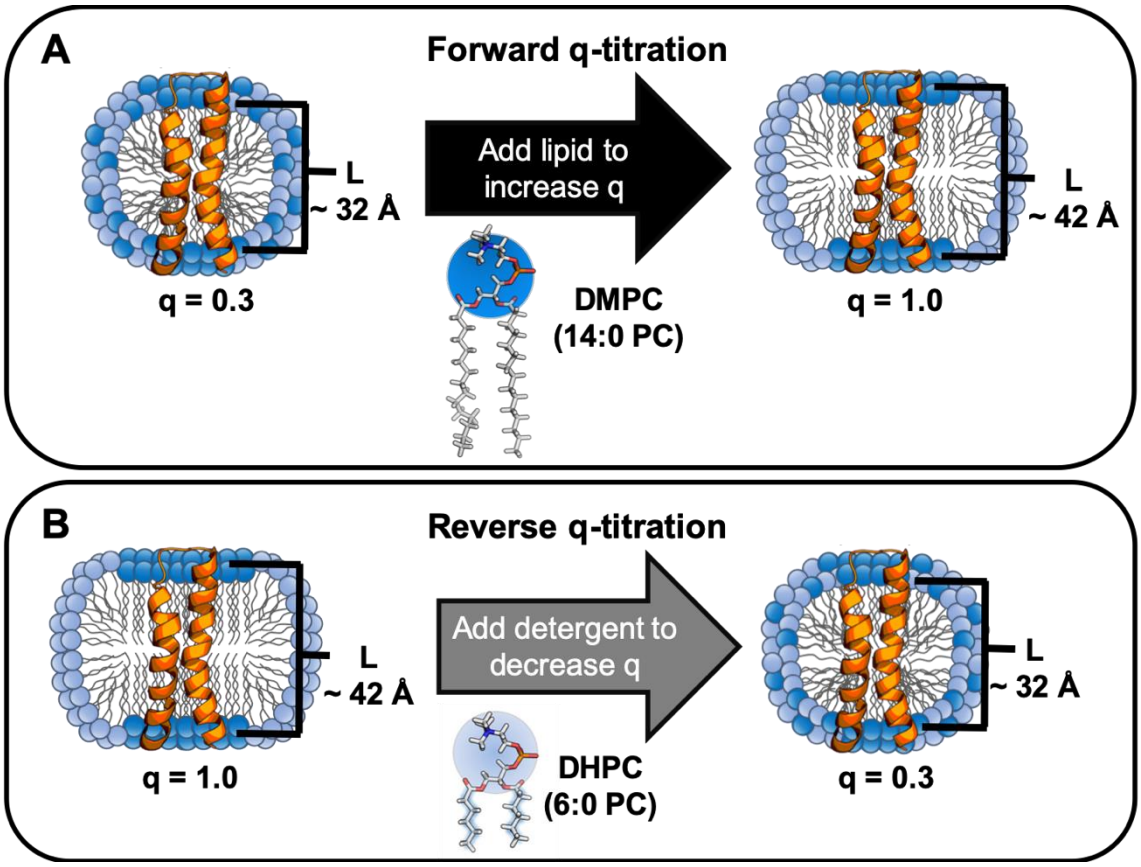


Figure 4.7: Schematic of TM0026 bicelle q-titration experiments. General schematic of "forward" (A) and "reverse" (B) q-titration experiments for TM0026 bicelles. Depicted bicelle assemblies are formed with DMPC lipid (dark blue) and DHPC detergent (light blue) in the example titration experiments. Three-dimensional structures of DMPC (dark blue) and DHPC (light blue) are rendered as sticks, with carbon-hydrogen bonds (gray) shown as well as oxygen (red), nitrogen (blue), and phosphorous (orange) atoms. Lipid/detergent localization and the headgroup to headgroup thickness, L, for q = 0.3 and 1.0 bicelles reflect published results for empty DMPC/DH_6PC from Caldwell *et al.* [4] and Chapter 3.

4.5 Results and discussion

4.5.1 TM0026 fold in DMPC/DH_6PC bicelles and DH_6PC micelles

To assess the overall fold of TM0026 in bicelles, ^{15}N, ^{1}H-HSQC spectra were recorded using a reverse q-titration experiment (Figure 4.7B) for TM0026 in DMPC/DH_6PC (q = 1.0, 0.7, 0.5, 0.4, and 0.3; Figure 4.8). Backbone NMR assignments for TM0026 in FC10/DDM [14] were used for preliminary resonance assignments in bicelles and mapped onto to the low-resolution TM0026 structure (Figure 4.9). The HSQC spectrum of TM0026 in DMPC/DH6PC, q = 1.0 (Figure 4.8) contain resonances only from highly mobile C-terminal residues (Figure 4.9). As the bicelle q-value is lowered to q = 0.7, side chain indole resonances appear from tryptophan residues (also mobile). The first instance of tertiary contact residues appearing in HSQC spectra was observed when q = 0.5 (Figure 4.9). Progressively reducing the bicelle q-value increased the number of resonances up until q = 0.4, at which point 70 out of 78 expected resonances were observed. Resonance intensities increased when q = 0.3 (Figure 4.8), and an additional resonance (F51) appeared in the HSQC spectrum.

TM0026 in both q = 0.4 and q = 0.3 DMPC/DH_6PC bicelles produce spectra with a number of well-dispersed resonances that is consistent with a single TM0026 conformer. Peak chemical shifts are almost identical between the two bicelle conditions, which suggests that the TM0026 fold is maintained. Furthermore, the limited chemical shift changes imply that the surrounding environment does not produce observable differences in the spectra. Although the number of resonances for TM0026 in DMPC/DH_6PC, q = 0.5

is lower (50 resonances) than q = 0.4 and q = 0.3, the resonances are both well-dispersed and similar in intensity (Figure 4.8). The resulting spectrum for q = 0.5 is not expected for a protein undergoing conformational exchange. Therefore, the lower number of observable TM0026 resonances for q = 0.5 bicelles may be due to unfavorable tumbling properties rather than protein unfolding and/or conformational exchange. If NMR spectral quality is due to unfavorable bicelle tumbling at higher q-values, the resulting HSQC spectra for q = 0.7 and q = 1.0 are consistent with this hypothesis. However, there are an insufficient number of TM0026 resonances when q = 0.7 and q = 1.0 to exclude conformational exchange processes and/or partial protein unfolding occurring within the protein.

To compare with TM0026 in bicelles, the HSQC spectrum for TM0026 in DH_6PC micelles (L ~ 22 Å) was also recorded (Figure 4.10). Unlike TM0026 in bicelles when $q \leq 0.5$, the HSQC spectrum for TM0026 in DH_6PC shows approximately 45% of the expected resonances. Furthermore, the resonances are not well-dispersed and contain unequal intensities, indicating that TM0026 is either partially unfolded and/or undergoing conformational exchange. The result for TM0026 in DH_6PC is consistent with previous observations [11, 12]. From NMR experiments with TM0026 in bicelles and micelles, it is clear that TM0026 adopts a single fold in DMPC/DH_6PC bicelles when q = 0.3 (L ~ 32 Å) and q = 0.4 (L ~ 36 Å) but not in pure DH_6PC micelles (L ~ 22 Å). These results agree with the micelle thickness preference of L ~ 34 Å for TM0026 (Figure 4.5). However, the chemical shift dispersion and resonance intensities of TM0026 in DMPC/DH_6PC q = 0.5 (L ~ 42 Å) suggest that the protein may also be adopting a single conformer in this condition. As HSQC experiments could not be used to assess the fold of TM0026 across

the entire range of bicelle q-values, the remainder of experiments in this study utilized EPR spectroscopy.

4.5.2 TM0026 tertiary helical contacts in micelles and bicelles

The TM0026 tertiary helical contact site A13R1 was prepared in DMPC/DH_6PC q = 0.3 and q = 1.0 for forward and reverse q-titration experiments (Figure 4.7). The resulting CW EPR spectra for TM0026 A13R1 in DMPC/DH_6PC (q = 0.3, 0.4, 0.5, 0.6, 0.7, and 1.0) are almost identical and contain line shapes with multiple components (Figure 4.11A). Differences between spectra are that the β component decreases and central linewidth splitting increases with higher q-values. Similarly, the β component increases and central linewidth splitting decreases with lower q-values during a reverse q-titration experiment (Figure 4.11B). These changes may reflect a spin label that is slightly more mobile at lower bicelle q-values due to the surrounding lipid/detergent dynamics in different conditions. Overall, the CW EPR q-titration experiments for TM0026 A13R1 in DMPC/DH_6PC demonstrate a rigid, tertiary helical contact is maintained for $0.3 \leq q \leq 1.0$ (32 Å $\leq L \leq$ 42 Å; Figure 4.5). The α and β components are furthermore narrow, which is not expected for a protein undergoing conformational exchange.

To demonstrate that TM0026 titration experiments with CW EPR recapitulates information from the overall protein fold, TM0026 A13R1 was prepared in DDM (L ~ 39 Å) and FC10 (L ~ 28 Å) micelles (Figure 4.12A). FC10 was then added to TM0026 in DDM (Figure 4.12C), and DDM was added to TM0026 in FC10 (Figure 4.12D), to produce TM0026 A13R1 in mixed FC10/DDM micelles. Similar “rescue” experiments with

TM0026 in pure FC10 and DDM have been performed with NMR spectroscopy [12], but the method has not been demonstrated with CW EPR in previous reports. The resulting CW EPR spectra indicate that for TM0026 in DDM, addition of FC10 produces sharper α and β components that reflect a reduction in conformational exchange processes occurring for TM0026 in pure DDM. For TM0026 in FC10, addition of DDM drastically changes the CW EPR spectrum from a line shape reflecting partially unfolded protein (Figure 4.4B) to a line shape representative of the helical contact shown in Figure 4.4A. Combined, results from TM0026 titration experiments in detergent demonstrate that CW EPR is appropriate for evaluating tertiary helical contacts during bicelle q-titration experiments.

To compare the TM0026 tertiary helical contact in bicelles formed with shorter and longer-chain PC lipids, TM0026 A13R1 was also prepared in q = 1.0 DLPC (12:0)/DH_6PC and DPPC (16:0)/DH_6PC bicelles (Figure 4.13A). The q = 1.0 condition was selected because PC bicelles without protein generally show evidence of a lipid core in this regime. The L values for DLPC/DH_6PC, DMPC/DH_6PC, and DPPC/DH_6PC when q = 1.0 are approximately 37 Å, 42 Å, and 50 Å, respectively (Figure 4.13B). Each CW EPR spectrum shows line shapes with multiple components but is different between bicelle environments, which is consistent with different lipids in the core. For example, the β component increases for DLPC/DH_6PC, and the central linewidth splitting is most apparent in the DMPC/DH_6PC condition. Both the α and β components broaden in DPPC/DH_6PC bicelles, suggesting that conformational exchange may be present even though the helical contact is maintained.

Chapter 3 demonstrated that the detergent component influences the headgroup to headgroup thickness across the core of bicelles formed with the same PC lipid, without protein present. To evaluate if detergent-dependent changes in the bicelle core affect the fold of a membrane protein, TM0026 A13R1 was prepared in q = 1.0 DMPC/FC10 bicelles (L ~ 48 Å). Compared to bicelles formed with DH_6PC, the CW EPR spectrum of TM0026 A13R1 in DMPC/FC10 bicelles is noticeably different (Figure 4.14A). The L parameter of DMPC/FC10 bicelles is closest to DPPC/DH_6PC bicelles (Figure 4.13B and Figure 4.14B), but the EPR spectrum of TM0026 A13R1 in DMPC/FC10 is not similar to any of the bicelle conditions with DH_6PC. This result suggests that the observed TM0026 A13R1 line shapes are not solely due to the bicelle thickness. The central linewidth of TM0026 A13R1 in DMPC/FC10 decreases, and the β component is broad, intense, and inconsistent with the β component in DMPC/DH_6PC. Changes in the line shape for the DMPC/FC10 condition may be due to a small population of unfolded protein relative to the folded TM0026 conformer present in DMPC/DH_6PC bicelles.

A reverse q-titration experiment with TM0026 A13R1 in DMPC/FC10 shows almost identical CW EPR spectra (Figure 4.14C). Trends are additionally similar to DMPC/DH_6PC titrations reporting increased A13R1 mobility with lower q-values. Collectively, CW EPR results for TM0026 A13R1 in different bicelle compositions with DMPC lipid support that the environment surrounding the protein is not the same. This finding is fundamentally different from the current understanding of DMPC bicelles in which the DMPC core should, intuitively, be consistent across different rim components when q = 1.0. The different core environments importantly appear to affect the fold of an

embedded membrane protein, complicating the use of bicelles beyond establishing the presence of a lipid core.

4.5.3 Additional TM0026 residues evaluated with EPR

To test if EPR line shapes change for additional TM0026 mutants in bicelles, V15R1, F34R1, and V52R1 mutants (Figure 4.15A) were prepared in q = 1.0 DMPC/DH_6PC. Reverse q-titration experiments were performed on each sample (q = 1.0, 0.7, 0.5, and 0.3), and the CW EPR spectra are shown in Figure 4.15. Across all sites, the line shapes reflect an increase in spin label mobility with lower bicelle q-values. Future experiments are required to determine the phenomenon causing the inverse dependence between q-value and spin label mobility. Finally, the lipid/detergent-facing residue V15R1 was prepared in q = 1.0 DMPC/FC10 bicelles to compare with q = 1.0 DMPC/DH_6PC (Figure 4.16A). Minor differences are observed between CW EPR spectra in each condition. Both conditions produce EPR spectral components indicating two side chain rotamers are present (Figure 4.6), and they are consistent with EPR spectra for lipid/detergent-facing residues in other membrane proteins [21]. For TM0026 V15R1 in both DMPC/DH_6PC and DMPC/FC10, reverse q-titration experiments demonstrate that the second spectral component disappears at lower q-values (Figure 4.15B and Figure 4.16B).

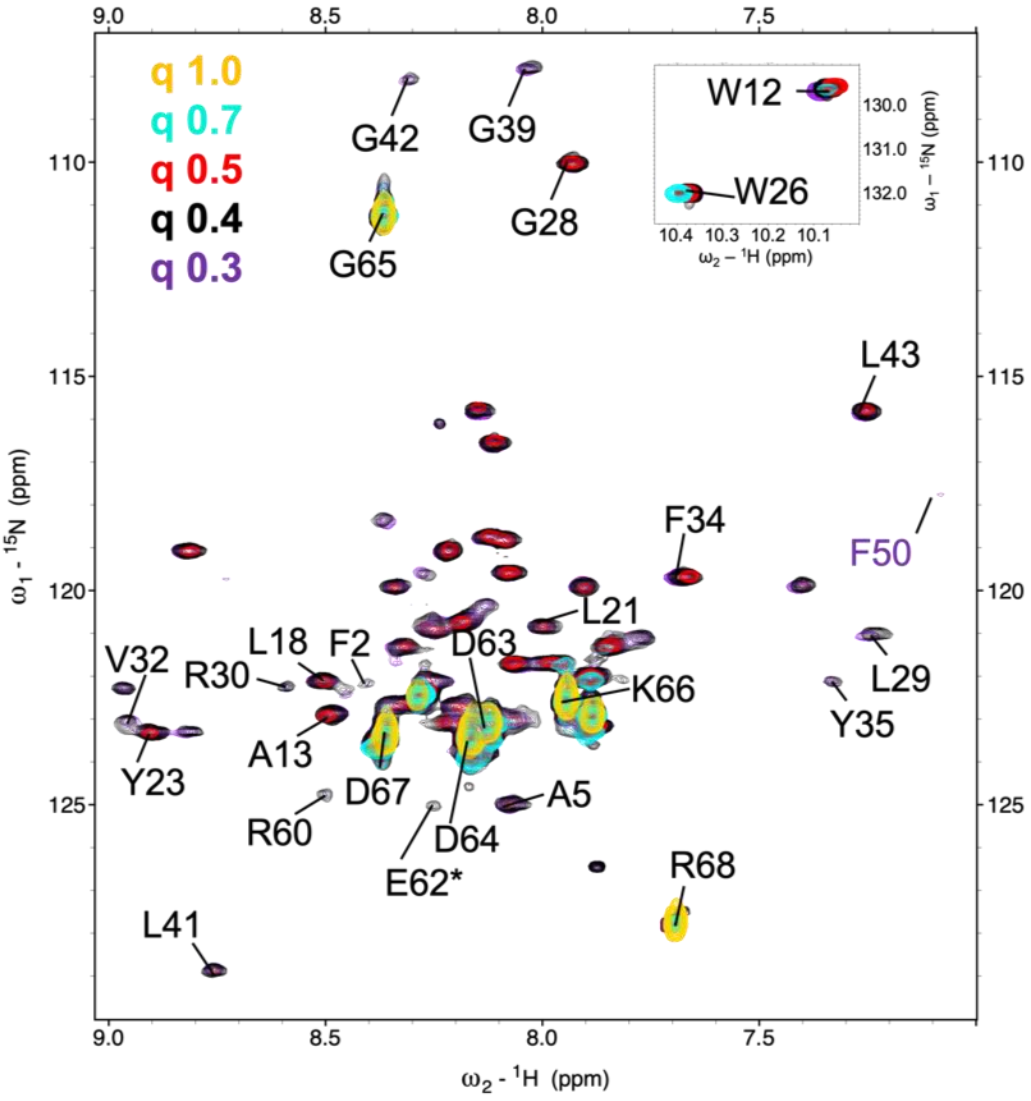


Figure 4.8: ^{15}N, ^{1}H-HSQC spectra of reverse q-titration for TM0026 in DMPC/DH$_6$PC. Overlaid HSQC spectra of TM0026 in DMPC/DH$_6$PC bicelles were partially assigned using the data reported by Kroncke and Columbus [14]. Spectra are colored according to the bicelle q-value as indicated by the legend in the top left corner. Tryptophan indole resonances are included in the inset. F50 is colored purple to denote an additional resonance that appears for the DMPC/DH$_6$PC q = 0.3 condition.

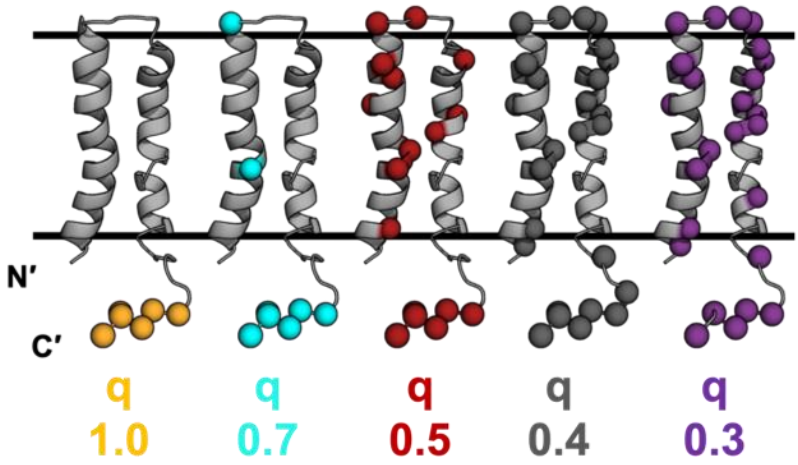


Figure 4.9: Location of resonances observed during TM0026 DMPC/DH$_6$PC reverse q-titration. Partial assignments from Figure 4.8 were mapped onto the low-resolution structure of TM0026, rendered as spheres. Spheres are colored according to the q-value color scheme in Figure 4.8.

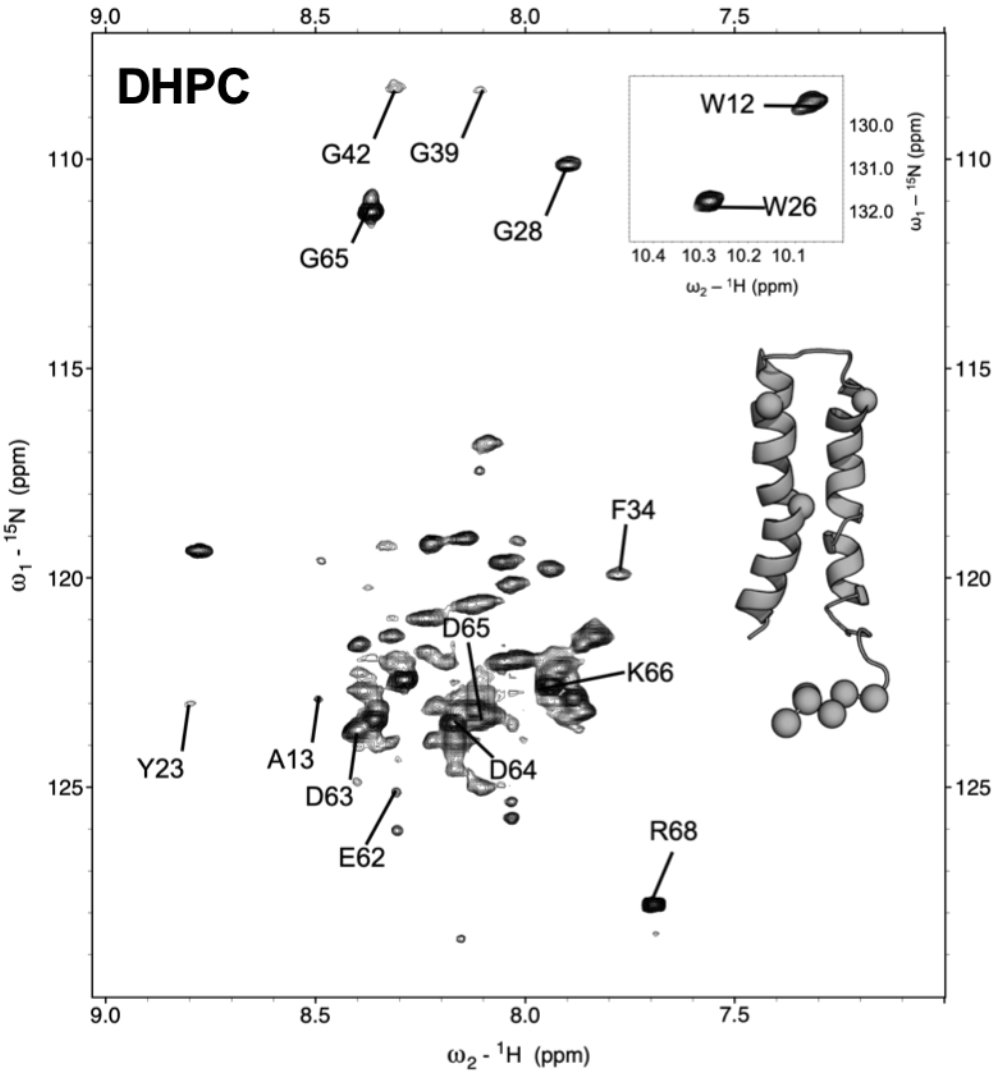


Figure 4.10: ^{15}N, ^{1}H-HSQC spectrum of TM0026 in DHPC. The HSQC spectrum of TM0026 in DH_6PC was partially assigned using the data reported by Kroncke and Columbus [14]. Tryptophan indole resonances are included in the top inset. Partial assignments were mapped onto the low-resolution structure of TM0026 and rendered as spheres (middle inset).

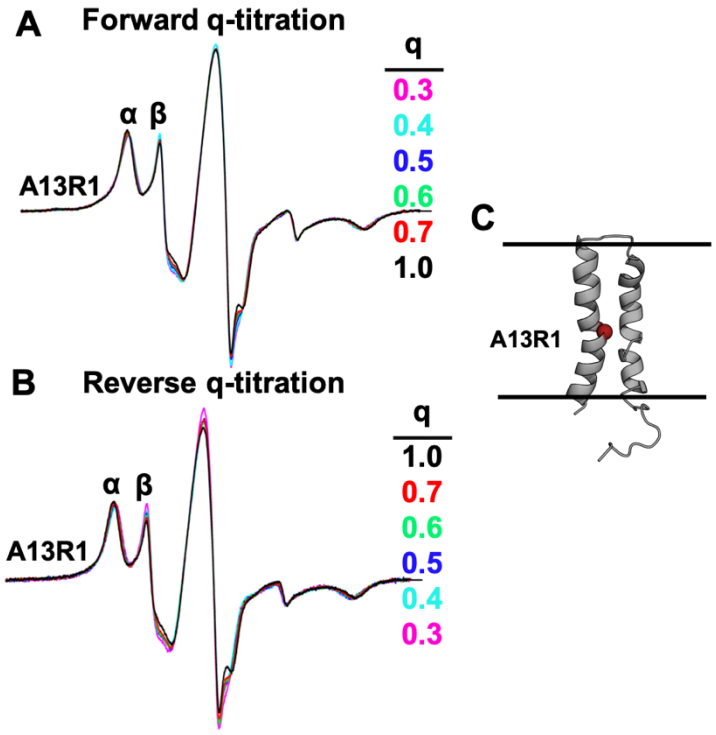


Figure 4.11: CW EPR spectra of TM0026 A13R1 during q-titration experiments. (A) CW EPR spectra from forward q-titration of TM0026 A13R1 prepared in q = 0.3 DMPC/DH_6PC bicelles. (B) CW EPR spectra from reverse q-titration of TM0026 A13R1 prepared in q = 1.0 DMPC/DH_6PC bicelles. Spectra are colored according to bicelle q-value, and spectral components are labeled α and β. (C) TM0026 A13R1 is rendered as a sphere in the low-resolution structure.

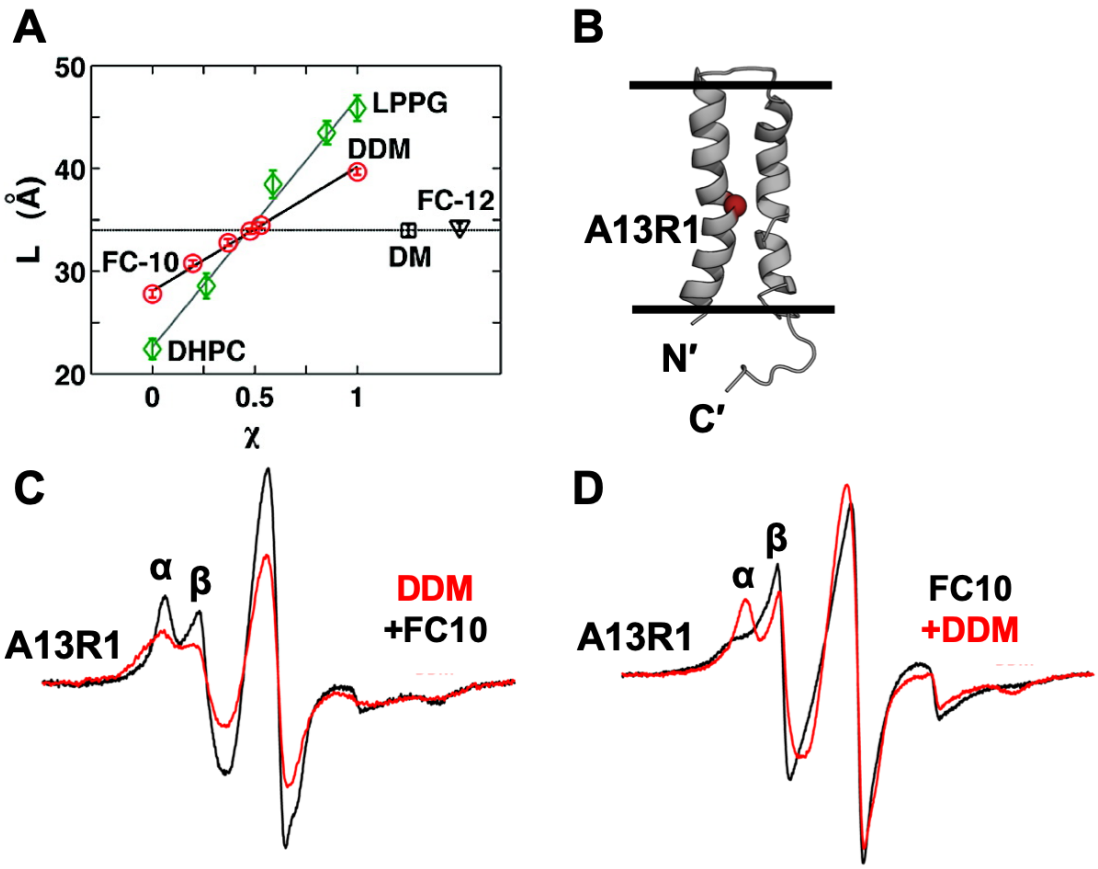


Figure 4.12: Detergent "rescue" experiments with TM0026 using CW EPR. (A) *Reprinted from [12] with permission*. Detergents FC10/DDM and DHPC/LPPG can be mixed at molar ratios that match the thickness, L, of detergents that stabilize one conformer of TM0026 (DM and FC12). (B) TM0026 A13R1 is rendered as a sphere in the low-resolution structure. (C) CW EPR spectra from addition of FC10 (black trace) to TM0026 A13R1 prepared in DDM (red trace). (D) CW EPR spectra from addition of DDM (red trace) to TM0026 A13R1 prepared in FC10 (black trace). Spectral components are labeled α and β.

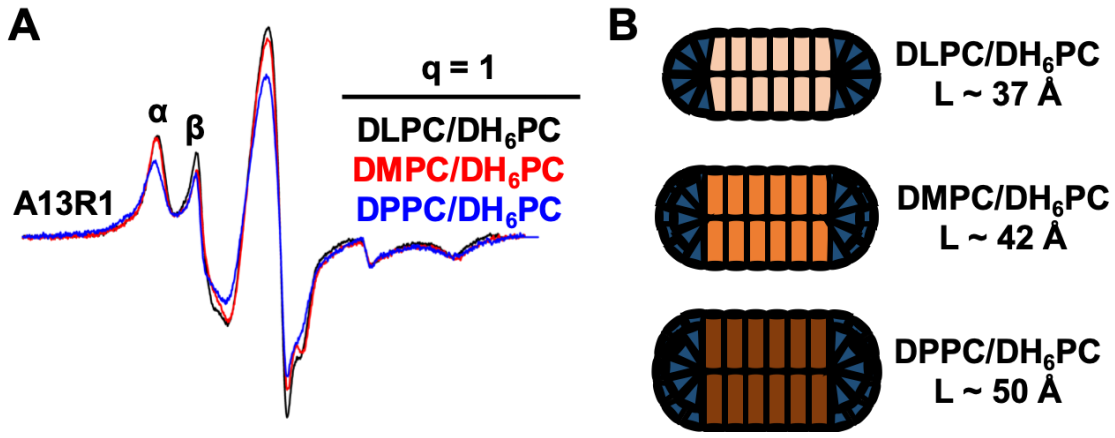


Figure 4.13: CW EPR spectra of TM0026 A13R1 in PC bicelles with q = 1.0. (A) CW EPR spectra of TM0026 A13R1 prepared in DLPC/DH_6PC (black trace), DMPC/DH_6PC (red trace), and DPPC/DH_6PC bicelles (blue trace) when q = 1.0. Spectra are colored by bicelle condition. Spectral components are labeled α and β. (B) L parameters for the empty bicelles used in panel A when q = 1.0 according to results in Chapter 3. DLPC (light orange), DMPC (orange), and DPPC (dark orange) molecules are depicted as cylinders, and DH_6PC (dark blue) is represented as a cone.

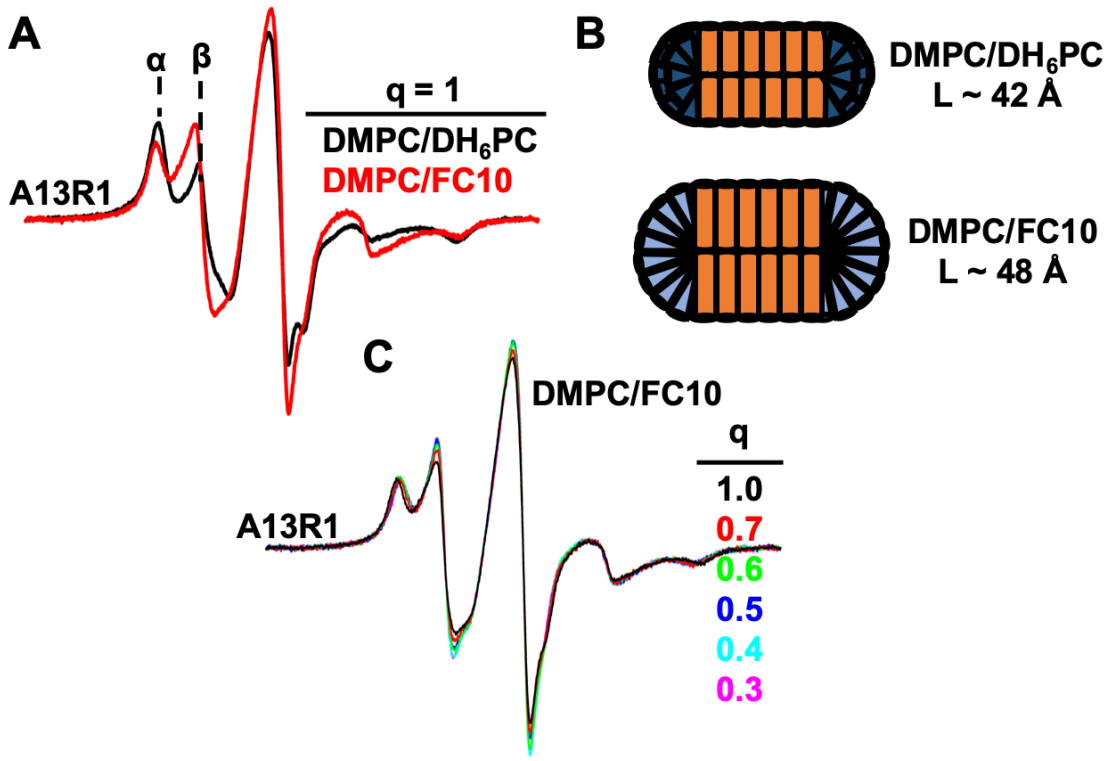


Figure 4.14: CW EPR spectra of A13R1 in DMPC/DH$_6$PC and DMPC/FC10 bicelles. (A) CW EPR spectra of TM0026 A13R1 prepared in DMPC/DH$_6$PC (black trace) and DMPC/FC10 (red trace) when q = 1.0. Spectral components in the DMPC/DH$_6$PC condition are labeled α and β. (B) L parameters for the empty bicelles used in panel A when q = 1.0 according to results in Chapter 3. DMPC molecules are depicted as cylinders, and DH$_6$PC (dark blue) and FC10 (light blue) molecules are represented as cones. (C) CW EPR spectra from reverse q-titration of TM0026 A13R1 prepared in q = 1.0 DMPC/FC10 bicelles. Spectra are colored according to bicelle q-value.

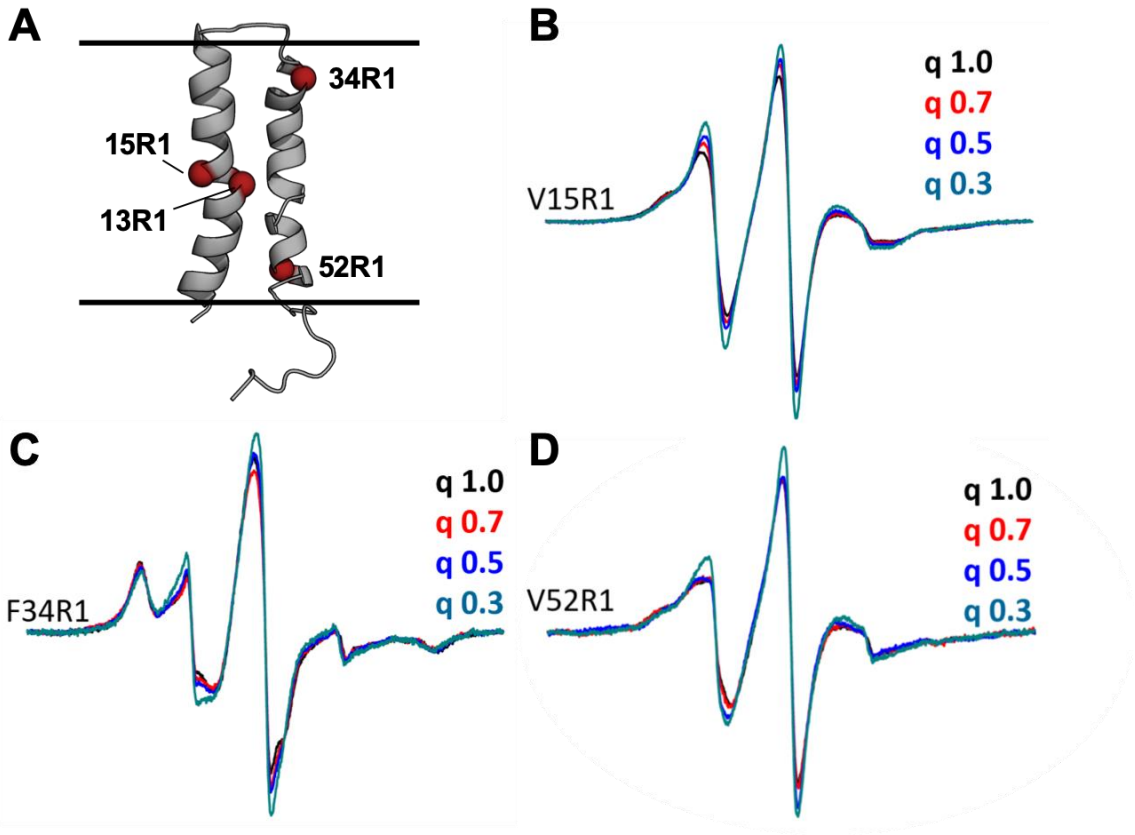


Figure 4.15: Reverse q-titration experiments and CW EPR spectra for TM0026 V15R1, F34R1, and V52R1. (A) V15R1, F34R1, and V52R1 spin labeled sites mapped onto the low-resolution structure of TM0026, rendered as spheres. A13R1 is included for reference. Panels B-D show CW EPR spectra of reverse q-titrations for the corresponding TM0026 mutants prepared in DMPC/DH_6PC, q = 1.0. Spectra are colored according to bicelle q-value.

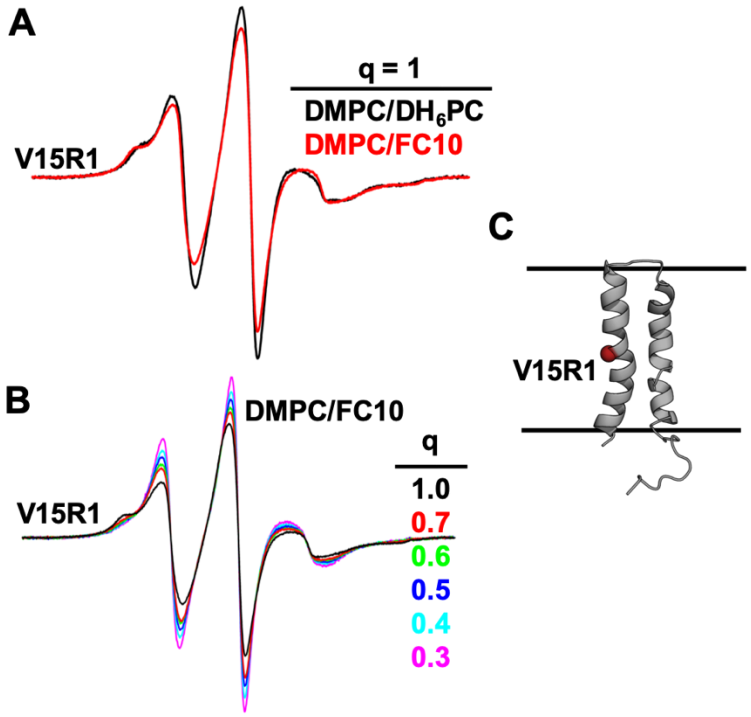


Figure 4.16: CW EPR spectra of V15R1 in DMPC/DH$_6$PC and DMPC/FC10 bicelles. (A) CW EPR spectra of TM0026 V15R1 prepared in DMPC/DH$_6$PC (black trace) and DMPC/FC10 (red trace) when q = 1.0. (B) CW EPR spectra from reverse q-titration of TM0026 V15R1 prepared in q = 1.0 DMPC/FC10 bicelles. Spectra are colored according to bicelle q-value. (C) TM0026 V15R1 is rendered as a sphere in the low-resolution structure.

4.6 Concluding remarks and future directions

Although this work is ongoing, experiments with TM0026 performed to date have set the foundation to test several different hypotheses regarding the observed fold of TM0026 in bicelle environments. Current hypotheses regarding bicelle properties that stabilize the fold of TM0026 are outlined in the following sections, and experiments to evaluate each hypothesis are discussed.

4.6.1 A lipid boundary layer around TM0026 stabilizes the protein fold

To determine if a lipid boundary layer forms around TM0026 (Figure 4.17A), the protein will be ^{15}N, ^{2}H-labeled and initially prepared in DMPC/DH_6PC q = 1.0 bicelles. If possible, separate samples with deuterated lipid/protonated detergent and protonated lipid/deuterated detergent will be prepared. For each sample, a reverse q-titration will be performed, and ^{15}N-edited NOESY spectra will be recorded to identify contacts between TM0026 amide protons and lipid and/or detergent resonances (Figure 4.17B). Similar experiments have been performed with TM0026 in detergent micelles [11, 12]. Isotopically-labeled TM0026 has been prepared for these experiments. Optimization of TM0026 purification and NMR data collection with perdeuterated lipid and detergents are in progress.

4.6.2 The bicelle detergent component affects TM0026 conformational exchange processes

To establish conformational exchange and changes in TM0026 helical packing (Figure 4.18A), double cysteine mutants of TM0026 (Figure 4.18B) have been cloned for

double electron-electron resonance (DEER) experiments. Two pairs of TM0026 cysteine mutants, T3R1/F34R1 and W26R1/V52R1, are designed to probe distance distributions between each pair in detergent micelle and bicelle environments described in this chapter. Both pairs fall within the typical distance range amenable to DEER experiments (~2 – 6 nm; Chapter 2) and are located at the top and bottom of the opposite TM0026 helix. Together, distance distributions from both pairs may be used to assess TM0026 conformational exchange and helical packing in the tested environments. DEER experiments are prioritized for DMPC/DH_6PC and DMPC/FC10 bicelles (q = 1.0) to determine if the A13R1 EPR line shape in DMPC/FC10 bicelles is due to conformational heterogeneity within the protein.

4.6.3 Bicelle morphology may change to accommodate TM0026 hydrophobic mismatch

The final hypothesis for TM0026 stability across a range of bicelle q-values that vary in hydrophobic thickness and lipid segregation properties is that bicelles may overcome hydrophobic mismatch. Recently, Chen *et al.* proposed that the viroporin transmembrane protein p7 causes membrane thinning in isotropic DMPC/DH_6PC (q = 0.6) bicelles (Figure 4.19A) [6]. The authors used solvent and lipophilic paramagnetic relaxation enhancement (PRE) experiments to estimate the depth of the environment surrounding the p7 protein, which they found was noticeably less than the expected depth of a DMPC bilayer. Although the authors propose that distorted lipids produce the thinned membrane, it is just as likely that lipid and detergent molecules may rearrange within the bicelle to accommodate a matched hydrophobic surface area around the protein (Figure 4.19B). The study described by Chen *et al.* utilized deuterated DH_6PC, which prevented

identification of NOEs between DH_6PC and the protein backbone from ^{15}N-edited NOESY experiments performed by the authors [6].

Distinguishing between hypotheses presented in Figure 4.19 requires identification of specific lipid and/or detergent contacts with TM0026 as well as establishing the structure of the protein-bicelle complex in different DMPC/DH_6PC conditions. Lipid/detergent contacts may be identified using the approach described in section 4.6.1. The ideal method for determining the morphology of protein-bicelle complexes is with small-angle scattering experiments. Until recently, studying amphipathic complexes containing protein was incredibly challenging because scattering from empty (without protein) assemblies and protein-containing complexes must be separated [12, 18]. This major bottleneck in small-angle scattering experiments with membrane proteins was solved by coupling size-exclusion chromatography (SEC) to SAXS, which records the scattering of protein assemblies as they elute from an SEC column [26]. Empty micelles elute later than protein-detergent complexes, allowing separation of the scattering between empty and protein-containing assemblies [27, 28]. SEC-SAXS is now available for general users at several national laboratories around the world, including the Advanced Photon Source (APS) at Argonne National Laboratory. APS beamlines also offer in-line SEC coupled to multi-angle light scattering, dynamic light scattering, and refractive index detectors in addition to the SAXS flow-cell to provide accurate molecular weight measurements. TM0026-detergent and bicelle complexes will be characterized by these methods in the future.

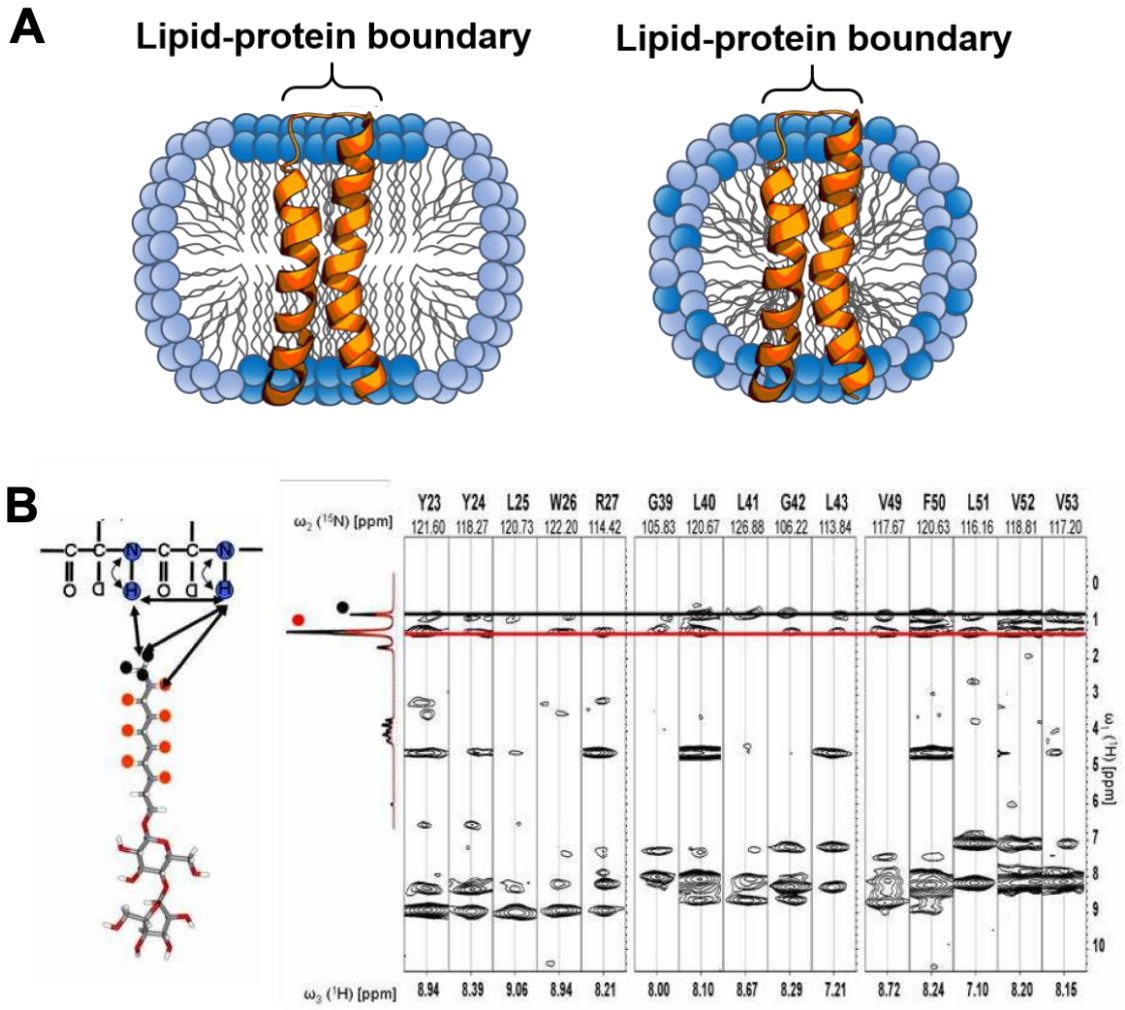


Figure 4.17: Hypothesis for lipid boundary that stabilizes TM0026 fold. (A) Proposed model of a lipid boundary maintained around TM0026 (orange) in conditions that represent segregated and mixed lipid (dark blue)/detergent (light blue). (B) *Adapted from [11] with permission*. Results from ^{15}N-edited NOESY spectra of TM0026 in DM detergent. The chemical structure of DM is shown on the left, with a simplified backbone protein structure above. In the experiment, magnetization is transferred between ^{15}N-^{1}H amide spins (blue circles) to ^{1}H spins within ~ 5 Å of amide protons. Example transfer is shown between backbone amide protons and DM terminal methyl (black circles) and methylene (red

circles) protons. Resulting strips from the ^{15}N-edited NOESY experiment (right) demonstrate contacts between assigned TM0026 backbone resonances and DM methyl/methylene protons. A similar approach may be used to identify lipid and detergent contacts in TM0026 bicelle samples.

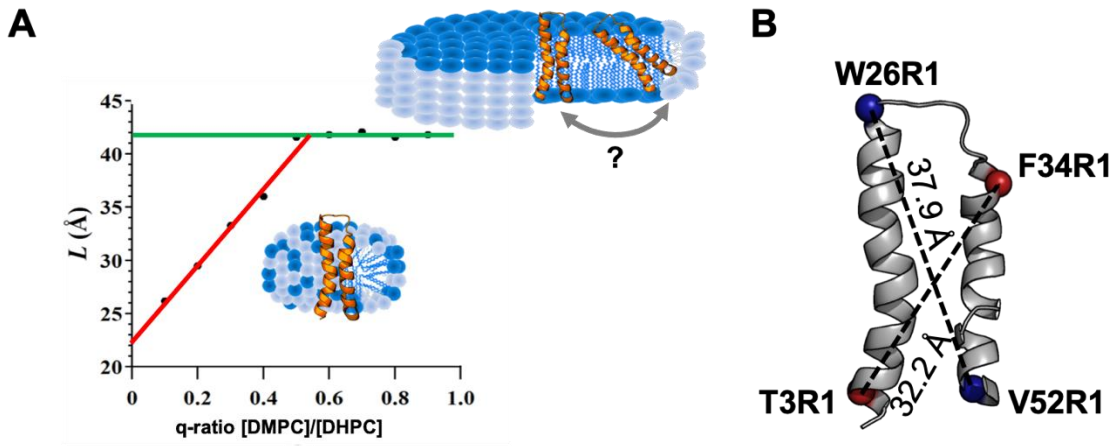


Figure 4.18: Hypothesis for TM0026 conformational exchange in different bicelle compositions. (A) Conformational exchange is possible for TM0026 prepared in bicelles with q-values above the phase transition point for a given bicelle composition. Example is shown for TM0026 in DMPC/DH_6PC based on the plot shown in Figure 4.5 [4]. The red line represents the linear fit for q < 0.5, and the green line represents the approximate L parameter of a pure DMPC bilayer. TM0026 with one conformer in mixed lipid/detergent is shown in the bottom inset, and TM0026 undergoing conformational exchange in segregated lipid/detergent bicelles is shown in the top inset. (B) TM0026 double cysteine mutant pairs designed to evaluate conformational exchange with DEER experiments are shown on the low-resolution structure of TM0026. Distances between Cβ atoms in the mutant pairs W26R1/V52R1 (blue spheres; ~ 38 Å) and T3R1/F34R1 (red spheres; ~ 32 Å) were calculated in PyMol.

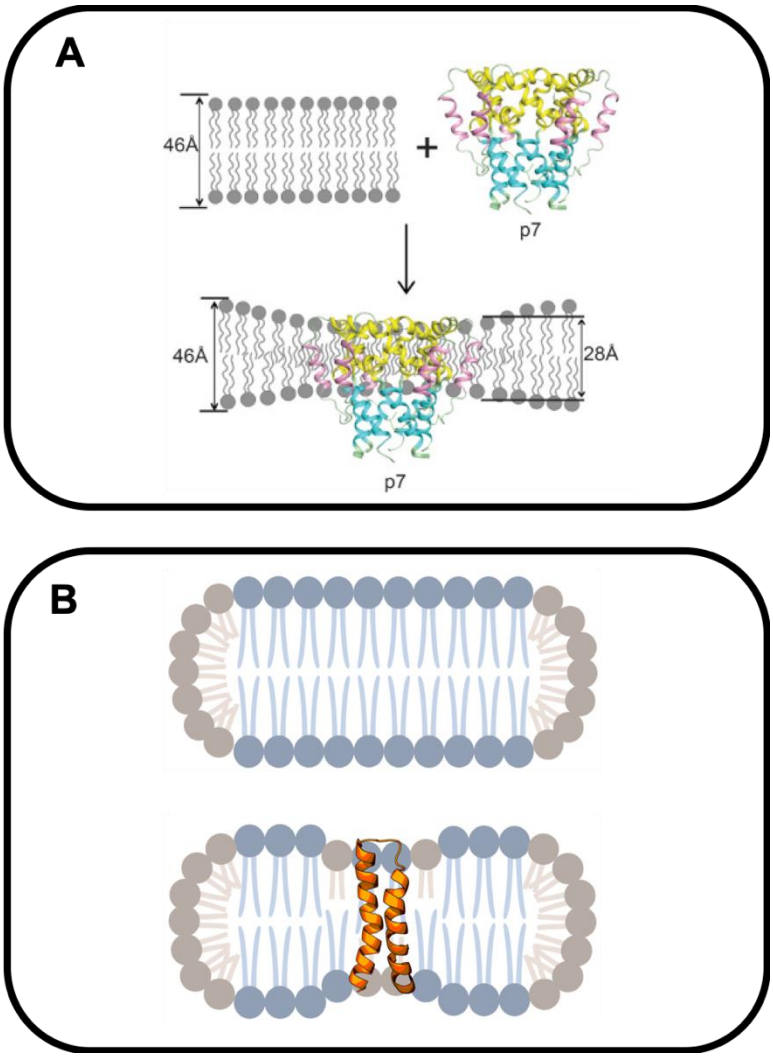


Figure 4.19: Hypothesis for bicelle morphology changes that avoid hydrophobic mismatch. (A) *Reprinted from [6] with permission*. Proposed membrane thinning of DMPC/DH_6PC bicelles induced by the viroporin protein p7. Authors represent the bilayer thickness between the ends of opposing head groups and show distorted lipids to produce the thinning effect. (B) Proposed membrane thinning effect caused by local recruitment of short-chain detergent molecules (tan) to satisfy hydrophobic mismatch of TM0026 (orange) rather than distortion of lipid molecules (blue).

4.7 Acknowledgments

I would like to thank Prof. Brett Kroncke and Dr. Ryan Lo for helpful discussions about TM0026 preparation over the years. They provided me with all of the resources that they could even after being away from the Columbus lab for several years, and I remain grateful to both of them. I would also like to thank Prof. Chuck Sanders at Vanderbilt University for guidance regarding protein purification and spin labeling into bicelles. Finally, although many of the TM0026 NMR experiments run over the years are missing from this dissertation, I am indebted to Dr. Jeff Ellena, Dr. Binyong Liang, Dr. Chuck Schmidt, and Dr. Adam Boulton for helping me troubleshoot any NMR-related issues. This work was supported by the NSF (MCB 1817735 to Prof. Linda Columbus; DDGE 1315231 to Nicole Swope) and the NIH (T32 GM008715 to Nicole Swope).

4.8 References

1. Dürr, U.H.N., M. Gildenberg, and A. Ramamoorthy, *The magic of bicelles lights up membrane protein structure.* Chemical reviews, 2012. **112**(11): p. 6054-6074.
2. Wu, H., et al., *Assessing the size, stability, and utility of isotropically tumbling bicelle systems for structural biology.* Biochimica et biophysica acta, 2010. **1798**(3): p. 482-488.
3. Lu, Z., et al., *Bicelles at low concentrations.* Molecular pharmaceutics, 2012. **9**(4): p. 752-761.
4. Caldwell, T.A., et al., *Low-q Bicelles Are Mixed Micelles.* The Journal of Physical Chemistry Letters, 2018. **9**(15): p. 4469-4473.
5. Lee, D., et al., *Bilayer in Small Bicelles Revealed by Lipid−Protein Interactions Using NMR Spectroscopy.* Journal of the American Chemical Society, 2008. **130**(42): p. 13822-13823.
6. Chen, W., et al., *The Unusual Transmembrane Partition of the Hexameric Channel of the Hepatitis C Virus.* Structure (London, England : 1993), 2018. **26**(4): p. 627-634.e4.
7. Piai, A., et al., *Optimal Bicelle Size q for Solution NMR Studies of the Protein Transmembrane Partition.* Chemistry (Weinheim an der Bergstrasse, Germany), 2017. **23**(6): p. 1361-1367.

8. Morrison, E.A. and K.A. Henzler-Wildman, *Reconstitution of integral membrane proteins into isotropic bicelles with improved sample stability and expanded lipid composition profile.* Biochimica et biophysica acta, 2012. **1818**(3): p. 814-820.
9. Laguerre, A., et al., *From Nanodiscs to Isotropic Bicelles: A Procedure for Solution Nuclear Magnetic Resonance Studies of Detergent-Sensitive Integral Membrane Proteins.* Structure (London, England : 1993), 2016. **24**(10): p. 1830-1841.
10. Lesley, S.A., et al., *Structural genomics of the Thermotoga maritima proteome implemented in a high-throughput structure determination pipeline.* Proceedings of the National Academy of Sciences, 2002. **99**(18): p. 11664.
11. Columbus, L., et al., *Expression, purification, and characterization of Thermotoga maritima membrane proteins for structure determination.* Protein science : a publication of the Protein Society, 2006. **15**(5): p. 961-975.
12. Columbus, L., et al., *Mixing and Matching Detergents for Membrane Protein NMR Structure Determination.* Journal of the American Chemical Society, 2009. **131**(21): p. 7320-7326.
13. Kroncke, B.M. and L. Columbus, *Identification and removal of nitroxide spin label contaminant: Impact on PRE studies of α-helical membrane proteins in detergent.* Protein Science, 2012. **21**(4): p. 589-595.
14. Kroncke, B.M. and L. Columbus, *Backbone 1H, 13C and 15N resonance assignments of the α-helical membrane protein TM0026 from Thermotoga maritima.* Biomolecular NMR Assignments, 2013. **7**(2): p. 203-206.
15. Lo, R.H., et al., *Mapping membrane protein backbone dynamics: a comparison of site-directed spin labeling with NMR 15N-relaxation measurements.* Biophysical journal, 2014. **107**(7): p. 1697-1702.
16. Kroncke, B.M. and L. Columbus, *Solution NMR Structure Determination of Polytopic α-Helical Membrane Proteins: A Guide to Spin Label Paramagnetic Relaxation Enhancement Restraints*, in *Membrane Proteins—Engineering, Purification and Crystallization*, A.K. Shukla, Editor. 2015, Elsevier.
17. Frock, A.D., S.R. Gray, and R.M. Kelly, *Hyperthermophilic Thermotoga species differ with respect to specific carbohydrate transporters and glycoside hydrolases.* Applied and environmental microbiology, 2012. **78**(6): p. 1978-1986.
18. Lipfert, J., et al., *Analysis of small-angle X-ray scattering data of protein-detergent complexes by singular value decomposition.* Journal of Applied Crystallography, 2007. **40**(s1): p. s235-s239.
19. 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. 7692-704.
20. Son, W.S., et al., *'q-Titration' of long-chain and short-chain lipids differentiates between structured and mobile residues of membrane proteins studied in bicelles by solution NMR spectroscopy.* Journal of magnetic resonance (San Diego, Calif. : 1997), 2012. **214**(1): p. 111-118.

21. Kroncke, B.M., P.S. Horanyi, and L. Columbus, *Structural origins of nitroxide side chain dynamics on membrane protein α-helical sites.* Biochemistry, 2010. **49**(47): p. 10045-10060.
22. Mineev, K.S., et al., *Characterization of Small Isotropic Bicelles with Various Compositions.* Langmuir, 2016. **32**(26): p. 6624-6637.
23. Klock, H.E. and S.A. Lesley, *The Polymerase Incomplete Primer Extension (PIPE) Method Applied to High-Throughput Cloning and Site-Directed Mutagenesis*, in *High Throughput Protein Expression and Purification: Methods and Protocols*, S.A. Doyle, Editor. 2009, Humana Press: Totowa, NJ. p. 91-103.
24. Delaglio, F., et al., *NMRPipe: A multidimensional spectral processing system based on UNIX pipes.* Journal of Biomolecular NMR, 1995. **6**(3): p. 277-293.
25. Lee, W., M. Tonelli, and J.L. Markley, *NMRFAM-SPARKY: enhanced software for biomolecular NMR spectroscopy.* Bioinformatics, 2014. **31**(8): p. 1325-1327.
26. Malaby, A.W., et al., *Methods for analysis of size-exclusion chromatography-small-angle X-ray scattering and reconstruction of protein scattering.* Journal of applied crystallography, 2015. **48**(Pt 4): p. 1102-1113.
27. Berthaud, A., et al., *Modeling Detergent Organization around Aquaporin-0 Using Small-Angle X-ray Scattering.* Journal of the American Chemical Society, 2012. **134**(24): p. 10080-10088.
28. Perez, J. and A. Koutsioubas, *Memprot: a program to model the detergent corona around a membrane protein based on SEC-SAXS data.* Acta Crystallographica Section D, 2015. **71**(1): p. 86-93.

Chapter 5: Phosphoglucose isomerase function and mechanism of TM1385 from *Thermotoga maritima*

5.1 Overview

The final chapter of this dissertation describes completed work from the "known structure, unknown function" initiative in the Columbus laboratory and is adapted from the forthcoming manuscript, "TM1385 from *Thermotoga maritima* functions as a phosphoglucose isomerase via *cis*-enediol-based mechanism with active site redundancy." The protein target in this chapter, TM1385 from *Thermotoga maritima* (PDB ID: 2Q8N), was crystallized from a high-throughput structural genomics study performed by the former Joint Center for Structural Genomics (JCSG) [1]. While the JCSG identified crystallization conditions for 23% of the *Thermotoga maritima* proteome, many of the protein functions, including TM1385, remained unconfirmed. TM1385 is putatively characterized as a phosphoglucose isomerase (PGI) based on structural similarity to established PGI crystal structures. PGIs belong to a class of enzymes that catalyze the reversible isomerization of glucose-6-phosphate (G6P) to fructose-6-phosphate (F6P). PGIs are crucial in glycolysis and gluconeogenesis pathways and proposed as serving additional extracellular functions in eukaryotic organisms. Comprehensive analysis of a PGI from a thermophilic bacterium with early evolutionary origins such as *Thermotoga maritima* [2] may provide important clues about PGI evolution in bacterial species. Identification of a thermostable PGI may additionally be advantageous for industrial applications, as bioengineering thermostable enzymes is of broad interest [3].

Multiple sequence and structure alignments between TM1385 and known PGIs support that TM1385 functions as a phosphoglucose isomerase. Kinetic and colorimetric assays combined with ^{1}H nuclear magnetic resonance (NMR) spectroscopy experimentally confirm that TM1385 is a phosphoglucose isomerase, which we have renamed TmPGI. Evidence of solvent exchange in ^{1}H NMR spectra supports that TmPGI isomerization proceeds through a *cis*-enediol-based mechanism. To determine which amino acid residues are critical for TmPGI catalysis, putative active site residues were mutated with alanine and screened for activity. Results support that E281 is most important for TmPGI function, and the presence of either H310 or K422 may be required for catalysis, similar to previous observations from homologous PGIs [4-6]. However, only TmPGI E281A/Q415A and H310A/K422A double mutations abolished activity, suggesting that Q415 may participate in sugar phosphate isomerization upon E281 mutation. Combined, I propose that TmPGI E281 participates directly in the *cis*-enediol intermediate step, and either H310 or K422 may facilitate sugar ring opening and closure.

5.2 Introduction

Phosphoglucose isomerases (PGIs) represent a class of enzymes responsible for catalyzing the reversible isomerization between glucose-6-phosphate (G6P) and fructose-6-phosphate (F6P) (Figure 5.1). This reaction is important *in vivo* as a step in the glycolysis and gluconeogenisis pathways, enabling sugar metabolism in cells [7]. Eukaryotic PGI has additional functions when secreted outside of the cell, functioning as a "moonlighting protein." For example in rabbits, PGI, neuroleukin, autocrine motility factor, and differentiation and maturation mediator are all the same protein, albeit with different names depending on its function, pathway, and location inside or outside of the cytoplasm [8]. In humans, the medical importance of PGI may be observed in individuals who develop nonspherocytic hemolytic anemia, a genetic disorder in which PGI mutations cause either destabilization of the functional protein dimer and/or active site [8], resulting in red blood cells that prematurely undergo hemolysis [9]. Serum PGI elevation may also indicate metastasis in cancer patients [10]. An understanding of the enzymatic function of PGI is therefore important to the study of various biological and clinical applications of this protein.

To date, there are over fifty PGI crystal structures reported in the Protein Data Bank (PDB). While most PGIs with available crystal structures have either bound substrates or *in vitro* kinetic assays to support their functions, protein TM1385 (PDB ID: 2Q8N) from *Thermotoga maritima* represents a proposed PGI based solely on its apo structure. TM1385 was crystallized as part of a JCSG initiative to crystallize the *T. maritima* proteome [1], but its putative designation as a PGI and functional mechanism had yet to be confirmed

biochemically. Characterization of a thermostable PGI from a deeply-branched organism such as *T. maritima* [2] may provide insight into PGI evolution, as the TM1385 sequence is homologous to PGIs spanning eukaryotic, bacterial, and archaeal species (Figure 5.6). Interestingly, PGI derived from several thermophilic bacterial and archaeal species possess bifunctionality and also serve as phosphomannose isomerases (PMIs) [11]. It is possible that TM1385 may similarly exhibit PGI/PMI bifunctionality. Thermostable PGIs are additionally advantageous for engineering enzymes suitable for use in bioreactors [3], underscoring the technological importance of understanding *T. maritima* PGI.

Mechanistic studies of PGI from human, rabbit, and *Pyrococcus furiosus* have been published [6, 12, 13]. While a direct hydride shift PGI mechanism (Figure 5.2) has been proposed and supported by structural analysis [14], functional assays and co-crystal structures with PGI inhibitors provide strong evidence for an acid-base catalysis mechanism proceeding through a substrate *cis*-enediol intermediate for the forward and reverse isomerization reaction between G6P and F6P (Figure 5.3) [4, 13, 15]. Proton donation by a PGI residue is thought to be responsible for acid-catalysis that results in the opening and closing of the glucose or fructose ring. A PGI residue is similarly thought to function as a catalytic base to form the *cis*-enediol intermediate [4-6, 16].

The PGI function and proposed *cis*-enediol mechanism for protein TM1385 from *Thermotoga maritima* (TmPGI) reported in this study is based on kinetic, colorimetric, and ^{1}H NMR assays. To determine which TmPGI residues are critical for catalysis, alanine mutations were introduced at conserved active site residues E281, H310, Q415, and K422. E281 is important for TmPGI function and may be the major residue catalyzing formation

of the *cis*-enediol intermediate, supporting previous studies with TmPGI homologs [4-6]. Activity assays additionally show that H310, Q415, and K422 mutations reduce function and play a role in optimal activity. Across functional studies with TmPGI mutants, only E281A/Q415A and H310A/K422A double mutations were consistently inactive. Based on mutations of these neighboring residue pairs, both faces of the TmPGI active site play direct roles in catalysis. H310 or K422 is thus proposed to open the sugar ring, with Q415 potentially able to form the *cis*-enediol intermediate in the absence of E281. Compared to PGIs found in other bacterial thermophiles, TmPGI functions solely as a PGI with no dual PMI activity and is the first non-metal-bound thermophilic PGI with functional evidence of *cis*-enediol-based isomerization.

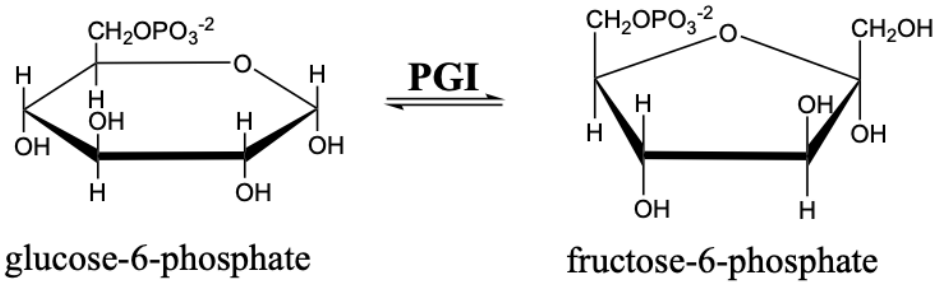


Figure 5.1: Phosphoglucose isomerases (PGIs) catalyze glucose-6-phosphate and fructose-6-phosphate isomerization. Glucose-6-phosphate (G6P) undergoes isomerization to fructose-6-phosphate (F6P) in a reversible process.

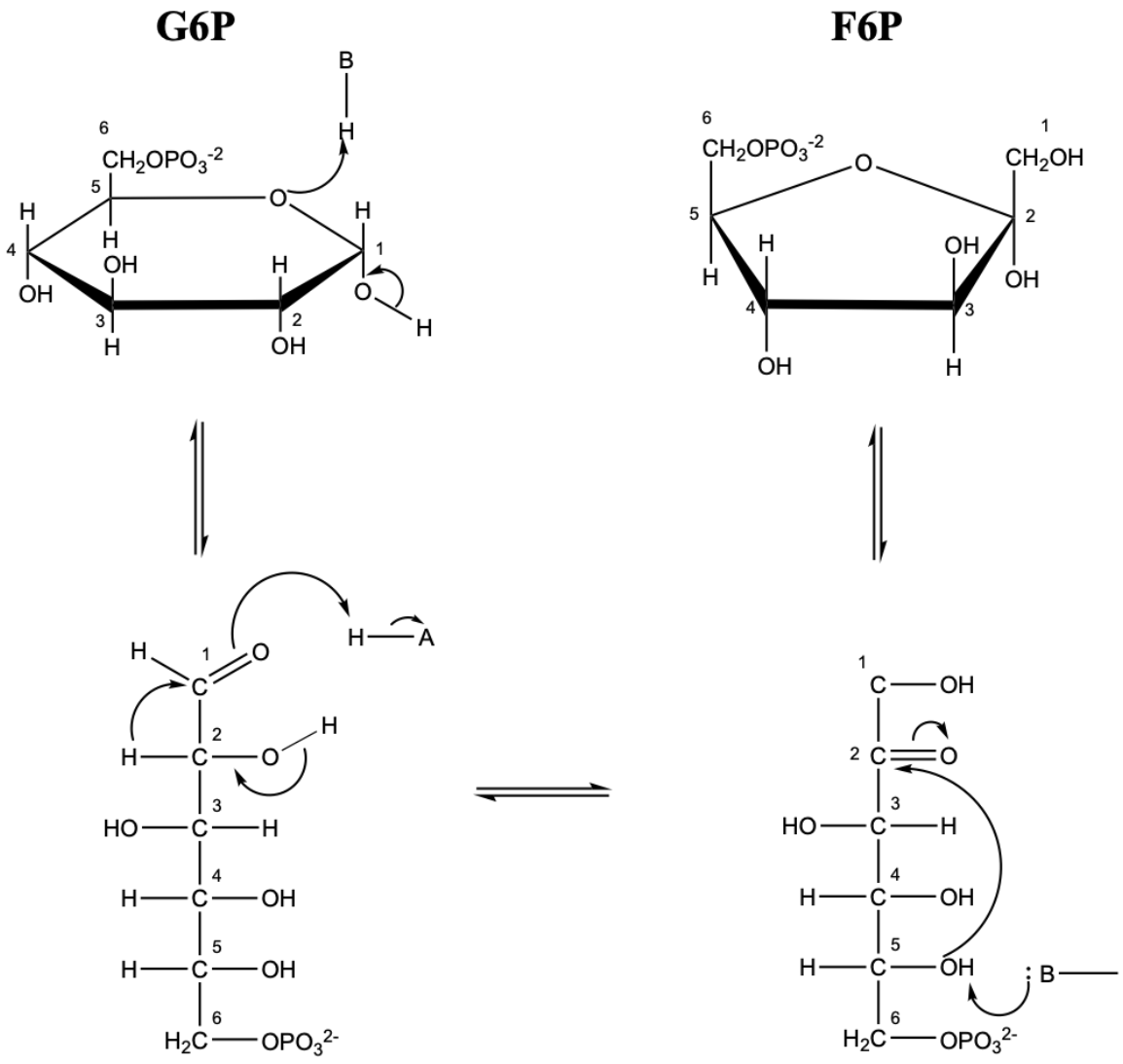


Figure 5.2: PGI-catalyzed isomerization of G6P to F6P via a direct hydride shift. G6P to F6P isomerization catalyzed by a PGI using a direct hydride shift mechanism. Electron movement is depicted as curved arrows. PGI residues undergoing acid and base catalysis are labeled as "A" and "B," respectively.

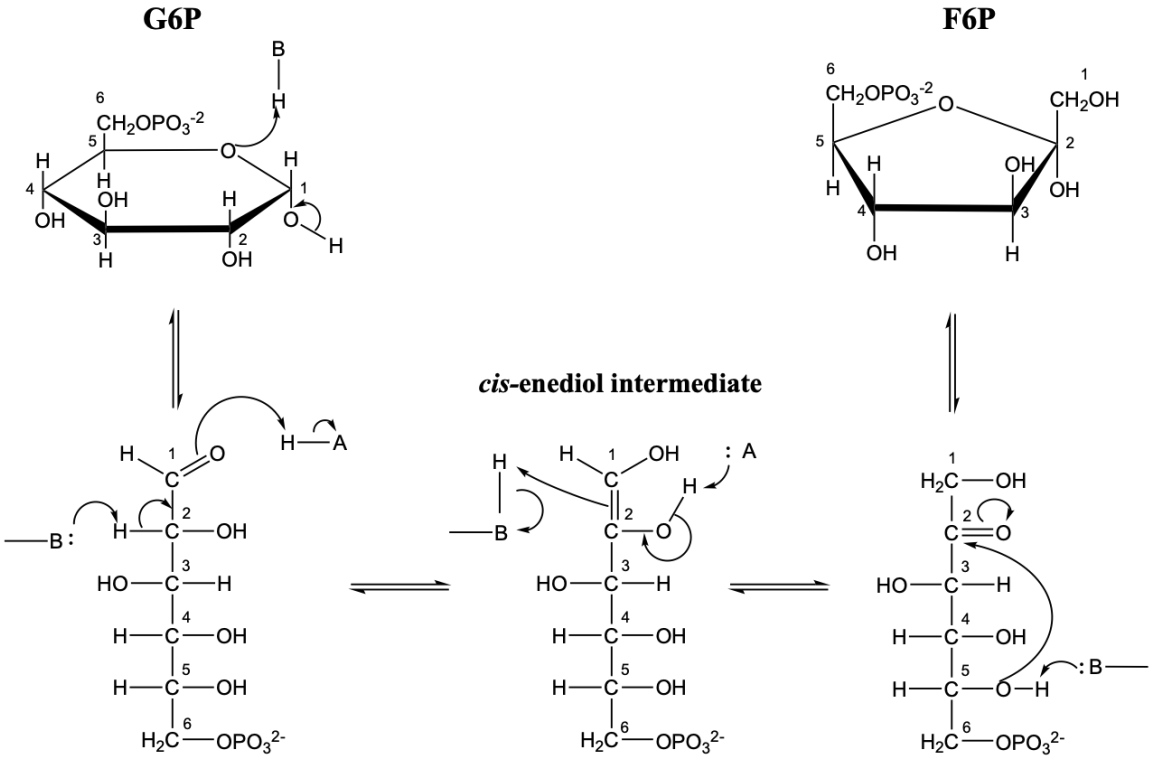


Figure 5.3: PGI-catalyzed isomerization of G6P to F6P via a *cis*-enediol intermediate. G6P to F6P isomerization catalyzed by PGI with a *cis*-enediol-based mechanism. Electron movement is depicted as curved arrows. PGI residues undergoing acid and base catalysis are labeled as "A" and "B," respectively.

5.3 Materials and methods

5.3.1 TM1385 structure and multiple sequence alignments

Global structure alignments of the TM1385 apo dimer (PDB ID: 2Q8N) and twenty classified PGIs were performed with Chimera [17] (Table 5.1). Selected PGIs were chosen based on availability of crystal structures in the PDB. *Pyrobaculum aerophilum* PGI (PDB ID: 1X9I [18]), which possesses dual phosphomannose isomerase (PMI) functionality (PaPGI/PMI) [11], was selected for final rendering due to similarity to the TM1385 structure, co-crystallized linear G6P in the active site, and thermophilic species origin. To identify conserved residues within the putative PGI active site, a multiple sequence alignment was generated with Expresso [19] and ESPript [20] using amino acid sequences from aligned structures.

5.3.2 TM1385 mutagenesis, expression, and purification

The pMH1 vector encoding the recombinant TM1385 protein with a N-terminal 6x-His fusion tag was obtained from the JCSG. To generate TM1385 with alanine point mutations, PCR using the Polymerase Incomplete Primer Extension (PIPE) method [21] or QuikChange Lightning Site-Directed Mutagenesis kit (Agilent Technologies) was performed with a Bio-Rad C1000 Thermal Cycler. After PCR, *E. coli* Top 10 cells were transformed with the mutagenized plasmid. Supercoiled mutagenized DNA was purified using the QiaPrep Spin Miniprep Kit (Qiagen). Protein sequences were confirmed by Sanger Sequencing (GENEWIZ) with pBAD forward and backward primers.

For protein expression, TM1385 wild-type and mutant plasmids were transformed into HK100 *E. coli* cells. One- liter cultures in LB media supplemented with 100 µg/mL ampicillin were grown in shake flasks at 37°C, 225 rpm until OD_{600} reached 0.8 – 1.0. Protein expression was induced with 0.02% arabinose for four hours at 37°C, 225 rpm. Cells were harvested by centrifugation at 4000 rpm, 4°C for 20 minutes. For lysis, cells were resuspended in cell lysis buffer (50 mM Tris pH 8.0 and 150 mM NaCl) containing EDTA-free protease inhibitor cocktail (Roche) and passed twice through a Nano DeBEE high pressure homogenizer (BEE International) at 25,000 psi. The lysate containing the soluble recombinant protein was loaded onto 1 mL bed volume of Chelating Sepharose Fast Flow resin (GE Life Sciences) charged with 100 mM $CoCl_2$ solution. The resin was washed with 20 column volumes (CV) of wash buffer (20 mM phosphate buffer, pH 7.8, 150 mM NaCl and 20 mM imidazole). Protein bound to the resin was eluted with 10 CV of elution buffer (wash buffer supplemented to 600 mM imidazole). Purification performance was monitored with SDS-PAGE analysis. The purified protein from the elution fraction was dialyzed twice at room temperature for 4 hours each against 4 L of Buffer A (50 mM Tris, 150 mM NaCl, pH 7.0). Protein concentration was determined by absorbance at 280 nm. For experiments using high TM1385 concentrations, dialyzed samples were concentrated with Amicon filtration devices using a 30,000 Da molecular weight cutoff (Millipore).

5.3.3 Kinetic assay using coupled Glucose-6-Phosphate dehydrogenase

The catalytic isomerization activity of TM1385 was determined spectrophotometrically by NADPH absorbance at 340 nm. The coupled enzyme assay to

measure TM1385 activity was adapted from Sigma-Aldrich using glucose-6-phosphate dehydrogenase (G6PDH from *Saccharomyces cerevisiae*, Sigma-Aldrich) [22] and the substrate fructose-6-phosphate (Sigma-Aldrich). Assays were performed in Buffer A at room temperature in reactions containing 60 nM TM1385, 4.7 U/mL G6PDH, 10 mM $MgCl_2$, and 1 mM $NADP^+$ with varying F6P concentrations (1-25 mM for wild type, 25-1500 mM for mutants). Coupled assay reactions were carried out in 96-well plates using a 200 µL reaction volume. Absorbance readings were recorded at 340 nm for 30 seconds with the Epoch microplate spectrophotometer (BioTek) to quantify NADPH formation upon G6PDH reaction with G6P produced by TM1385. The change in NADPH concentration was determined by dividing the change in absorbance at 340 nm by the extinction coefficient of NADPH (6,220 $M^{-1}cm^{-1}$) and path length within the microplate (0.5 cm) to obtain TM1385 catalysis rates. Mutant TM1385 kinetic assays required high concentrations of commercially-available F6P to observe activity, resulting in G6P impurities at a concentration sufficient for G6PDH catalysis. Therefore, controls were run with excess $NADP^+$, G6PDH, 10 mM $MgCl_2$, and 1500 mM F6P, the maximum F6P concentration used in mutant kinetic studies, prior to mutant TM1385 addition. These studies resulted in plate reader absorbance values beyond the linear range of the instrument. Thus, F6P reactions without TM1385 for each F6P concentration were used in background subtraction for determining reaction rates. Michaelis-Menten kinetic parameters were generated based on rate of NADPH formation (V) versus F6P concentration ([F6P]) and fit to the following equation in Origin Pro software:

$$V = \frac{V_{max}[F6P]}{K_M + [F6P]}$$

Equation 5.1

Phosphogluconate inhibition assays were performed using the G6PDH-coupled assay conditions described for F6P substrate. Michaelis-Menten curves were generated from reacting 1.0 mM 6-phosphogluconic acid (6P-gluconate) with 60 nM of wild-type TM1385. To exclude the possibility of 6P-gluconate acid acting primarily on the G6PDH coupling enzyme, assays were conducted with the G6PDH-catalyzed conversion of G6P to 6-phosphogluconolactone using 0.6 U of G6PDH in the absence of TM1385. The inhibition constant (K_i) of 6P-gluconate was determined from the concentration of 6P-gluconate (I), apparent K_M with 6P-gluconate, and K_M (Equation 5.1) according to the equation for a competitive inhibitor:

$$K_i = \frac{[I]}{\frac{K_{M,apparent}}{K_M} - 1}$$

Equation 5.2

5.3.4 NMR spectroscopy

All NMR experiments were performed on a Bruker 600 MHz spectrometer at 25°C using 500 µL sample volumes. Samples containing 60 nM TM1385, 10 mM substrate (G6P or F6P), and Buffer A in 90% H_2O/10% D_2O and 15% H_2O/85% D_2O were prepared and incubated for 2 hours at 60 °C. Incubation at 60 °C allowed efficient TM1385 equilibration of substrate, likely owing to the enzyme's native thermophilic environment. One-

dimensional ^{1}H NMR spectra were recorded using water presaturation with composite pulses (*zgcppr* pulse sequence) and 16 transients. The NMR spectra of substrate samples without TM1385 were used to assign the proton resonances of G6P and F6P and serve as negative controls (Figure 5.4 and Figure 5.5).

5.3.5 TM1385 computational docking with linear G6P substrate

The SwissDock server [23] was utilized to dock linear G6P ligand to the TM1385 structure. All non-standard ligands (nonaethylene glycol, H_2O, SO_4^{2-}, Cl^-) as well as chain C, which does not contribute to the active site, were removed from the PDB file prior to docking analysis. Location parameters were unchanged to prevent biased docking to the putative active site. Docking type was set to accurate, a setting allowing for the most accurate docking results at the expense of speed. Flexibility of ligand side chains was allowed up to 3.0 Å. From the SwissDock output, resulting ligand poses were ranked according to the lowest Gibbs free energy, which allows for a non-biased ranking system of ligand positions and sites that G6P is most likely to be interacting with TM1385. The 15 docked G6P poses with lowest Gibbs free energy scores were further analyzed to establish presence in the active site and relative orientation of the ligand phosphate group. Root-mean-square deviation (RMSD) calculations were performed between the selected 15 ligand poses relative to the pose with the lowest Gibbs free energy. To compare docking results with available PGI co-crystal structures containing linear G6P, TM1385 with docked G6P was structurally aligned with PaPGI/PMI. RMSD values were then calculated between the SwissDock poses and aligned G6P from the PaPGI/PMI co-crystal structure.

5.3.6 Colorimetric (Seliwanoff) assay

End-point assays to determine F6P and G6P substrate equilibria from TM1385 catalysis were performed in 96-well plates at 100 µL volumes. Reactions were tested with the low concentration of TM1385 used in G6PDH-coupled assays (60 nM) or high TM1385 concentration (2.3 µM), 10 mM substrate (F6P, G6P, or mannose-6-phosphate [M6P]) and Buffer A. Equilibration was achieved by incubating reactions for 2 hours at 60 °C. Wells without TmPGI but known F6P concentrations (0, 2, 4, 6, 8, and 10 mM) were included on the same plate to generate a standard curve. 10 mM G6P and M6P controls in Buffer A ensured that the presence of aldose sugars did not interfere with quantification of F6P. After the incubation period, 100 µL of Seliwanoff's reagent (4 N HCl, 0.1% (w/v) resorcinol) was added to samples. A microplate seal (VWR) was applied, and the plate was placed on glass supports in a 100°C convection oven for 20 minutes. Upon incubation at 100°C, a red product forms with reaction of F6P and Seliwanoff's reagent [24]. Following the incubation period, the microplate seal was removed, and absorbances were measured at 480 nm using the iMark microplate absorbance reader (BioRad). F6P concentrations in TmPGI reactions were determined by the F6P standard curve, and G6P (or M6P) concentrations were determined from subtracting the final F6P concentration from the initial F6P concentration (10 mM).

Table 5.1: Apo PGI structures deposited in the Protein Data Bank

Species	Domain	PDB ID	Uniprot ID
Thermotoga maritima	bacteria	2Q8N	Q9X1A5
Francisella tularensis	bacteria	3LJK	Q5NFC4
Brucella melitensis	bacteria	4EM6	Q8YF86
Escherichia coli	bacteria	3NBU	P0A6T1
Vibrio cholerae	bacteria	3HJB	Q9KUY4
Staphylococcus aureus	bacteria	3FFL	Q5HHC2
Mycobacterium tuberculosis	bacteria	2WU8	P9WN69
Bacillus anthracis	bacteria	3IFS	Q81K75
Geobacillus stearothermophilus	bacteria	2PGI	P13376
Thermus thermophilus	bacteria	1WIW	Q5SIM3
Pyrobaculum aerophilum	archaea	1TZB	Q8ZWV0
Colias eurytheme	eukarya	4WMJ	G9BZJ0
Trypanosoma cruzi	eukarya	4QFH	Q4E5N1
Toxoplasma gondii	eukarya	3UJH	S8FBB6
Plasmodium falciparum	eukarya	3QKI	Q8ILA4
Trypanosoma brucei	eukarya	2O2D	P13377
Mus musculus	eukarya	1U0E	P06745
Oryctolagus cuniculus	eukarya	1HM5	Q9N1E2
Leishmania mexicana	eukarya	1Q50	P42861
Homo sapiens	eukarya	1JLH	P06744
Sus scrofa	eukarya	1GZD	P08059

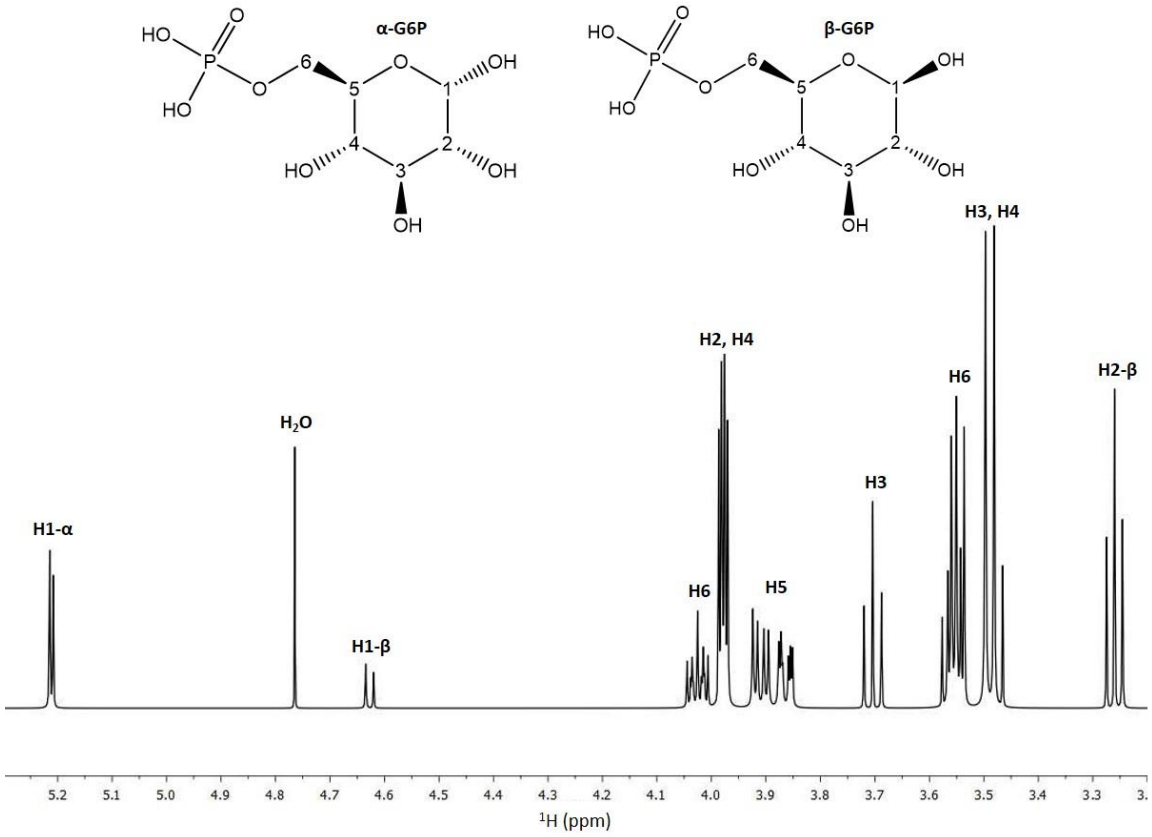


Figure 5.4: Proton resonance assignments of G6P from one-dimensional NMR spectrum. The one-dimensional ^{1}H NMR spectrum of 10 mM G6P in 10% D_2O was recorded to assign distinguishable α-G6P and β-G6P resonances. Chemical shift frequencies are given in ppm. The chemical structures of G6P anomers are shown above the spectrum, and carbon positions are labeled numerically. All proton resonances are labeled according to the carbon number/position.

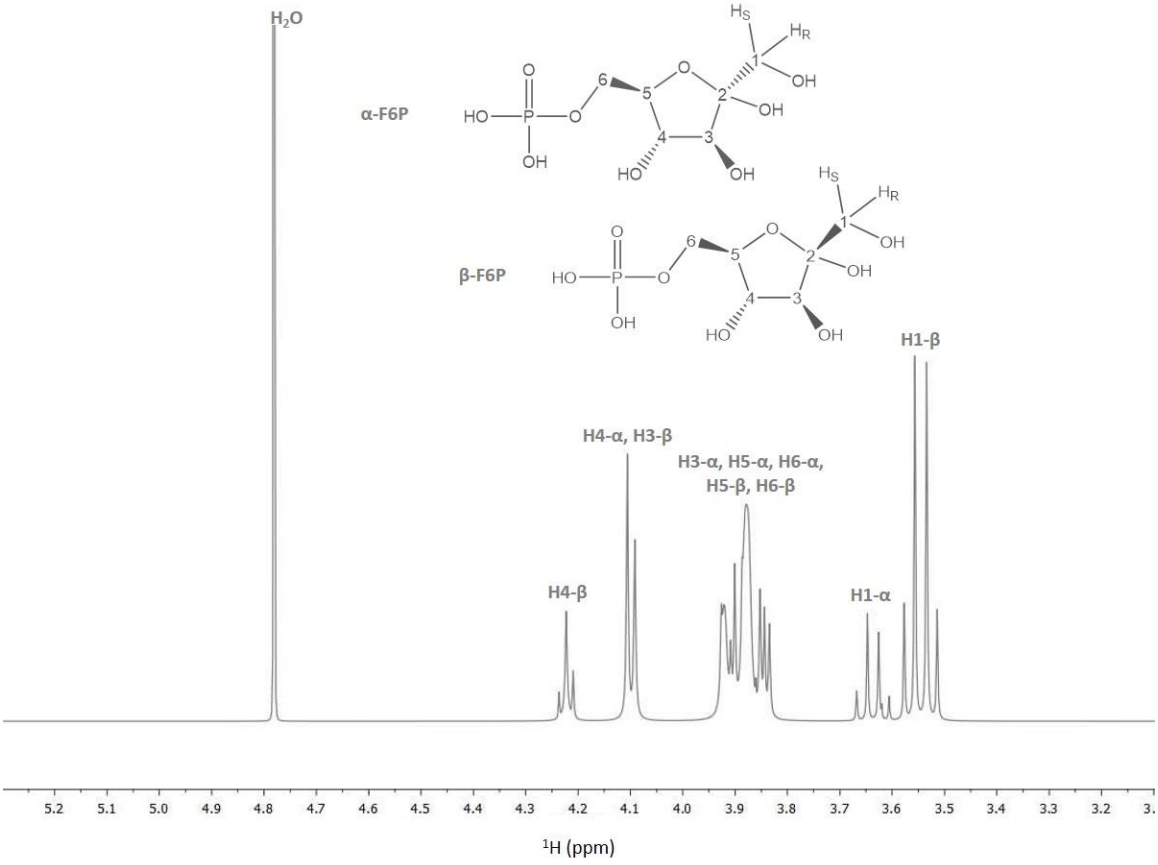


Figure 5.5: Proton resonance assignments of F6P from one-dimensional NMR spectrum. . The one-dimensional ^{1}H NMR spectrum of 10 mM F6P in 10% D_2O was recorded to assign distinguishable α-F6P and β-F6P resonances. Chemical shift frequencies are given in ppm. The chemical structures of F6P anomers are shown above the spectrum, and carbon positions are labeled numerically. All proton resonances are labeled according to the carbon number/position.

5.4 Results and discussion

5.4.1 Structure alignments to homologous PGIs and substrate docking supports TM1385 PGI function

To compare the TM1385 protein structure to classified PGIs, structure alignments were generated using twenty PGI crystal structures spanning bacterial, eukaryotic, and archaeal species. Aligning PGIs with known structure and function allowed for comparison of homologous PGI active site residues to conserved residues within the TM1385 sequence (Figure 5.6). A representative structure alignment of TM1385 with an archaeal PGI homolog with dual phosphomannose isomerase (PMI) functionality from *Pyrobaculum aerophilum* [11] (PaPGI/PMI, PDB ID: 1X9I [18]) is shown in Figure 5.7. The reported PaPGI/PMI crystal structure includes linear G6P substrate and identification of active site residues. Most of the PaPGI/PMI active site residues are highly or strictly conserved across PGI homologs and TM1385. A multiple sequence alignment with TM1385 and known PGIs demonstrates that S77, S136, T141, R198, E281, H310, and K422 are strictly conserved, while position 138 is highly conserved as a serine or threonine. Positions 197 and 415 are glycine and glutamine, respectively, among aligned PGIs except for PaPGI/PMI (Figure 5.6). Furthermore, conserved residues within the TM1385 crystal structure show a remarkable similarity to the PaPGI/PMI active site architecture in Figure 5.7. We thus propose that for TM1385, residues S77, S136, S138, T141, G197, R198, E281, Q415, and K422 in the first subunit and H310 in the second subunit form the putative active site for PGI function. Homologous residues to E281, H310, and K422 have all been proposed as serving direct roles in PGI catalysis, with R198 and Q415 thought to stabilize

the negative charge of E281 and/or *cis*-enediol(ate) intermediate and participate in hydrogen bonding with the substrate, respectively [5, 15, 25]. The hydroxyl groups of serine and threonine residues have previously been reported to stabilize the substrate phosphate group [26]. Combined, the structure and sequence alignments of TM1385 with known PGIs lends convincing support that TM1385 is likely in the phosphoglucose isomerase family.

5.4.2 TM1385 (TmPGI) catalyzes fructose-6-phosphate to glucose-6-phosphate isomerization

To test the proposed PGI function of TM1385, a coupled enzymatic assay with glucose-6-phosphate dehydrogenase (G6PDH) [22] was designed to measure TM1385 activity in the F6P to G6P isomerization direction. Upon F6P isomerization to G6P, G6PDH converts G6P to 6-phosphogluconolactone using $NADP^+$ as a cofactor. Thus, formation of NADPH was measured by spectrophotometry at 340 nm. Increasing F6P concentration resulted in increasing reaction rates, following Michaelis Menten kinetics, with a K_M of 0.31 mM and k_{cat} of 17.4 s^{-1} (Figure 5.8; Table 5.2), confirming that TM1385 is a PGI (TmPGI). The K_M is in line with reported values for *Pyrobaculum aerophilum* (0.30 mM) [11], *Mycobacterium tuberculosis* (0.27 mM) [27], and *Carassius auratus* (0.36 mM) [28], while the k_{cat} is most similar to those reported for *Pyrococcus furiosus* (8.4 s^{-1}) [13] and *Carassius auratus* (18.0 s^{-1}) [28]. Furthermore, addition of the known PGI inhibitor 6P-gluconate [29] decreased TmPGI activity (K_i = 44 µM). Inhibition was not observed for G6PDH activity with G6P substrate.

5.4.3 TmPGI catalysis proceeds through a *cis*-enediol intermediate pathway

Kinetic studies may be used to designate TM1385 as a PGI; however, kinetics alone do not give information on the mechanism of accomplishing substrate isomerization. There are two proposed mechanisms for substrate isomerization: (1) a direct hydride shift (Figure 5.2) [14] or (2) a mechanism involving a *cis*-enediol intermediate (Figure 5.3) [13]. In the direct hydride shift mechanism, a hydride ion is removed from carbon 2 (C2) of G6P and directly donated to carbon 1 (C1) in order to form F6P without exchange with the solvent. In the *cis*-enediol mechanism, a residue in the PGI enzyme active site acts as a base catalyst to remove a proton from C2 of G6P, and the substrate forms a *cis*-enediol intermediate. The proton that is removed by the PGI base catalyst may exchange with protons from the solvent before being donated to carbon 1 to form the product [13]. Proton exchange between TmPGI residues and substrate is demonstrated in Figure 5.9.

In order to distinguish between these two mechanisms for the PGI reaction, NMR experiments were used to determine whether the proton transferred from C2 to C1 of the sugar substrate is able to exchange with protons of the solvent during the course of the reaction [13]. The ^{1}H NMR spectra of the sugar-phosphate substrates after reaction with PGI may be compared when the reaction takes place in a solvent containing mostly H_2O or D_2O. When D_2O solvent is used, deuteron incorporation may occur after formation of the *cis*-enediol intermediate, when the proton removed by a TmPGI residue exchanges with a deuteron of the solvent (as D_2O). A deuteron is then incorporated into the substrate to complete the isomerization. If deuterons are incorporated into the sugar during the PGI reaction in high D_2O solvent (at either of the H-1 positions in F6P or at the H-1 or H-2

position in G6P), a decrease in intensity or disappearance of the resonances corresponding to these protons in a ^{1}H NMR spectrum will be observed (Figure 5.10A). Conversely, because the direct hydride shift mechanism does not allow for the transferred hydrogen atom to exchange with the solvent, a decrease in resonance intensity corresponding to proton positions in the substrate that are involved in the reaction will not be observed (Figure 5.10B). Reactions performed with G6P substrate and TmPGI in 10% and 85% D_2O (Figure 5.11) show that the G6P H-1α, H1-β, H2-β, and H3 resonances disappear or significantly decrease, supporting that TmPGI catalysis proceeds through a *cis*-enediol intermediate.

5.4.4 Substrate docking to identify important residues for TmPGI function

As the TmPGI crystal structure does not contain bound substrate to clarify which residues may be important for substrate binding and/or catalysis via a *cis*-enediol mechanism, computational docking between the TmPGI structure and linear G6P substrate was performed via the SwissDock server [23]. The 15 docked G6P poses with the lowest Gibbs free energy ranged from -9.41 to -9.73 kcal/mol and were classified into groups based on active site (A/B) or surface locations (Table 5.3). Active site A is defined as the active site comprised of residues primarily in chain A, while active site B contains residues from chain B, with the exception of H310 in each case. Among the 15 poses with lowest Gibbs free energy, 8 are found in active site B (Figure 5.12), 4 in active site A (Figure 5.13), and 3 on the exterior surface of the protein. RMSD values were calculated for the 15 lowest energy poses relative to the lowest energy pose (Table 5.3). The average RMSD between lowest-energy docked poses was 19.2 ± 18.9 Å due to differences in protein

localization. However, separate RMSD calculations for poses in active site A or B resulted in average RMSDs of 2.4 ± 2.1 Å and 3.5 ± 1.1 Å, respectively.

The most noticeable difference between G6P docking results was relative orientation of the substrate phosphate group. Within the lowest-energy G6P poses found in active sites (12 total), 6 are found with the G6P phosphate group angled toward TmPGI residue R198 (Figure 5.12B). The remaining 6 poses possess a phosphate orientation near TmPGI residue T141 (Figure 5.12C), which is a similar G6P conformation compared to the PaPGI/PMI co-crystal structure (Figure 5.14B). All PGIs co-crystallized with sugar-phosphate substrates to date closely associate the phosphate group with homologous threonine residues, which suggests the substrate orientation in Figure 5.12B is likely a byproduct of the docking protocol. The RMSD between T141-oriented poses in active site B and aligned G6P from the PaPGI/PMI structure was 4.8 ± 0.2 Å.

Although PGI crystal structures with substrates and inhibitors have allowed detailed analysis of active site residues, different structural interpretations have led to the identification of various residues that may participate in the *cis*-enediol reaction mechanism. For example, it has been proposed that substrate ring opening and closing may occur with residues analogous to H310 [4, 5, 16, 30] or K422 [4-6, 15, 30], both of which can act as an acid/base catalyst according to the mechanism in Figure 5.9. Importantly, the positive charge of K422 may stabilize the phosphate group of the substrate and orient the molecule within the active site without playing a direct role in catalysis [4]. H310 has additionally been proposed to act as a base catalyst that forms the *cis*-enediol substrate intermediate [8, 25, 30]. A separate hypothesis for *cis*-enediol formation is that E281

instead acts as the base catalyst [4-6, 31], perhaps forming an acid/base pair with H310 [25, 30] that increases reaction efficiency. H310, E281, and K422 are all strictly conserved across the PGI family, highlighting their potential importance for PGI function. Q415 represents another highly conserved residue adjacent to E281 that could act as a TmPGI base catalyst. However, the glutamine residue would likely be a weaker proton acceptor than the nearby glutamate.

Distances of G6P substrate to E281, H310, Q415, and K422 are all less than 5 Å in the PaPGI/PMI crystal structure, suggesting crucial catalytic and/or stabilization roles for these residues. However, docked G6P in the TmPGI crystal structure contains a slightly different substrate orientation from PaPGI/PMI, resulting in longer distances (~3.5 – 7.0 Å) between proposed catalytic residues and docked substrate. E281 and Q415 remain less than 5 Å to docked G6P, while H310 and K422 are positioned ~5 – 7 Å away. Longer distances observed between TmPGI and docked G6P does not preclude E281, H310, Q415, and K422 from catalytic roles, as conformational changes not captured by docking may occur in the active site during ligand binding and/or catalysis. For example, two induced-fit conformational rearrangement steps are proposed for the rabbit PGI (rPGI) *cis*-enediol mechanism [4]. The first rPGI active site conformational change occurs upon ligand binding. Residues analogous to S136, S138, and T141 shift into the binding pocket to participate in phosphoryl interactions with cyclic F6P. Similarly, the rPGI helix containing homologous H310 experiences both a minor shift toward cyclic F6P and large shifts with complexed linear substrates, supporting the role of H310 in ring-opening. The second major conformational change in the rPGI active site is observed between ring-opening and

isomerization steps, in which the helix containing homologous K422 shifts ~3 Å toward the linearized substrate to reposition the residue homologous to E281. Movement of E281 toward the substrate suggests that E281 is directly involved in catalyzing the isomerization step in the rPGI mechanism. Future studies with TmPGI co-crystallized with cyclic and linear forms of F6P and/or G6P will be needed to confirm similar conformational rearrangements during TmPGI catalysis.

5.4.5 Functional analysis of proposed TmPGI catalytic residues

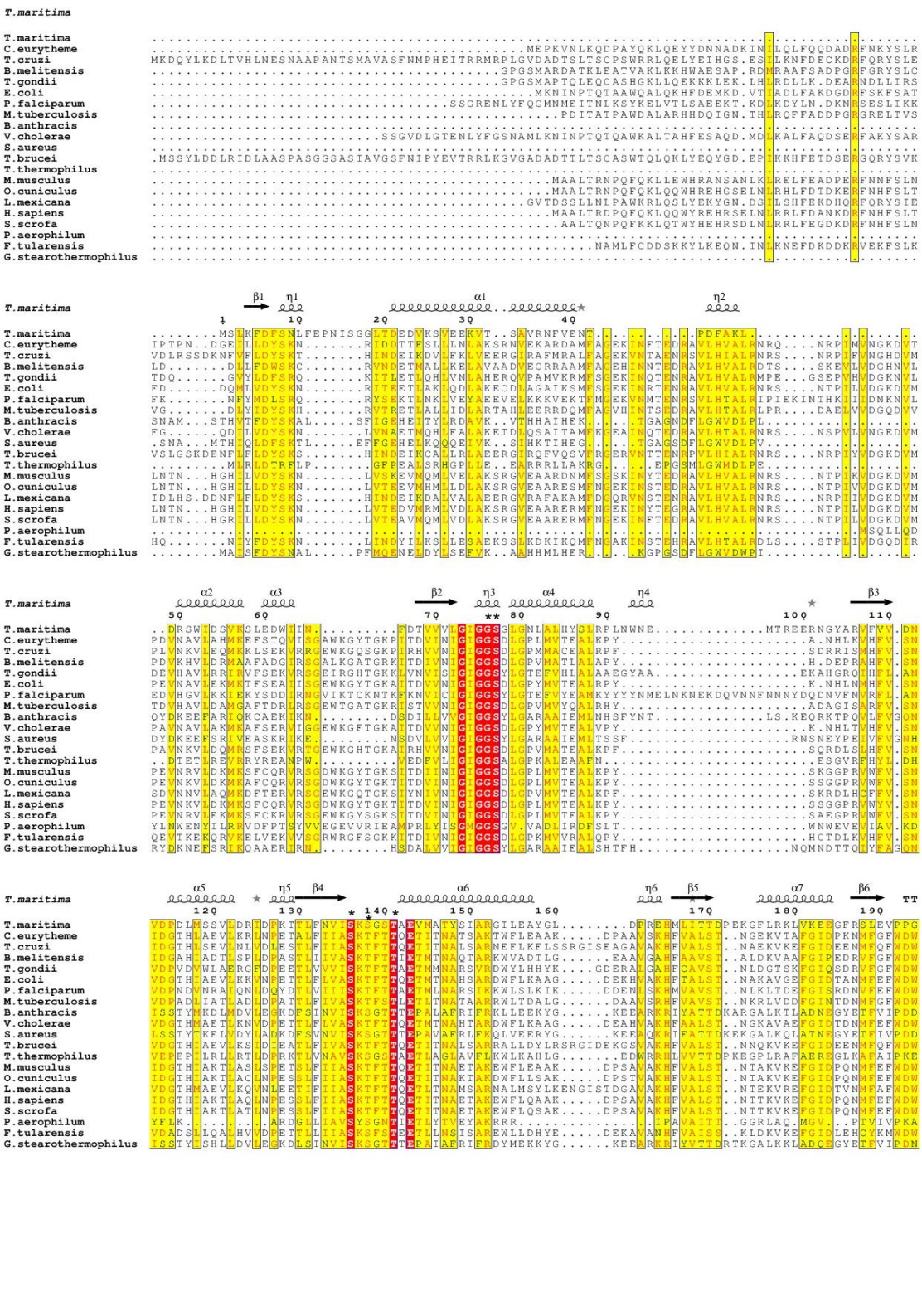
To establish TmPGI residues crucial for G6P/F6P isomerization based on activity, the G6PDH-coupled assay was performed with E281A, H310A, Q415A, and K422A TmPGI mutants. Steady-state kinetic parameters of each mutant were determined from fitting reaction rates to a hyperbolic function and compared with wild-type TmPGI (Table 5.2). An additional E281S mutation was made, as we hypothesized that the hydroxyl group may be sufficient as a base catalyst if the side chain remains near C1 and C2 substrate atoms.

Both E281 mutations abolished TmPGI activity, suggesting that the conserved glutamate residue at this position is required for F6P isomerization (Figure 5.15A). In contrast, H310A, Q415A, and K422A exhibited low catalytic efficiency. Rather than acting as a base catalyst in the isomerization step, the primary role of Q415 is instead proposed as positioning E281 [5] or the substrate [6, 25, 30] within the active site. As either H310 or K422 could open or close the substrate ring (Figure 5.15A and C), it is possible that these residues can compensate for each other in the event of a mutation. In support of H310

and K422 acting in compensatory roles, mutating both residues to alanine (H310A/K422A) resulted in no activity (Figure 5.15D). However, the double mutation, along with the E281A/Q415A double mutation, may destabilize the active site and subsequent protein fold, which would similarly lead to loss of function.

The G6PDH-coupled assay allows TmPGI functional assessment in the reverse (F6P to G6P), but not forward, reaction. Furthermore, we wished to develop a faster approach to determine general functionality of each TmPGI mutant based on final substrate equilibria without relying on a coupling enzyme. Commercially available F6P contains 2% G6P impurity, making TmPGI mutant assay conditions and background subtraction for the G6PDH-coupled reaction an arduous task with the high F6P concentrations required for turnover by mutants. Difficulty controlling for G6P impurities likely resulted in the large errors associated with steady-state kinetic parameters. Instead, the colorimetric Seliwanoff assay commonly used to distinguish ketose and aldose sugars [32] was adapted to measure relative percentages of F6P (ketose sugar) and G6P (aldose sugar) after incubation with TmPGI variants. The Seliwanoff assay may also be used to screen PGIs for dual phosphomannose isomerase (PMI) activity in which M6P (aldose sugar) is converted to F6P. F6P concentrations ranging from 0 – 10 mM were optimal for generating standard curves. As these substrate concentrations are too low to detect TmPGI mutant activity, only wild-type activity was observed for F6P and G6P substrates in initial TmPGI experiments with the Seliwanoff assay. M6P substrate did not form F6P in any assays performed in this study, demonstrating that TmPGI does not possess dual PMI activity.

TmPGI concentrations were increased to amplify the enzymatic reaction rates, which allowed for detecting isomerase activity of negatively-affected TmPGI variants (Table 5.4). Overall, the trends in TmPGI function assayed by the Seliwanoff method generally agreed with the G6PDH-coupling assay. However, both E281A and E281S had activity at high enzyme concentrations, indicating that E281 is important but not essential for TmPGI activity. Activity of E281 mutants detected in the Seliwanoff assay but not the coupled assays suggests that Q415 or another neighboring residue may facilitate TmPGI residual activity in the event of E281 mutations (Figure 5.15B). TmPGI E281 mutants showing activity only at high enzyme concentrations is recapitulated in the proposed acid/base pair mutant E281A/H310A. The partial function of the E281A/H310A mutant supports the presence of compensatory residues in the TmPGI active site. However, TmPGI function was abolished in both E281A/Q415A and H310A/K422A based on both G6PDH-coupled and Seliwanoff assays. Combined, the G6PDH-coupled and Seliwanoff assays indicate that E281 and H310 may be compensated by Q415 and K422, respectively, in the event of mutations (Figure 5.15). Mutation of both residues in either pair leads to loss of TmPGI function.


T.maritima
T.maritima
C.eurytheme
T.cruzi
B.melitensis
T.gondii
E.coli
P.falciparum
M.tuberculosis
B.anthracis
V.cholerae
S.aureus
T.brucei
T.thermophilus
M.musculus
O.cuniculus
L.mexicana
H.sapiens
S.scrofa
P.aerophilum
F.tularensis
G.stearothermophilus

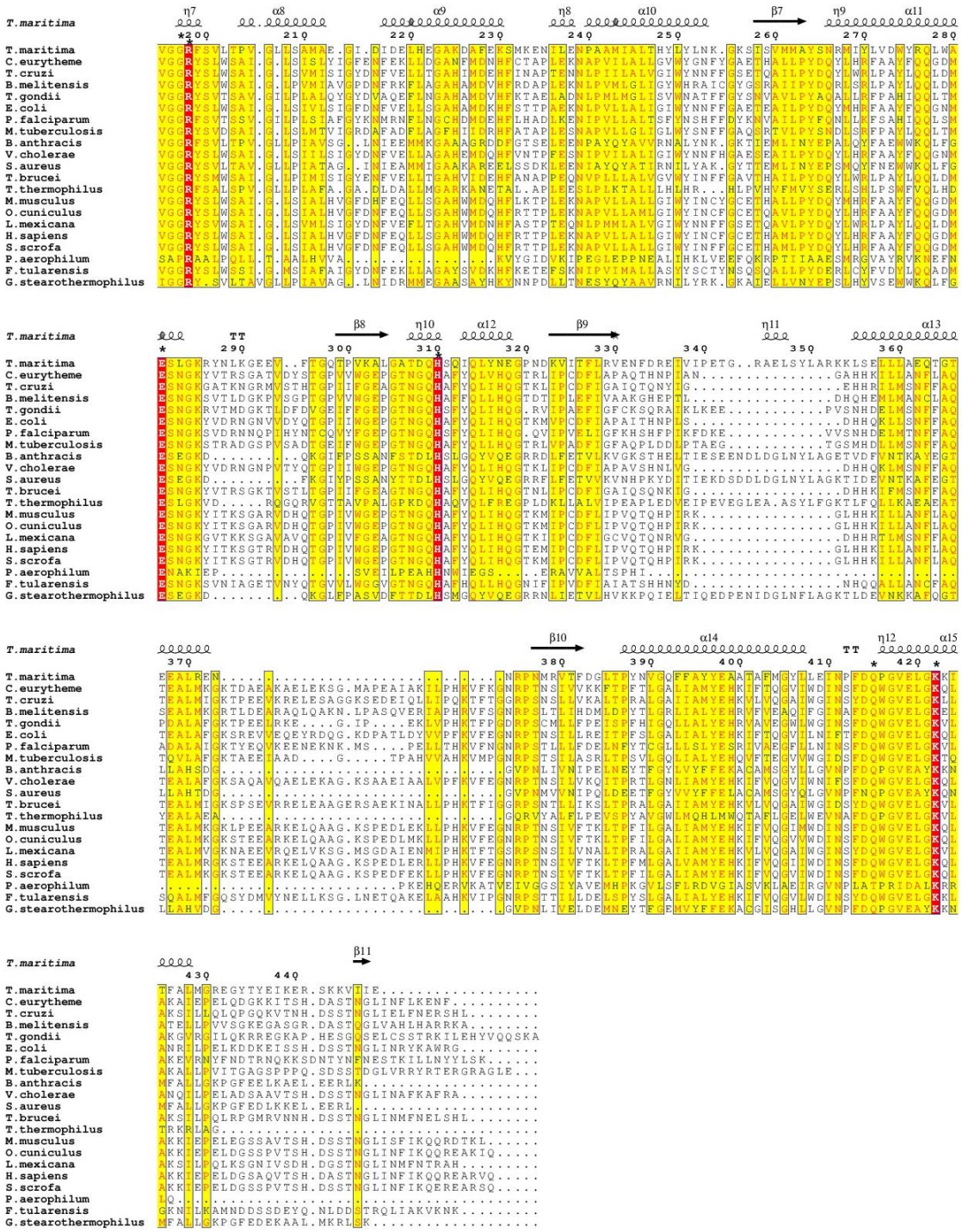


Figure 5.6: Multiple sequence alignment of TM1385 from *T. maritima* with PGIs from eukaryotic, bacterial, and archaeal species(continued from previous page). All PGIs in

this sequence alignment have crystal structures deposited in the PDB (Table 5.1). Expresso [19] was used to generate the multiple sequence alignment and includes additional information based on the three-dimensional structure of TM1385. Strict β-turns are represented by the characters TT, and loops appear for helical structures, represented by η (310-helices) and α (α-helices). Gray stars are shown above residues with alternate conformations in the crystal structure. Expected TM1385 active site residues include black asterisks above their positions in the alignment. Strict identity of residues in the alignment are rendered as red boxes with white characters, while similarities within a group are depicted as red characters. Finally, similarities across groups are framed in blue and filled in with yellow.

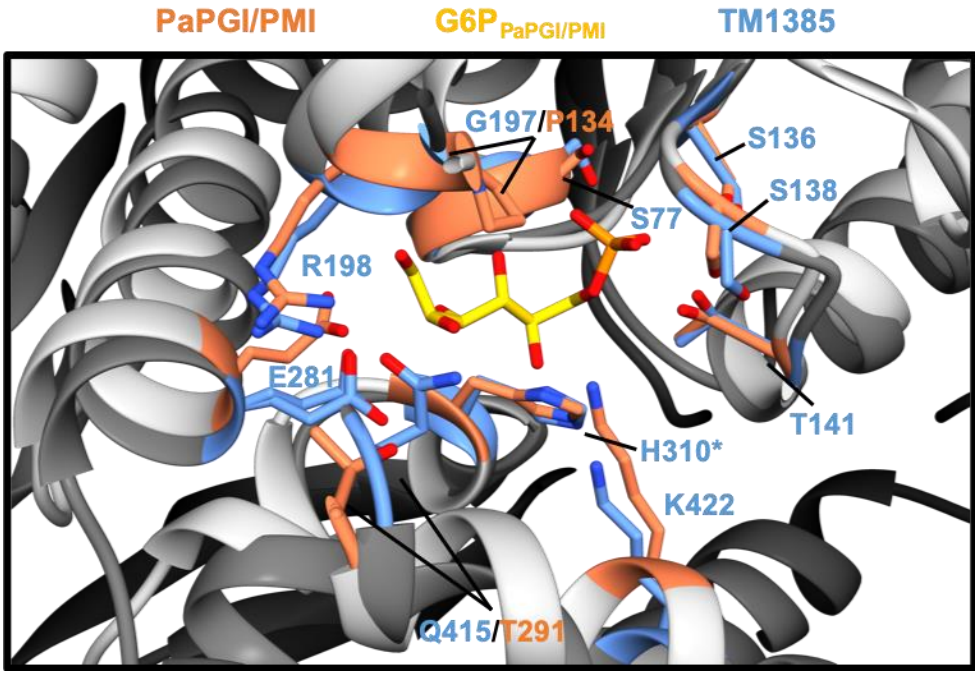


Figure 5.7: Structural evidence for TM1385 phosphoglucose isomerase function based on alignment with *Pyrobaculum aerophilum* PGI/PMI (phosphomannose isomerase). (A) TM1385 (PDB ID: 2Q8N, dark gray and blue) and PaPGI/PMI co-crystallized with glucose-6-phosphate (G6P, yellow) substrate (PDB ID:1X9I [18], light gray and coral) were structurally aligned using Chimera software [17]. Conserved PGI active site residues within the TM1385 sequence are labeled with TM1385 residue numbers, with non-conserved residues labeled for both proteins. Asterisk denotes a residue located in the second subunit.

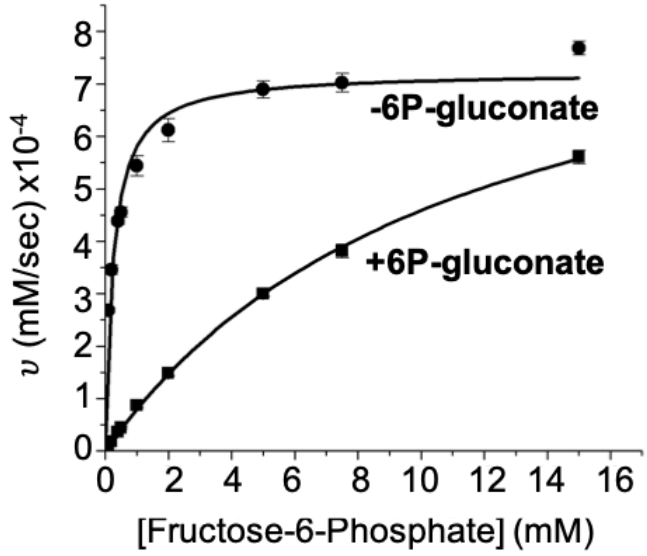


Figure 5.8: Steady-state kinetics using glucose-6-phosphate dehydrogenase coupling enzyme confirms PGI function of TM1385. TM1385 (TmPGI) catalysis was assayed based on F6P to G6P isomerization coupled to glucose-6-phosphate dehydrogenase (G6PDH) activity. Reaction rates were measured based on formation of NADPH (mM) for 30 seconds. Addition of the competitive PGI inhibitor 6P-gluconate [29] decreases TmPGI activity. Error bars represent the standard deviation of triplicates.

Table 5.2: Michaelis-Menten kinetic parameters of TmPGI variants with fructose-6-phosphate (G6PDH-Coupled Assay)

TmPGI Variant	K_M (mM)	k_{cat} (s^{-1})	k_{cat}/K_M ($mM^{-1}s^{-1}$)
WT	0.31 ± 0.03	17.4 ± 0.4	58.8 ± 6.2
E281S	NA*	NA	NA
E281A	NA	NA	NA
H310A	233 ± 100	65 ± 12	0.3 ± 0.2
Q415A	376 ± 150	70 ± 11	0.2 ± 0.1
K422A	478 ± 180	86 ± 13	0.2 ± 0.1
E281A/H310A	NA	NA	NA
E281A/Q415A	NA	NA	NA
H310A/K422A	NA	NA	NA

*No activity

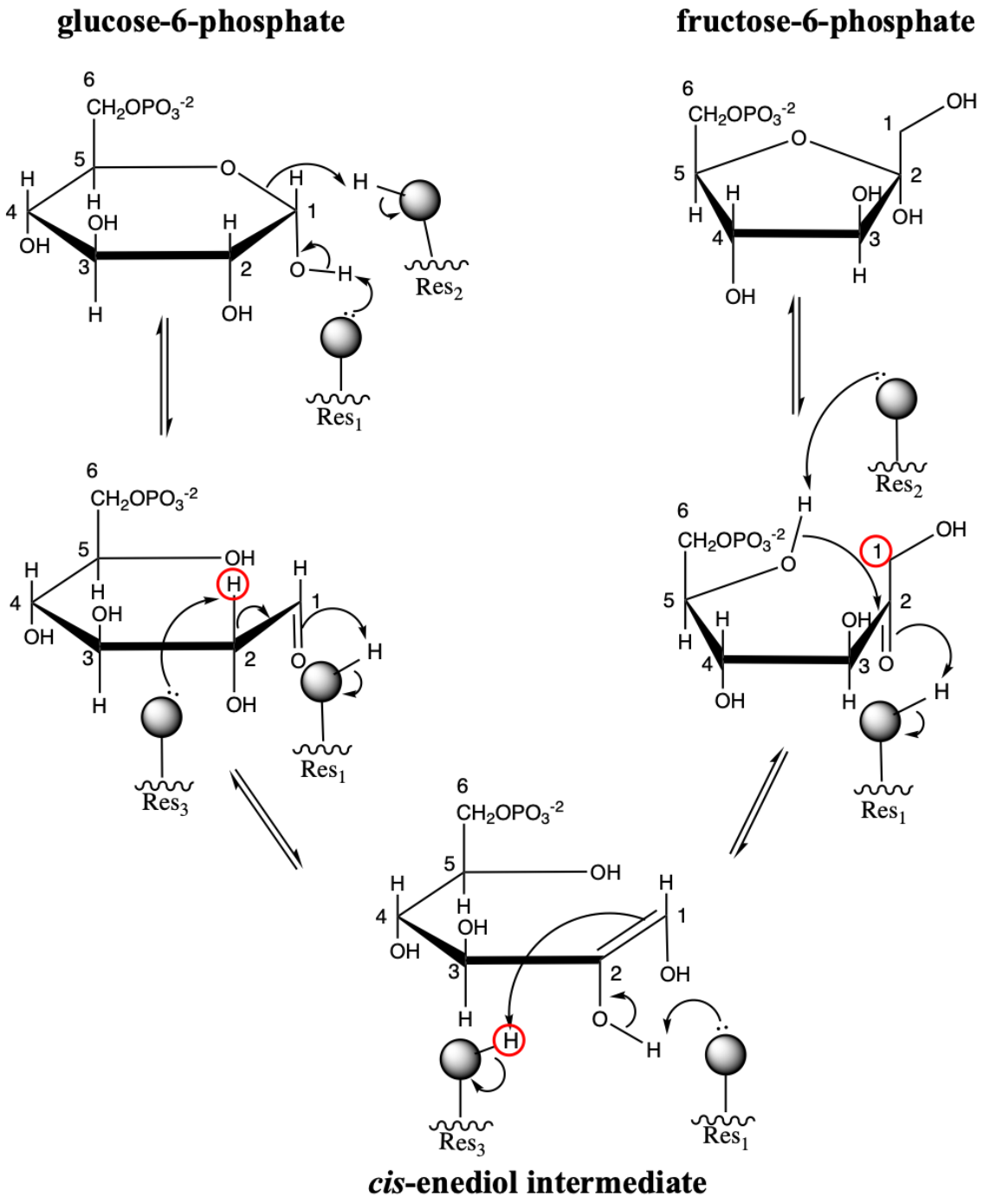


Figure 5.9: TmPGI proton exchange is facilitated by catalytic active site residues posed in the *cis*-enediol-based mechanism. G6P ring opening is the first step in the hypothesized *cis*-enediol mechanism for TmPGI, initiated by residue 1 (Res_1). Residue 2

(Res_2) donates a proton to the oxygen atom in the sugar ring to form a hydroxyl at carbon 5, facilitating a linearized substrate conformation. The *cis*-enediol intermediate occurs upon proton abstraction (red circle) at C2 by residue 3 (Res_3), with Res_1 accepting the leaving proton as part of an acid/base residue pair with Res_3. Isomerization to linear F6P is complete after a proton is donated back to the substrate C1 position by Res_3. Finally, Res_1 and Res_2 return the substrate to its cyclic form. Proposed mechanism is based on previous work from [4, 6, 12, 13].

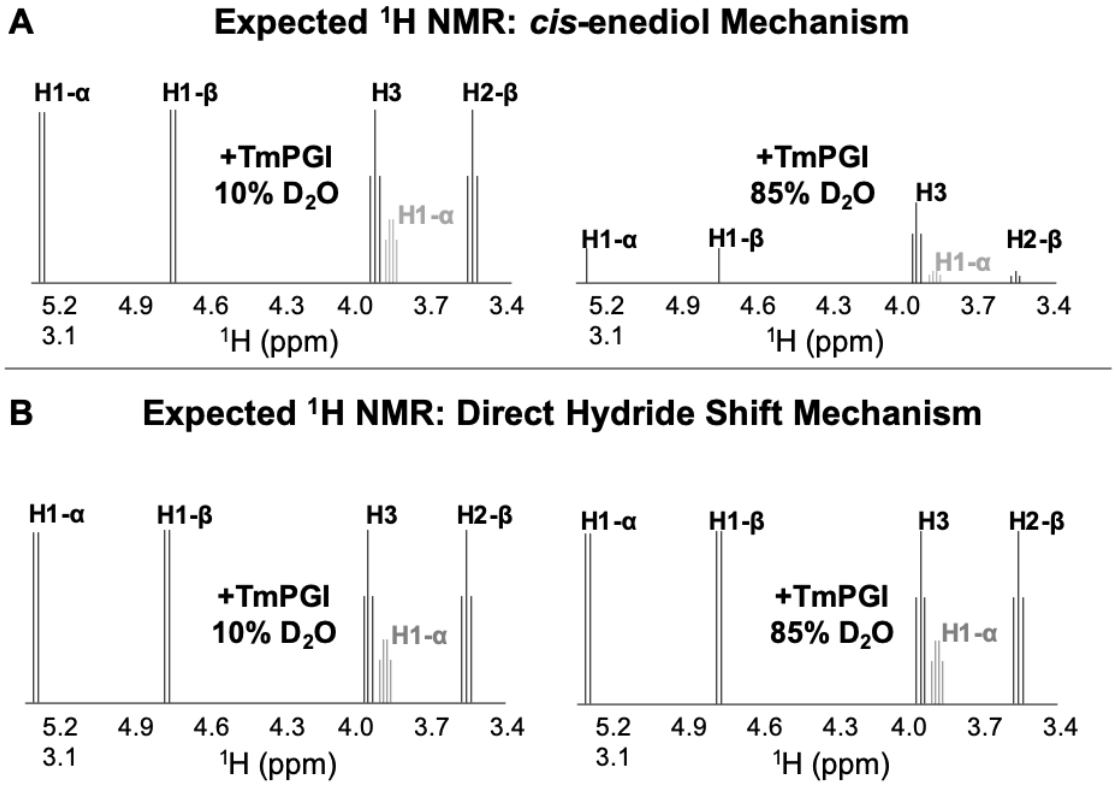


Figure 5.10: Theoretical ^{1}H NMR spectra for TmPGI-catalyzed substrate isomerization with *cis*-enediol and direct hydride shift mechanisms. (A) Expected one-dimensional ^{1}H NMR spectra of G6P incubated with TmPGI in 10% D_2O and 85% D_2O if a *cis*-enediol mechanism occurs. G6P resonances (black sticks) and F6P resonances (light gray sticks) will both be present in spectra where TmPGI catalyzes isomerization. Resonances shown are limited to protons used to monitor solvent exchange, as intensity decreases are expected in D_2O solvent with the *cis*-enediol mechanism (Figure 5.9). The F6P H1-β resonance is not included due to spectral crowding. G6P H1 resonance coupling has previously been shown to change from a doublet to a singlet in cases where G6P H1 resonances are still visible [13]. The G6P H3 resonance can additionally be used to identify solvent exchange when deuterium is incorporated at the C2 position. (B) It is expected that

there will be no significant change in resonance intensities if a direct hydride shift occurs since proton exchange is not present in the reaction mechanism (Figure 5.2).

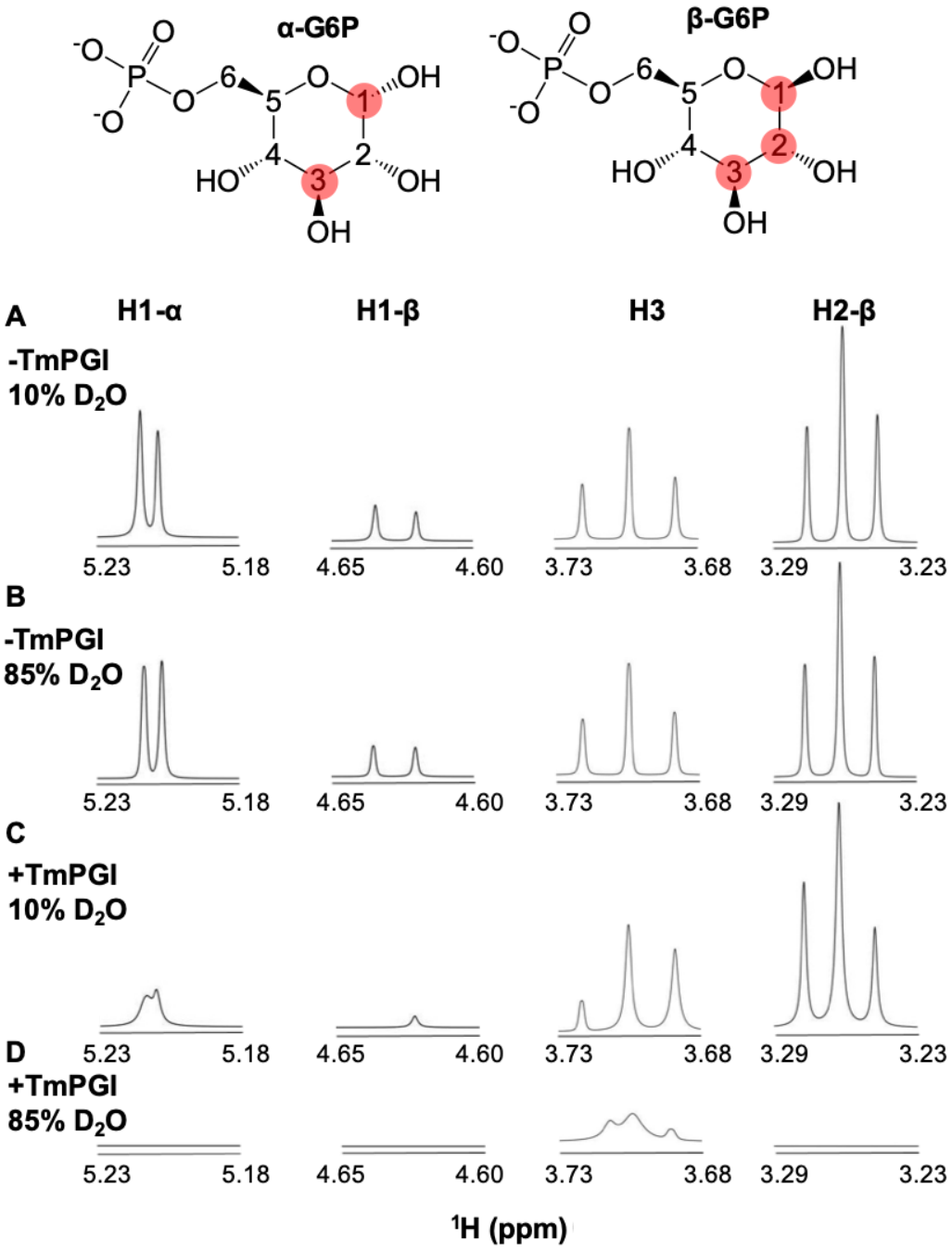


Figure 5.11: TmPGI catalyzes G6P to F6P isomerization via a *cis*-enediol intermediate. Experimental 1D ^{1}H NMR spectra of (A) G6P in 10% D_2O, (B) G6P in 85% D_2O, (C) TmPGI and G6P in 10% D_2O and (D) TmPGI and G6P in 85% D_2O. G6P α and

β anomers are both present and produce unique ^{1}H NMR signals. Loss of G6P H1-α, H1-β, H2-β, and H3 signals in reactions with TmPGI and 85% D_2O solvent indicates that TmPGI catalyzes G6P and F6P isomerization through the *cis*-enediol intermediate pathway. Corresponding protons are labeled in red on the G6P anomer structures. The H3 resonance could not be assigned to a specific anomer (Figure 5.4).

Table 5.3: Gibbs free energy and RMSD values of top 15 docked G6P poses with TmPGI

Pose	Energy (kcal/mol)	Location	RMSD to pose 1 (Å)
1	-9.73	Active Site B	0
2	-9.70	Surface	38.3
3	-9.67	Surface	38.0
4	-9.58	Active Site B	5.0
5	-9.56	Active Site B	3.5
6	-9.54	Active Site B	5.0
7	-9.53	Surface	39.1
8	-9.53	Active Site B	0.8
9	-9.51	Active Site B	5.0
10	-9.48	Active Site A	41.3
11	-9.48	Active Site A	41.3
12	-9.46	Active Site B	2.7
13	-9.42	Active Site B	2.1
14	-9.41	Active Site A	40.2
15	-9.41	Active Site A	40.7

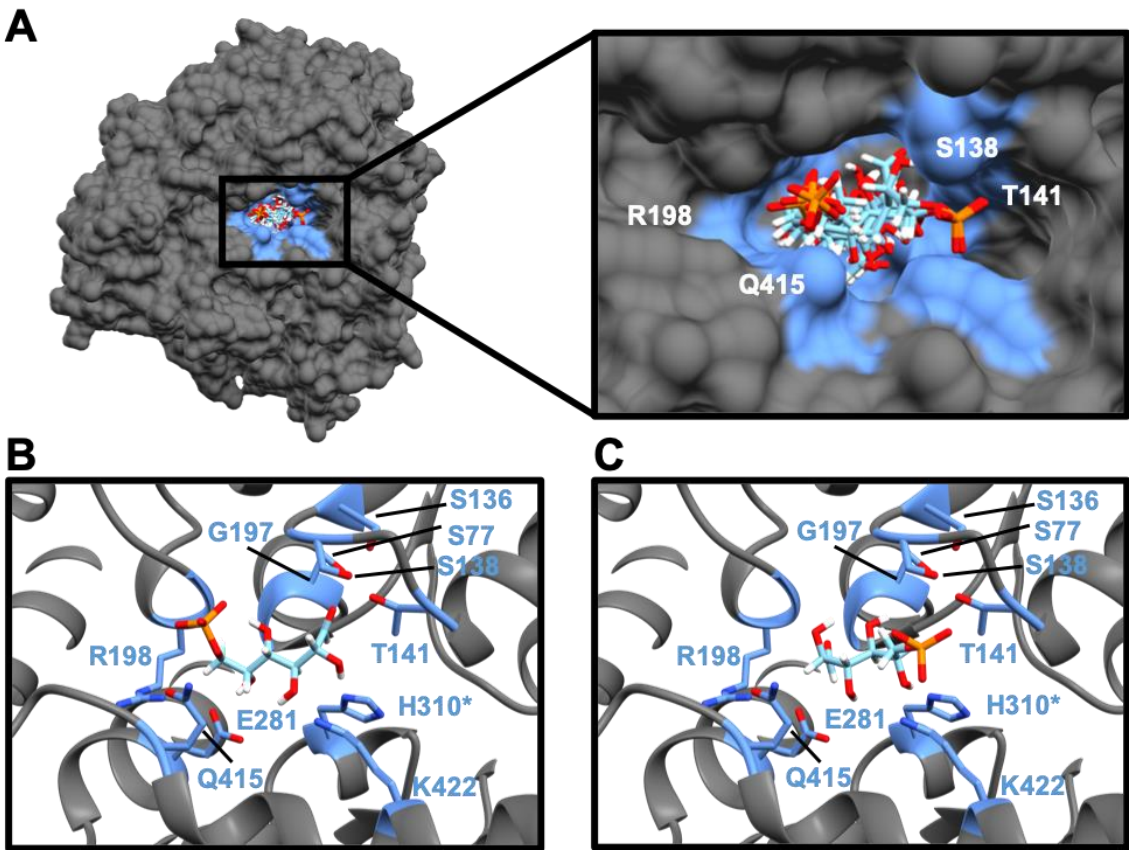


Figure 5.12: TmPGI computational docking with linear G6P substrate supports that E281, Q415, H310, and K422 may be important for function. (A) Computational docking with SwissDock [23] was performed with the apo TmPGI crystal structure (dark gray) and linear G6P (cyan sticks). The substrate binding pocket agrees well with the location of proposed active site residues from Figure 5.7, highlighted in light blue on the TmPGI surface representation. Eight of the 15 ligand poses with the lowest Gibbs free energy are found in the active site formed primarily by residues in chain B and H310 in chain A (active site B). The 8 ligand poses in active site B are superimposed, with S138, T141, R198, and Q415 labeled on the active site to clarify ligand orientation. Docked poses differed based on orientation of the phosphate group in the active site, demonstrated in

cartoon representations for (B) and (C). Docked G6P in panel B represents the pose with the lowest Gibbs free energy (-9.73 kcal/mol), while the pose in panel C has a Gibbs free energy of -9.58 kcal/mol.

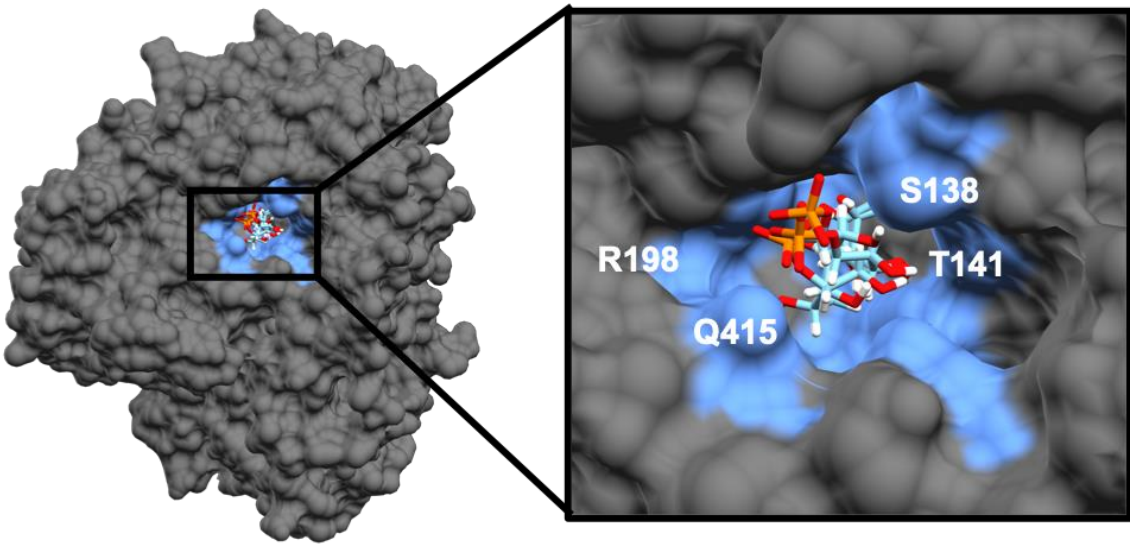


Figure 5.13: Docked G6P poses within TmPGI active site A. The top 15 G6P poses from computational docking with the apo TmPGI structure (Table 5.3) included 4 poses (cyan sticks) found in active site A, formed primarily by residues in chain A and H310 in chain B. The surface representation of TmPGI is shown in dark gray, and proposed active site residues are highlighted in light blue. The 4 ligand poses in active site A are superimposed, with S138, T141, R198, and Q415 labeled on the active site to clarify ligand orientation.

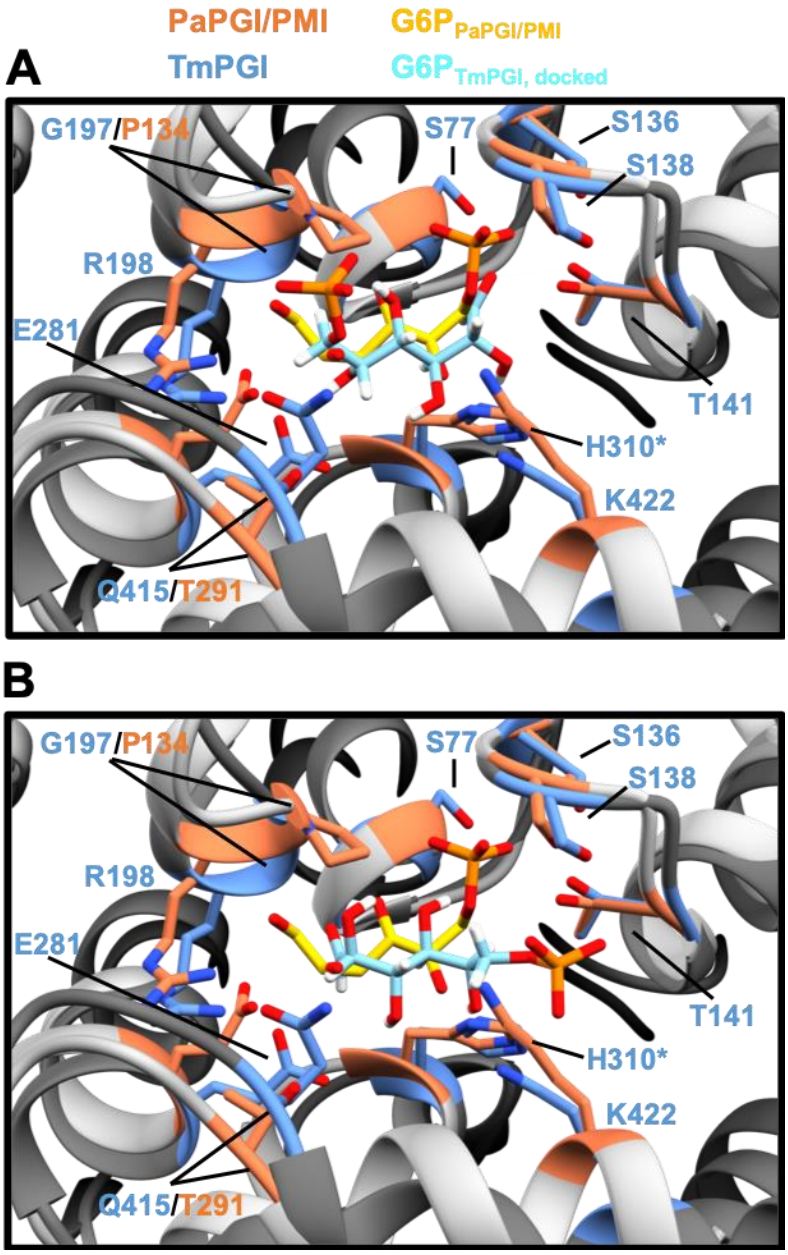


Figure 5.14: Docked G6P substrate with TmPGI compared to structural alignment with PaPGI/PMI co-crystallized with linear G6P. (A) The G6P docking pose with the lowest Gibbs free energy (pose 1, Table 5.3) in TmPGI active site B is shown with TmPGI structurally aligned to PaPGI/PMI. Conserved PGI active site residues within the TmPGI

sequence are labeled with TmPGI residue numbers, with non-conserved residues labeled for both proteins. Asterisk denotes a residue located in the second subunit. (B) G6P docking pose 4 (Table 5.3) shares a similar orientation to G6P in the PaPGI/PMI crystal structure with the substrate phosphoryl group located near T141. The free energy change from pose 1 to pose 4 is +0.15 kcal/mol.

Table 5.4: Substrate equilibria of TmPGI variants with F6P substrate (Seliwanoff Assay)

TmPGI Variant*	% F6P	% G6P
WT	30	70
E281S	64	36
E281A	55	45
H310A	47	53
Q415A	36	64
K422A	51	49
E281A/H310A	60	40
E281A/Q415A	100	0
H310A/K422A	100	0

*Results shown are based on 40X TmPGI concentration relative to G6PDH-coupled assays.

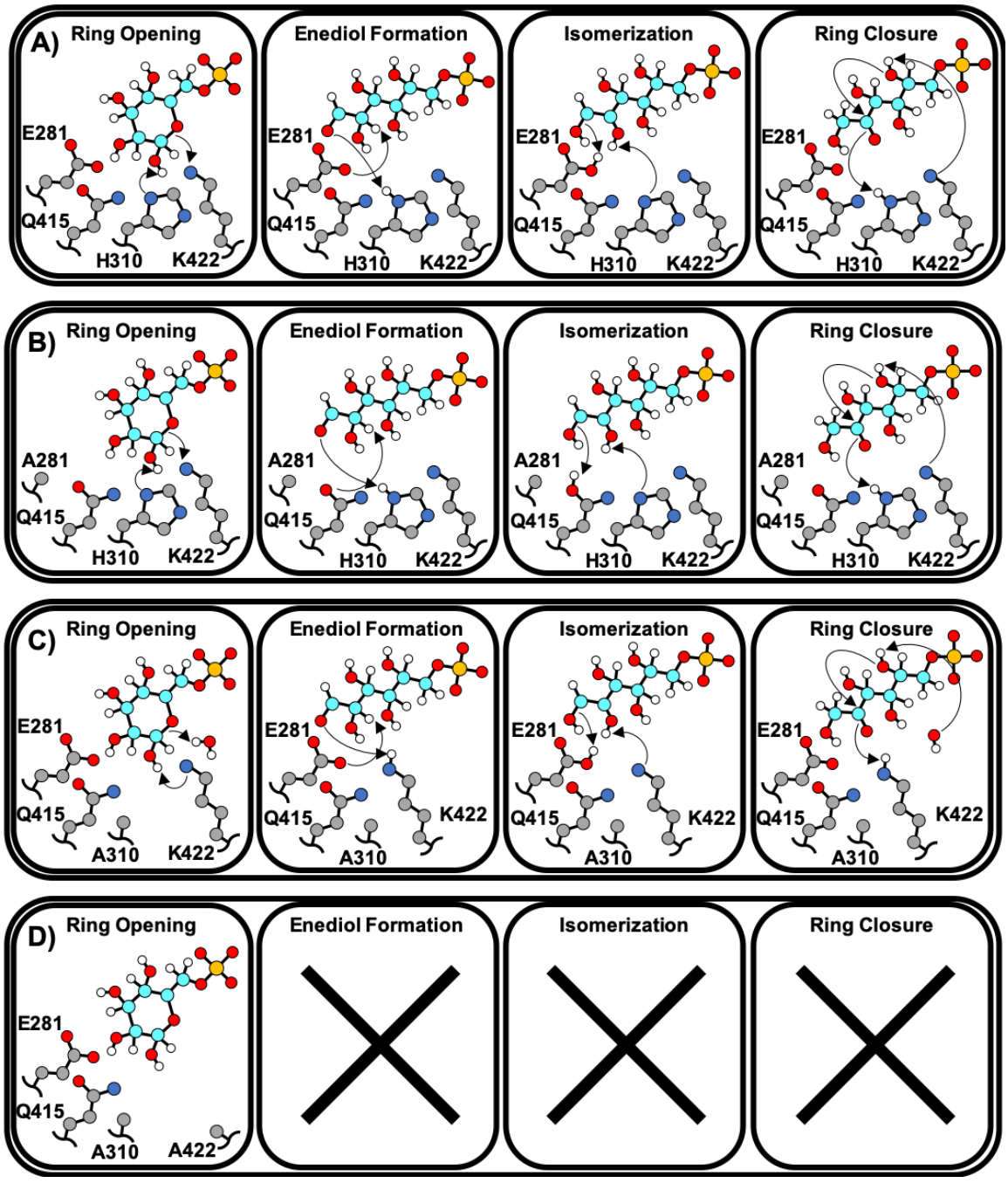


Figure 5.15: Proposed steps of TmPGI catalysis and compensatory residue functions. Simplified isomerization mechanism for (A) wild-type, (B) E281A, (C) H310A, and (D) H310A/K422A TmPGI variants. Substrate (cyan) and TmPGI residue side chains (gray)

are represented as ball-and-stick models. Explicit hydrogens were eliminated for TmPGI side chains unless relevant to the enzyme mechanism for simplicity. Arrows indicate proton exchange between substrate and TmPGI residues/water as well as the fructose ring cyclization step. Detailed electron flow may be referenced in Figure 5.9.

5.5 Concluding remarks

In summary, kinetic and colorimetric assays confirm that TM1385 from *Thermotoga maritima* functions as a PGI (now TmPGI). The TmPGI reaction mechanism proceeds through a *cis*-enediol intermediate, similar to experimental observations based on homologous PGIs with known crystal structures. TmPGI E281 is important for enzyme function and is likely the residue facilitating substrate isomerization in the *cis*-enediol mechanism. The roles of Q415, H310, and K422 are less clear. However, Q415 is proposed to position E281 or the substrate in the TmPGI active site and could act as a base catalyst in the absence of E281. Similarly, H310 or K422 could catalyze ring opening/closing of G6P and F6P, and mutation of one residue may be compensated by the other. Mutation of H310 and K422 simultaneously abolishes TmPGI function, supporting the hypothesis that either H310 or K422 must be present to facilitate conversion between ring and linear substrate forms. Finally, TmPGI activity persisting through multiple mutations of key active site residues highlights the robustness of the PGI enzyme class and may provide clues about diverse PGI functionality both inside and outside of the cell.

5.6 Acknowledgments

I would like to thank the following individuals for help with the initial characterization of TM1385 as a PGI: Taylor Koerner, Jing Liu, Sun Young Park, Robert Schenck, Amanda Taylor, Ryan Lo, Carol Price, and Cameron Mura. I would also like to thank all members of the Columbus laboratory and colleagues for manuscript revisions. This work was supported by the NSF (MCB 1817735 to Prof. Linda Columbus; DDGE 1315231 to Nicole Swope), NIH (T32 GM008715 to Nicole Swope), and the Double Hoo

Research Grant funded by the Jefferson Trust at the University of Virginia (Katherine Lake and Nicole Swope).

5.6.1 Author contributions

Nicole Swope, Golda Harris, Daniel Yu, and Prof. Linda Columbus conceived this study. Nicole Swope, Katherine Lake, and Linda Columbus coordinated and designed this study. Nicole Swope designed, conducted, and analyzed the experiments shown in Figure 5.4, Figure 5.5, Figure 5.6, Figure 5.7, Figure 5.8, Figure 5.11, Table 5.2, and Table 5.4. Katherine Lake designed, conducted, and analyzed the experiments shown in Figure 5.7, Figure 5.12, Figure 5.13, Figure 5.14, Table 5.2, Table 5.3, and Table 5.4. The manuscript was drafted and edited by Nicole Swope, Katherine Lake, and Prof. Linda Columbus.

5.7 References

1. Lesley, S.A., et al., *Structural genomics of the Thermotoga maritima proteome implemented in a high-throughput structure determination pipeline.* Proceedings of the National Academy of Sciences, 2002. **99**(18): p. 11664.
2. Logsdon, J.M. and D.M. Faguy, *Evolutionary genomics: Thermotoga heats up lateral gene transfer.* Current Biology, 1999. **9**(19): p. R747-R751.
3. Rigoldi, F., et al., *Review: Engineering of thermostable enzymes for industrial applications.* APL Bioengineering, 2018. **2**(1): p. 011501.
4. Arsenieva, D., et al., *The crystal structure of rabbit phosphoglucose isomerase complexed with 5-phospho-d-arabinonohydroxamic acid.* Proceedings of the National Academy of Sciences, 2002. **99**(9): p. 5872-5877.
5. Jeffery, C.J., R. Hardré, and L. Salmon, *Crystal Structure of Rabbit Phosphoglucose Isomerase Complexed with 5-Phospho-d-Arabinonate Identifies the Role of Glu357 in Catalysis.* Biochemistry, 2001. **40**(6): p. 1560-1566.
6. Read, J., et al., *The crystal structure of human phosphoglucose isomerase at 1.6 Å resolution: implications for catalytic mechanism, cytokine activity and haemolytic anaemia11Edited by R. Huber.* Journal of Molecular Biology, 2001. **309**(2): p. 447-463.
7. Achari, A., et al., *Glucose - 6 - phosphate isomerase.* Philosophical Transactions of the Royal Society of London. B, Biological Sciences, 1981. **293**(1063): p. 145-157.

8. Jeffery, C.J., et al., *Crystal Structure of Rabbit Phosphoglucose Isomerase, a Glycolytic Enzyme That Moonlights as Neuroleukin, Autocrine Motility Factor, and Differentiation Mediator.* Biochemistry, 2000. **39**(5): p. 955-964.
9. Repiso, A., et al., *Glucose phosphate isomerase deficiency: enzymatic and familial characterization of Arg346His mutation.* Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease, 2005. **1740**(3): p. 467-471.
10. Funasaka, T., et al., *Down-Regulation of Phosphoglucose Isomerase/Autocrine Motility Factor Results in Mesenchymal-to-Epithelial Transition of Human Lung Fibrosarcoma Cells.* Cancer Research, 2007. **67**(9): p. 4236.
11. Hansen, T., C. Urbanke, and P. Schönheit, *Bifunctional phosphoglucose/phosphomannose isomerase from the hyperthermophilic archaeon Pyrobaculum aerophilum.* Extremophiles, 2004. **8**(6): p. 507-512.
12. Lee, J.H. and C.J. Jeffery, *The crystal structure of rabbit phosphoglucose isomerase complexed with D-sorbitol-6-phosphate, an analog of the open chain form of D-glucose-6-phosphate.* Protein science : a publication of the Protein Society, 2005. **14**(3): p. 727-734.
13. Berrisford, J.M., et al., *Evidence Supporting a cis-enediol-based Mechanism for Pyrococcus furiosus Phosphoglucose Isomerase.* Journal of Molecular Biology, 2006. **358**(5): p. 1353-1366.
14. Swan, M.K., et al., *Structural Evidence for a Hydride Transfer Mechanism of Catalysis in Phosphoglucose Isomerase from Pyrococcus furiosus.* Journal of Biological Chemistry, 2003. **278**(47): p. 47261-47268.
15. Dyson, J.E.D. and E.A. Noltmann, *The Effect of pH and Temperature on the Kinetic Parameters of Phosphoglucose Isomerase: PARTICIPATION OF HISTIDINE AND LYSINE IN A PROPOSED DUAL FUNCTION MECHANISM.* Journal of Biological Chemistry, 1968. **243**(7): p. 1401-1414.
16. Cordeiro, A.T., et al., *Crystal structure of human phosphoglucose isomerase and analysis of the initial catalytic steps.* Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics, 2003. **1645**(2): p. 117-122.
17. Pettersen, E.F., et al., *UCSF Chimera—A visualization system for exploratory research and analysis.* Journal of Computational Chemistry, 2004. **25**(13): p. 1605-1612.
18. Swan, M.K., et al., *Structural Basis for Phosphomannose Isomerase Activity in Phosphoglucose Isomerase from Pyrobaculum aerophilum: A Subtle Difference between Distantly Related Enzymes.* Biochemistry, 2004. **43**(44): p. 14088-14095.
19. Armougom, F., et al., *Expresso: automatic incorporation of structural information in multiple sequence alignments using 3D-Coffee.* Nucleic Acids Research, 2006. **34**(suppl_2): p. W604-W608.
20. Robert, X. and P. Gouet, *Deciphering key features in protein structures with the new ENDscript server.* Nucleic Acids Research, 2014. **42**(W1): p. W320-W324.
21. Klock, H.E. and S.A. Lesley, *The Polymerase Incomplete Primer Extension (PIPE) Method Applied to High-Throughput Cloning and Site-Directed Mutagenesis*, in *High Throughput Protein Expression and Purification: Methods and Protocols*, S.A. Doyle, Editor. 2009, Humana Press: Totowa, NJ. p. 91-103.

22. Noltmann, E.A., C.J. Gubler, and S.A. Kuby, *Glucose 6-Phosphate Dehydrogenase (Zwischenferment) : I. ISOLATION OF THE CRYSTALLINE ENZYME FROM YEAST.* Journal of Biological Chemistry, 1961. **236**(5): p. 1225-1230.
23. Grosdidier, A., V. Zoete, and O. Michielin, *SwissDock, a protein-small molecule docking web service based on EADock DSS.* Nucleic acids research, 2011. **39**(Web Server issue): p. W270-W277.
24. Sánchez-Viesca, F. and R. Gómez, *Reactivities Involved in the Seliwanoff Reaction.* 2018.
25. Chou, C.-C., et al., *The Crystal Structure of Phosphoglucose Isomerase/Autocrine Motility Factor/Neuroleukin Complexed with Its Carbohydrate Phosphate Inhibitors Suggests Its Substrate/Receptor Recognition.* Journal of Biological Chemistry, 2000. **275**(30): p. 23154-23160.
26. Yamamoto, H., H. Miwa, and N. Kunishima, *Crystal Structure of Glucose-6-Phosphate Isomerase from Thermus thermophilus HB8 Showing a Snapshot of Active Dimeric State††H.Y. performed structural determination and biochemical and biophysical experiments, and wrote this paper. H.M. contributed to large-scale protein production. N.K. supervised this work.* Journal of Molecular Biology, 2008. **382**(3): p. 747-762.
27. Mathur, D. and L.C. Garg, *Functional phosphoglucose isomerase from Mycobacterium tuberculosis H37Rv: Rapid purification with high yield and purity.* Protein Expression and Purification, 2007. **52**(2): p. 373-378.
28. Sun, L.-C., et al., *Glucose-6-phosphate Isomerase Is an Endogenous Inhibitor to Myofibril-Bound Serine Proteinase of Crucian Carp (Carassius auratus).* Journal of Agricultural and Food Chemistry, 2009. **57**(12): p. 5549-5555.
29. Gaitonde, M.K., E. Murray, and V.J. Cunningham, *Effect of 6-Phosphogluconate on Phosphoglucose Isomerase in Rat Brain In Vitro and In Vivo.* Journal of Neurochemistry, 1989. **52**(5): p. 1348-1352.
30. Meng, M., et al., *Functions of the conserved anionic amino acids and those interacting with the substrate phosphate group of phosphoglucose isomerase.* FEBS Letters, 2001. **499**(1): p. 11-14.
31. Berrisford, J.M., et al., *Crystal Structure of Pyrococcus furiosus Phosphoglucose Isomerase: IMPLICATIONS FOR SUBSTRATE BINDING AND CATALYSIS.* Journal of Biological Chemistry, 2003. **278**(35): p. 33290-33297.
32. Sánchez-Viesca, F. and R. Gómez, *Reactivities Involved in the Seliwanoff Reaction.*

Appendix I

R source code to combine scattering data for plotting

```
#---
#Program: combines individual SAXS profiles for plotting in R
#author: Nicole Swope
#date: March 11, 2020
#version: 5
#Source code is being developed for general use
#---

#Clear all objects from environment

rm(list = ls(all.names = TRUE))

#Change bicelle composition, temp, and output file extension

bicelle <- "POPC-DH7PC"
temp <- "25C"
output_extension <- ".txt"

# Set working directory to path where text files are stored

averaged_path <-
"/Users/nkswope/Dropbox/bicelles_SAXS/APS_data/20191129/Averaged
"
bsub <- "bsub_1.1"
mypath <- paste(averaged_path, bicelle, temp, bsub, sep="/")

setwd(mypath)

#Create output file string

bicelle_temp <- paste(bicelle, temp, sep="_")
output_file <- paste(bicelle_temp, output_extension, sep="")

#Delete output file if it already exists
old_file <- paste(mypath, output_file, sep="/")

if (file.exists(old_file)){
  file.remove(old_file)
}

#Create list of files to combine
file_list <- list.files(path=mypath, pattern="*.txt")

for (file in file_list){

  # if the merged dataset doesn't exist, create it
  if (!exists("SAXS_data")){
    SAXS_data <- read.table(file, header=TRUE, sep="")
```

```
  }

  # if the merged dataset does exist, append to it
  if (exists("SAXS_data")){
    temp_data <-read.table(file, header=TRUE, sep="")
    SAXS_data <-cbind(SAXS_data, temp_data)
    rm(temp_data)
  }

}

#Remove duplicated first entry
SAXS_data <- SAXS_data[-c(1:4)]

#Save dataframe to file
write.table(SAXS_data, file=output_file, sep=" ", dec=".",
row.names=FALSE)
```

R source code to plot scattering profiles

```
#---
#Program: SAXS profile plotting in R
#author: Nicole Swope
#date: March 11, 2020
#version: 10
#Source code is being developed for general use
#---

#Plots q 0.1 to 1.0 for lipid-detergent bicelle scattering

#Import libraries

library(dplyr)
require(ggplot2)
require(reshape2)

#Clear all objects from environment

rm(list = ls(all.names = TRUE))

#Change bicelle composition, temp, and input file extension

bicelle <- "DLPC-CHAPS"
temp <- "25C"
input_extension <- ".txt"

# Set working directory to path where text files are stored

averaged_path <-
"/Users/nkswope/Dropbox/bicelles_SAXS/APS_data/20191129/Averaged
"
bsub <- "bsub_1.1"
mypath <- paste(averaged_path, bicelle, temp, bsub, sep="/")

setwd(mypath)

#Create input file string

bicelle_temp <- paste(bicelle, temp, sep="_")
input_file <- paste(bicelle_temp, input_extension, sep="")

#Read in file
SAXS_data <- read.delim(input_file, header=TRUE, sep="")

#Save data frame with X, Y coordinates only

c <- 1:ncol(SAXS_data)
SAXS_data <- SAXS_data[, (c%%4==1:2)]
```

```
#Save new data frame with first X coordinates and all Y
c <- 1:ncol(SAXS_data)

#Hard-coded columns, needs optimization
SAXS_plot <- SAXS_data[c(1,4,6,8,10,12,14,16,18,20,22)]

#Define Q range to find max intensity
min_Q <- 0.1
max_Q <- 0.4

#Subset I(Q) range to find max intensity
Q_subset <- subset(SAXS_plot, (X.X.>min_Q & X.X.<max_Q))

#Function to find maximum intensity value within data frame
column
colMax <- function(x) sapply(x, max, na.rm = TRUE)

#Store list of positional values for column numbers to iterate
through
#Will exclude first column of dataframe (Y columns only) (i.e.
2,3...10)
n <- seq(2,length(Q_subset))

#Store second max intensities as a list and remove names
max_y <- colMax(Q_subset[,2:length(Q_subset)])
names(max_y) <- NULL

#Store list of number sequence to iterate through max intensity
values (i.e.1,2...9)
max_y_length <- seq(length(max_y))

#Store sequence of number of intensity points (i.e.
1,2,3,4,5...2000)
npoints <- seq(length(SAXS_plot[[1]]))

#Iterates through each value in a fixed plot column
#and divides that value by the corresponding max intensity (from
max_y)
for (ypos in max_y_length){
  for (point in npoints){
    SAXS_plot[point,n[ypos]] <-
SAXS_plot[point,n[ypos]]/max_y[ypos]
  }
}

#Rename X column to Q (SAXS) and Y columns to bicelle q-ratios
colnames(SAXS_plot) <- c("Q", "0.1", "0.2", "0.3", "0.4", "0.5",
"0.6", "0.7", "0.8", "0.9", "1.0")
```

```
#Melt dataframe for plotting - output is X values, Y name, Y
values
SAXS_plot <- melt(SAXS_plot, id.vars='Q')

#Rename variable column to "q-ratio" and values column to I(Q)
colnames(SAXS_plot)[2:3] <- c("q", "Intensity")

#Plot graph
SAXS_plot %>%
  ggplot(aes(x=Q, y=Intensity)) +
  theme(panel.grid.major = element_blank(), panel.grid.minor =
element_blank(),
        panel.background = element_blank(), axis.line =
element_line(colour = "black"),
        axis.text = element_text(size = 20),  axis.title =
element_text(size = 28),
        legend.text = element_text(size = 20), legend.title =
element_text(size = 28)) +
  xlab(bquote('Q'~ (ring(A)^-1))) +
  ylab('I(Q)') +
  xlim(0.01,0.5) +
  ylim(0,1.1) +
  geom_point(aes(color=q, fill=q), size=1) +
  scale_color_brewer(palette="Spectral")

#Include below for image export when satisfied with result
#tiff("q-range_SAXS.tiff", units="in", width=5, height=5,
res=300))
#dev.off()
```

R source code to calculate L parameter from SAXS profiles

```
#---
#Program: Calculates SAXS "L" parameter for bicelles
#author: Nicole Swope
#date: March 11, 2020
#version: 2
#Source code is being developed for general use
#---

#Calculates L from SAXS profiles for lipid-detergent mixtures
when 0.0 < q < 1.0

#Clear all objects from environment
rm(list = ls(all.names = TRUE))

#Change bicelle composition, temp, and input file extension

bicelle <- "DLPC-Cyclofos6"
temp <- "25C"
input_extension <- ".txt"

# Set working directory to path where text files are stored

averaged_path <-
"/Users/nkswope/Dropbox/bicelles_SAXS/APS_data/20191129/Averaged
"
bsub <- "bsub_1.1"
mypath <- paste(averaged_path, bicelle, temp, bsub, sep="/")

setwd(mypath)

#Create input file string

bicelle_temp <- paste(bicelle, temp, sep="_")
input_file <- paste(bicelle_temp, input_extension, sep="")

#Read in file
SAXS_data <- read.delim(input_file, header=TRUE, sep="")

#Save data frame with X, Y coordinates only
c <- 1:ncol(SAXS_data)
SAXS_data <- SAXS_data[, (c%%4==1:2)]

#Save new data frame with first X coordinates and all Y
c <- 1:ncol(SAXS_data)
#Hard-coded even columns, needs optimization
SAXS_plot <- SAXS_data[c(1,2,4,6,8,10,12,14,16,18,20,22)]

#Rename X column to Q (SAXS) and Y columns to bicelle q-ratios
```

```
colnames(SAXS_plot) <- c("Q", "0.0", "0.1", "0.2", "0.3", "0.4",
"0.5", "0.6", "0.7", "0.8", "0.9", "1.0")

#Empty vectors to store L parameters and maximum I(Q) positions
all_L <- c()
y_max <- c()


#Determine Q-range to find second maxima
min_Q <- 0.1
max_Q <- 0.25

#Subset I(Q) range to determine second maxima
Q_subset <- subset(SAXS_plot, (Q>min_Q & Q<max_Q))
n <- seq(2,length(Q_subset))

#Finds position of the second maximum for each dataset
for (pos in n){
    max <- which.max(Q_subset[[pos]])
    y_max <- c(y_max, max)
}

#Converts Q-value of second maximum for each dataset
for (value in y_max){
  max <- (Q_subset[[1]][value])
  L <- 2*pi/max
  all_L <- c(all_L, L)
}

#prints list of L parameters for each dataset by q-ratio
all_L
```