Structure-function relationships in monotopic phosphoglycosyl transferases

by

Sonya Entova

M.S. Biochemistry
Brandeis University, 2013

Submitted to the Department of Biology in
Partial Fulfillment of the Requirements for the
Degree of Doctor of Philosophy at the

Massachusetts Institute of Technology

June 2019

Signature redacted

Signature of Author: _____

Department of Biology
April 5th, 2019

Signature redacted

Certified by: ___ _____

Barbara Imperiali
Class of 1922 Professor of Biology and Chemistry
Thesis Supervisor

Signature redacted

Accepted by: _____ _____

Stephen Bell
Uncas and Helen Whitaker Professor of Biology
Investigator, Howard Hughes Medical Institute
Co-Director, Biology Graduate Committee

1

Structure-function relationships in monotopic phosphoglycosyl transferases

by
Sonya Entova

Submitted to the Department of Biology
on April 5th, 2019 in Partial Fulfillment of the
Requirements for the Degree of Doctor of Philosophy

Abstract

Complex glycans play essential roles in prokaryotic and eukaryotic biology. While this ubiquitous post-translational modification takes a diversity of forms, many glycoconjugate biosynthesis pathway across domains of life follows a common logic. Glycan assembly is initiated by a phosphoglycosyl transferase (PGT) that transfers a phosphosugar from a nucleotide donor to a polyprenol phosphate (PrenP) chain embedded in the membrane. The PrenPP-sugar product is elaborated by downstream glycosyltransferases, transferred across the membrane and ultimately appended to various acceptor molecules.

The PGTs initiating glycan assembly adopt diverse membrane architectures. An extensive superfamily of PGTs, elucidated in part by this thesis, is exemplified by PglC from the Gram-negative pathogen, *Campylobacter jejuni*. PglC comprises a globular cytosolic domain and an N-terminal membrane-resident domain. Recent structural and biochemical analyses determined that this domain forms a helix-break-helix motif, termed the reentrant membrane helix (RMH), that enters and exits on the same face of the membrane, resulting in a monotopic topology. The RMH anchors the PglC fold in the membrane in a manner not previously observed among other monotopic membrane proteins. This thesis focuses on structure-function relationships in the RMH and associated domains. Two conserved motifs are shown to drive formation of a reentrant topology for PglC, and to exemplify common principles of topology determination among diverse monotopic proteins. These principles are further applied to the identification of reentrant domains in an extensive superfamily of monotopic lipid A acyltransferases previously thought to be membrane-spanning.

The next section of the thesis explores the highly conserved role of PrenP in complex glycan biosynthesis. The significance of PrenP geometry in mediating substrate binding and modulating the local membrane environment is presented. Additionally, a conserved proline residue in the PglC RMH is determined to drive PrenP binding and specificity. Molecular insights from this study shed new light on the roles of PrenP in facilitating diverse glycoconjugate biosynthesis pathways. Finally, a cell-free methodology for expression of PglC directly into model membrane lipid Nanodiscs is described. This system has valuable applications for the study of interactions between PglC and downstream glycosyltransferase enzymes, and for further structural characterization of PglC in a membrane environment.

Thesis Supervisor: Barbara Imperiali
Class of 1922 Professor of Biology and Chemistry

Acknowledgements

4

Table of Contents

9

List of Figures

Chapter 5

List of Tables

List of Abbreviations

Standard 3-letter and 1-letter codes are used for the 20 natural amino acids.
Standard 1-letter codes are used for the 4 common DNA bases.

CEF	cell-envelope fraction
DDM	n-dodecyl β-D-maltoside
DecP	decaprenol phosphate
diNAcBac	N,N'-diacetylbacillosamine
DMPC	1,2-dimyristoyl-sn-glycero-3-phosphocholine
DMPG	1,2-dimyristoyl-sn-glycero-3-phosphoglycerol
DMSO	dimethylsulfoxide
GT	glycosyltransferase
LB	lysogeny broth
LPS	lipopolysaccharide
MSP	molecular scaffolding protein
Ni-NTA	nickel nitrilotriacetic acid
OST	oligosaccharyltransferase
Pgl	Protein glycosylation
PGT	phosphoglycosyl transferase
PIRS	polyisoprenol recognition sequence
POPE	1-palmitoyl-2-oleoyl-sn-glycero-3-phosphoethanolamine
POPG	1-palmitoyl-2-oleoyl-sn-glycero-3-phosphoglycerol
PrenP	polyprenol phosphate
PrenPP	polyprenol diphosphate
RMH	reentrant membrane helix
SMA	styrene-maleic acid
SMALP	SMA-lipoparticle
SUMO	small ubiquitin-like modifier
SolP	solanesol phosphate
TMHMM	transmembrane hidden Markov model
UDP	uridine diphosphate
UMP	uridine monophosphate
UndP	undecaprenol phosphate

Chapter 1: PglC as a platform for the study of monotopic topology determination and polyprenol phosphate substrate interactions

INTRODUCTION

Glycosylation is a complex post-translational modification essential to all domains of life. Glycoconjugates can take the form of many important molecules, including glycoproteins, glycolipids and proteoglycans, as well as many bacterial cell wall components. Eukaryotic glycoproteins are often secreted or displayed on the cell surface (**Figure 1-1**); such glycans mediate important interactions involved in processes as diverse as development,[1] host-pathogen interactions,[2] immune response,[3] and cancer progression.[4] Bacterial cell surface glycoconjugates are key modulators of bacterial physiology[5] and infectivity[6] (**Figure 1-2**). Thus, elucidation of this ubiquitous post-translational modification carries great importance not only for our fundamental knowledge of biology, but also for our understanding of the many human diseases caused or signified by defects in glycosylation.
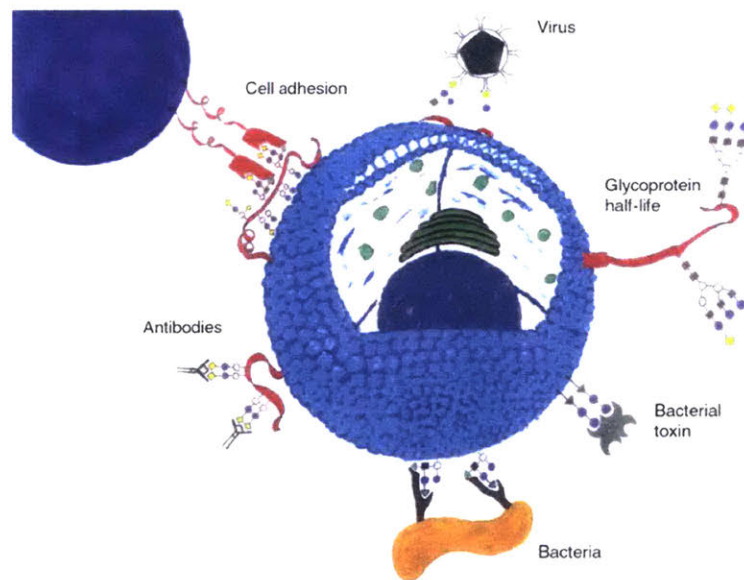
Figure 1-1 Eukaryotic cell surface glycoproteins mediate many important interactions. Glycoproteins on the cell surface mediate interactions with neighboring cells, pathogens, and extracellular proteins. Glycans also extend the half-life of secreted proteins. Figure reproduced from Holgersson et al., 2005.[7]
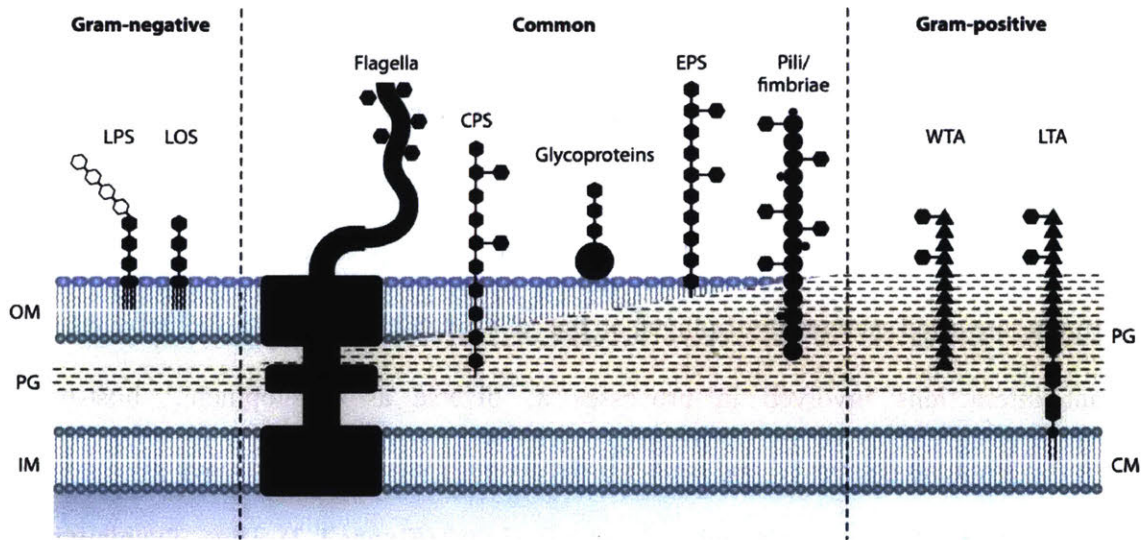
Figure 1-2 Bacterial cell walls display many essential glycoconjugates.
Bacterial cell wall glycoconjugates include peptidoglycan (PG), capsular polysaccharide (CPS), exopolysaccharide (EPS), as well as glycoproteins and glycosylated pili and flagella. Gram-negative bacteria additionally display lipopolysaccharide (LPS) and lipooligosaccharide (LOS), while Gram-positive bacteria display wall teichoic acid (WTA) and lipoteichoic acid (LTA). IM = inner membrane, OM = outer membrane, CM = cell membrane. Figure reproduced from Tytgat and Lebeer, 2014.[8]

Glycoconjugate biosynthesis pathways across domains share many key features. For example, the assembly of complex glycans most commonly occurs via a series of peripheral and integral membrane-associated enzymes. Glycoconjugate biosynthesis in prokaryotes occurs at the inner membrane (in Gram-negative bacteria) or the cell membrane (in Gram-positive bacteria and archaea), whereas in eukaryotes it occurs at the endoplasmic reticulum membrane. In known bacterial and eukaryotic pathways, glycan assembly is initiated by a phosphoglycosyl transferase (PGT), which transfers a phosphosugar from an activated nucleotide-sugar donor in the cytoplasm to an amphiphilic polyprenol phosphate (PrenP) chain in the membrane (**Figure 1-3**). Due to their important role in catalyzing the first membrane-committed step in complex glycan assembly, PGTs

act as the "gatekeepers" of glyconjugate biosynthesis.[9] The PrenPP-sugar PGT product is elaborated by downstream glycosyltransferases (GTs), and the completed PrenPP-linked glycan is transferred across the membrane and finally appended *en bloc* to acceptor molecules to become one of many essential glycoconjugates. In subsequent steps (not shown), PrenPP is dephosphorylated to regenerate PrenP.
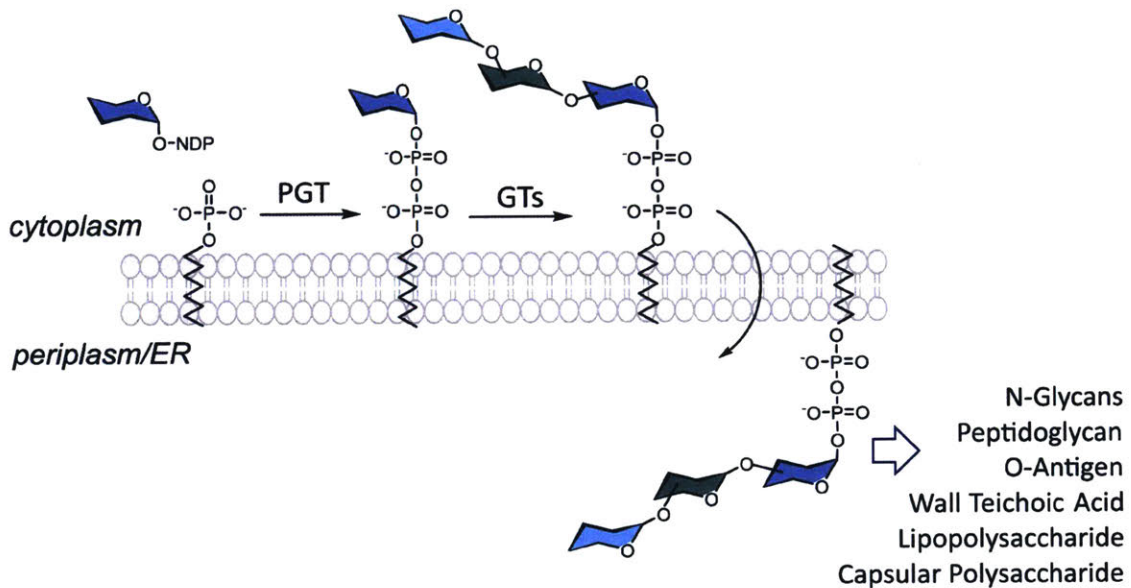
Figure 1-3 General scheme for complex glycan assembly at the membrane.
Glycans are assembled on a polyprenol phosphate chain (PrenP) embedded in the membrane. Assembly in bacterial, eukaryotic and certain archaeal pathways is initiated by a phophoglycosyl transferase (PGT) and continued by the action of several glycosyltransferases (GTs). Figure adapted from Lukose et al., 2017.[9]

We note that biochemical and bioinformatics analyses of archaeal glycoconjugate assembly pathways have found that many of these pathways, unlike those in bacteria and eukaryotes, are initiated by PrenP-glycosyltransferase instead of PGTs. This results in a monophosphate PrenP-sugar linkage that is then similarly elaborated to downstream GTs, transferred across the membrane and appended to acceptor substrates.[10-11]

All PrenPs used in complex glycan assembly take the generalized form of a polyisoprene chain with multiple *cis* isoprene units at the functionalized terminus, although the precise molecular structure of PrenPs can vary by species (**Figure 1-4**). The length of PrenPs can vary from eight isoprene units in prokaryotes to over 20 in eukaryotes. Bacterial PrenPs are fully unsaturated,[12-13] while archaeal and eukaryotic PrenPs are α-saturated,[14] and archaeal PrenPs can be additionally unsaturated at other isoprene units.[15-16] Despite these variations in structure, the broader conservation of the unique isoprene geometry in PrenPs across all domains of life suggests that the particular molecular structure of the PrenP plays a tailored role in glycoconjugate assembly, beyond that of an amphiphilic membrane-associated carrier on which glycan assembly can occur. Diverse roles for PrenPs in mediating substrate-enzyme binding and multi-enzyme complex formation, as well as in modulating the biophysical properties of the membrane, have been proposed previously.[17-22]
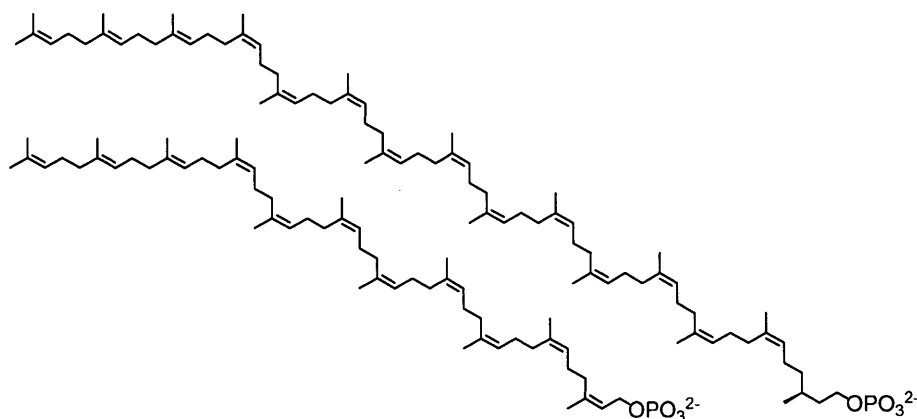
Figure 1-4 Structures of PrenP glycan carriers.
Structures of the C_{75} dolichol, utilized in eukaryotic glycoconjugate biosynthesis (*top*), and bacterial undecaprenol phosphate (UndP, *bottom*).

N-linked protein glycosylation in *Campylobacter jejuni*

Campylobacter jejuni is a Gram-negative commensal in the gastrointestinal tracks of poultry and cattle, but is pathogenic to humans and one of the leading causes of gastroenteritis worldwide.[23] The N-linked protein glycosylation (pgl) pathway in *C. jejuni* was the first N-linked protein glycosylation pathway identified in bacteria;[24] prior to its discovery, N-linked glycosylation was believed to occur exclusively in the eukaryotic and archaeal domains. The *C. jejuni* pathway is the most extensively characterized bacterial N-linked glycosylation pathway to date; however, since its discovery, similar N-linked pathways have been identified in several additional bacterial species,[25] many of which are also pathogenic. The *C. jejuni* N-linked glycan is hypothesized to be involved in key host-pathogen interactions at the cell surface, and the ability to synthesize and append assembled glycans from the pgl pathway to target proteins has been linked to colonization and virulence in *C. jejuni*.[26-28] Thus, a deeper understanding of the pgl pathway is of interest to human health insofar as it may inform the development of therapeutics against this gastrointestinal pathogen.[29-31]

The pgl pathway follows the general pattern of complex glycoconjugate assembly described in **Figure 1-3**. Glycan assembly is initiated by the PGT, PglC, which transfers an unusual bacterial sugar phosphate, phospho-*N,N'*-diacetylbacillosamine (diNAcBac), to PrenP[32] (**Figure 1-5**). The PrenPP-diNAcBac is elaborated by four downstream GTs (PglA, PglJ, PglH and PglI, **Figure 1-6**) to give the complete heptasaccharide[33] (**Figure 1-7**). The flippase PglK transfers the PrenPP-linked heptasaccharide across the inner membrane to the periplasm. The oligosacchryltransferase (OST), PglB, appends the heptasaccharide to acceptor proteins via Asn residues in target D/E-X-N-X-S/T sequons (where X is any residue except proline).
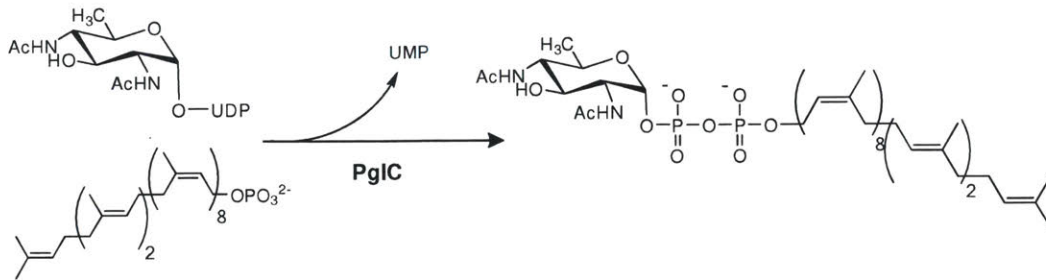
Figure 1-5 The PglC reaction.
The PGT, PglC, transfers phospho-diNAcBac from a soluble UDP donor to a PrenP chain embedded in the membrane to yield PrenPP-diNAcBac. UMP is released in the process.
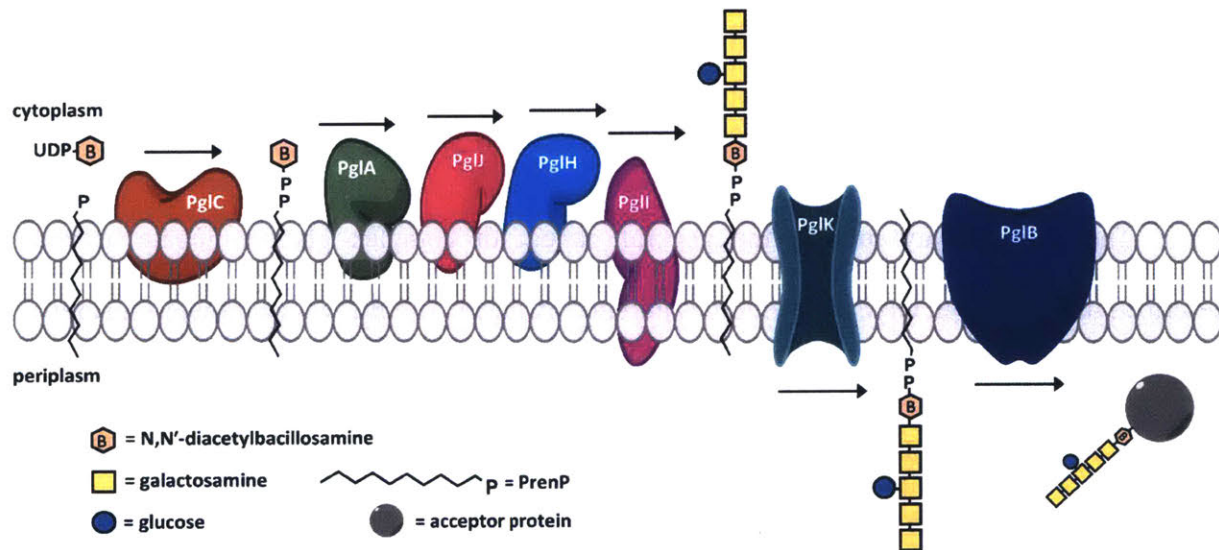
Figure 1-6 Schematic representation of the pgl pathway from *C. jejuni*.
Glycan assembly is initiated by the PGT, PglC, and completed sequentially by several GTs (PglA, PglJ, PglH and PglI). PglK is a flippase that transfers the completed glycan to the periplasm, where Asn residues on target proteins are glycosylated by the OST, PglB.
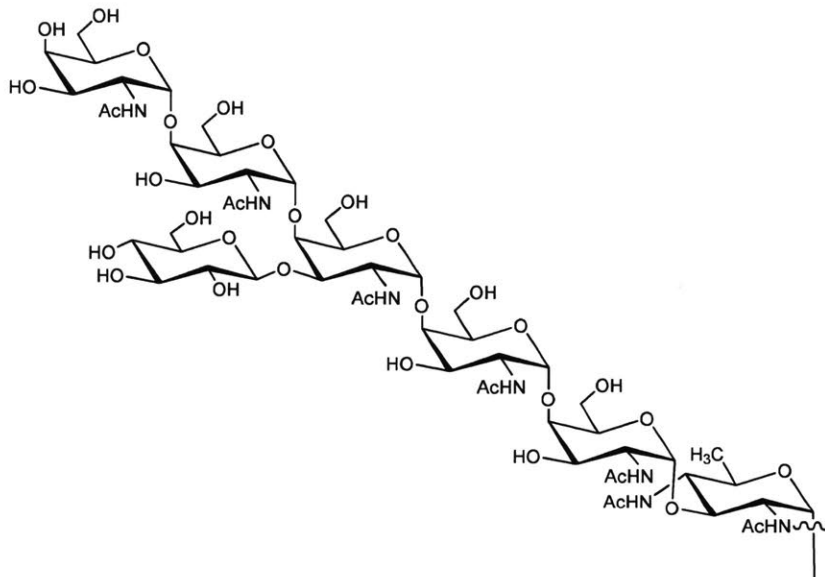
Figure 1-7 The heptasaccharide synthesized by the pgl pathway.
Five N-acetylgalactosamines and a branching glucose are appended to an unusual bacterial monosaccharide, *N,N'*-diacetylbacillosamine (diNAcBac), found at the reducing end of glycan.

The well-characterized pgl pathway also serves as a prototype for the study of many aspects of both prokaryotic and eukaryotic glycoconjugate biosynthesis. For example, the OST from the pgl pathway, PglB, is homologous to the archaeal OST AglB,[34-35] and to STT3, the catalytic subunit of the multi-subunit OST in eukaryotes.[36] Indeed, a recently reported structure of the OST complex from *Saccharomyces cerevisiae* underscored many key structural similarities between STT3 from *S. cerevisiae*, PglB homolog *Campylobacter lari*, and AglB from *Archaeoglobus fulgidus*.[36-38] In addition, the pgl pathway PGT, PglC, is a representative member of an extensive superfamily of PGTs acting at the initiation of the biosynthesis of many essential bacterial cell-wall glycoconjugates;[39] this superfamily is the lesser known of two PGT superfamilies, described

in more detail below. Thus, elucidation of PglC structure and function also sheds light on many additional glycoconjugate biosynthesis pathways. Finally, the pgl pathway serves as a more tractable system, relative to complex eukaryotic pathways, for investigations into additional conserved themes in glycoconjugate biosynthesis, such as the use of membrane-associated pathways for glycan assembly and the utility of PrenP substrates as glycan carriers. Due to the many commonalities between prokaryotic and eukaryotic glycan assembly (**Figure 1-3** and **Figure 1-4**), answers to such questions are anticipated to be broadly applicable to many diverse glycoconjugate biosynthesis pathways across all domains of life.

PglC is a representative member of the monotopic PGT superfamily

An inherent property of any integral membrane protein fold is its topology, which is defined by how many times the protein spans the membrane[40] (**Figure 2-1**). Polytopic membrane proteins span the membrane multiple times, while bitopic proteins span the membrane once. A third (and lesser known) category comprises monotopic proteins, which do not span the membrane, but are embedded via a reentrant domain that enters and exits on a single face of the membrane. Notably, monotopic membrane proteins are integrally associated with the membrane and typically penetrate past the headgroup region into the hydrophobic core of the lipid bilayer; as such, they are distinct from peripheral membrane proteins.
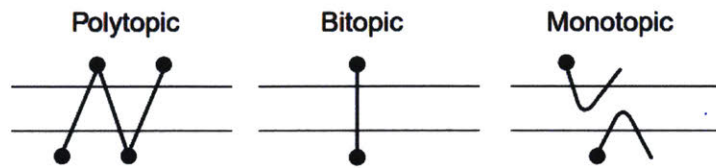
Figure 1-8 Classification of membrane protein topologies.

Integral membrane proteins can be classified by membrane topology: those with multiple transmembrane helices are termed polytopic, those with a single transmembrane helix are termed bitopic, and those that associate with, but do not span, the membrane are termed monotopic. Figure adapted from Krebs et al., 2004.[41]

The PGTs that act at the initiation of a diversity of glycoconjugate biosynthesis pathways can be grouped broadly into two superfamilies based on membrane topology. One superfamily is composed of polytopic PGTs (**Figure 1-9**) and is exemplified by bacterial MraY and WecA, which have 10- and 11- transmembrane helices, respectively, and active sites crafted from extended cytoplasmic inter-TM loops.[42] PGTs initiating the dolichol pathway for N-linked protein glycosylation in eukaryotes, such as Alg7 in *S. cerevisiae* and GPT in humans, belong to this superfamily.[43]
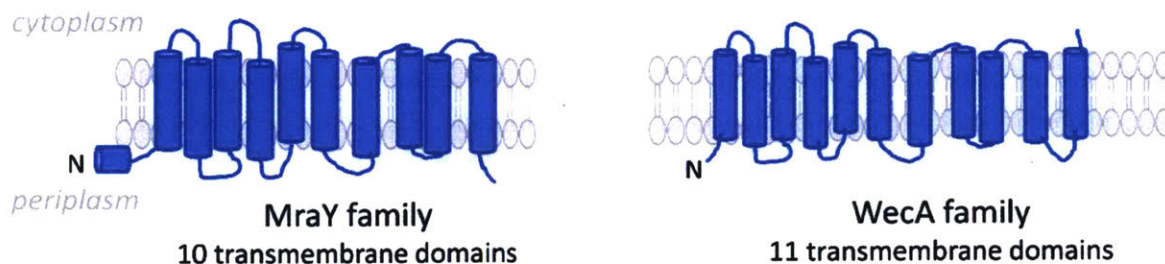
Figure 1-9 The polytopic superfamily of PGTs.

The polytopic superfamily of PGTs is exemplified by MraY and WecA, with 10- and 11-transmembrane segments, respectively.

PglC from the *C. jejuni* pgl pathway exemplifies a second superfamily of PGTs.[24, 32] The PGTs in this superfamily share a common functional core which is homologous to PglC and comprises a single N-terminal membrane-inserted domain and a small globular domain.[39] The superfamily also includes WbaP from *Escherichia coli*, which features a PglC-like core elaborated with four N-terminal transmembrane helices that are not necessary for catalytic activity,[44] and the bifunctional enzyme PglB from *Neisseria gonorrhoeae*, which includes a C-terminal acetyltransferase domain.[45] Based on a hydropathy analysis using a transmembrane hidden Markov model (TMHMM),[46] the N-terminal hydrophobic domain of PglC (and the analogous domain in other superfamily members) had previously been predicted to form a transmembrane helix, resulting in a bitopic topology for PglC (**Figure 1-10**).[9, 47]
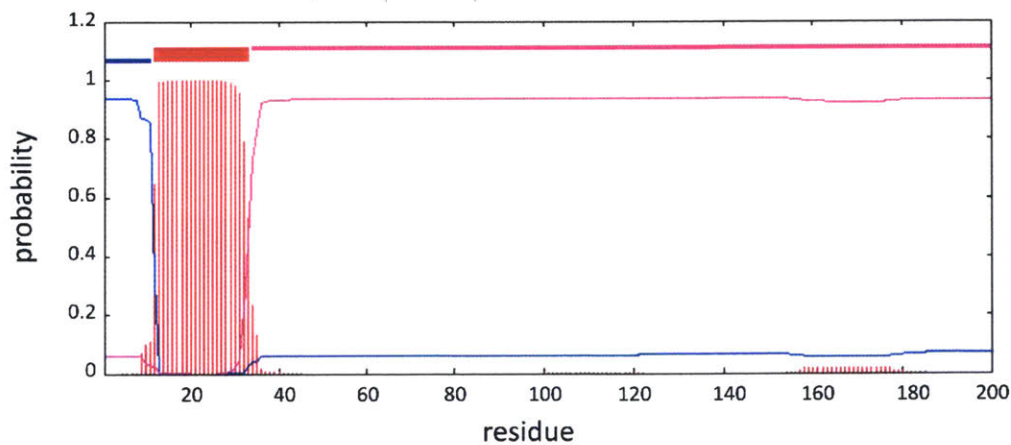
Figure 1-10 Hydropathy plot of PglC from *C. jejuni*.
A hydropathy analysis of PglC predicted a transmembrane N-terminal domain. Red coloring indicates probability of a residue being found in the membrane. Plot generated using the TMHMM[46] Server v. 2.0 (http://www.cbs.dtu.dk/services/TMHMM/).

Despite this predicted transmembrane topology, a biochemical analysis of WcaJ, a WbaP homolog, supported a reentrant topology, in which the domain enters and exits the membrane on the cytoplasmic side, rather than a membrane-spanning one.[47] We recently reported the x-ray crystal structure of PglC from *Campylobacter concisus*;[48] to date, the reported structure of PglC is the only experimentally-determined structure available for any member of the monotopic PGT superfamily. Both biochemical analyses and the reported structure of PglC confirmed a reentrant, monotopic topology for this minimalistic fold, suggesting that this topology is a shared feature of all the PGTs in the PglC-like superfamily. This PGT superfamily has thus been dubbed the "monotopic" PGT family in reference to the monotopic common catalytic unit (**Figure 1-11**), and to distinguish it from the polytopic PGT superfamily represented by MraY and WecA.
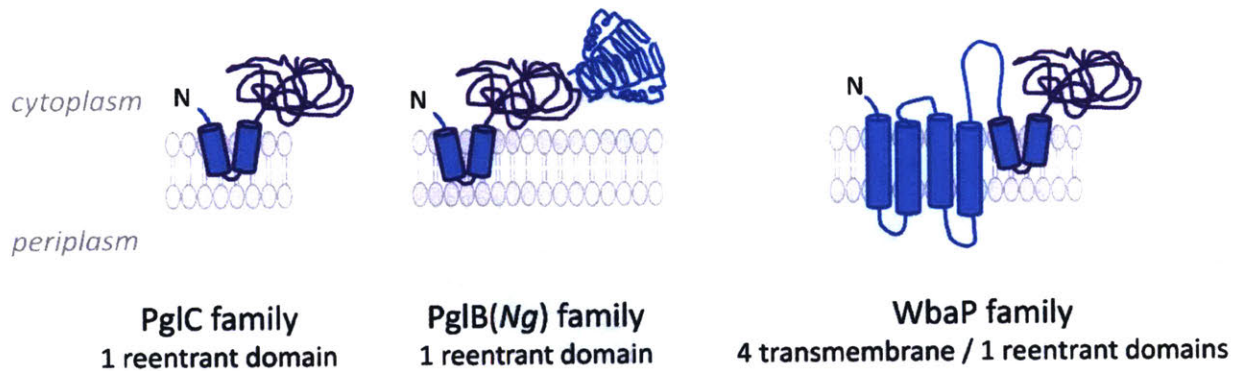
Figure 1-11 The monotopic superfamily of PGTs.

PglC is a representative member of a second, "monotopic" superfamily of PGTs, which shares a monotopic catalytic core, outlined in purple, homologous to PglC. The catalytic core is composed of an N-terminal reentrant membrane-resident domain and a C-terminal globular domain.

PglC exemplifies a novel mode of monotopic membrane association

The determined structure of PglC reveals a mode of membrane association not previously observed in known structures of monotopic membrane proteins.[48] The N-terminal hydrophobic domain forms a helix-break-helix motif, termed the reentrant membrane helix (RMH), that anchors the fold at a depth of 14 Å in the cytosolic face of the inner membrane (**Figure 1-12**). Three additional amphiphilic helices, one of which contains the conserved Asp-Glu catalytic dyad, interact with the RMH to position the active site of PglC at the membrane interface. Placement of the active site at the membrane interface is advantageous as it facilitates the transfer of phospho-*N,N'*-diNAcBac from a soluble nucleotide donor in the cytoplasm to the amphiphilic PrenP carrier embedded in the membrane (**Figure 1-5**).

14 Å
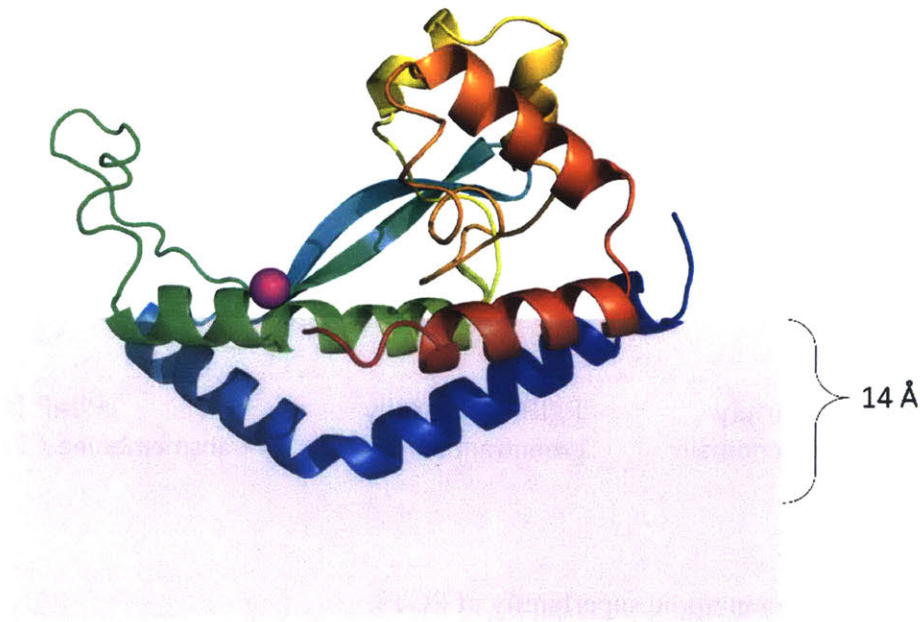
Figure 1-12 The structure of PglC.
PglC is shown colored from N-terminus (*blue*) to C-terminus (*red*) and positioned in the membrane (*gray*). Mg^{2+}, shown as a magenta sphere, denotes the position of the active site. The RMH (*blue*) interacts with three amphipathic helices to position the active site at the membrane interface.

The current understanding of modes of monotopic membrane association is limited by the scarcity of available structural information about monotopic proteins: a recent analysis found that monotopic membrane proteins account for only ~0.06% of nonredundant structures in the Protein Data Bank (PDB).[49] However, within the scope of known monotopic protein structures, the mode of membrane association exemplified by PglC is distinct from those commonly employed by other monotopic folds. The membrane association modes of several representative monotopic membrane proteins is illustrated in **Figure 1-13** and summarized in **Table 1-1**. Common modes include the use of amphipathic helices, hydrophobic loops, hydrophobic patches and electrostatic interactions with lipid head groups to embed folds shallowly (typically at a depth of <10 Å) in the lipid bilayer. Although monotopic membrane proteins typically catalyze reactions involving a membrane-resident substrate, in many cases, these folds include large, soluble domains that share homology with soluble proteins that perform a similar reaction on aqueous substrates. This suggests a strategy wherein soluble folds have evolved hydrophobic domains in order to facilitate function at the membrane. Indeed, such monotopic membrane proteins often use a similarly positioned active site as their soluble counterparts, even at the high energetic cost of having to extract hydrophobic substrates from the membrane via a hydrophobic "channel" to an active site located distally from the membrane.[49-50]
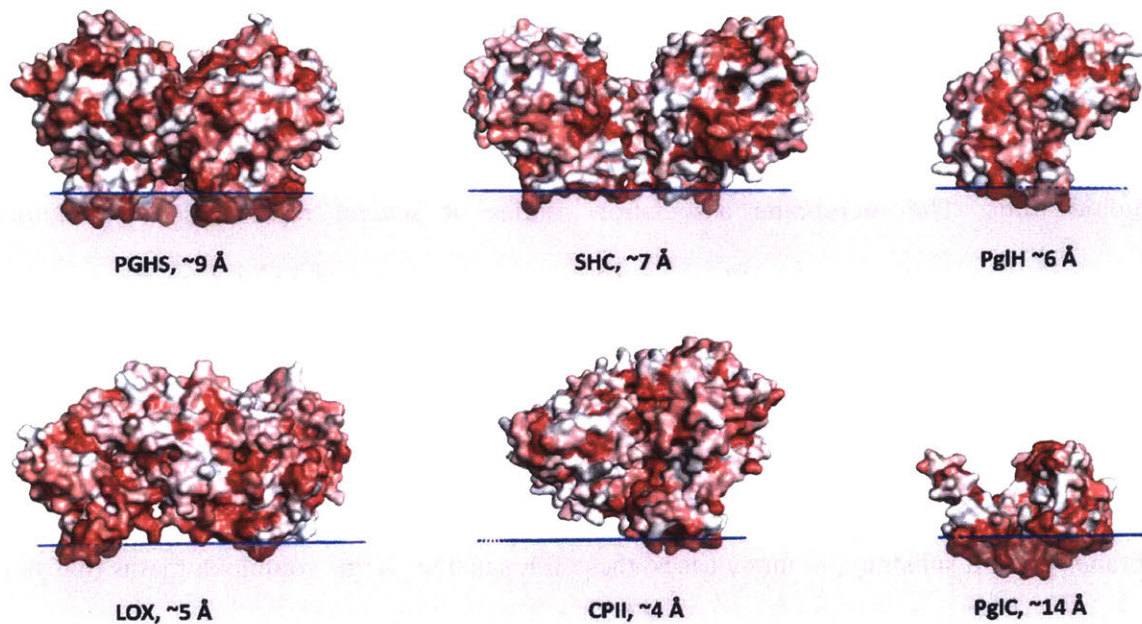
Figure 1-13 Common modes of monotopic membrane association.

The membrane insertion modes for several representative monotopic proteins and PglC are shown. Structures are shown in surface view colored by hydrophobicity from hydrophobic (*red*) to hydrophilic (*white*): prostaglandin-H synthase (PGHS, PDB 1CGE); squalene-hopene synthase (SHC, PDB 2SQC); the glycosyltransferase PglH (PDB 6EJI); 15-lipoxygenase (LOX, PDB 4NRE); carnitine palmitoyltransferase II (CTII, PDB 2H4T), PglC (PDB 5W7L). Structures were obtained in aqueous conditions, but for this figure have been positioned in a model membrane using the PPM server.[51] The membrane is represented in gray. The depth to which each fold is embedded in the membrane is shown.

Characteristic	Exemplified by...
Amphipathic helices	PGHS, CTII, **PglC**
Long hydrophobic loop	LOX
Hydrophobic patch surrounded by positively-charged lipids	SHC, PglH
Hydrophobic "channel" to active site	PGHS, SHC, CTII
Homology to soluble folds	PGHS, SHC, PglH, CTII

Table 1-1 Common modes of monotopic membrane association.
The various modes of membrane association exemplified in **Figure 1-13** are summarized.

By contrast, PglC has no known soluble homologs. Although PglC utilizes amphipathic helices similarly to some other monotopic folds, the use of a highly-ordered and deeply-penetrating reentrant helical domain, such as the RMH, for membrane association is unprecedented in the PDB. The use of an RMH and coordinating amphipathic helices to position the active site of PglC at the membrane interface also obviates the need for extraction of the PrenP substrate from the membrane, making catalysis more energetically favorable. Finally, unlike the hydrophobic domains of other monotopic membrane proteins, which primarily act in tethering their respective folds to the membrane, the PglC RMH has been hypothesized to play a significant role in PrenP binding.[17-18] Thus, PglC represents a novel mode of monotopic membrane association that expands the known structure-function relationship for monotopic membrane proteins.

PglC as a model for reentrant topology formation in diverse monotopic membrane proteins

PglC is the first structurally-characterized example of a monotopic protein embedding in the membrane via an RMH domain. However, several additional examples exist in the literature of biochemically-characterized monotopic proteins that similarly lack soluble homologs and appear to have an extensive, α-helical reentrant domain that supports function. For example, the structural membrane protein caveolin-1, the principle component of caveolae in eukaryotes, contains a single hydrophobic domain that adopts a reentrant α-helical hairpin;[52-53] this "wedge" structure likely results in the membrane curvature necessary for formation of caveolae. The eukaryotic membrane protein stomatin, which regulates ion channels in a cholesterol-dependent manner, is similarly anchored in the membrane via a hydrophobic hairpin domain;[54-55] this domain is thought to function in cholesterol binding.

The hydrophobic domains of PglC, caveolin and stomatin share certain sequence features. Most notably, each contains a single conserved proline residue that is important for both the topology and the function of the respective monotopic protein (**Figure 1-14**). In PglC, Pro24 was found to be necessary for catalytic activity, and is hypothesized to be involved in PrenP substrate binding.[56] In Chapter 2, we show that Pro24 is part of a two-residue motif responsible for the formation of a reentrant topology in PglC. In caveolin, mutation of Pro110 to Ala resulted in a membrane-spanning, rather than the native reentrant, topology and decreased cellular lipid accumulation and lipid droplet formation.[57] Later NMR studies further identified both Pro110 and the preceding Ile109 as residues necessary for the formation of the helix-break-helix structure of the hydrophobic domain.[53] Mutation of the conserved Pro47 in the hydrophobic domain of stomatin to Ala also resulted in a membrane-spanning topology and loss of cholesterol binding relative to wild-type.[55]

Figure 1-14 Similarities between the reentrant domains of PglC, caveolin, and stomatin. The RMH of PglC is shown with the α-helical peptide backbone in sticks (*top*). The PglC RMH and the putative reentrant domains of caveolin and stomatin (*bottom*) share conserved prolines (*orange*) reported to be important for both function and formation of the respective reentrant helix-break-helix domains. Human caveolin-1 UniProt Q03135, mouse stomatin UniProt P54116.

As proline is known to break α-helices,[58] it is probable that proline, with additional contributions from neighboring residues in some cases, facilitates the formation of reentrant topologies for the hydrophobic domains of PglC, caveolin and stomatin. As all three hydrophobic domains have been biochemically characterized to be reentrant α-helices of similar length and broken by a conserved proline residue, we postulate that the PglC, caveolin and stomatin folds may ultimately be found to share certain structural features. Furthermore, the additional

significance of all three conserved prolines for activity indicates that the structure of these reentrant domains is also intimately tied to function. However, as the only such fold structurally characterized to date, PglC serves as a prototype for structure-function analyses of RMH formation. In Chapter 2, we describe an in-depth analysis of two motifs in the RMH that act together to determine a reentrant topology for PglC.

Besides PglC, caveolin and stomatin, few examples exist in the literature of monotopic membrane proteins that have been structurally and/or biochemically determined to embed in the membrane via similar reentrant helix-break-helix domains. However, it is possible that this mode of membrane association may be more common than previously thought. It has been suggested that almost half of all integral membrane proteins in the human proteome have a bitopic membrane topology.[59] However, as topology assignment by current topology prediction tools is largely based on the assumption that hydrophobic sequence of a certain length are membrane-spanning,[46, 60] it is likely that among the abundance of predicted bitopic folds, there exist reentrant hydrophobic domains that have been erroneously annotated as membrane-spanning. Indeed, the hydrophobic domains of PglC, caveolin and stomatin were all initially assumed to be transmembrane helices[9, 52, 61] prior to biochemical validation that all three adopted reentrant topologies.

Monotopic membrane proteins have proven recalcitrant to structural characterization;[49] however, with recent advances in structure determination methodologies, including cryo-electron microscopy and the development of lipidic-cubic phase for x-ray crystallography, such structures may become more accessible. We anticipate that with continued biochemical and structural analyses of additional monotopic membrane proteins, mode of membrane association exemplified by the recently-reported structure of PglC will be found to be employed more broadly. In Chapter 3, we demonstrate that the principles of reentrant topology formation identified for PglC can be

leveraged towards the identification of reentrant helix-break-helix domains in additional monotopic folds previously predicted, based on their protein sequence, to have a bitopic topology.

PglC as a model for PrenP-substrate interactions

The polytopic and monotopic PGT superfamilies (**Figure 1-9** and **Figure 1-11**) represent two very different membrane topologies used by diverse glycoconjugate assembly pathways to execute the same phosphosugar transfer reaction. Both superfamilies of PGTs must recognize and bind PrenP substrate for catalysis; intriguingly, the monotopic PGTs must accomplish this using a single membrane-resident RMH domain. In addition, a common PrenP pool is used in the biosynthesis of multiple bacterial cell wall glycoconjugates (**Figure 1-2**), such that multiple bacterial PGTs must compete for PrenP for complex glycan assembly. As the minimal catalytic unit for the monotopic PGT superfamily, PglC serves as a valuable platform for investigations into substrate recognition and binding by the RMH domains of these PGTs.

Notably, the use of PrenP as a substrate in membrane-associated glycoconjugate biosynthesis pathways is conserved across all domains of life (**Figure 1-4**), and the unique isoprene geometry of PrenP has been hypothesized to be important for substrate binding.[17-18] In addition, a variety of roles for PrenP in facilitating multi-enzyme complex assembly and in modulating biophysical properties of the membrane have also been proposed.[17-22] The pgl pathway serves as a tractable system for investigations into the conserved role of PrenP in glycoconjugate assembly across domains. In Chapter 4, we interrogate the molecular basis for interactions between PrenP, the local membrane environment, and PglC; such insights shed new light on the roles of PrenP in facilitating diverse prokaryotic and eukaryotic glycoconjuage biosynthesis pathways.

It has been proposed that the enzymes in many bacterial glycan assembly pathways, such as those involved in peptidoglycan and teichoic acid biosynthesis, act as large multi-enzyme complexes surrounding a shared molecule of PrenP.[62-63] The enzymes of the pgl pathway were similarly suggested to function as a macromolecular complex.[64] In Chapter 5 we describe efforts to implement cell-free expression of PglC into lipid Nanodisc model membranes as a platform for probing multi-enzyme complex assembly in the pgl pathway, and for further investigations of PglC, or multi-enzyme complex, structure in a native-like membrane environment.

CONCLUSIONS

The biosynthesis pathways leading to many essential prokaryotic and eukaryotic glycoconjugates share a common logic and many conserved features, including initiation by "gatekeeper" PGTs and the subsequent assembly of complex glycans on a membrane-embedded PrenP carrier molecule. Although the universality of such features in glycoconjugate biosynthesis across domains of life is intriguing, a deeper understanding of the molecular basis for interactions between polyprenol phosphates, the local membrane environment, and the glycan assembly machinery remains elusive. The well-characterized pgl pathway from *C. jejuni* is of interest not only as a target for the development of therapeutics against this gastrointestinal pathogen, but also as a tractable functional homolog for the more intricate glycoconjugate biosynthesis pathways in eukaryotes.

The PGT that acts in initiation of the pgl pathway, PglC, is also the representative member of a large superfamily of bacterial PGTs that share a common PglC-like monotopic core. PglC was previously predicted to have a bitopic topology, but was recently shown to embed in the membrane via a reentrant helix-break-helix domain and three amphipathic helices. PglC is the first

structurally-characterized example of a protein with such a monotopic membrane-association modality. We anticipate that similar reentrant domains exist in a variety of membrane proteins, but are misannotated as transmembrane helices by current topology prediction methodologies. As such, PglC presents an opportunity to investigate the drivers of reentrant topology determination in monotopic membrane proteins; such insights are invaluable for the development of more sophisticated topology prediction algorithms.

In addition, PglC serves as a model for delineating the relationship between PrenP, the local membrane environment, and PrenP-binding enzymes, most notably PGTs. Such studies have the potential to address many unanswered questions regarding the molecular basis for PGT-PrenP interactions, as well as additional proposed roles for PrenP in complex glycan assembly. A deeper understanding of how a single PrenP pool is shared by multiple bacterial cell-wall glycoconjugate biosynthesis pathways is particularly relevant to the study of bacterial physiology, while the molecular interactions underpinning PrenP substrate specificity and binding within the context of the local membrane environment have broad significance for both prokaryotic and eukaryotic glycobiology.

REFERENCES

(1) Haltiwanger, R. S.; Lowe, J. B., Role of glycosylation in development. *Annu Rev Biochem* **2004**, *73*, 491-537.

(2) Hooper, L. V.; Gordon, J. I., Glycans as legislators of host-microbial interactions: spanning the spectrum from symbiosis to pathogenicity. *Glycobiology* **2001**, *11* (2), 1R-10R.

(3) Ley, K., The role of selectins in inflammation and disease. *Trends Mol Med* **2003**, *9* (6), 263-8.

(4) Christiansen, M. N.; Chik, J.; Lee, L.; Anugraham, M.; Abrahams, J. L.; Packer, N. H., Cell surface protein glycosylation in cancer. *Proteomics* **2014**, *14* (4-5), 525-46.

(5) Messner, P.; Schaffer, C.; Kosma, P., Bacterial cell-envelope glycoconjugates. *Adv Carbohydr Chem Biochem* **2013**, *69*, 209-72.

(6) Ray, A.; Cot, M.; Puzo, G.; Gilleron, M.; Nigou, J., Bacterial cell wall macroamphiphiles: pathogen-/microbe-associated molecular patterns detected by mammalian innate immune system. *Biochimie* **2013**, *95* (1), 33-42.

(7) Holgersson, J.; Gustafsson, A.; Breimer, M. E., Characteristics of protein-carbohydrate interactions as a basis for developing novel carbohydrate-based antirejection therapies. *Immunol Cell Biol* **2005**, *83* (6), 694-708.

(8) Tytgat, H. L.; Lebeer, S., The sweet tooth of bacteria: common themes in bacterial glycoconjugates. *Microbiol Mol Biol Rev* **2014**, *78* (3), 372-417.

(9) Lukose, V.; Walvoort, M. T. C.; Imperiali, B., Bacterial phosphoglycosyl transferases: initiators of glycan biosynthesis at the membrane interface. *Glycobiology* **2017**, *27* (9), 820-833.

(10) Eichler, J.; Guan, Z., Lipid sugar carriers at the extremes: The phosphodolichols Archaea use in N-glycosylation. *Biochim Biophys Acta Mol Cell Biol Lipids* **2017**, *1862* (6), 589-599.

(11) Eichler, J.; Imperiali, B., Stereochemical divergence of polyprenol phosphate glycosyltransferases. *Trends Biochem Sci* **2018**, *43* (1), 10-17.

(12) Higashi, Y.; Strominger, J. L.; Sweeley, C. C., Structure of a lipid intermediate in cell wall peptidoglycan synthesis: a derivative of a C55 isoprenoid alcohol. *Proc Natl Acad Sci U S A* **1967**, *57* (6), 1878-84.

(13) Adair, W. L., Jr.; Cafmeyer, N., Characterization of the Saccharomyces cerevisiae cis-prenyltransferase required for dolichyl phosphate biosynthesis. *Arch Biochem Biophys* **1987**, *259* (2), 589-96.

(14) Eggens, I.; Chojnacki, T.; Kenne, L.; Dallner, G., Separation, quantitation and distribution of dolichol and dolichyl phosphate in rat and human tissues. *Biochim Biophys Acta* **1983**, *751* (3), 355-68.

(15) Guan, Z.; Naparstek, S.; Kaminski, L.; Konrad, Z.; Eichler, J., Distinct glycan-charged phosphodolichol carriers are required for the assembly of the pentasaccharide N-linked to the Haloferax volcanii S-layer glycoprotein. *Mol Microbiol* **2010**, *78* (5), 1294-303.

(16) Kuntz, C.; Sonnenbichler, J.; Sonnenbichler, I.; Sumper, M.; Zeitler, R., Isolation and characterization of dolichol-linked oligosaccharides from Haloferax volcanii. *Glycobiology* **1997**, *7* (7), 897-904.

(17) Hartley, M. D.; Imperiali, B., At the membrane frontier: a prospectus on the remarkable evolutionary conservation of polyprenols and polyprenyl-phosphates. *Arch Biochem Biophys* **2012**, *517* (2), 83-97.

(18) Zhou, G. P.; Troy, F. A., 2nd, Characterization by NMR and molecular modeling of the binding of polyisoprenols and polyisoprenyl recognition sequence peptides: 3D structure of the complexes reveals sites of specific interactions. *Glycobiology* **2003**, *13* (2), 51-71.

(19) Kern, N. R.; Lee, H. S.; Wu, E. L.; Park, S.; Vanommeslaeghe, K.; MacKerell, A. D., Jr.; Klauda, J. B.; Jo, S.; Im, W., Lipid-linked oligosaccharides in membranes sample conformations that facilitate binding to oligosaccharyltransferase. *Biophys J* **2014**, *107* (8), 1885-1895.

(20) Valtersson, C.; van Duyn, G.; Verkleij, A. J.; Chojnacki, T.; de Kruijff, B.; Dallner, G., The influence of dolichol, dolichol esters, and dolichyl phosphate on phospholipid polymorphism and fluidity in model membranes. *J Biol Chem* **1985**, *260* (5), 2742-51.

(21) Wang, X.; Mansourian, A. R.; Quinn, P. J., The effect of dolichol on the structure and phase behaviour of phospholipid model membranes. *Mol Membr Biol* **2008**, *25* (6-7), 547-56.

(22) Janas, T.; Chojnacki, T.; Swiezewska, E.; Janas, T., The effect of undecaprenol on bilayer lipid membranes. *Acta Biochim Pol* **1994**, *41* (3), 351-8.

(23) Allos, B. M., Campylobacter jejuni Infections: update on emerging issues and trends. *Clin Infect Dis* **2001**, *32* (8), 1201-6.

(24) Szymanski, C. M.; Yao, R.; Ewing, C. P.; Trust, T. J.; Guerry, P., Evidence for a system of general protein glycosylation in Campylobacter jejuni. *Mol Microbiol* **1999**, *32* (5), 1022-30.

(25) Nothaft, H.; Szymanski, C. M., Protein glycosylation in bacteria: sweeter than ever. *Nat Rev Microbiol* **2010**, *8* (11), 765-78.

(26) Kelly, J.; Jarrell, H.; Millar, L.; Tessier, L.; Fiori, L. M.; Lau, P. C.; Allan, B.; Szymanski, C. M., Biosynthesis of the N-linked glycan in Campylobacter jejuni and addition onto protein through block transfer. *J Bacteriol* **2006**, *188* (7), 2427-34.

(27) Karlyshev, A. V.; Everest, P.; Linton, D.; Cawthraw, S.; Newell, D. G.; Wren, B. W., The Campylobacter jejuni general glycosylation system is important for attachment to human epithelial cells and in the colonization of chicks. *Microbiology* **2004**, *150* (Pt 6), 1957-64.

(28) Szymanski, C. M.; Burr, D. H.; Guerry, P., Campylobacter protein glycosylation affects host cell interactions. *Infect Immun* **2002**, *70* (4), 2242-4.

(29) Walvoort, M. T. C.; Lukose, V.; Imperiali, B., A Modular Approach to Phosphoglycosyltransferase Inhibitors Inspired by Nucleoside Antibiotics. *Chemistry-a European Journal* **2016,** *22* (11), 3856-3864.

(30) De Schutter, J. W.; Morrison, J. P.; Morrison, M. J.; Ciulli, A.; Imperiali, B., Targeting bacillosamine biosynthesis in bacterial pathogens: development of inhibitors to a bacterial amino-sugar acetyltransferase from Campylobacter jejuni. *J Med Chem* **2017,** *60* (5), 2099-2118.

(31) Madec, A. G. E.; Schocker, N. S.; Sanchini, S.; Myratgeldiyev, G.; Das, D.; Imperiali, B., Facile solid-phase synthesis and assessment of nucleoside analogs as inhibitors of bacterial UDP-sugar processing enzymes. *ACS Chem Biol* **2018,** *13* (9), 2542-2550.

(32) Glover, K. J.; Weerapana, E.; Chen, M. M.; Imperiali, B., Direct biochemical evidence for the utilization of UDP-bacillosamine by PglC, an essential glycosyl-1-phosphate transferase in the Campylobacter jejuni N-linked glycosylation pathway. *Biochemistry* **2006,** *45* (16), 5343-50.

(33) Glover, K. J.; Weerapana, E.; Imperiali, B., In vitro assembly of the undecaprenylpyrophosphate-linked heptasaccharide for prokaryotic N-linked glycosylation. *Proc Natl Acad Sci U S A* **2005,** *102* (40), 14255-9.

(34) Larkin, A.; Chang, M. M.; Whitworth, G. E.; Imperiali, B., Biochemical evidence for an alternate pathway in N-linked glycoprotein biosynthesis. *Nat Chem Biol* **2013,** *9* (6), 367-73.

(35) Magidovich, H.; Eichler, J., Glycosyltransferases and oligosaccharyltransferases in Archaea: putative components of the N-glycosylation pathway in the third domain of life. *FEMS Microbiol Lett* **2009,** *300* (1), 122-30.

(36) Lizak, C.; Gerber, S.; Numao, S.; Aebi, M.; Locher, K. P., X-ray structure of a bacterial oligosaccharyltransferase. *Nature* **2011,** *474* (7351), 350-5.

(37) Wild, R.; Kowal, J.; Eyring, J.; Ngwa, E. M.; Aebi, M.; Locher, K. P., Structure of the yeast oligosaccharyltransferase complex gives insight into eukaryotic N-glycosylation. *Science* **2018,** *359* (6375), 545-550.

(38) Matsumoto, S.; Shimada, A.; Nyirenda, J.; Igura, M.; Kawano, Y.; Kohda, D., Crystal structures of an archaeal oligosaccharyltransferase provide insights into the catalytic cycle of N-linked protein glycosylation. *Proc Natl Acad Sci U S A* **2013,** *110* (44), 17868-73.

(39) Lukose, V.; Luo, L. Q.; Kozakov, D.; Vajda, S.; Allen, K. N.; Imperiali, B., Conservation and covariance in small bacterial phosphoglycosyltransferases identify the functional catalytic core. *Biochemistry* **2015,** *54* (50), 7326-7334.

(40) Blobel, G., Intracellular protein topogenesis. *Proc Natl Acad Sci U S A* **1980,** *77* (3), 1496-500.

(41) Krebs, M. P.; Noorwez, S. M.; Malhotra, R.; Kaushal, S., Quality control of integral membrane proteins. *Trends Biochem Sci* **2004,** *29* (12), 648-55.

(42) Anderson, M. S.; Eveland, S. S.; Price, N. P., Conserved cytoplasmic motifs that distinguish sub-groups of the polyprenol phosphate:N-acetylhexosamine-1-phosphate transferase family. *FEMS Microbiol Lett* **2000**, *191* (2), 169-75.

(43) Burda, P.; Aebi, M., The dolichol pathway of N-linked glycosylation. *Biochim Biophys Acta* **1999**, *1426* (2), 239-57.

(44) Saldias, M. S.; Patel, K.; Marolda, C. L.; Bittner, M.; Contreras, I.; Valvano, M. A., Distinct functional domains of the Salmonella enterica WbaP transferase that is involved in the initiation reaction for synthesis of the O antigen subunit. *Microbiology* **2008**, *154* (Pt 2), 440-53.

(45) Hartley, M. D.; Morrison, M. J.; Aas, F. E.; Borud, B.; Koomey, M.; Imperiali, B., Biochemical characterization of the O-linked glycosylation pathway in Neisseria gonorrhoeae responsible for biosynthesis of protein glycans containing N,N'-diacetylbacillosamine. *Biochemistry* **2011**, *50* (22), 4936-48.

(46) Krogh, A.; Larsson, B.; von Heijne, G.; Sonnhammer, E. L., Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J Mol Biol* **2001**, *305* (3), 567-80.

(47) Furlong, S. E.; Ford, A.; Albarnez-Rodriguez, L.; Valvano, M. A., Topological analysis of the Escherichia coli WcaJ protein reveals a new conserved configuration for the polyisoprenyl-phosphate hexose-1-phosphate transferase family. *Sci Rep* **2015**, *5*, 9178.

(48) Ray, L. C.; Das, D.; Entova, S.; Lukose, V.; Lynch, A. J.; Imperiali, B.; Allen, K. N., Membrane association of monotopic phosphoglycosyl transferase underpins function. *Nat Chem Biol* **2018**, *14*, 538-541.

(49) Allen, K. N.; Entova, S.; Ray, L. C.; Imperiali, B., Monotopic membrane proteins join the fold. *Trends Biochem Sci* **2019**, *44* (1), 7-20.

(50) Bracey, M. H.; Cravatt, B. F.; Stevens, R. C., Structural commonalities among integral membrane enzymes. *FEBS Lett* **2004**, *567* (2-3), 159-65.

(51) Lomize, M. A.; Pogozheva, I. D.; Joo, H.; Mosberg, H. I.; Lomize, A. L., OPM database and PPM web server: resources for positioning of proteins in membranes. *Nucleic Acids Res* **2012**, *40*, D370-6.

(52) Monier, S.; Parton, R. G.; Vogel, F.; Behlke, J.; Henske, A.; Kurzchalia, T. V., VIP21-caveolin, a membrane protein constituent of the caveolar coat, oligomerizes in vivo and in vitro. *Mol Biol Cell* **1995**, *6* (7), 911-27.

(53) Lee, J.; Glover, K. J., The transmembrane domain of caveolin-1 exhibits a helix-break-helix structure. *Biochim Biophys Acta* **2012**, *1818* (5), 1158-64.

(54) Salzer, U.; Ahorn, H.; Prohaska, R., Identification of the phosphorylation site on human erythrocyte band 7 integral membrane protein: implications for a monotopic protein structure. *Biochim Biophys Acta* **1993**, *1151* (2), 149-52.

(55) Kadurin, I.; Huber, S.; Grunder, S., A single conserved proline residue determines the membrane topology of stomatin. *Biochem J* **2009**, *418* (3), 587-94.

(56) Lukose, V.; Luo, L.; Kozakov, D.; Vajda, S.; Allen, K. N.; Imperiali, B., Conservation and Covariance in Small Bacterial Phosphoglycosyltransferases Identify the Functional Catalytic Core. *Biochemistry* **2015**, *54* (50), 7326-34.

(57) Aoki, S.; Thomas, A.; Decaffmeyer, M.; Brasseur, R.; Epand, R. M., The role of proline in the membrane re-entrant helix of caveolin-1. *J Biol Chem* **2010**, *285* (43), 33371-80.

(58) Piela, L.; Nemethy, G.; Scheraga, H. A., Proline-induced constraints in alpha-helices. *Biopolymers* **1987**, *26* (9), 1587-600.

(59) Almen, M. S.; Nordstrom, K. J.; Fredriksson, R.; Schioth, H. B., Mapping the human membrane proteome: a majority of the human membrane proteins can be classified according to function and evolutionary origin. *BMC Biol* **2009**, *7*, 50.

(60) Tsirigos, K. D.; Peters, C.; Shu, N.; Kall, L.; Elofsson, A., The TOPCONS web server for consensus prediction of membrane protein topology and signal peptides. *Nucleic Acids Res* **2015**, *43* (W1), W401-7.

(61) Stewart, G. W.; Hepworth-Jones, B. E.; Keen, J. N.; Dash, B. C.; Argent, A. C.; Casimir, C. M., Isolation of cDNA coding for an ubiquitous membrane protein deficient in high Na+, low K+ stomatocytic erythrocytes. *Blood* **1992**, *79* (6), 1593-601.

(62) Anderson, R. G.; Hussey, H.; Baddiley, J., The mechanism of wall synthesis in bacteria. The organization of enzymes and isoprenoid phosphates in the membrane. *Biochem J* **1972**, *127* (1), 11-25.

(63) Leaver, J.; Hancock, I. C.; Baddiley, J., Fractionation studies of the enzyme complex involved in teichoic acid synthesis. *J Bacteriol* **1981**, *146* (3), 847-52.

(64) Hartley, M. D.; Schneggenburger, P. E.; Imperiali, B., Lipid bilayer nanodisc platform for investigating polyprenol-dependent enzyme interactions and activities. *Proc Natl Acad Sci U S A* **2013**, *110* (52), 20863-70.

Chapter 2: Two key structural motifs determine the membrane topology of PglC

The work presented in this chapter was performed in collaboration with Dr. Jean-Marc Billod, a graduate student in the lab of Prof. Sonsoles Martín-Santamaría at the CIB-CSIS in Madrid, Spain. Prof. Martín-Santamaría and Dr. Billod performed all of the molecular dynamics experiments described herein.

INTRODUCTION

Membrane proteins represent an essential and diverse component of the proteome. Our understanding of how integral membrane proteins are folded and inserted into the membrane continues to evolve with the development of more sophisticated structural, biochemical, and computational analytical tools. The topology of integral membrane proteins is defined by how many times, and in which direction, the sequence spans the lipid bilayer: polytopic membrane proteins span the membrane multiple times, bitopic membrane proteins span the membrane a single time, and monotopic membrane proteins do not span the membrane, but instead are embedded in the membrane via a reentrant domain that enters and exits on a single face of the membrane (**Figure 2-1**).[1]
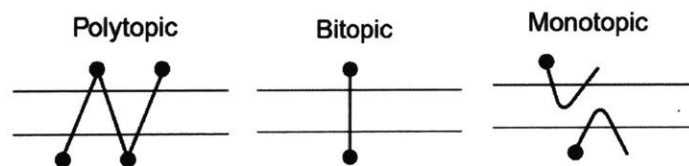
Figure 2-1 Classification of membrane protein topologies.
Integral membrane proteins can be classified by membrane topology. Proteins with multiple membrane-spanning helices are termed polytopic, those with one are termed bitopic, and those that do not span the membrane are termed monotopic. Figure adapted from Krebs et al., 2004.[2]

Current bioinformatics approaches[3-5] enable a relatively reliable prediction of transmembrane helix topology in polytopic and bitopic membrane proteins on the basis of hydrophobicity, homology with known protein structures, and the "positive-inside rule," by which membrane proteins have been empirically determined to preferentially adopt topologies that

position positively charged residues at the cytoplasmic face of the membrane,[6-8] among other parameters. By comparison, current knowledge of monotopic topologies is relatively limited. In particular, the key determinants that distinguish reentrant domains in monotopic proteins from transmembrane helices in bitopic proteins, to result in the two distinct topologies, are poorly understood. In this chapter we present an in-depth analysis of two conserved motifs that are key determinants of a monotopic topology in PglC, a membrane-bound phosphoglycosyl transferase (PGT) enzyme with a single reentrant domain. We anticipate that these two motifs embody common themes in reentrant domain formation across diverse monotopic membrane proteins, and in Chapter 3 we will demonstrate how they can used to distinguish such domains from transmembrane helices on a protein sequence level.

PGTs are integral membrane proteins that initiate a wide variety of essential glycoconjugate biosynthesis pathways, including peptidoglycan and N-linked glycan biosynthesis, by catalyzing the transfer of a phosphosugar from a sugar nucleoside diphosphate donor to a membrane-resident polyprenol phosphate. PGTs can be grouped broadly into two superfamilies based on their membrane topology.[9] One superfamily is exemplified by bacterial MraY and WecA, (**Figure 1-9**) and is composed of polytopic PGTs with 10- and 11- transmembrane helices respectively.[10] This family also includes the PGTs responsible for initiating the dolichol pathway for N-linked protein glycosylation in eukaryotes.[11] A second superfamily is exemplified by the monotopic enzyme, PglC, from the Gram-negative bacterium *Campylobacter jejuni*.[12-13] The PGTs in this superfamily share a common functional core, which is homologous to PglC and comprises a single N-terminal membrane-inserted domain and a small globular domain.[14] The superfamily also includes WbaP, which features a PglC-like core elaborated with four N-terminal transmembrane helices that are not necessary for catalytic activity,[15] and the bifunctional enzyme

PglB from *Neisseria gonorrhoeae*, which has an additional C-terminal acetyltransferase domain[16] (**Figure 1-11**). Topology predictions using multiple algorithms suggested that the N-terminal hydrophobic domain of PglC, and the analogous domain in other superfamily members, forms a transmembrane helix, such that the N-terminus is in the periplasm and the globular domain in the cytoplasm.[9, 17] However, biochemical analysis of WcaJ, a WbaP homolog, supported a model wherein both termini of the corresponding domain of WcaJ are in the cytoplasm, forming a reentrant topology, rather than a membrane-spanning one.[17]

Recently, we reported the X-ray structure of PglC from *Campylobacter concisus* to a resolution of 2.7 Å (PDB 5W7L).[18] The structural analysis complements previous biochemical studies on PglC from *C. jejuni*; the homologs share 72% sequence identity. In the reported structure, the N-terminal domain of PglC forms a reentrant helix-break-helix structure, termed the reentrant membrane helix (RMH), such that both the N-terminus and the globular domain, which includes the active site, are on the cytoplasmic face of the inner membrane (**Figure 2-2**). The RMH, as the only membrane-resident domain of PglC, plays a crucial role in anchoring PglC in the membrane. The RMH also interacts with a coplanar triad of amphipathic helices to position the active site of PglC at the membrane-water interface, thereby enabling efficient phosphosugar transfer from a soluble nucleotide diphosphate donor to the membrane-resident polyprenol phosphate acceptor.

cytoplasm

periplasm
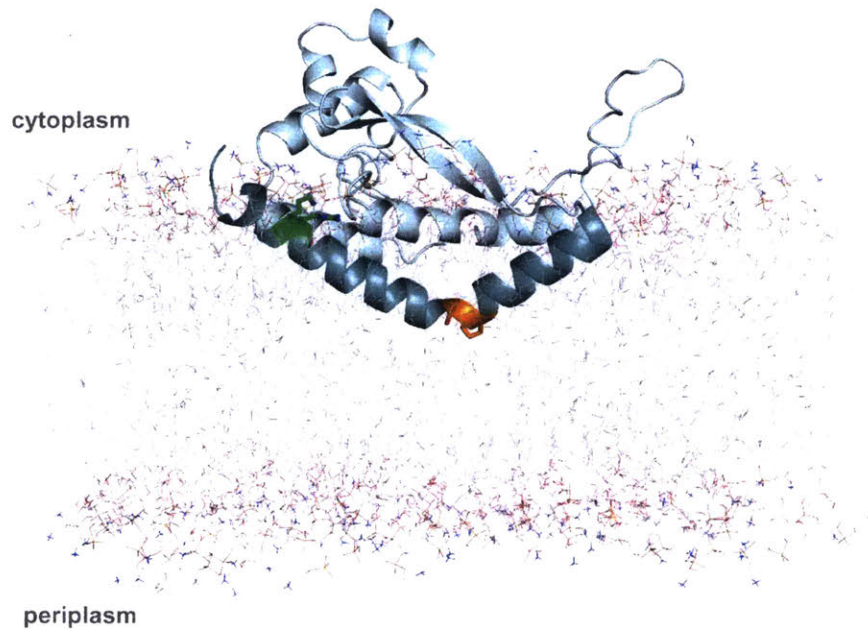
Figure 2-2 Overview of PglC highlighting the two conserved motifs in the RMH. Model of PglC from *C. concisus* in a 1-palmitoyl-2-oleoyl-*sn*-glycero-3-phosphoethanolamine (POPE) lipid bilayer. The RMH is shown in teal, Lys7 and Arg8 as green sticks, Ser23 and Pro24 as orange sticks.

The RMH is composed of an α-helix broken at a 118° angle by a conserved Ser23-Pro24 dyad (**Figure 2-3**). In the reported structure, Pro24 disrupts the hydrogen-bonding network of the RMH backbone, creating a break in the helix. This break is stabilized in turn by the orientation of the Ser23 side chain, which forms a 2.6-Å hydrogen bond with the backbone carbonyl of Ile20, thus satisfying one of the backbone hydrogen bonds lost due to Pro24. Pro24 is highly conserved among PglC homologs, and has been previously shown to be essential for PglC activity.[14] At the N-terminus of PglC there is a similarly conserved Lys7-Arg8 dyad. Lys7 makes short-range contacts with the C-terminus of the globular domain via residue Asp169, and Arg8 interacts with the headgroup of a co-purified phospholipid molecule (**Figure 2-3**).
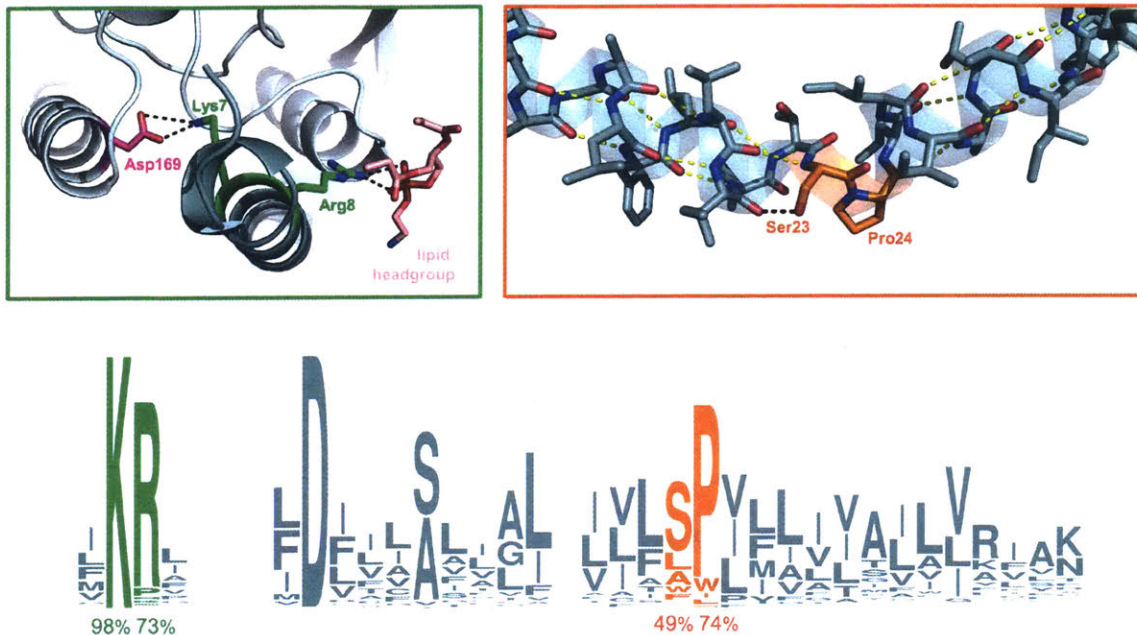
Figure 2-3 Detailed view of two conserved motifs in the PglC RMH.
Top left: Lys7 forms a salt bridge with Asp169, shown in magenta. Arg8 interacts with the headgroup of a copurified lipid, shown in salmon. *Top right:* Pro24 disrupts the hydrogen-bonding network (*yellow dashes*) of the RMH backbone. A 2.6 Å-hydrogen bond (*black dashes*) is formed between the hydroxyl side group of Ser23 and the backbone carbonyl of Ile20. *Bottom:* a sequence logo shows conservation in the RMH domain among PglC homologs. Percent conservation is noted below each residue of interest. Logo generated using weblogo.berkeley.edu.

Most structurally-characterized examples of ordered reentrant domains occur in polytopic membrane proteins in which topology is largely defined by the presence of multiple transmembrane helices; often times, such domains lack secondary structure[19] (**Figure 2-4**). In contrast, the RMH of monotopic PglC is the sole determinant of the reentrant topology and entirely α-helical. Thus, PglC provides a unique opportunity to identify structural motifs that influence helix geometry such that membrane-inserted domains favor a reentrant topology over a transmembrane one. Many of the structurally-characterized monotopic membrane proteins

associate with the membrane through hydrophobic loops or amphipathic helices[20] (**Figure 1-13**). However, PglC is currently the only example of a structurally-characterized monotopic membrane protein that associates with the membrane via a highly-ordered, reentrant helix-break-helix motif that significantly penetrates the hydrophobic membrane core.
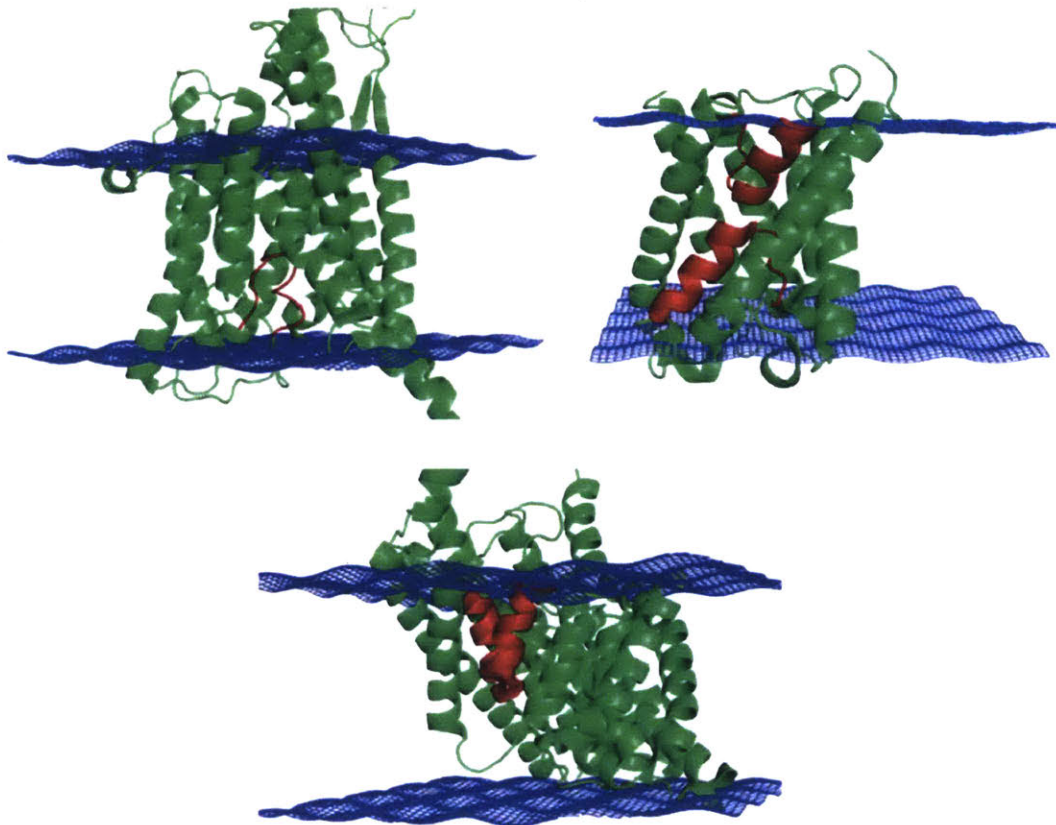
Figure 2-4 Reentrant domains in polytopic membrane proteins often lack secondary structure. Reentrant domains in polytopic membrane proteins, shown in red, typically lack secondary structure completely (*top left*) or partially (*top right*). Instances of entirely α-helical reentrant domains (*bottom*) are less common. Topology determination in such folds is driven primarily by the flanking transmembrane helices. Blue mesh indicates the location of the membrane. Structures shown: Sec61 (*top left*, PDB 1RHZ), aquaporin Z (*top right*, 1RC2), and ClC chloride channel (*bottom*, 1OTS). Figure adapted from Viklund et al., 2006.[19]

Current membrane topology prediction algorithms typically default to the assumption that hydrophobic α-helices of a certain length range are membrane-spanning,[3-5] making the prediction of monotopic topologies with reentrant α-helices challenging. Thus, delineating drivers of reentrant topology would enable accurate predictions of this topology among membrane proteins, and would provide insight into other membrane proteins similarly identified as having a single membrane-inserted domain, including the eukaryotic scaffolding protein caveolin-1, known to be monotopic,[21-22] and the mammalian membrane-bound enzyme diacylglycerol kinase ε.[23] Diacylglycerol kinase ε, like PglC, has a single N-terminal hydrophobic domain, but its membrane topology remains unclear.[24-25] In the current study, we apply *in vivo*, biochemical, and computational methodologies to identify two conserved motifs that both drive formation of the RMH in PglC and contribute to the stability of the final fold, enforcing a monotopic topology for PglC. The significance of proper RMH formation for catalytic activity is also discussed.

RESULTS

Determining native PglC topology using the substituted cysteine accessibility method

Prior to the availability of structural information about PglC from x-ray crystallography, we sought

to determine the topology of the N-terminal hydrophobic domain *in vivo* using the substituted

cysteine accessibility method (SCAM).[26] Although the N-terminal domain was predicted to form

a transmembrane helix, previous analysis using of the corresponding domain in a WbaP monotopic

PGT homolog had suggested a reentrant topology.[17] Our understanding of the native topology of

PglC later provided context for the physiological relevance of the determined structure.[18]

For topology analysis by SCAM, the subcellular localization of unique cysteines

introduced into a target protein is determined by whether such cysteines react *in vivo* with the cell-

permeable reagent N-ethylmaleimide (NEM) or with 2-sulfonatoethyl methanethiosulfonate

(MTSES), permeable only to the outer membrane (**Figure 2-5**). Reaction with either reagent

"protects" the cysteine from subsequent labeling by methoxypolyethylene glycol maleimide (PEG-

mal, MW 5 kDa) following cell lysis. PEGylation is determined by a 5-10 kDa shift to a higher

molecular weight band in Western blot analysis relative to the native protein. Cysteines located in

the periplasm are thus distinguished from cytoplasmic cysteines by protection of the former from

PEGylation by both NEM and MTSES, whereas cytoplasmic cysteines are PEGylated following

treatment with MTSES but not following treatment with NEM. PglC from *C. jejuni* strain 11168

was used for SCAM analysis and *in vitro* assays described herein; corresponding residues in PglC

from *C. jejuni* and *C. concisus* are listed in Experimental Methods, **Table 2-1**. Unique cysteines

were introduced either N-terminally (K4C, F6C) or C-terminally (S88C, S186C) to the membrane-

associated domain at non-conserved sites predicted previously to be surface-exposed.[14]
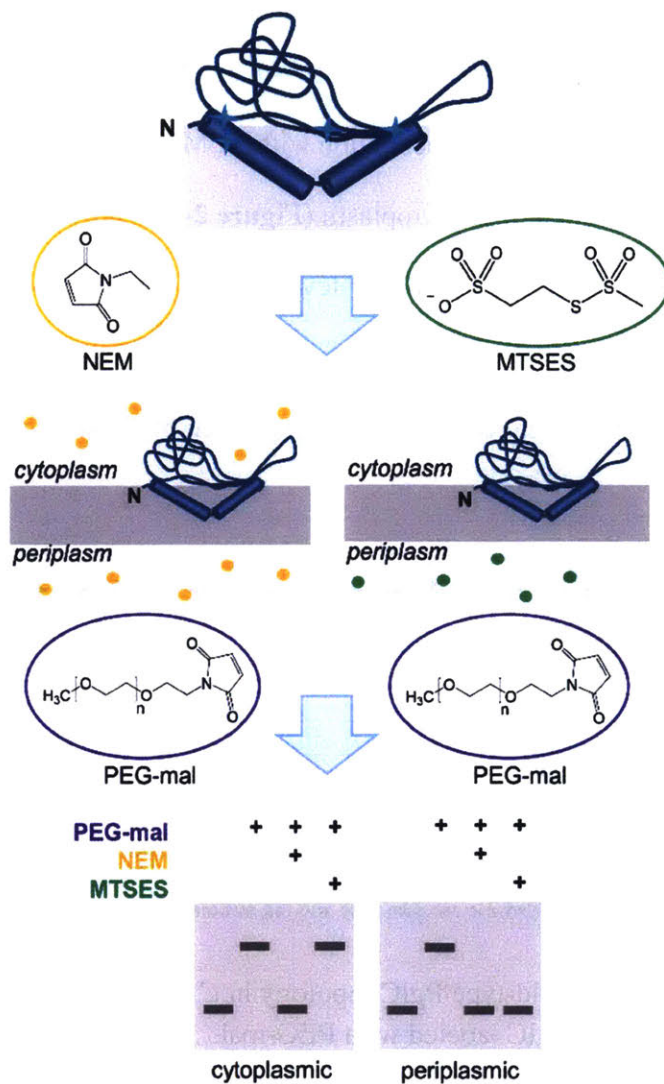
Figure 2-5 The substituted cysteine accessibility method (SCAM).
A schematic representation of the SCAM analysis used to assess the native topology of PglC.
Cyan starbursts (*top*) indicate the location of unique cysteines introduced into PglC.

In preliminary SCAM analyses, PglC was expressed heterologously in *C. jejuni* strain 81-176 under a native promoter, and subjected to labeling as described above. Wild-type PglC has no native cysteines, and thus was not labeled with PEG-mal. All four cysteine-substituted variants were blocked from PEG-mal labeling by incubation with NEM but not by incubation with MTSES, indicating that all four are located in the cytoplasm (**Figure 2-6**). These results suggest a reentrant topology for PglC; however, we noted that low levels of PglC expression in *C. jejuni* and poor PEG-mal labeling efficiency significantly complicated the SCAM analysis. Thus, analysis of PglC was also performed on variants heterologously overexpressed in *Escherichia coli*.
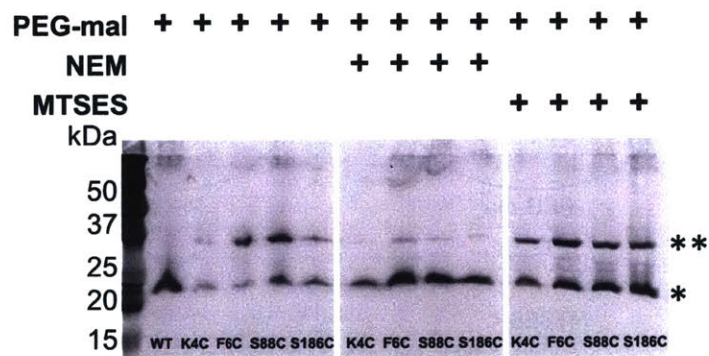
Figure 2-6 SCAM analysis of wild-type PglC topology in *C. jejuni*.
Legend: * = native PglC; ** = PglC labeled with PEG-mal.
All SCAM experiments were performed in duplicate or more. Representative Western blots are shown.

The high expression levels of PglC overexpressed in *E. coli* simplified analysis and also appeared to improve the efficiency of labeling by PEG-mal. All four cysteine-substituted variants were again blocked from PEG-mal labeling by incubation with NEM but not by incubation with MTSES, indicating a cytoplasmic location (**Figure 2-7**). A unique cysteine variant of a periplasmic protein, PEB3 A204C, served as a positive control for thiol-blocking of cysteines in the periplasm of *E. coli* (**Figure 2-8**). Thus, SCAM analysis confirms a reentrant topology for PglC expressed both in *C. jejuni* and in *E. coli*. All subsequent SCAM analyses were performed on variants overexpressed in *E. coli*.

Figure 2-7 SCAM analysis of wild-type PglC topology in *E. coli*.
Legend: * = native PglC; ** = PglC labeled with PEG-mal; C = control, no PEG-mal labeling.
All SCAM experiments were performed in duplicate or more. Representative Western blots are shown.

Figure 2-8 PEB3 A204C positive control from SCAM analysis.
Legend: * = native PglC; ** = PglC labeled with PEG-mal; C = control, no PEG-mal labeling.
All SCAM experiments were performed in duplicate or more. Representative Western blots are shown.

The *in vitro* experiments described in this chapter utilize SUMO-PglC variants. Incorporation of an N-terminal SUMO tag has been previously reported to aid greatly in purification of PglC[14, 18, 27] and SUMO-PglC has been confirmed to be catalytically active.[14, 27] For this reason, it was important to determine whether incorporation of the N-terminal SUMO tag alters PglC topology. SCAM analysis of SUMO-PglC was performed using the same four cysteine substitutions as for wild-type PglC. SUMO-PglC was found to adopt a native reentrant topology (**Figure 2-9**); thus, incorporation of the SUMO-tag does not alter PglC topology.

Figure 2-9 SCAM analysis of SUMO-PglC topology.
Legend: * = native PglC; ** = PglC labeled with PEG-mal; C = control, no PEG-mal labeling.
All SCAM experiments were performed in duplicate or more. Representative Western blots are shown.

58

Catalytic activity of all four cysteine-substituted PglC variants used in SCAM analysis was also confirmed using SUMO-PglC variants. Importantly, all four SUMO-PglC variants were found to retain 10-50% catalytic activity relative to wild-type SUMO-PglC (**Figure 2-10**), indicating that the four cysteine substitutions used for SCAM analysis permit near-native folding of PglC.
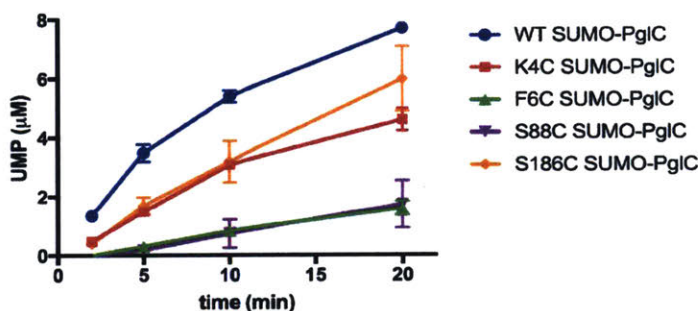
Figure 2-10 Activity assays of PglC variants used for SCAM analysis.
Activity assays of wild-type SUMO-PglC and K4C, F6C, S88C and S186C SUMO-PglC using UMP-Glo® (Promega, Madison, WI) to monitor UMP release. Error bars are given for mean ± SD, n = 2.

The reported structure of PglC is stable in a membrane environment

The crystal structure of PglC suggests that the N-terminal RMH anchors the fold in the membrane.[18] However, as the reported structure was generated using detergent-solubilized PglC in an aqueous environment, we applied molecular dynamics (MD) simulations to investigate whether the structure would be stable in a more native membrane-like environment. A model of the reported structure of PglC from *C. concisus* in a 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphoethanolamine (POPE) membrane was generated computationally[28] (**Figure 2-2**). The resulting model confirms the interaction of PglC with the membrane via the RMH, which

penetrates the membrane through one of the lipid leaflets. This model was further submitted to

MD simulations to assess the stability of PglC in a membrane environment. Over 400 ns of

simulation time, RMSD values within ~1 Å were measured for PglC (**Figure 2-11**), indicating that

the fold is stable in a simulated membrane. These analyses support that the conformation of PglC

observed in the crystal structure reflects a native state in a lipid bilayer environment.
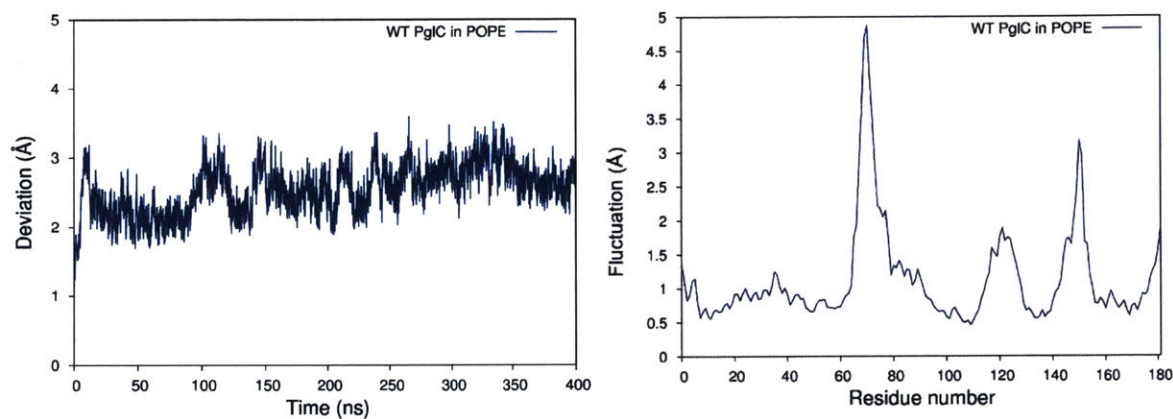
Figure 2-11 PglC is stable in a model POPE lipid bilayer.
RMSD (root mean square deviation) values measured over 400 ns of MD simulations of PglC in a model POPE membrane (*left*) indicated deviations within ~1 Å for the fold, suggesting that the reported PglC fold is stable in a membrane environment. RMSF (root mean square fluctuation) averaged over the 400 ns of MD simulations (*right*) indicated that most of the fold, with the exception of a few flexible loop regions, experiences little movement.

A conserved Ser-Pro motif is a key determining factor of reentrant topology in PglC

The conservation of the Ser-Pro motif and its location at the break in the RMH suggested that these residues might play an important role in establishing reentrant topology in PglC. Accordingly, peptides spanning the RMH from *C. jejuni* were submitted to the PepLook peptide folding prediction tool.[29] Peptides representing the RMH of wild-type, S23A and P24A PglC all adopted helix-break-helix conformations suggestive of a reentrant topology (**Figure 2-12**). An overlay of all three peptides found very little difference in their overall conformations, with the exception of a loss of secondary structure between Leu18 and Thr22 in S23A PglC.

RIFDILGALFLLILT<u>SP</u>IIIATAIFIK
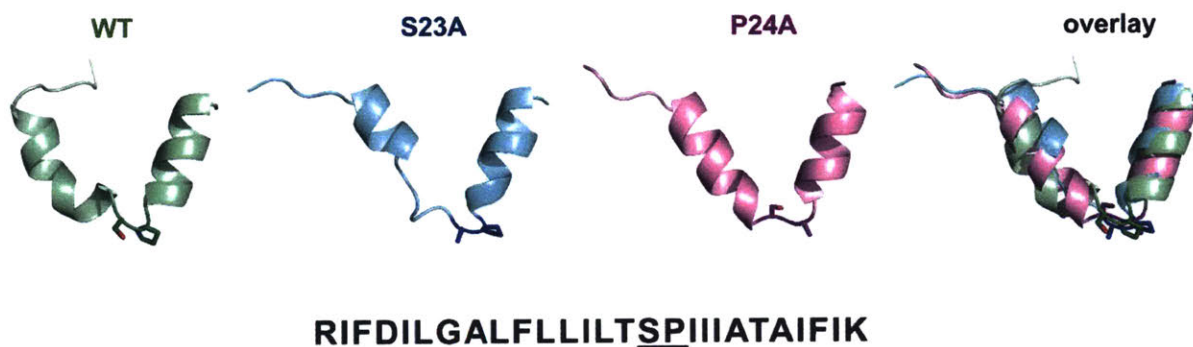
Figure 2-12 PepLook folding of wild-type, S23A, and P24A PglC RMH peptides.
All three peptides adopted similar helix-break-helix conformations (*overlay*). Ser23 and Pro24 in each peptide (or alanine substitutions thereof) are shown as sticks and highlighted in a darker color. The wild-type RMH peptide sequence is shown with Ser23 and Pro24 underlined.

By contrast, PepLook folding of a S23A/P24A peptide predicted a very different conformation from wild-type and single alanine-substituted peptides (**Figure 2-13**). Rather than a helix-break-helix, the S23A/P24A peptide was predicted to fold into a continuous α-helix, reminiscent of a transmembrane helix. These preliminary findings suggested that Ser23 and Pro24 might be working synergistically to establish a helix-break-helix conformation for the PglC RMH.
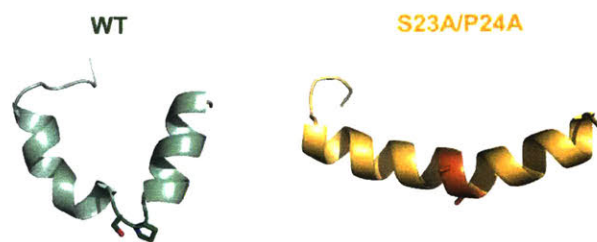
Figure 2-13 PepLook folding of an S23A/P24A PglC RMH peptide.
Ser23 and Pro24 in each peptide (or alanine substitutions thereof) are shown as sticks and highlighted in a darker color.

The significance of Ser23 and Pro24 in topology determination was further evaluated *in vivo* by the same SCAM analysis used previously to assess reentrant topology in PglC and in the elaborated PglC superfamily member, WcaJ.[17] Similar to wild-type PglC, both variants were found to form reentrant topologies with both termini in the cytoplasm (**Figure 2-14**). However, in a S23A/P24A PglC variant, all four cysteines were significantly protected from PEGylation by both NEM and MTSES, suggesting that all four cysteines were in the periplasm. Although the precise nature of the S23A/P24A PglC topology cannot be determined from these data, it is clear that the folding of this variant is substantially perturbed by mutation of the Ser-Pro dyad to Ala-Ala. In agreement with the PepLook peptide folding prediction, these results suggest that Ser23 and Pro24

act synergistically to establish a reentrant topology that positions both the N- and C-terminus of PglC in the cytoplasm.
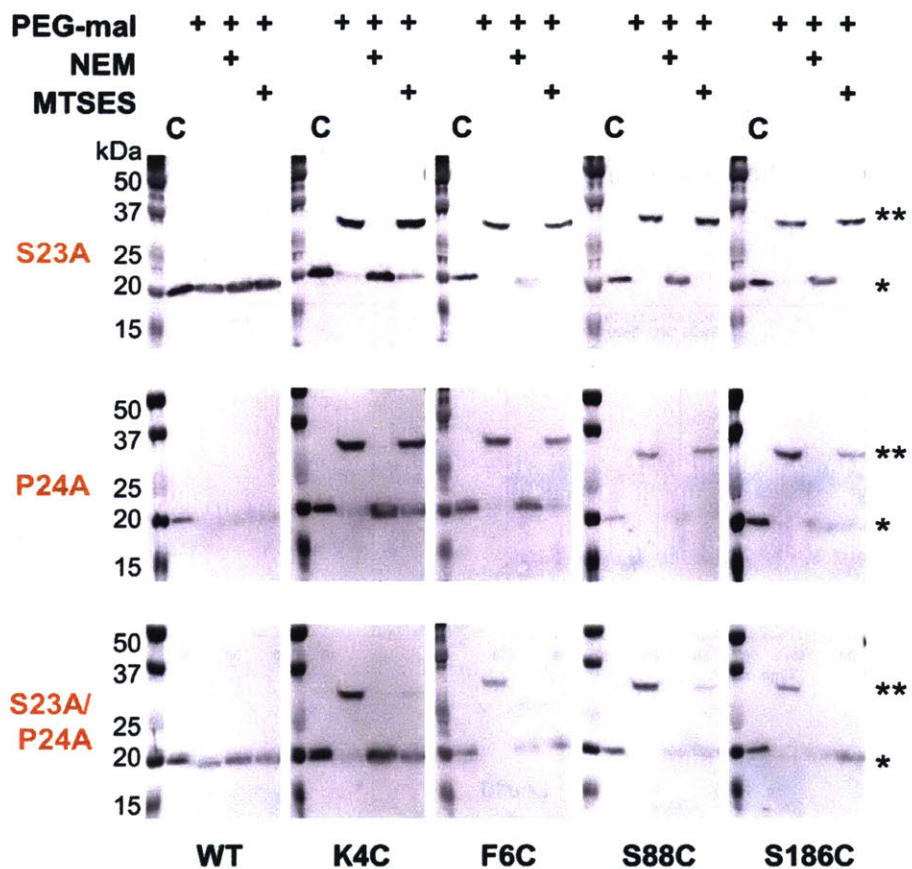
Figure 2-14 SCAM analysis of S23A, P24A and S23A/P24A PglC variant topologies. Legend: * = native PglC; ** = PglC labeled with PEG-mal; C = control, no PEG-mal labeling. All SCAM experiments were performed in duplicate or more. Representative Western blots are shown.

Whereas the crystal structure of PglC provides a "snapshot" of molecular interactions at the helix break in the RMH, MD simulations supplied a more dynamic view. Indeed, in simulations of PglC performed in a POPE membrane, it was observed that the break imposed by the Ser-Pro motif is additionally maintained by hydrogen bonding alternately between the side chain hydroxyl group and backbone amide of Ser23 and the backbone carbonyls of Leu19 and Ile20 (**Figure 2-15**). This suggests that the break in the RMH is stabilized by an extensive network of dynamic hydrogen bonds between Leu19, Ile20 and Ser23 that compensate for the absence of a hydrogen bond donor in Pro24 and enforce a helix-break-helix topology.
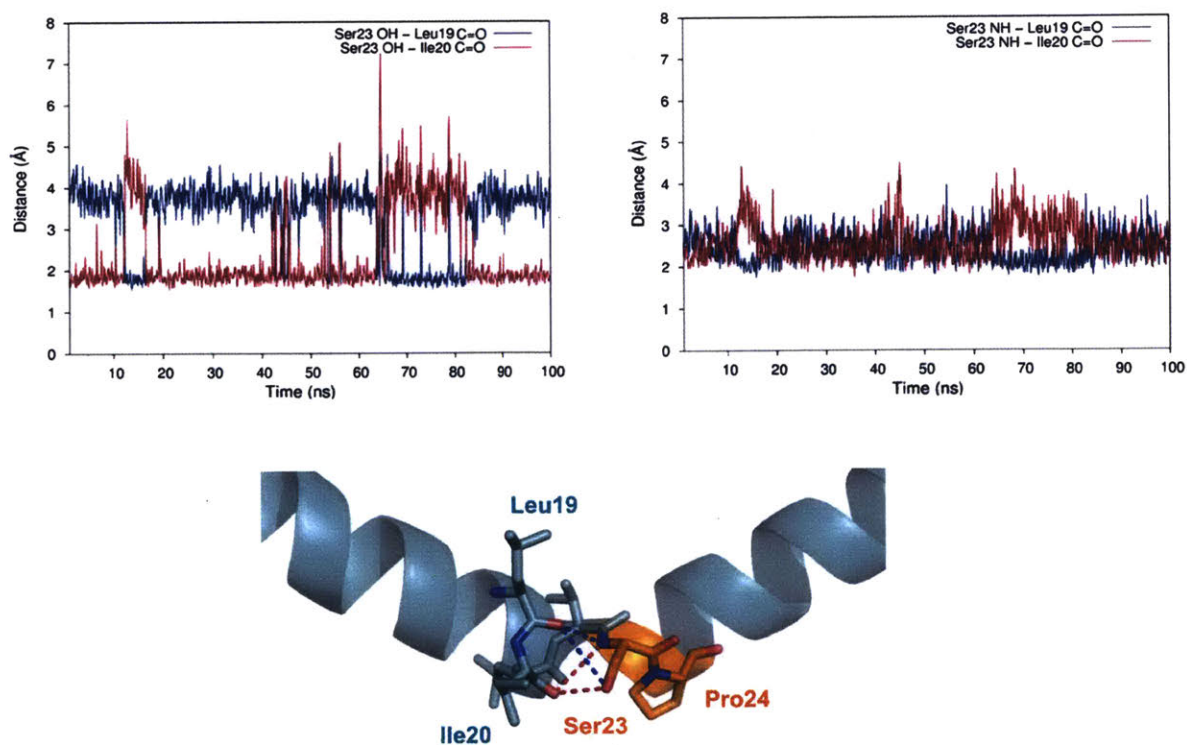
Figure 2-15 Dynamic hydrogen bonding stabilizes the break in the PglC RMH.
Dynamic hydrogen bonding between the side chain hydroxyl (*top left*) or backbone amide-NH (*top right*) of Ser23 and the backbone amide carbonyl of Leu19 and Ile20, was measured during MD simulations of PglC. The hydrogen bonding network is illustrated in the context of the RMH (*bottom*) with blue dashes for Ser23-Leu19 bonds and red dashes for Ser23-Ile20 bonds.

The high overall hydrophobicity of the RMH domain near the N-terminus of PglC likely results in targeting of the nascent polypeptide to the membrane early in translation.[30] Therefore, we hypothesized that the break in the RMH, enforced by the Ser-Pro motif and the resulting hydrogen-bonding network described above, could allow the N-terminal RMH to form independently of the C-terminal globular domain and insert the membrane in a reentrant topology. To investigate further the role that this motif might play in such early folding events, peptides representing the RMH of both wild-type and S23A/P24A PglC were examined in MD folding simulations (**Figure 2-16**). Both peptides were initially generated in an extended conformation and allowed to fold for 100 ns in water to simulate early co-translational folding events. Next, a shift to a more hydrophobic medium (20% isopropanol in water) for a further 1.5 μs provided an environment more closely resembling the membrane. Significant differences in folding behavior between the two peptides were observed over the full 1.6 μs of simulation time: in the peptide corresponding to wild-type PglC, residues Leu19 through Pro24, encompassing the Ser-Pro motif and surrounding hydrogen bonds, remained mostly unstructured, while the N-terminus and the sequence following the Ser-Pro motif rapidly adopted an α-helical conformation (**Figure 2-16** and **Figure 2-17**). In contrast, in the peptide corresponding to S23A/P24A PglC, residues Leu19 through Ile33 folded into a continuous α-helix, and this secondary structure appeared much later in the folding process.
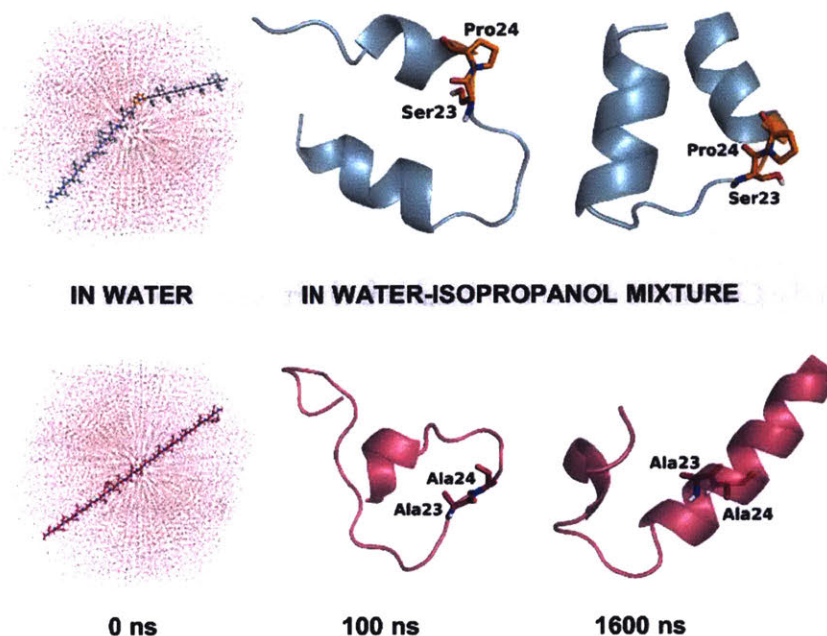
Figure 2-16 The Ser-Pro motif facilitates early folding of the RMH domain.
Peptides corresponding to the RMH domains of wild-type (*top*) and S23A/P24A PglC (*bottom*)
were folded from an extended conformation (*left panel*) for 100 ns in water (*middle panel*),
followed by an additional 1500 ns in 20% isopropanol/water (*right panel*).

These simulations indicate that the Ser-Pro motif drives formation of the helix-break-helix

structure observed in the RMH of wild-type PglC, while the corresponding domain of S23A/P24A

PglC has an intrinsic tendency to form a single, uninterrupted helix. The difference in the simulated

folding of the two peptides underscores the significance of the Ser-Pro motif for RMH formation,

and further supports a model by which this motif facilitates proper folding and membrane insertion

of the N-terminal domain in a reentrant topology during an early co-translational folding event.
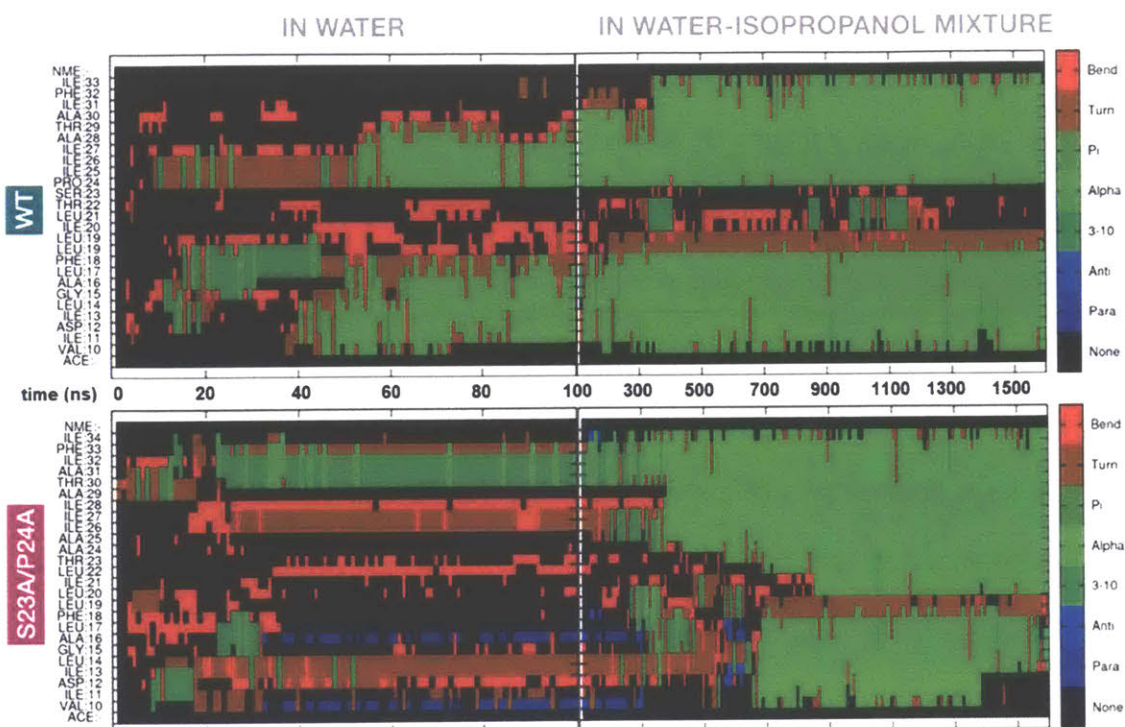
Figure 2-17 The Ser-Pro motif facilitates formation of secondary structure in the RMH domain. Appearance of secondary structure in peptides modeling the RMH of wild-type (*top*) and S23A/P24A PglC (*bottom*) during folding simulations in water (for 100 ns) followed by 20% isopropanol/water (for an additional 1500 ns). Secondary structure for the S23A/P24A PglC RMH peptide appeared much later relative to the wild-type RMH, and indicated the formation of a continuous α-helix.

67

The Ser-Pro motif contributes to the overall stability of the PglC fold

The SCAM analyses and MD simulations suggest a role for the Ser-Pro motif in early formation of the N-terminal RMH domain. However, additional analyses were needed to elucidate the contribution of the motif to stabilizing the overall PglC fold. To illustrate the contribution of the Ser-Pro dyad to the overall stability of PglC, *in vitro* thermal stability measurements were performed on purified wild-type and S23A/P24A SUMO-PglC variants. Thermal stability could not be measured by circular dichroism (CD), as it was observed that S23A/P24A SUMO-PglC does not purify to homogeneity (**Figure 2-18**); therefore, CD analysis on such a sample would be confounded by signal from co-purifying contaminants. It has also been noted that unfolding of α-helical membrane proteins is often driven by loss of tertiary contacts rather than secondary structure,[31-33] making CD spectroscopy less informative for thermal stability analysis. Thus, a thermal shift assay that specifically reports on stability as a function of resistance to precipitation upon heating, recently applied to the **polytopic membrane** protein dolichylphosphate mannose synthase,[34] was used.
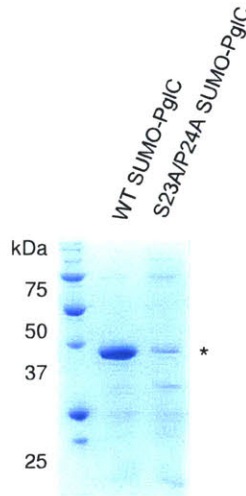
Figure 2-18 SDS-PAGE analysis of wild-type and S23A/P24A SUMO-PglC. Comparative SDS-PAGE analysis of wild-type and S23A/P24A SUMO-PglC, Coomassie stain; * = SUMO-PglC. Sample loading was normalized by UV absorbance at 280 nm.

Following heating, precipitated protein was removed by centrifugation and the soluble fraction quantified by gel densitometry to determine a T_m for both variants. It was found that the S23A/P24A SUMO-PglC variant, with a T_m of 42.6 ± 2.0 °C, is significantly less stable to heating relative to wild-type SUMO-PglC, which had a T_m of 49.9 ± 1.5 °C (**Figure 2-19**). The ΔT_m of 7.3 ± 2.5 °C between the wild-type and S23A/P24A SUMO-PglC indicates a loss in thermal stability upon mutation of the Ser-Pro motif to Ala-Ala. Notably, the Ala-Ala substitution control variants I26A/L27A SUMO-PglC and K187A/E188A SUMO-PglC showed only slight decreases in stability relative to wild-type, with T_m values of 47.3 ± 0.9 and 46.6 ± 1.1 °C, respectively (**Figure 2-20**). Thus, the Ser-Pro motif is not only necessary for formation of the RMH domain, but additionally has a strong influence on the stability of the entire PglC structure.
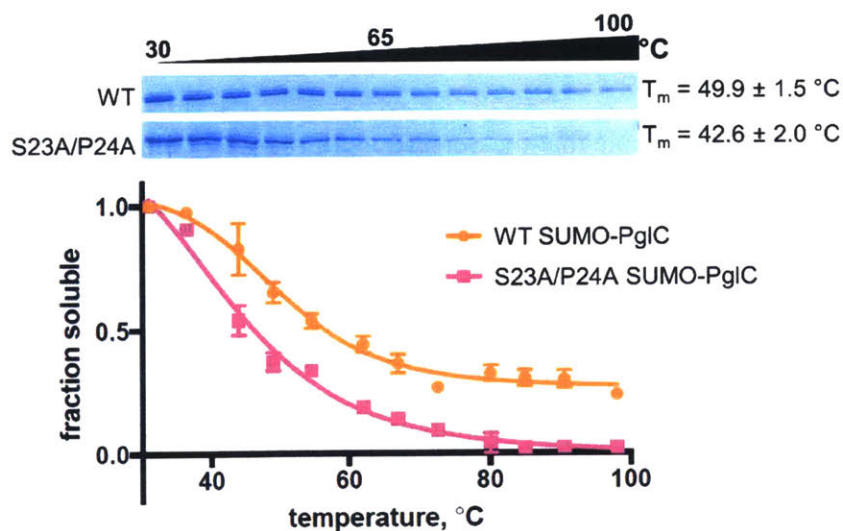
Figure 2-19 The Ser-Pro motif contributes to stability of the PglC fold.
Thermal shift analysis of wild-type and S23A/P24A SUMO-PglC. Error bars are given for mean ± SEM, n = 3.
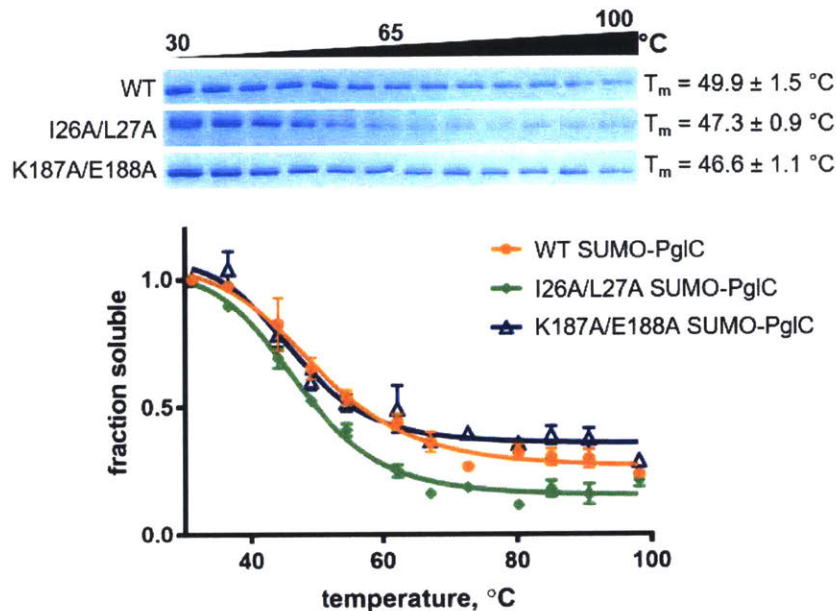
Figure 2-20 Thermal shift assays of control SUMO-PglC variants.
Thermal shift analysis of wild-type, I26A/L27A and K187A/E188A SUMO-PglC. Error bars are given for mean ± SEM, n = 3.

To elucidate further the role that the Ser-Pro motif plays in maintaining the stability of the native PglC fold, MD simulations were performed on both wild-type PglC and a S23A/P24A variant, generated *in silico* by substituting the Ser-Pro motif with Ala-Ala. Whereas wild-type PglC was stable over 400 ns of simulations in a POPE membrane, the interior of S23A/P24A PglC, containing the putative polyprenol phosphate binding site proximal to the active site, appeared to collapse. Specifically, in S23A/P24A PglC, the positions of residues Leu21 on the RMH, and of Leu90 and Val180 on two amphipathic helices at the membrane interface, were all found to move closer to each other over the course of 400 ns, significantly reducing the volume of the protein interior (**Figure 2-21** and

70

Figure 2-22). The positions of these same residues remained relatively unchanged in wild-type PglC. The collapse of the S23A/P24A PglC tertiary structure upon substitution of the Ser-Pro motif suggests that the motif plays an important role not only in forming the RMH, but also in maintaining the RMH and the associated amphipathic helices in the correct conformation in the PglC fold.
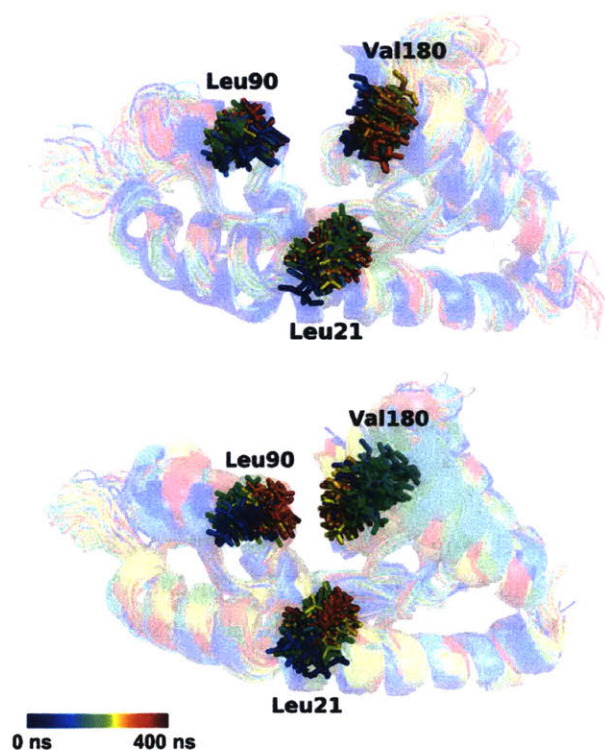
Figure 2-21 Mutation of the Ser-Pro motif in silico causes a "collapse" of the PglC fold interior. Superimposition of frames, taken at 10 ns intervals, along MD simulations of wild-type (*top*) and S23A/P24A (*bottom*) PglC. Colored from blue, t = 0 ns to red, t = 400 ns. PglC is represented as a semi-transparent cartoon and residues Leu21, Leu90 and Val180 as sticks.
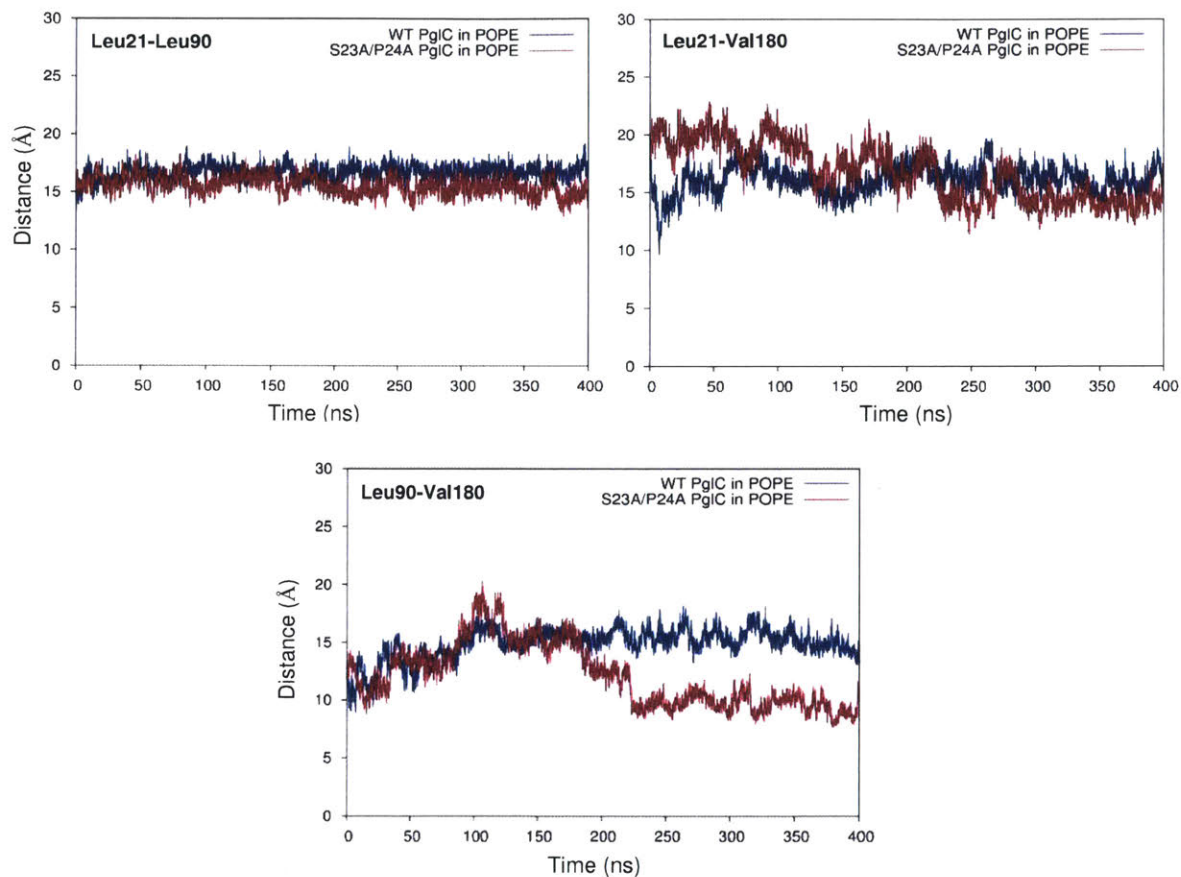
Figure 2-22 Monitoring distances during the "collapse" of the S23A/P24A PglC-fold interior. C_α-C_α distances between Leu21 and Leu90 (*top left*), Leu21 and Val180 (*top right*) and Leu90 and Val180 (*bottom*), measured over 400 ns of MD simulations of wild-type and S23A/P24A PglC in a POPE membrane.

Additionally, it was observed that the collapse of the S23A/P24A PglC fold appeared to hinder the passage of lipids into the interior of the fold relative to wild-type PglC. At the start of MD simulations of both wild-type and S23A/P24A PglC, two phospholipid molecules were observed to occupy the fold interior. Both phospholipids continued to occupy this inner cavity during most of the simulation of wild-type PglC, whereas over the course of the simulation of S23A/P24A PglC, one phospholipid appeared to leave the cavity (**Figure 2-23**). As the fold interior contains the putative polyprenol phosphate binding site, lipid access into the PglC-fold interior is crucial for substrate binding. Thus, the role played by the Ser-Pro motif in maintaining the PglC fold also has immediate significance for catalytic activity.
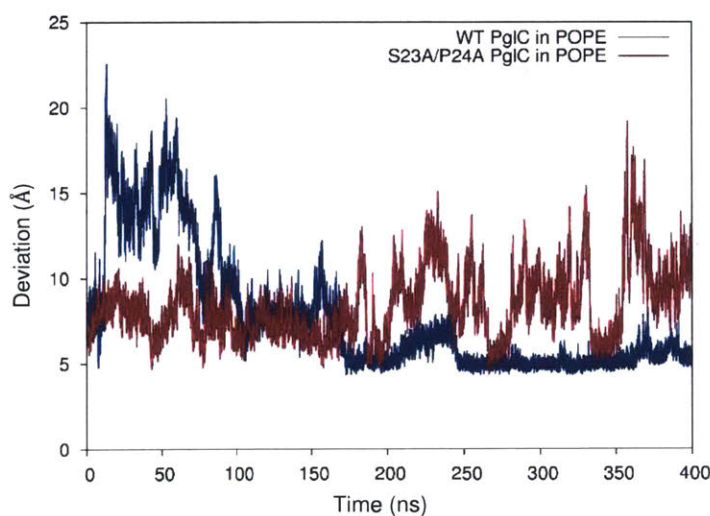
Figure 2-23 Mutation of the Ser-Pro motif reduces lipid occupancy in the PglC fold interior. Distance between the geometric centers of the two lipids in the PglC fold interior, measured over 400 ns of MD simulations of wild-type and S23A/P24A PglC in a POPE membrane.

Tryptophan as an alternative to Ser23 and Pro24

It was previously noted that among monotopic PGTs that do not have a proline at the conserved position, the most common substitution occurring natively is tryptophan; a P24W substitution occurs in ~3% of all PGTs in this superfamily. Whereas P24A SUMO-PglC was catalytically inactive, a P24W variant retained some activity, although significantly less than wild-type.[14] SCAM analysis of a P24W PglC variant determined that it adopted the same native, reentrant topology as P24A PglC (**Figure 2-24**).
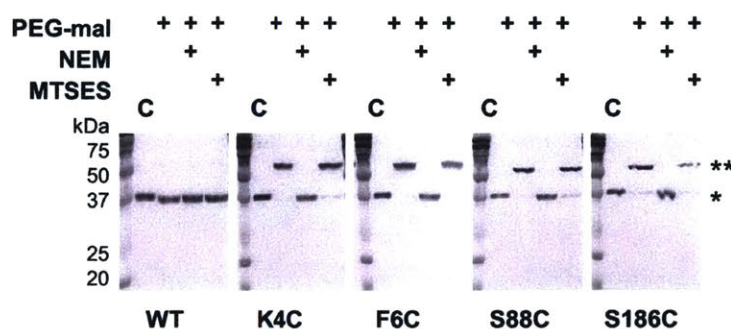

Figure 2-24 SCAM analysis of a P24W PglC variant.
Legend: * = native PglC; ** = PglC labeled with PEG-mal; C = control, no PEG-mal labeling.
All SCAM experiments were performed in duplicate or more. Representative Western blots are shown.

Peptide folding predictions performed with the PepLook tool[29] similarly suggested that the RMH of a P24W PglC variant would adopt a helix-break-helix topology like wild-type and P24A PglC, although the P24W PglC predicted folding did appear to lose some secondary structure (**Figure 2-25**).

WT   **P24A**   **P24W**   **overlay**

RIFDILGALFLLILTSPIIIATAIFIK
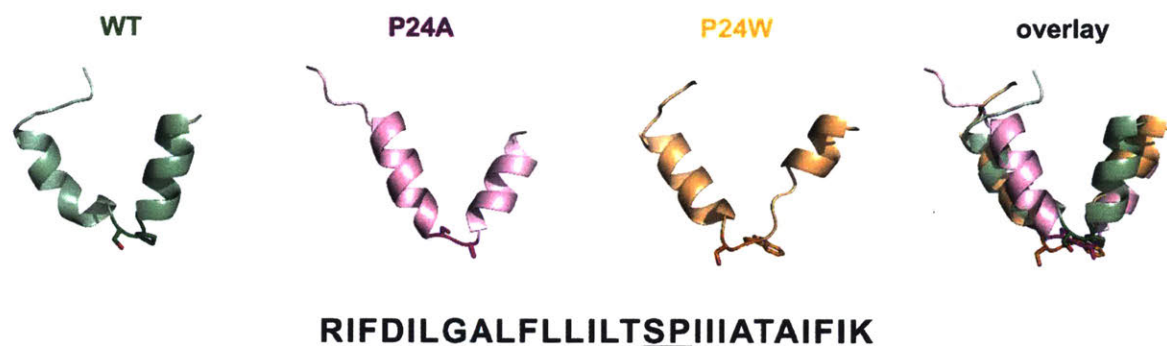
Figure 2-25 PepLook folding of wild-type, P24A, and P24W PglC RMH peptides. All three peptides adopted similar helix-break-helix conformations (*overlay*). Ser23 and Pro24 in each peptide (or substitutions thereof) are shown as sticks and highlighted in a darker color. Wild-type peptide sequence is shown with Ser23 and Pro24 underlined.

In the course of further analysis of the monotopic PGTs superfamily, it was discovered in an alignment of the monotopic PGT superfamily[14] that among homologs with a natively-occurring tryptophan at residue 24, the most commonly occurring residue at position 23 is also a tryptophan. In fact, of all homologs with a Trp24, ~40% also carry a Trp23, while only ~20% carry the otherwise more common Ser23. This suggests that a Trp-Trp motif may be a more favorable substitution for the entire Ser-Pro motif that the singular P24W substitution, which had been the subject of previous studies. Furthermore, unlike the S23A/P24A RMH peptide, which was predicted by PepLook to adopt a single, continuous helix, a S23W/P24W peptide was predicted to fold into a helix-break-helix (**Figure 2-26**). Like the predicted fold of the P24W PglC RMH (**Figure 2-25**), the S23W/P24W PglC RMH fold appeared to lose some secondary structure. These early findings suggest that, while less favorable than the conserved Ser-Pro motif, a Trp-Trp substitution at the same position is able to enforce a native, reentrant topology in monotopic PGTs. A preference for Trp substitution of *both* Ser and Pro indicates that, like Ser23 and Pro24, the Trp23 and Trp24 residues act synergistically to create a helix-break-helix.
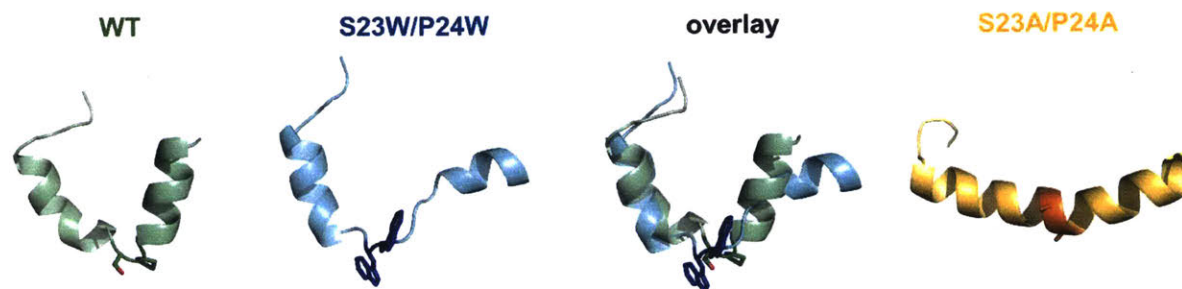
Figure 2-26 PepLook folding of an S23W/P24W PglC RMH peptide.
Ser23 and Pro24 in each peptide (or substitutions thereof) are shown as sticks and highlighted in a darker color.

A conserved positively charged motif is also a determinant of reentrant topology

In addition to the Ser-Pro motif, two residues at the N-terminus of PglC, Lys7 and Arg8, are also highly conserved and participate in electrostatic interactions with the globular domain and surrounding phospholipids.[18] We hypothesized that this Lys-Arg motif might also contribute to the reentrant topology of PglC. Thus, the topology of K7A and R8A PglC variants were evaluated by SCAM (**Figure 2-27**). Mutation of either Lys7 or Arg8 to Ala did not have a significant effect on the final topology of PglC. However, when both conserved residues were mutated in the K7A/R8A PglC variant, it was observed that the N-terminus (represented by K4C and F6C) became partially localized to the periplasm. This suggested that the K7A/R8A PglC population was split between proteins adopting the native reentrant topology, in which the N-terminus is in the cytoplasm, and those adopting a non-native membrane-spanning topology. Notably, a K7A/R8A/P24A PglC variant adopted the same non-native, all-periplasmic topology previously observed for S23A/P24A PglC, suggesting that, as with S23A/P24A PglC, the native folding

process of the RMH had been significantly perturbed. Taken together, these analyses indicate that the positively charged Lys-Arg motif and the helix-breaking Ser-Pro motif act cooperatively to enforce a reentrant topology for PglC.
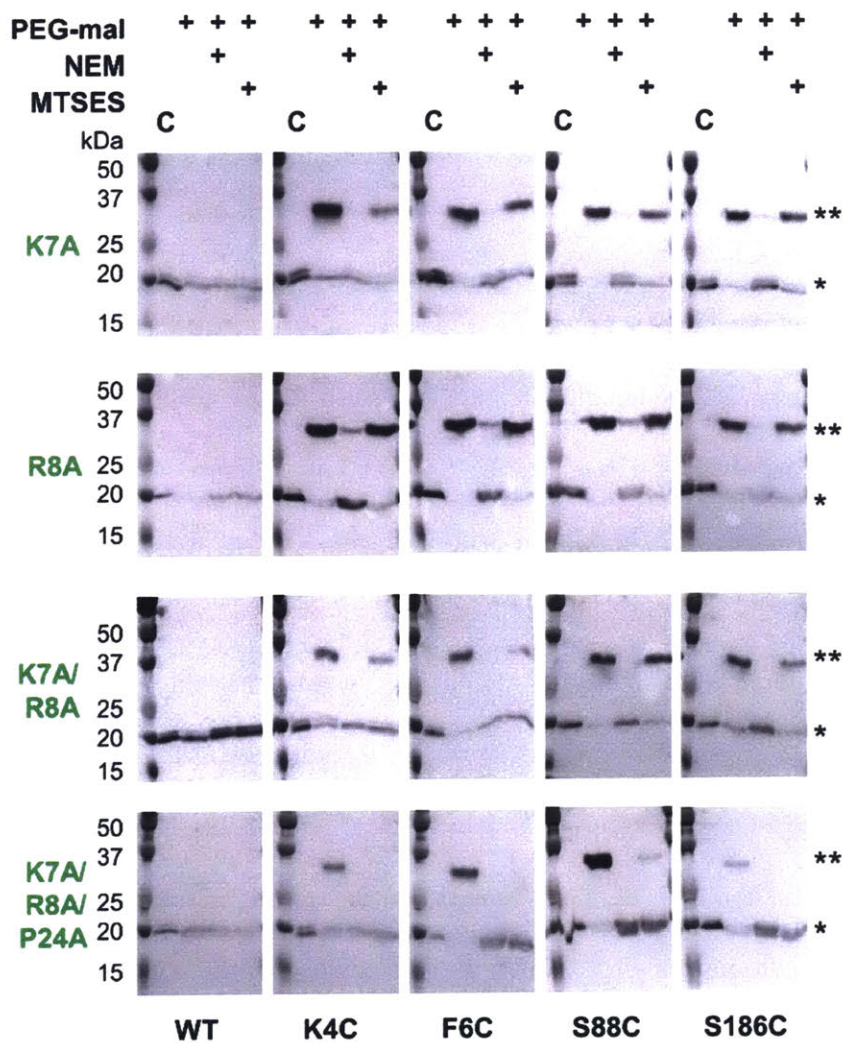
Figure 2-27 SCAM analysis of K7A, R8A, K7A/R8A and K7A/R8A/P24A PglC variants. Legend: * = native PglC; ** = PglC labeled with PEG-mal; C = control, no PEG-mal labeling. All SCAM experiments were performed in duplicate or more. Representative Western blots are shown.

RMH geometry is held stable during catalysis

Although the *in vivo* SCAM analyses and MD simulations are consistent with the reported structure of PglC, we noted that the inter-helical angle of the RMH in the structure (118°) (**Figure 2-2**) is significantly more open than that observed in the peptide modeling (**Figure 2-16**) and in reentrant helices found in reported polytopic membrane protein structures.[19] Therefore, we performed cysteine crosslinking studies to determine whether the conformation of the RMH in the reported structure reflects the native PglC fold and to further probe the significance of this conformation in catalysis.

Two cysteines were introduced into PglC, one at the N-terminus and one in the globular domain, at positions Glu3 and Ile163 respectively. If the reported structure reflects a native conformation of the RMH, the two residues at these positions have a C_β-C_β distance of ~5.2 Å in the PglC fold (**Figure 2-28**). Although oxidation to a disulfide was not observed, the two cysteines could be crosslinked in the E3C/I163C SUMO-PglC variant with dibromobimane (bBBr), a short, bifunctional thiol-specific labeling reagent (**Figure 2-28**). An increase in fluorescence is reported in the literature for bBBr upon thiol-thiol crosslinking.[35-36] Indeed, bBBr-treated E3C/I163C SUMO-PglC was significantly more fluorescent than were either single cysteine variants (E3C SUMO-PglC and I163C SUMO-PglC), or a control variant (E3C/S88C SUMO-PglC) with two cysteine residues too far apart for crosslinking by bBBr (**Figure 2-29**). Crosslinking was quantified by fluorescence densitometry of bands corresponding to monomeric PglC by SDS-PAGE, ensuring that only intramolecular crosslinking was included in the analysis. These crosslinking studies suggest that the extended RMH observed in the crystal structure reflects a native conformation of the PglC fold.

78
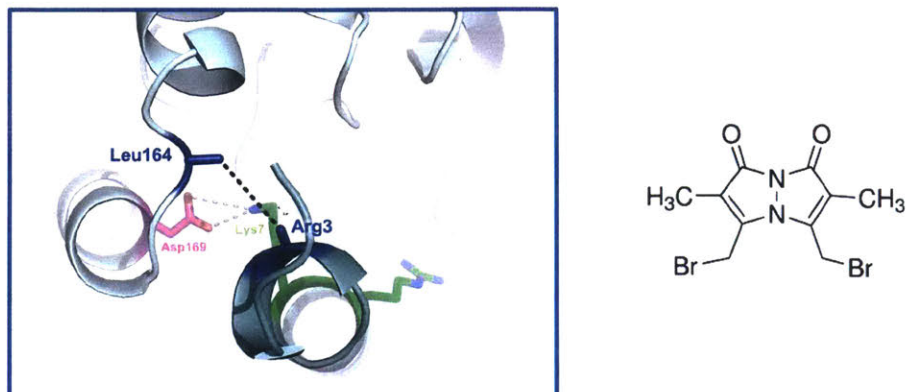
Figure 2-28 Location of PglC crosslinking by dibromobimane (bBBr).
Left: Detailed view showing the location of bBBr crosslinking (black dashes). The C_α and C_β of Arg3 and Leu164 in the structure of PglC from *C. concisus* are shown as dark blue sticks (the remainder of the side chains is omitted for clarity). The corresponding residues Glu3 and Ile163 in PglC from *C. jejuni* were substituted with Cys for bBBr crosslinking studies. *Right:* structure of bBBr cross-linker.
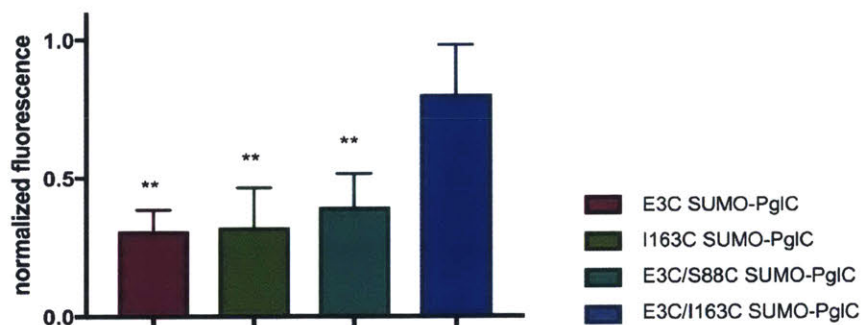
Figure 2-29 An E3C/I163C SUMO-PglC variant is crosslinked by bBBr.
Fluorescence of SUMO-PglC variants following crosslinking with bBBr, normalized to fluorescence of DTT-quenched samples (quenched samples represent maximum possible fluorescence). Error bars are given for mean $\pm$ SD, n = 4 (**$p < 0.01$, Student's t-test; *p*-values for each variant are 0.0022 (E3C), 0.0055 (I163C), 0.0091 (E3C/S88C)).

To confirm that crosslinking was capturing a stable conformation of the RMH in E3C/I163C SUMO-PglC and not a transient intermediate, the position of the N-terminus relative to the globular domain was further investigated by MD. In simulations of PglC in a POPE membrane it was observed that the distance between these two residue positions remains invariable over 400 ns (**Figure 2-30**). This supports a model of PglC in which the RMH is constrained in the position observed in the crystal structure, with little conformational freedom of the N-terminus relative to the globular domain. Importantly, it was found that intramolecular crosslinking of E3C/I163C SUMO-PglC by bBBr did not impact catalytic activity (**Figure 2-31**). This indicates that the crosslinked conformation of the RMH domain, which places the N-terminus in close proximity to the globular domain, is compatible with catalysis.
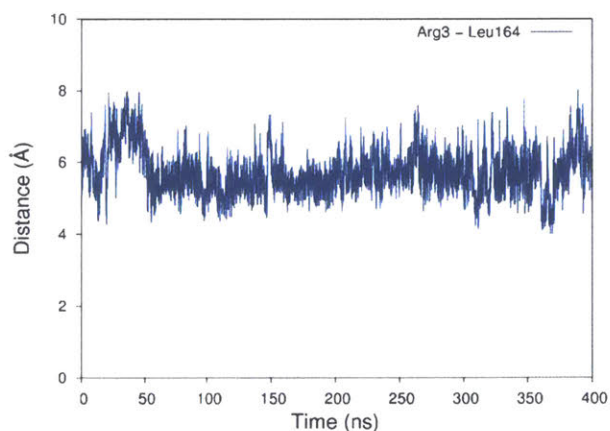
Figure 2-30 The N-terminus of PglC remains fixed relative to the globular domain. Distance between Arg3 and Leu164 (measured from the centroid of each residue), over 400 ns of MD simulations of PglC in a POPE membrane.
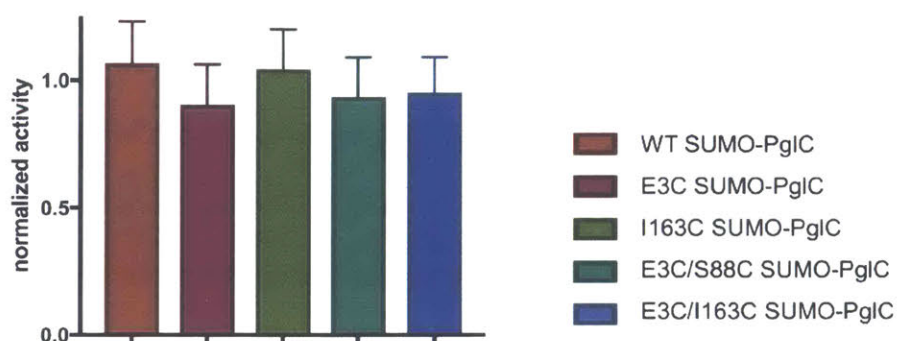
Figure 2-31 Crosslinked PglC is compatible with catalysis.
Activity of SUMO-PglC variants following crosslinking with bBBr, normalized to activity following treatment with vehicle. Error bars are given for mean ± SD, n = 3.

DISCUSSION

Wild-type PglC adopts a reentrant topology

In complement to the recently reported crystal structure,[18] *in vivo* SCAM analyses determined that PglC adopts a reentrant topology in the membrane in both in *C. jejuni* and when overexpressed in *E. coli*. The finding contradicts many previous predictions, described in detail Chapter 3, that the N-terminus of PglC forms a transmembrane helix. Notably, the same reentrant topology was previously reported for the corresponding domain of WcaJ, a homolog of the elaborated WbaP monotopic PGT.[17] This suggests that the topology and membrane association enforced by the PglC RMH is a conserved feature of all the PGTs in this extensive superfamily.

Two conserved motifs contribute to RMH formation

The recently-reported structure of the monotopic PGT, PglC, describes a mode of membrane association reliant on a single reentrant helix-break-helix domain inserted in the membrane.[18] Formation of this RMH is largely driven by two highly conserved motifs, elucidated in this work, that contribute to both early folding events that facilitate membrane insertion of the RMH in the proper topology, and late folding events that stabilize the folded PglC in the membrane.

The *in vivo* SCAM and *in silico* peptide folding experiments presented herein suggest a synergistic effect between the Lys7-Arg8 and Ser23-Pro24 motifs in determining proper RMH formation. The synergy observed between Ser23 and Pro24 echoes the intramolecular interactions observed both in the crystal structure and MD simulations of PglC, wherein Pro24 disrupts the backbone hydrogen-bonding network of the RMH helix to create the characteristic break, and the backbone amide and side chain of Ser23 stabilize this break by forming a compensatory dynamic hydrogen bond with the backbone carbonyls of Leu19 and Ile20.

Prolines are known to break α-helices by perturbing the peptide backbone hydrogen-bonding network that dictates α-helicity.[37] Hydrogen bonding of the peptide backbone is thought to drive folding and membrane insertion of hydrophobic peptides by greatly reducing the energetic cost of partitioning peptide bonds into the membrane,[38-39] wherein unsatisfied backbone hydrogen bonds are estimated to carry a free energy cost of 0.4 kcal mol^{-1} per residue.[40] Consequently, the interruption in backbone hydrogen bonding introduced by Pro24 has a destabilizing effect on the RMH, which is largely mitigated by auxiliary hydrogen bonding between Ser23 and the peptide backbone carbonyl groups of Ile19 and Ile20. Polar residues such as serine in hydrophobic peptides were also identified previously as characteristic features of reentrant helices.[41] Thus, the two residues of the Ser-Pro motif act cooperatively to define a modified hydrogen-bonding network in

the RMH that yields the observed helix-break-helix structure. The absence of these key residues in S23A/P24A PglC accordingly biases the N-terminus towards formation of a continuous α-helical secondary structure (**Figure 2-16**), and could explain the aberrant topology observed for S23A/P24A PglC by SCAM analysis (**Figure 2-14**).

Notably, a conserved Ile-Pro motif was previously identified as a key determinant of the reentrant loop topology of caveolin-1,[22, 42] and while Ser23 is 49% conserved among PglC homologs, Pro24 is otherwise often preceded by one of the branched-chain amino acids (BCAAs): Ile, Leu, or Val. This suggests that BCAA-proline motifs may also be capable of supporting helix-break-helix formation among monotopic membrane proteins with reentrant helix domains.

Lys7 and Arg8 also act jointly with each other and with the Ser-Pro motif to enforce a reentrant topology, with the K7A/R8A and the K7A/R8A/P24A PglC variants adopting increasingly perturbed topologies. The observation that substitution of both conserved positive charges caused the N-terminus to lose some of its proclivity for cytoplasmic localization is in good agreement with previous reports of topology determination by the positive-inside rule,[7, 25] but the observation that a significant portion of the K7A/R8A PglC population continued to adopt the native reentrant topology suggests that while Lys7 and Arg8 contribute to proper topology formation, they are not the primary drivers. More likely, the native reentrant PglC topology results from the cumulative effect of Lys7, Arg8, Ser23 and Pro24: each of these four conserved residues exerts a unique influence on formation and maintenance of the RMH domain to collectively result in the observed topology.

RMH formation as an early folding event

Due to the hydrophobicity of the N-terminus of PglC, it may act as an uncleaved signal

sequence[30] to target the nascent polypeptide to the membrane early during PglC translation.

Thus, formation and membrane insertion of the RMH domain could occur while the globular

domain of PglC is still being translated (**Figure 2-32**). We propose that the two conserved motifs

identified herein each play critical roles in this early folding event. The Lys7 and Arg8 at the N-

terminus each contribute positive charges that disfavor translocation of the N-terminus through

the membrane to the periplasm,[30] and, in agreement with the positive-inside rule, retain these

residues at the cytoplasmic face of the inner membrane. As translation of the RMH continues,

the modified backbone hydrogen-bonding pattern of the Ser-Pro motif facilitates formation of the

characteristic helix-break-helix of the RMH and directs translation and folding of the globular

domain to proceed on the cytoplasmic side of the membrane. Thus, under the combined

influence of the Lys-Arg and Ser-Pro motifs, the hydrophobic domain of PglC inserts into the

membrane as a reentrant helix, establishing a monotopic topology for the final fold. Accordingly,

alanine-substitution of these key motifs likely leads to aberrant RMH formation and resulted in

the non-native topologies observed by SCAM analysis (**Figure 2-14** and **Figure 2-27**). Indeed,

alanine-substitution results in such a disruption in the native PglC membrane insertion process

that the entire construct appears to be erroneously translocated into the periplasm in a non-native

conformation. A similar translocation of mistranslated proteins into the periplasm has previously

been reported as a possible mechanism of action for aminoglycosidase toxicity in *E. coli*.[43]
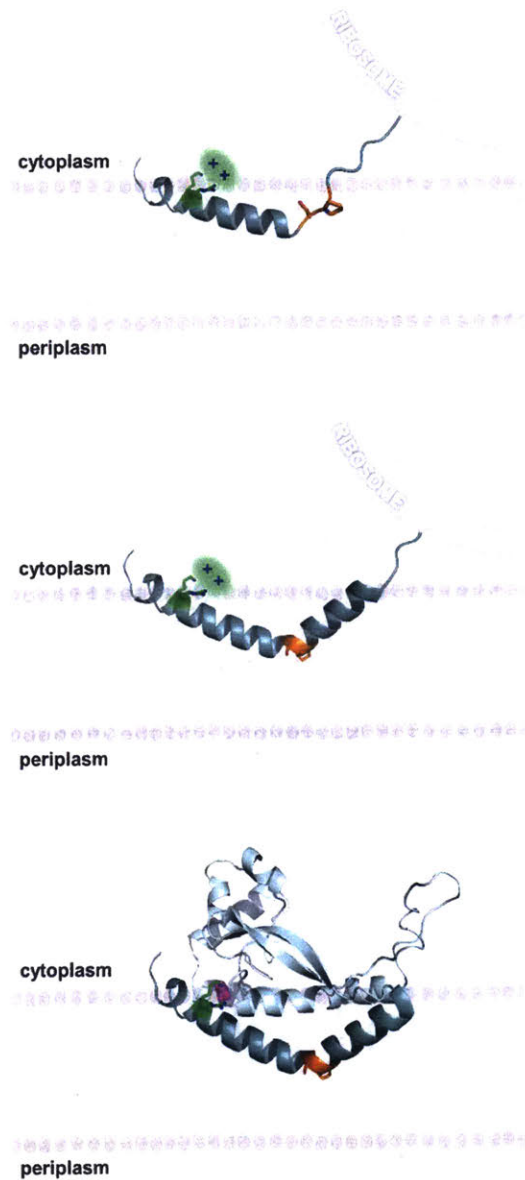
Figure 2-32 Model of RMH folding and membrane insertion.
The Lys-Arg (*green*) and the Ser-Pro (*orange*) motifs facilitate formation of the RMH and insertion into the membrane in an early co-translational event. The positively-charged Lys-Arg motif favors localization of the N-terminus to the cytoplasm (*top panel*). The Ser-Pro motif creates the characteristic break in the RMH (*middle panel*), resulting in insertion of the RMH into the membrane in a reentrant topology. Following translation of the globular domain, both motifs further contribute to the overall stability of the PglC fold (*bottom panel*).

85

PglC structure and function are supported by the two conserved motifs

Several lines of evidence suggest that both the Lys-Arg and Ser-Pro motifs additionally make significant contributions to maintaining the stability of the RMH within the context of fully-folded PglC to support function. Crosslinking of the N-terminus to the globular domain in the E3C/I163C SUMO-PglC variant indicates that the N-terminus can be captured in close proximity to the globular domain and that it can remain in such a conformation during catalysis (**Figure 2-29** and **Figure 2-31**). MD simulations additionally demonstrate that the N-terminus has little conformational freedom, supporting the hypothesis that crosslinking captures a stable conformation of the RMH. Taken together, these results strongly suggest that the reported structure represents a native and active conformation of the RMH in the PglC fold. In the reported structure of PglC, Lys7 forms a salt bridge with Asp169, a residue which is 98% conserved among PglC homologs, at the interface between the N-terminus and the globular domain (**Figure 2-3**). The corresponding Asp in the PglC homolog from *C. jejuni*, Asp168, was reported previously to be necessary for catalytic activity.[14] Interestingly, SCAM analysis of the inactive D168A PglC (from *C. jejuni*) determined that the variant adopts a native reentrant topology (**Figure 2-33**).
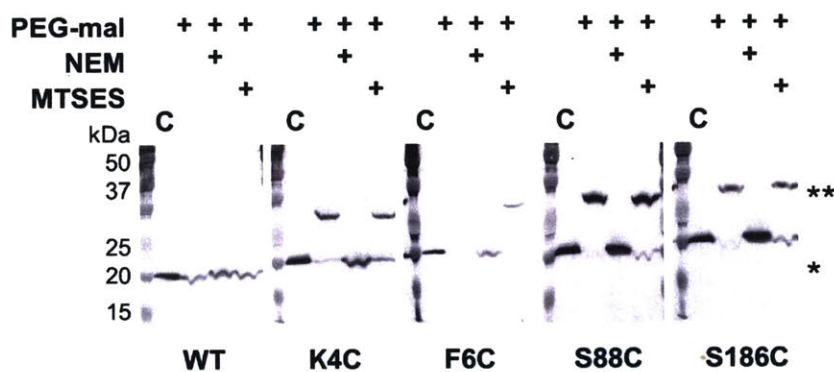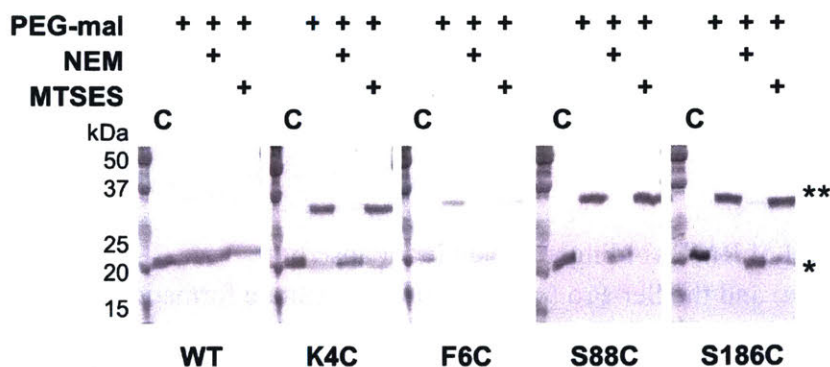

Figure 2-33 SCAM analysis of a D168A PglC variant.
Legend: * = native PglC; ** = PglC labeled with PEG-mal; C = control, no PEG-mal labeling.
All SCAM experiments were performed in duplicate or more. Representative Western blots are shown.

The K7A SUMO-PglC variant was similarly found to be catalytically inactive, despite adopting a reentrant topology (**Figure 2-34**). Both K7A/R8A and K7A/R8A/P24A SUMO-PglC variants were also inactive. This suggests that the salt bridge formed between Lys7 and Asp169 is itself not essential for formation of a reentrant topology, but *is* necessary for catalytic activity, likely due do its critical role in constraining the N-terminus in close proximity to the globular domain. Thus, in addition to informing early RMH folding, Lys7 plays a crucial structural role by stabilizing the RMH in the PglC fold via a key interaction with Asp169 on the globular domain and thereby promoting PglC activity.
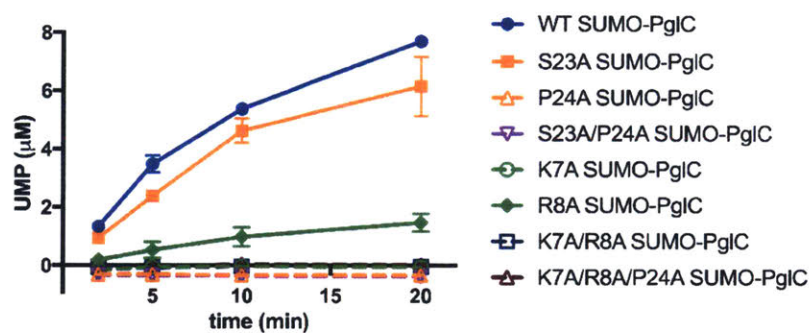
Figure 2-34 Individual resides differ in their importance for PglC function.
Activity assays of wild-type SUMO-PglC and variants using UMP-Glo® (Promega) to monitor UMP release. Error bars are given for mean ± SD, n = 2.

The contribution made by Arg8 to stability of PglC is more subtle. In the reported structure of PglC, a phosphatidylethanolamine head-group is coordinated to the guanidinium side chain of Arg8.[18] Previously, MD simulations of guanidinium ion translocation into the membrane reported a free energy minimum for the ion at the head-group region of the membrane-water interface.[44] This suggests that interactions between Arg8 and surrounding lipid headgroups help stabilize the RMH at the membrane interface. Unlike the inactive K7A SUMO-PglC variant, R8A SUMO-PglC was found to retain ~15% catalytic activity (**Figure 2-34**). Thus, Arg8 may play a less significant role in the PglC fold than does Lys7.

The stability of PglC is also influenced by the Ser-Pro motif. We propose that the network of hydrogen bonds enforced by the motif (**Figure 2-3** and **Figure 2-15**) acts as a rigidifying "staple" at the break in the RMH to restrict the conformational freedom of this domain. The rigidity conferred by the Ser-Pro motif would then allow proper positioning of the RMH domain with respect to the globular domain and formation of key intramolecular contacts (such as the salt bridge between Lys7 and Asp169), resulting in stabilization of the native PglC fold. The contribution made by the Ser-Pro motif is evidenced by the increase in thermal stability of the wild-type SUMO-PglC relative to the S23A/P24A SUMO-PglC variant (**Figure 2-19**), and by MD simulations in which S23A/P24A PglC experienced a significant collapse into the fold interior relative to wild-type PglC (**Figure 2-21**). Thus, the Lys-Arg and Ser-Pro motifs, in addition to influencing formation of the RMH domain, play important roles in positioning the RMH within PglC to stabilize the overall fold and support PglC function.

As the only membrane-inserted domain in PglC, the RMH is likely responsible for binding to the polyprenol phosphate substrate in the membrane; Pro24 in particular is hypothesized to play an important role.[14] Conserved prolines are found in polyisoprenol recognition sequences (PIRS)

in various enzymes,[45] and were previously proposed to contribute to polyprenol binding, possibly as a result of their disruption of peptide backbone hydrogen bonding.[46] Indeed, a P24A SUMO-PglC variant was previously reported to be catalytically inactive.[14] Notably, although both S23A and P24A SUMO-PglC variants retain native-like reentrant topologies, the S23A variant exhibits near-wild type activity, whereas the P24A and S23A/P24A variants are inactive (**Figure 2-34**). This suggests that in addition to contributing to RMH formation, Pro24 in particular additionally plays a crucial role in mediating polyprenol phosphate binding. Thus, in contributing to formation and stabilization of the RMH, the Lys-Arg and Ser-Pro motifs may enable function by positioning Pro24 to recognize polyprenol phosphate and promote binding in a catalytically-relevant configuration relative to the PglC active site.

A Trp-Trp motif can substitute for Ser-Pro

In an elaboration on previous findings that Pro24 is substituted with tryptophan in some monotopic PGTs,[14] is has been determined that a double Trp-Trp substitution for the Ser-Pro dyad is preferred over the single P24W substitution. Unlike the S23A/P24A PglC variant, a S23W/P24W is predicted to adopt a native reentrant topology (**Figure 2-26**). Like serine and proline residues, tryptophan residues are generally disfavored in transmembrane helices (not including at the lipid/water interface).[8] Additionally, like serine, tryptophan is able to play a similar role in disrupting the α-helical peptide backbone and creating new hydrogen-bonding contacts to stabilize a helix-break-helix conformation in the RMH.

CONCLUSIONS

The RMH domain is central to the structure and function of PglC: it anchors PglC in the membrane, interacts with several amphipathic helices to position the active site of PglC at the membrane-water interface, and mediates binding of the polyprenol phosphate substrate. The essentiality of the RMH to PglC structure and function, and the lack of homology between PglC and any known soluble protein folds, indeed suggest evolution of the fold at the membrane interface precisely to catalyze such transfer reactions. In this chapter, we identify four conserved residues that each make specific contributions both to the early formation of the RMH and to the stability of the final PglC fold, creating a clear preference for a reentrant topology for PglC and facilitating PglC function.

The proposed model for RMH formation in PglC (**Figure 2-32**) builds on existing principles of membrane protein topogenesis; both the conserved positive charges of the Lys-Arg motif and the helix-breaking "staple" of the Ser-Pro motif agree well with published observations of the positive-inside rule[6-8] and peptide backbone hydrogen bonding[38-39] as driving forces for folding and topology determination of membrane proteins. These key principles are also well-illustrated in the context of the PglC structure. A previous study into the topologies of caveolin and diacylglycerol kinase ε suggested a fine balance between reentrant and membrane-spanning topologies for both proteins, and noted that both positive flanking charges and conserved proline residues in these proteins could play decisive roles in topology determination;[25] the formation of the PglC RMH similarly depends on both. However, unlike PglC, caveolin and diacylglycerol kinase ε have not been structurally characterized to date. The current study is significant as it is the first to frame these effects in the context of the experimentally-determined structure.

ACKNOWLEDGEMENTS

EXPERIMENTAL PROCEDURES

PglC variants

Wild-type PglC from *C. jejuni* strain 11168 was cloned into the pET24a vector using the NdeI and XhoI restriction sites to insert a C-terminal His$_6$-tag or into the pE-SUMO vector as reported previously.[14] Unique cysteines at the K4, F6, S88 and S186 positions (for SCAM) and at the E3 and I163 positions (for bBBr crosslinking) were introduced using QuikChange II Site-Directed Mutagenesis (Agilent Technologies, Santa Clara, CA) according to manufacturer's instructions. Alanine and tryptophan substitutions at K7, R8, S23 and P24, as indicated, were all similarly introduced by Quikchange. Corresponding residues in PglC from *C. jejuni* and *C. concisus* are listed in **Table 2-1**. All primers used for subcloning and mutagenesis are listed in **Table 2-2**.

Table 2-1 – Corresponding key residues in *C. jejuni* and *C. concisus* PglC		
C. jejuni PglC	*C. concisus* PglC (PDB ID 5W7L)	Purpose in this study
Glu3	Arg3	Cys substitution for bBBr crosslinking
Lys4	Asn4	Cys substitution for SCAM
Phe6	Leu6	Cys substitution for SCAM
Lys7	Lys7	Conserved residue
Arg8	Arg8	Conserved residue
Ser23	Ser23	Conserved residue
Pro24	Pro24	Conserved residue
Ile26	Ile26	Control Ala mutation for thermal shift assay
Leu27	Ile27	Control Ala mutation for thermal shift assay
Ser88	Ser89	Cys substitution for SCAM
Ile163	Leu164	Cys substitution for bBBr crosslinking
Asp168	Asp169	Conserved residue
Ser186	Ser187	Cys substitution for SCAM
Lys187	Lys188	Control Ala mutation for thermal shift assay
Glu188	Glu189	Control Ala mutation for thermal shift assay

Table 2-2 – Primers used for cloning and mutagenesis of PglC and PEB3 variants	
PglC_F_NdeI	5'-AAAAAACATATGTATGAAAAAGTTTTTAAAAGAATTTTTG-3'
PglC_R_XhoI	5'-AAAAAACTCGAGGTTCTTGCCATTAAATTTCTCTG-3'
K4C_F	5'-GGAGATATACATATGTATGAATGCGTTTTTAAAAGAATTTTTG-3'
K4C_R	5'-CAAAAATTCTTTTAAAAACGCATTCATACATATGTATATCTCC-3'
SUMO_K4C_F	5'-GGTGGGATGTATGAATGCGTTTTTAAAAGAATTTTTG-3'
SUMO_K4C_R	5'-CAAAAATTCTTTTAAAAACGCATTCATACATCCCACC-3'
F6C_F	5'-CATATGTATGAAAAAGTTTGCAAAGAATTTTTGATTTTATTTTAGC-3'
F6C_R	5'-GCTAAAATAAAATCAAAAATTCTTTTGCAAACTTTTTCATACATATG-3'
SUMO_F6C_F	5'-GGGATGTATGAAAAAGTTTGCAAAGAATTTTTGATTTTATTTTAGC-3'
SUMO_F6C_R	5'-GCTAAAATAAAATCAAAAATTCTTTTGCAAACTTTTTCATACATCCC-3'
S88C_F	5'-GGAAAAATCGTTAGATGCTTAAGTTTGGATGAGCTTTTGC-3'
S88C_R	5'-GCAAAAGCTCATCCAAACTTAAGCATCTAACGATTTTTCC-3'
S186_F	5'-GGTTTTAAAACGAAGTGGGGTATGCAAAGAAGGCCATGTTAC-3'
S186_R	5'-GTAACATGGCCTTCTTTGCATACCCCACTTCGTTTTAAAACC-3'
PglC_F_BamHI	5'-AAAAAAGGATCCATGTATGAAAAAGTTTTTAAAAGAATTTTTG-3'
PglC_K4C_BamHI_F	5'-AAAAAAGGATCCATGTATGAATGCGTTTTTAAAAGAATTTTTG-3'
PglC_F6C_BamHI_F	5'-AAAAAAGGATCCATGTATGAAAAAGTTTGCAAAGAATTTTTG-3'
PglC_R_XhoI_2	5'-AAAAAACTCGAGGTTCTTGCCATTAAATTTC-3'
S23A_F	5'-GCTTTAGTGCTTTTAGTGCTTTTTGCTCCGGTGATTTTAATC-3'
S23A_R	5'-GATTAAAATCACCGGAGCAAAAAGCACTAAAAGCACTAAAGC-3'
P24A_F	5'-GATTAAAATCACCGCAGAAAAAAGCACTAAAAGCACTAAAGC-3'
P24A_R	5'-GCTTTAGTGCTTTTAGTGCTTTTTGCTCCGGTGATTTTAATC-3'
P24W_F	5'-GCTTTAGTGCTTTTAGTGCTTTTTTCTTGGGTGATTTTAATC-3'
P24W_R	5'-GATTAAAATCACCCAAGAAAAAAGCACTAAAAGCACTAAAGC-3'
S23A/P24A_F	5'-GCTTTAGTGCTTTTAGTGCTTTTTGCTGCGGTGATTTTAATC-3'
S23A/P24A_R	5'-GATTAAAATCACCGCAGCAAAAAGCACTAAAAGCACTAAAGC-3'
I26A/L27A_F	5'-GTGCTTTTTTCTCCGGTGGCTGCAATCACTGCTTTACTTTTAAAAATCAC-3'
I26A/L27A_R	5'-GTGATTTTTAAAAGTAAAGCAGTGATTGCAGCCACCGGAGAAAAAAGCAC-3'
K187A/E188A_F	5'-CGAAGTGGGGTAAGCGCAGCAGGCCATGTTACAACAGAG-3'
K187A/E188A_R	5'-CTCTGTTGTAACATGGCCTGCTGCGCTTACCCCACTTCG-3'
K7A_F	5'-GTATGAAAAAGTTTTTGCAAGAATTTTTGATTTTATTTTAGCTTTAGTGC-3'
K7A_R	5'-GCACTAAAGCTAAAATAAAATCAAAAATTCTTGCAAAAACTTTTTCATAC-3'
K4C/K7A_F	5'-GTATGAATGCGTTTTTGCAAGAATTTTTGATTTTATTTTAGCTTTAGTGC-3'
K4C/K7A_R	5'-GCACTAAAGCTAAAATAAAATCAAAAATTCTTGCAAAAACGCATTCATAC-3'

93

F6C/K7A_F	5'-GTATGAAAAAGTTTGTGCAAGAATTTTTGATTTTATTTTAGCTTTAGTGC-3'
F6C/K7A_R	5'-GCACTAAAGCTAAAATAAAATCAAAAATTCTTGCACAAACTTTTTCATAC-3'
R8A_F	5'-GTATGAAAAAGTTTTTAAAGCAATTTTTGATTTTATTTTAGCTTTAGTGC-3'
R8A_R	5'-GCACTAAAGCTAAAATAAAATCAAAAATTGCTTTAAAAACTTTTTCATAC-3'
K4C/R8A_F	5'-GTATGAATGCGTTTTTAAAGCAATTTTTGATTTTATTTTAGCTTTAGTGC-3'
K4C/R8A_R	5'-GCACTAAAGCTAAAATAAAATCAAAAATTGCTTTAAAAACGCATTCATAC-3'
F6C/R8A_F	5'-GTATGAAAAAGTTTGTAAAGCAATTTTTGATTTTATTTTAGCTTTAGTGC-3'
F6C/R8A_R	5'-GCACTAAAGCTAAAATAAAATCAAAAATTGCTTTACAAACTTTTTCATAC-3'
K7A/R8A_F	5'-GTATGAAAAAGTTTTTGCAGCAATTTTTGATTTTATTTTAGCTTTAGTGC-3'
K7A/R8A_R	5'-GCACTAAAGCTAAAATAAAATCAAAAATTGCTGCAAAAACTTTTTCATAC-3'
K4C/K7A/R8A_F	5'-GTATGAATGCGTTTTTGCAGCAATTTTTGATTTTATTTTAGCTTTAGTGC-3'
K4C/K7A/R8A_R	5'-GCACTAAAGCTAAAATAAAATCAAAAATTGCTGCAAAAACGCATTCATAC-3'
F6C/K7A/R8A_F	5'-CATATGTATGAAAAAGTTTGTGCAGCAATTTTTGATTTTATTTTAGC-3'
F6C/K7A/R8A_R	5'-GCTAAAATAAAATCAAAAATTGCTGCACAAACTTTTTCATACATATG-3'
SUMO_E3_F	5'-GGTGGGATGTATTGCAAAGTTTTTAAAAGAATTTTTGATTTTATTTTAGC-3'
SUMO_E3_R	5'-GCTAAAATAAAATCAAAAATTCTTTTAAAAACTTTGCAATACATCCCACC-3'
I163C_F	5'-GAACTTGATGTGTATTATGTGAAAAATTGCTCTTTTCTGCTTGATTTAAAAATC-3'
I163C_R	5'-GATTTTTAAATCAAGCAGAAAAGAGCAATTTTTCACATAATACACATCAAGTTC-3'
D168A_F	5'-GTGAAAAATATTTCTTTTCTGCTTGCGTTAAAAATCATGTTTTTAACAGCTTTAAAGG-3'
D168A_R	5'-CCTTTAAAGCTGTTAAAAACATGATTTTTAACGCAAGCAGAAAAGAAATATTTTTCAC-3'
PEB3_A204C_F	5'-CTGACATAGGAACTGCCGTATGCATAGAAAAGATTTGG-3'
PEB3_A204C_R	5'-CCAAATCTTTTTCTATGCATACGGCAGTTCCTATGTCAG-3'

94

SCAM analysis

SCAM analysis methods were based on previously published protocols.[17, 26]

For analysis in *C. jejuni*, PglC variants were cloned out of pET24a into the pCE11128H::rbs expression vector[47] using the BamHI and XhoI restriction sites to insert at C-terminal His_6-tag. Constructs were transformed into *E. coli* DH5α cells carrying the helper plasmid RK212.2 for conjugation into *C. jejuni* 81-176 as described previously.[47] Expression of target constructs was confirmed by Western blot against the C-terminal His_6-tag as described below. Cultures were grown at 37 °C under microaerophilic conditions (10% CO_2, 5% O_2, 85% N_2) for 24 hours in 5 mL Mueller Hinton media with selection antibiotics (10 μg/mL trimethoprim, 15 μg/mL chloramphenicol). After 24 hours, cells were washed once with PBS (137 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, 1.8 mM KH_2PO_4, pH 7.4), adjusted to a final OD_{600} of 4 in 360 μL PBS containing 1 mM EDTA, and divided into three 120 μL aliquots. Labeling proceeded as described below.

For analysis in *E. coli*, BL21-CodonPlus (DE3)-RIL cells (Agilent Technologies) were transformed with the appropriate plasmid constructs. Cultures were grown at 37 °C overnight in 3 mL LB media with selection antibiotics (30 μg/mL kanamycin, 30 μg/mL chloramphenicol). Overnight cultures were diluted into 5 mL fresh LB (with the same selection antibiotic) to a final OD_{600} of 0.2. Cultures were grown for 1 hr at 37 °C, then moved to 16 °C and induced with 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) for 3-5 hr. Following expression, cells were washed once with PBS (137 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, 1.8 mM KH_2PO_4, pH 7.4), adjusted to a final OD_{600} of 1 in 200 μL PBS containing 1 mM EDTA, and divided into four 50 μL aliquots.

Aliquots of *C. jejuni* or *E. coli* cells were treated with 13 μL of 300 mM (*C. jejuni*), or 5 μL of 55 mM (*E. coli*), N-ethylmaleimide (NEM, Alfa-Aesar, cat. # 40526) or sodium(2-sulfonatoethyl)methanethiosulfonate (MTSES, Cayman Chemical, cat. # 16529), to a final concentration of 30 mM or 5 mM, respectively. Samples were incubated at room temperature in the dark for 1 hr. Cells were harvested by centrifugation at 16,000 x *g* for 3 min and washed with 100 μL of PBS. Pellets were resuspended in 22.5 μL lysis buffer (50 mM HEPES, 5% SDS, pH 7.5) and 7.5 μL of 25 mM PEG-mal (Sigma, cat. # 63187) in DMSO. Samples not treated with PEG-mal were treated with 7.5 μL DMSO. Samples were mixed and incubated at room temperature in the dark for 1-1.5 hr. Reactions were quenched by the addition of 30 μL of 2x loading buffer (100 mM Tris pH 6.8, 10% glycerol, 3% SDS, 6 M urea, 1% β-mercaptoethanol, 0.1% bromophenol blue) and frozen at -20 °C until needed.

Proteins were separated by SDS-PAGE on 12% acrylamide gels and transferred to nitrocellulose membranes for Western blot analysis against the His6-tag. Blots were blocked with 3% bovine serum albumin in TBST (10 mM Tris, 150 mM NaCl, 0.05% Tween-20, pH 7.5) for 1 hr. The primary antibody, mouse anti-His (LifeTein, cat. # LT0426) was applied in a 1:3,000 dilution for 1 hr. The secondary antibody, goat anti-mouse AP (Invitrogen, cat. # 31328) was applied in a 1:10,000 dilution for 1 hr. Immunoblots were developed using 1-Step NBT/BCIP Substrate Solution (ThermoFisher Scientific, cat. # 34042) and imaged using a Molecular Imager Gel Doc XR+ System (BioRad).

For the PEB3 positive control, the PEB3-His6 construct in the pET22b vector, described previously,[47] was used. A unique cysteine at the A204 position was introduced using QuikChange II Site-Directed Mutagenesis (Agilent Technologies). Primers used for mutagenesis are listed in **Table 2-2**. Expression in *E. coli*, labeling, and Western blot analysis proceeded as described above.

Molecular dynamics simulations of full-length PglC

All MD simulations were performed with Amber14 [48] using the deposited PglC coordinates (PDB 5W7L).

Membrane insertion model. The placement of PglC in relation to the membrane was predicted using the PPM server: http://opm.phar.umich.edu/server.php.[28]

Simulations of PglC in a POPE membrane. Steepest descent gradient algorithm was iterated for 5,000 steps followed by 5,000 iterations of conjugate gradient algorithm under no constraint. The system was then heated from 0 to 100 K for 2,500 steps in the NVT ensemble while the protein and the lipids were held by a 10 kcal mol^{-1} A^{-1} harmonic potential. In the subsequent step, the system was heated from 100 K to 303 K for 50,000 steps. In the membrane system, the dimensions of the box can change considerably during the first nanoseconds of simulation, thus to allow the program to recalculate them frequently the first 10 steps of the production run were performed for a maximum of 500 ps. In all steps the temperature was controlled by a Langevin thermostat. The heating phase and the production run were performed under an anisotropic NPT ensemble to account for different physical properties along the dimensions tangential to the membrane relative to the ones normal to the membrane.

Distance measurement over time. Distances were measured over time with cpptraj within AmberTools15.[49] Hydrogen bonding was evaluated based on the distance threshold criterion of a H-O distance less than 2.5 Å.

In silico *mutations.* The *in silico* mutations of the Ser-Pro motif in the PglC structure were introduced with Modeller[50] in two steps: Ser23 was mutated to Ala first, followed by mutation of Pro24 to Ala.

Comparison of wild-type and S23A/P24A PglC. Dynamics of wild-type and S23A/P24A PglC were compared in terms of backbone RMSD and RMS fluctuations per residues based on a minimum fit performed on the backbone of the protein. Snapshots of the structures along the simulations were also compared visually and rendered with PyMol 1.6.0.0.[51]

PepLook peptide fold predictions

The online PepLook[29] tool (http://peplook.gembloux.ulg.ac.be/) was used for preliminary peptide folding predictions. The sequence, RIFDILGALFLLILTS̲P̲IIIATAIFIK (from PglC from *C. jejuni*), was submitted used, with substitutions for Ser23 and Pro24 (underlined) as indicated. All sequences were capped with NH at the N-terminal and CO at the C-terminus and submitted for folding in a hydrophobic environment at pH 7.

Peptide folding using molecular dynamics

Two elongated peptide structures were generated using Leap, distributed within AmberTools15.[49] Peptide sequences from *C. concisus* were used so as to compliment molecular dynamics studies performed on the full-length PglC structure. Peptide sequences VIDILGALFLLILTSPIIIATAIFI and VIDILGALFLLILTAAIIIATAIFI were used for wild-type and S23A/P24A PglC, respectively. Leap was also used to generate solvent boxes and add counter ions. Peptides were placed into a water-solvated truncated octahedron simulation box with a minimum distance between the box border and the peptide of 5 Å. The aspartic acid was considered charged negatively, so the total charge was zeroed by adding a sodium ion. The two peptides were simulated for 100 ns each following the protocol described below. The last snapshot of each simulation was extracted and inserted into new truncated octahedron box solvated with a pre-equilibrated 20% isopropanol/water mixture

retrieved from the literature.[52] Both peptides were further simulated in this environment for an additional 1.5 μs.

Peptide folding in water and in 20% isopropanol/water. The system underwent 1,000 steps of steepest descent algorithm followed by 7,000 steps of a conjugate gradient algorithm under a 100 kcal mol^{-1} A^{-2} harmonic potential constraint applied to the protein. The conjugate gradient algorithm minimization continued while the harmonic potential is progressively lowered to 10, 5, 2.5 and 0 kcal mol^{-1} A^{-1} every 600 steps. The system was then heated from 0 K to 100 K using the Langevin thermostat in the canonical ensemble (NVT) while a 20 kcal mol^{-1} A^{-2} harmonic potential restraint was applied on the protein. Finally, the system was heated from 100 K to 300 K in the isothermal–isobaric ensemble (NPT) under the same restraint conditions as the previous step, followed by a simulation for 100 ps under no harmonic restraint. At this point, the system was ready for the production run, which was performed using the Langevin thermostat under NPT ensemble, at a 2 fs time step.

SUMO-PglC purification

SUMO-PglC variants were expressed in BL21-CodonPlus (DE3)-RIL *E. coli* cells (Agilent Technologies) using the Studier auto-induction method.[53] Overnight cultures grown in 3 mL MDG media (0.5% (w/v) glucose, 0.25 (w/v) % aspartate, 2 mM MgSO$_4$, 25 mM Na$_2$HPO$_4$, 25 mM KH$_2$PO$_4$, 50 mM NH$_4$Cl, 5 mM Na$_2$SO$_4$ and 0.2x trace metal mix (from 1000x stock, Teknova, Hollister, CA) with kanamycin and chloramphenicol (30 μg/mL each) were used to inoculate 0.5 L auto-induction media (1% (w/v) tryptone, 0.5% (w/v) yeast extract, 0.5% (v/v) glycerol, 0.05% (w/v) glucose, 0.2% (w/v) α-D-lactose, 2 mM MgSO$_4$, 25 mM Na$_2$HPO$_4$, 25 mM KH$_2$PO$_4$, 50 mM NH$_4$Cl, and 5 mM Na$_2$SO$_4$, 0.2x trace metal mix) containing kanamycin (90 μg/mL) and

chloramphenicol (30 µg/mL). Cells were grown for 4 hr at 37 °C followed by an additional 16-18 hr at 16 °C and then harvested at 3,700 x g for 30 min.

All protein purification steps were carried out at 4 °C. Cells were re-suspended in 10% original culture volume in lysis buffer (50 mM HEPES, 150 mM NaCl, pH 7.5), supplemented with 0.5 mg/mL lysozyme (Research Products International, Mount Prospect, IL), 1:1000 dilution of EDTA-free protease inhibitor cocktail (EMD Millipore, Burlington, MA) and 1 unit/mL DNase I (New England Biolabs, Ipswich, MA). Cells were lysed by sonication (Vibra-Cell, 50% amplitude, 1 sec ON – 2 sec OFF, 2 X 1.5 min; Sonics, Newtown, CT). Lysate was centrifuged at 9,000 x g for 45 min at 4 °C. The resulting supernatant was further centrifuged at 140,000 x g for 65 min at 4 °C to yield the cell envelope fraction (CEF), which was homogenized into 2.5% original culture volume of solubilization buffer (50 mM HEPES, 100 mM NaCl, 1% DDM (*n*-dodecyl β-D-maltoside, Anatrace, Maumee, OH), pH 7.5, and additional protease inhibitor (1:1000 dilution). The suspension was tumbled on a rotating mixer overnight at 4 °C, after which it was centrifuged at 150,000 x g for 65 min to remove any insoluble material.

The supernatant was incubated with Ni-NTA resin (1 mL per L original culture volume; Thermo Fisher Scientific, Waltham, MA) pre-treated with equilibration buffer (50 mM HEPES, 100 mM NaCl, 20 mM imidazole, 5% glycerol, pH 7.5) for 1 hr. The resin was washed with 20 column volumes of Wash I buffer (equilibration buffer + 0.03% DDM), followed by 20 column volumes of Wash II buffer (equilibration buffer + 0.03% DDM + 45 mM imidazole). The protein was eluted from the column in 2 column volumes of elution buffer (equilibration buffer + 0.03% DDM + 500 mM imidazole). Elution fractions were combined and immediately desalted using a 5 ml HiTrap desalting column (GE Healthcare, Marlborough, MA) that was pre-equilibrated with PglC buffer (50 mM HEPES, 100 mM NaCl, 0.03% DDM, 5% glycerol, pH 7.5). Fractions

containing SUMO-PglC were supplemented with addition DDM to a final concentration of 0.2% and flash frozen for storage at -80 °C.

Circular dichroism

Circular dichroism was performed on a JASCO Model J-1500 Circular Dichroism Spectrometer. Spectral scans were performed with 16 μM purified SUMO-PglC in phosphate buffer (20 mM Na_2HPO_4, 100 mM NaCl, 5% glycerol, 0.2% DDM, pH 7.5). Reads were taken at 4 °C from 190-250 nm at 0.5 nm intervals.

Thermal shift assay

Thermal shift assays were based on a protocol described previously.[34] Aliquots of 6-8 μM purified SUMO-PglC in PglC buffer (50 mM HEPES, 100 mM NaCl, 5% glycerol, 0.2% DDM, pH 7.5) was heated for 10 min at 30-99 °C in a PCR machine (MJ Mini Thermal Cycler; BioRad, Hercules, CA). Precipitate was immediately removed by centrifugation at 16,000 x g for 10 min at 4 °C. The resulting supernatant, containing protein that remained soluble, was analyzed by SDS-PAGE with Coomassie staining and quantified by gel densitometry using a Molecular Imager Gel Doc XR+ System with Image Lab software (BioRad). Data were fitted using Graphpad Prism 7 (GraphPad Software, La Jolla, CA). Technical replicates were performed on distinct aliquots taken from a common protein purification prep.

Dibromobimane crosslinking

For crosslinking experiments, 5 μM SUMO-PglC in PglC buffer (50 mM HEPES, 100 mM NaCl, 5% glycerol, 0.2% DDM, pH 7.5) was treated with 25 μM dibromobimane (Sigma-Aldrich, St. Louis, MO) from a 1 mM stock in 20% acetonitrile for 20 min in the dark at room temperature. For fluorescence analysis, some crosslinked samples were further treated with 50 mM DTT for 1 min, and all samples were then analyzed by SDS-PAGE and quantified by gel densitometry using a Molecular Imager Gel Doc XR+ System with Image Lab software (BioRad). Fluorescence was measured using the ethidium bromide excitation setting (302 nm) and normalized to band intensity after Coomassie staining. Relative fluorescence of SUMO-PglC variants is reported as fluorescence intensity, normalized to Coomassie staining, relative to DTT-quenched samples. Technical replicates were performed on distinct aliquots taken from a common protein purification prep.

For activity assays crosslinked SUMO-PglC prepared as above was diluted with PglC buffer to a 50 nM stock and assayed as described below. Activity of variants is reported relative to control samples that were treated with carrier (20% acetonitrile) and assayed in parallel with crosslinked samples. Data were plotted using Graphpad Prism 7 (GraphPad Software).

SUMO-PglC activity assay

SUMO-PglC activity assays were performed as described previously using the UMP/CMP-Glo assay (Promega).[27, 54] Assays were performed on a 10 μL scale at room temperature with 5 nM enzyme (from a 50 nM stock) and 20 μM of both UDP-N,N'-diacetylbacillosamine and Und-P substrates (from 200 μM stocks in water and DMSO, respectively) in assay buffer (50 mM HEPES, 100 mM NaCl, 5 mM MgCl2, 0.1% Triton X-100, pH 7.5; in assays of Cys-containing SUMO-

PglC variants, 7 mM β-mercaptoethanol was also added to the assay buffer). Following quenching

with 10 μL of UMP-Glo reagent, luminescence was read in a 96-well plate (Corning, Inc., Corning,

NY) using a SynergyH1 multimode plate reader (Biotek, Winooski, VT). The 96-well plate was

shaken inside the plate reader chamber at 237 cpm at 25 °C in the double orbital mode for 16 min,

followed by 44 min incubation at the same temperature, after which time the luminescence was

recorded (gain: 200, integration time: 0.5 s). Conversion of luminescence to UMP concentration

was carried out using a standard curve. Data were plotted using GraphPad Prism 7 (GraphPad

Software). Technical replicates were performed on distinct aliquots taken from a common protein

purification prep.

REFERENCES

(1) Blobel, G., Intracellular protein topogenesis. *Proc Natl Acad Sci U S A* **1980**, *77* (3), 1496-500.

(2) Krebs, M. P.; Noorwez, S. M.; Malhotra, R.; Kaushal, S., Quality control of integral membrane proteins. *Trends Biochem Sci* **2004**, *29* (12), 648-55.

(3) Krogh, A.; Larsson, B.; von Heijne, G.; Sonnhammer, E. L., Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J Mol Biol* **2001**, *305* (3), 567-80.

(4) Tsirigos, K. D.; Peters, C.; Shu, N.; Kall, L.; Elofsson, A., The TOPCONS web server for consensus prediction of membrane protein topology and signal peptides. *Nucleic Acids Res* **2015**, *43* (W1), W401-7.

(5) Elazar, A.; Weinstein, J. J.; Prilusky, J.; Fleishman, S. J., Interplay between hydrophobicity and the positive-inside rule in determining membrane-protein topology. *Proc Natl Acad Sci U S A* **2016**, *113* (37), 10340-5.

(6) von Heijne, G., The distribution of positively charged residues in bacterial inner membrane proteins correlates with the trans-membrane topology. *EMBO J* **1986**, *5* (11), 3021-7.

(7) Gafvelin, G.; von Heijne, G., Topological "frustration" in multispanning E. coli inner membrane proteins. *Cell* **1994**, *77* (3), 401-12.

(8) von Heijne, G., Membrane-protein topology. *Nat Rev Mol Cell Biol* **2006**, *7* (12), 909-18.

(9) Lukose, V.; Walvoort, M. T. C.; Imperiali, B., Bacterial phosphoglycosyl transferases: initiators of glycan biosynthesis at the membrane interface. *Glycobiology* **2017**, *27* (9), 820-833.

(10) Anderson, M. S.; Eveland, S. S.; Price, N. P., Conserved cytoplasmic motifs that distinguish sub-groups of the polyprenol phosphate:N-acetylhexosamine-1-phosphate transferase family. *FEMS Microbiol Lett* **2000**, *191* (2), 169-75.

(11) Burda, P.; Aebi, M., The dolichol pathway of N-linked glycosylation. *Biochim Biophys Acta* **1999**, *1426* (2), 239-57.

(12) Glover, K. J.; Weerapana, E.; Chen, M. M.; Imperiali, B., Direct biochemical evidence for the utilization of UDP-bacillosamine by PglC, an essential glycosyl-1-phosphate transferase in the Campylobacter jejuni N-linked glycosylation pathway. *Biochemistry* **2006**, *45* (16), 5343-50.

(13) Szymanski, C. M.; Yao, R.; Ewing, C. P.; Trust, T. J.; Guerry, P., Evidence for a system of general protein glycosylation in Campylobacter jejuni. *Mol Microbiol* **1999**, *32* (5), 1022-30.

(14) Lukose, V.; Luo, L. Q.; Kozakov, D.; Vajda, S.; Allen, K. N.; Imperiali, B., Conservation and covariance in small bacterial phosphoglycosyltransferases identify the functional catalytic core. *Biochemistry* **2015**, *54* (50), 7326-7334.

(15) Saldias, M. S.; Patel, K.; Marolda, C. L.; Bittner, M.; Contreras, I.; Valvano, M. A., Distinct functional domains of the Salmonella enterica WbaP transferase that is involved in the initiation reaction for synthesis of the O antigen subunit. *Microbiology* **2008,** *154* (Pt 2), 440-53.

(16) Hartley, M. D.; Morrison, M. J.; Aas, F. E.; Borud, B.; Koomey, M.; Imperiali, B., Biochemical characterization of the O-linked glycosylation pathway in Neisseria gonorrhoeae responsible for biosynthesis of protein glycans containing N,N'-diacetylbacillosamine. *Biochemistry* **2011,** *50* (22), 4936-48.

(17) Furlong, S. E.; Ford, A.; Albarnez-Rodriguez, L.; Valvano, M. A., Topological analysis of the Escherichia coli WcaJ protein reveals a new conserved configuration for the polyisoprenyl-phosphate hexose-1-phosphate transferase family. *Sci Rep* **2015,** *5,* 9178.

(18) Ray, L. C.; Das, D.; Entova, S.; Lukose, V.; Lynch, A. J.; Imperiali, B.; Allen, K. N., Membrane association of monotopic phosphoglycosyl transferase underpins function. *Nat Chem Biol* **2018,** *14,* 538-541.

(19) Viklund, H.; Granseth, E.; Elofsson, A., Structural classification and prediction of reentrant regions in alpha-helical transmembrane proteins: application to complete genomes. *J Mol Biol* **2006,** *361* (3), 591-603.

(20) Bracey, M. H.; Cravatt, B. F.; Stevens, R. C., Structural commonalities among integral membrane enzymes. *FEBS Lett* **2004,** *567* (2-3), 159-65.

(21) Okamoto, T.; Schlegel, A.; Scherer, P. E.; Lisanti, M. P., Caveolins, a family of scaffolding proteins for organizing "preassembled signaling complexes" at the plasma membrane. *J Biol Chem* **1998,** *273* (10), 5419-22.

(22) Aoki, S.; Thomas, A.; Decaffmeyer, M.; Brasseur, R.; Epand, R. M., The role of proline in the membrane re-entrant helix of caveolin-1. *J Biol Chem* **2010,** *285* (43), 33371-80.

(23) Epand, R. M.; So, V.; Jennings, W.; Khadka, B.; Gupta, R. S.; Lemaire, M., Diacylglycerol kinase-epsilon: properties and biological roles. *Front Cell Dev Biol* **2016,** *4,* 112.

(24) Decaffmeyer, M.; Shulga, Y. V.; Dicu, A. O.; Thomas, A.; Truant, R.; Topham, M. K.; Brasseur, R.; Epand, R. M., Determination of the topology of the hydrophobic segment of mammalian diacylglycerol kinase epsilon in a cell membrane and its relationship to predictions from modeling. *J Mol Biol* **2008,** *383* (4), 797-809.

(25) Nørholm, M. H.; Shulga, Y. V.; Aoki, S.; Epand, R. M.; von Heijne, G., Flanking residues help determine whether a hydrophobic segment adopts a monotopic or bitopic topology in the endoplasmic reticulum membrane. *J Biol Chem* **2011,** *286* (28), 25284-90.

(26) Nasie, I.; Steiner-Mordoch, S.; Schuldiner, S., Topology determination of untagged membrane proteins. *Methods Mol Biol* **2013,** *1033,* 121-30.

(27) Das, D.; Walvoort, M. T. C.; Lukose, V.; Imperiali, B., A rapid and eEfficient luminescence-based method for assaying phosphoglycosyltransferase enzymes. *Sci Rep* **2016,** *6,* 33412.

(28) Lomize, M. A.; Pogozheva, I. D.; Joo, H.; Mosberg, H. I.; Lomize, A. L., OPM database and PPM web server: resources for positioning of proteins in membranes. *Nucleic Acids Res* **2012**, *40*, D370-6.

(29) Thomas, A.; Deshayes, S.; Decaffmeyer, M.; Van Eyck, M. H.; Charloteaux, B.; Brasseur, R., Prediction of peptide structure: how far are we? *Proteins* **2006**, *65* (4), 889-97.

(30) Martoglio, B.; Dobberstein, B., Signal sequences: more than just greasy peptides. *Trends Cell Biol* **1998**, *8* (10), 410-5.

(31) Grinberg, A. V.; Gevondyan, N. M.; Grinberg, N. V.; Grinberg, V. Y., The thermal unfolding and domain structure of Na+/K+-exchanging ATPase. A scanning calorimetry study. *Eur J Biochem* **2001**, *268* (19), 5027-36.

(32) Stowell, M. H.; Rees, D. C., Structure and stability of membrane proteins. *Adv Protein Chem* **1995**, *46*, 279-311.

(33) Vogel, R.; Siebert, F., Conformation and stability of alpha-helical membrane proteins. 2. Influence of pH and salts on stability and unfolding of rhodopsin. *Biochemistry* **2002**, *41* (11), 3536-45.

(34) Gandini, R.; Reichenbach, T.; Tan, T. C.; Divne, C., Structural basis for dolichylphosphate mannose biosynthesis. *Nat Commun* **2017**, *8* (1), 120.

(35) Kim, J. S.; Raines, R. T., Dibromobimane as a fluorescent crosslinking reagent. *Anal Biochem* **1995**, *225* (1), 174-6.

(36) Kosower, N. S.; Kosower, E. M., Thiol labeling with bromobimanes. *Methods Enzymol* **1987**, *143*, 76-84.

(37) Piela, L.; Nemethy, G.; Scheraga, H. A., Proline-induced constraints in alpha-helices. *Biopolymers* **1987**, *26* (9), 1587-600.

(38) Cymer, F.; von Heijne, G.; White, S. H., Mechanisms of integral membrane protein insertion and folding. *J Mol Biol* **2015**, *427* (5), 999-1022.

(39) Mackenzie, K. R., Folding and stability of alpha-helical integral membrane proteins. *Chem Rev* **2006**, *106* (5), 1931-77.

(40) Almeida, P. F.; Ladokhin, A. S.; White, S. H., Hydrogen-bond energetics drive helix formation in membrane interfaces. *Biochim Biophys Acta* **2012**, *1818* (2), 178-82.

(41) Yan, C.; Luo, J., An analysis of reentrant loops. *Protein J* **2010**, *29* (5), 350-4.

(42) Lee, J.; Glover, K. J., The transmembrane domain of caveolin-1 exhibits a helix-break-helix structure. *Biochim Biophys Acta* **2012**, *1818* (5), 1158-64.

(43) Kohanski, M. A.; Dwyer, D. J.; Wierzbowski, J.; Cottarel, G.; Collins, J. J., Mistranslation of membrane proteins and two-component system activation trigger antibiotic-mediated cell death. *Cell* **2008**, *135* (4), 679-90.

(44) Schow, E. V.; Freites, J. A.; Cheng, P.; Bernsel, A.; von Heijne, G.; White, S. H.; Tobias, D. J., Arginine in membranes: the connection between molecular dynamics simulations and translocon-mediated insertion experiments. *J Membr Biol* **2011,** *239* (1-2), 35-48.

(45) Zhou, G. P.; Troy, F. A., 2nd, Characterization by NMR and molecular modeling of the binding of polyisoprenols and polyisoprenyl recognition sequence peptides: 3D structure of the complexes reveals sites of specific interactions. *Glycobiology* **2003,** *13* (2), 51-71.

(46) Zhou, G. P.; Troy, F. A., 2nd, NMR study of the preferred membrane orientation of polyisoprenols (dolichol) and the impact of their complex with polyisoprenyl recognition sequence peptides on membrane structure. *Glycobiology* **2005,** *15* (4), 347-59.

(47) Silverman, J. M.; Imperiali, B., Bacterial N-glycosylation efficiency is dependent on the structural context of target sequons. *J Biol Chem* **2016,** *291* (42), 22001-22010.

(48) Case, D. A.; Babin, V.; Berryman, J. T.; Betz, R. M.; Cai, Q.; Cerutti, D. S.; Cheatham III, T. E.; Darden, T. A.; Duke, R. E.; Gohlke, H.; Goetz, A. W.; Gusarov, S.; Homeyer, N.; Janowski, P.; Kaus, J.; Kolossvary, I.; Kovalenko, A.; Lee, T. S.; LeGrand, S.; Luchko, T.; Luo, R.; Madej, B.; Merz, K. M.; Paesani, F.; Roe, D. R.; Roitberg, A.; Sagui, C.; Salomon-Ferrer, R.; Seabra, G.; Simmerling, C. L.; Smith, W.; Swails, J.; Walker, R. C.; Wang, J.; Wolf, R. M.; Wu, X.; Kollman, P. A. *AMBER 2014*, University of California, San Francisco, 2014.

(49) Case, D. A.; Berryman, J. T.; Betz, R. M.; Cerutti, D. S.; Cheatham III, T. E.; Darden, T. A.; Duke, R. E.; Giese, T. J.; Gohlke, H.; Goetz, A. W.; Homeyer, N.; Izadi, S.; Janowski, P.; Kaus, J.; Kovalenko, A.; Lee, T. S.; LeGrand, S.; Li, P.; Luchko, T.; Luo, R.; Madej, B.; Merz, K. M.; Monard, G.; Needham, P.; Nguyen, H.; Nguyen, H. T.; Omelyan, I.; Onufriev, A.; Roe, D. R.; Roitberg, A.; Salomon-Ferrer, R.; Simmerling, C. L.; Smith, W.; Swails, J.; Walker, R. C.; Wang, J.; Wolf, R. M.; Wu, X.; York, D. M.; Kollman, P. A. *AMBER 2015*, University of California, San Francisco, 2015.

(50) Sali, A.; Blundell, T. L., Comparative protein modelling by satisfaction of spatial restraints. *J Mol Biol* **1993,** *234* (3), 779-815.

(51) Schrodinger, L. *The PyMOL Molecular Graphics System, Version 1.6,* : , 2015.

(52) Alvarez, D. Organic solvent boxes. http://www.ub.edu/cbdd/?q=content/organic-solvent-boxes.

(53) Studier, F. W., Protein production by auto-induction in high density shaking cultures. *Protein Expr Purif* **2005,** *41* (1), 207-34.

(54) Das, D.; Kuzmic, P.; Imperiali, B., Analysis of a dual domain phosphoglycosyl transferase reveals a ping-pong mechanism with a covalent enzyme intermediate. *Proc Natl Acad Sci U S A* **2017,** *114* (27), 7019-7024.

107

Chapter 3: Insights into determinants of membrane topology enable the identification of new monotopic folds

INTRODUCTION

Integral membrane proteins play many diverse and essential roles in prokaryotic and eukaryotic biology; yet, obtaining experimental structural information for membrane proteins has proven to be relatively difficult. For this reason, there has long been interest in the development of computational algorithms for the reliable prediction of membrane topology in protein sequences for which structural information is unavailable.

Two of the earliest attempts at such an algorithm were first reported almost four decades ago by Kyte and Doolittle[1] and by Eisenberg et al.[2] Such early methods strove to predict the location of a peptide sequence (i.e. globular protein interior, protein surface, transmembrane domain, etc.) based on the calculated hydrophobicity of the residues in the sequence. Subsequently, more refined approaches such as TopPredII[3] and MEMSAT[4] applied a sliding window to the evaluation of hydrophobicity along a protein sequence, and additionally incorporated the notion of the "positive-inside rule" to predict the orientation of transmembrane sequences in the membrane. The positive-inside rule is an empirically-determined principle that dictates that transmembrane sequences preferentially adopt the membrane topology which maximizes the localization of positive charges in the sequence to the cytoplasmic side of the membrane.[5-7]

The late 1990s saw the development of algorithms that applied hidden Markov models and more rigorous topological constraints to assess membrane protein topology globally, rather than locally, in protein sequences.[8-9] One such algorithm developed by Krogh and coworkers, the TMHMM,[10-11] was widely used throughout the early 2000s to predict protein topology, and is still in use today. The TMHMM algorithm was also used to predict that 20-30% of all annotated genes encode integral membrane proteins.[10] Many of the algorithms described in the last decade,

such as OCTOPUS[12] and TOPCONS,[13] have built on the use of hidden Markov models for topology prediction, additionally implementing artificial neural network methods[12-14] and incorporating information from homology to proteins of known structure. Such developments have improved the ability of algorithms to discern between genuine transmembrane segments and signal peptides and allowed for more refined annotations of interface and loop regions. Notably, one recently developed prediction tool, TopGraph, is not based on a hidden Markov model or homology to known structures; instead, this method uses experimentally derived free energies of membrane insertion for each of the 20 amino acids to search a queried sequence for the topology that results in the minimum membrane insertion energy.[15] The reported accuracy of current topology prediction algorithms ranges from 80-95% for the accurate prediction of transmembrane segments.[12-15]

By contrast, relatively little work has been done to address the prediction of reentrant domains in membrane proteins. One prediction tool, OCTOPUS, attempts to discern single transmembrane domains from helical hairpins (two transmembrane helices connected by a short loop), reentrant helices, and "membrane dips"; however, the authors note that such predictions are difficult, and the tool only predicts reentrant helices and "membrane dips" to an accuracy of ~20%.[12] Indeed, topology predictions of PglC and the eukaryotic reentrant protein caveolin-1[16-18] by several commonly-used algorithms results in inaccurate topology prediction (**Table 3-1** and **Table 3-2**). Most prediction algorithms incorrectly identify the boundaries of the hydrophobic domains of both folds, and none are able to identify the reentrant topologies, instead predicting transmembrane domains (or in one case, no hydrophobic domains at all). In addition, all the algorithms tested predict both folds to be in an orientation with the N-terminus in the cytoplasm and the C-terminus in the periplasm or extracellular space; for both PglC and caveolin, such a

conformation would not orient the protein correctly for its functional role. Thus, despite the significant progress made towards accurate prediction of membrane-spanning domains, our ability to predict reentrant hydrophobic domains is still relatively poor.

Table 3-1 – Topology predictions for PglC (*C. jejuni*), by algorithm			
algorithm	topology	residues	orientation*
Actual topology[19]	reentrant	8-37	N_{in}-C_{in}
TMHMM[10]	transmembrane	12-33	N_{in}-C_{out}
OCTOPUS[12]	transmembrane	12-32	N_{in}-C_{out}
SCOPTOPUS[14]	transmembrane	12-32	N_{in}-C_{out}
TOPCONS[13]	-	no TM regions predicted	-
TopGraph[15]	transmembrane	8-37	N_{in}-C_{out}

Table 3-2 – Topology predictions for caveolin-1 (*H. sapiens*), by algorithm			
algorithm	topology	residues	orientation*
Actual topology[18]	reentrant	96-136	N_{in}-C_{in}
TMHMM[10]	transmembrane	97-119	N_{in}-C_{out}
OCTOPUS[12]	transmembrane	101-131	N_{in}-C_{in}
SCOPTOPUS[14]	transmembrane	101-131	N_{in}-C_{out}
TOPCONS[13]	transmembrane	108-128	N_{in}-C_{out}
TopGraph[15]	transmembrane	102-123	N_{in}-C_{out}

Table 3-1 and Table 3-2 Topology predictions for PglC and caveolin-1 by multiple algorithms. Topology predictions by multiple commonly used algorithms are compared. The experimentally determined topology, residues involved, and orientation in the membrane are also shown. *"Orientation" refers to whether the N- and C-termini of the membrane-inserted domain are predicted to be in the cytoplasm ("in") or in the periplasm/extracellular space ("out").

The development of improved algorithms for the identification of reentrant membrane

domains would be of significant utility in the study of membrane proteins with diverse

topologies. Reentrant domains have been reported in the experimentally-determined structures of

many polytopic folds.[20-21] In polytopic proteins, the orientation in the membrane of any

transmembrane helix is constrained by the orientations of neighboring helices: if one helix is N_{in}-

C_{out}, the next must be N_{out}-C_{in}, and so on. Thus, in predicting the topology of a polytopic

membrane protein, the misannotation of a reentrant helix as membrane-spanning results in the

incorrect orientation assignment of the following transmembrane helices. Alternatively, in

membrane proteins containing a single membrane-inserted domain, such as PglC, proper

assignment of a reentrant helix distinguishes between a monotopic and a bitopic predicted

topology.

Although few reentrant monotopic topologies have been structurally characterized to

date, topology predictions have suggested that bitopic membrane proteins constitute 15-25% of

all integral membrane proteins in bacteria and are even more prevalent among eukaryotes,

accounting for almost half of all human membrane proteins.[10, 22] As demonstrated in **Table 3-1**

and **Table 3-2**, the topology prediction algorithms used to generate such estimates are likely to

annotate both "true" bitopic proteins, and monotopic proteins with a reentrant helix, as bitopic

transmembrane proteins. Furthermore, the membrane-inserted domains of predicted bitopic

proteins are reported to vary in length, hydrophobic thickness, and free energies of folding and of

membrane insertion.[23] Thus, it is possible that among the many, diverse predicted single-pass

transmembrane proteins, there exist reentrant, monotopic folds, like those of PglC and caveolin,

that have been erroneously identified as bitopic. A more thorough understanding of the key

114

features distinguishing reentrant domains from transmembrane helices has the potential to inform the creation of prediction algorithms capable of accurately identifying such folds.

RESULTS

PglC topogenesis studies inform the identification of novel monotopic protein folds

Insights from the in-depth study of motifs driving topogenesis in PglC enabled us to demonstrate that the co-occurrence of similar motifs could be employed to identify reentrant topologies in additional, unrelated membrane proteins.

The Membranome database of single-helix transmembrane proteins[23] contains curated data on >6,000 known and predicted bitopic membrane proteins from eukaryotic and prokaryotic genomes. We manually parsed the Membranome list of bacterial bitopic proteins, comprising 196 sequences from *Escherichia coli*, for hydrophobic domains that might be reentrant based on the following criteria: 1) the sequence contains an N-terminal hydrophobic domain that contains conserved helix-breaking residues and is preceded by conserved positive charges; 2) for candidates of known function, a reentrant topology would be consistent with the reported biological role; 3) where possible, covariance analysis confirms probable contacts between the N-terminal hydrophobic domain and the rest of the fold. Covariance analysis on sequence homologs is a powerful tool for identifying interacting pairs of residues within a structure,[24-25] and in some cases has even be used to model unknown protein folds.[26] A reentrant topology is significantly more likely than a membrane-spanning one to facilitate interactions between the hydrophobic and soluble domains of a fold: indeed, covariance analyses of PglC previously identified several residues in the reentrant membrane helix (RMH) that contact the globular

domain.[19, 27] Based on the above criteria, several "bitopic" proteins from *E. coli* were identified

to show evidence of reentrant N-terminal hydrophobic domains (**Table 3-3** and Appendix).

Name	Function	Sequence of N-terminus (residues 1-50)
HofN	Putative DNA utilization protein	MNPPINFLPWRQQRRTAFLRFWLLMFVAPLLLAVGITLILRLTGSAEARI
LpxL	Lipid A biosynthesis lauroyltransferase	MTNLPKFSTALLHPRYWLTWLGIGVLWLVVQLPYPVIYRLGCGLGKLALR
LpxM	Lipid A biosynthesis myristoyltransferase	METKKNNSEYIPEFDKSFRHPRYWGAWLGVAAMAGIALTPPKFRDPILAR
LpxP	Lipid A biosynthesis palmitoleoyltransferase	MFPQCKFSREFLHPRYWLTWFGLGVLWLWVQLPYPVLCFLGTRIGAMARP
YihG	Probable acyltransferase	MANLLNKFIMTRILAAITLLLSIVLTILVTIFCSVPIIIAGIVKLLLPVP

Table 3-3 Candidate reentrant domains identified among predicted bitopic proteins from *E. coli*. Sequences were selected from the Membranome database of 196 known and predicted bitopic folds from *E. coli*, based on the criteria described in the text. Sequences highlighted in red represent the predicted transmembrane domain (TMHMM Server v2.0, http://www.cbs.dtu.dk/services/TMHMM/).

Lipid A biosynthesis acyltransferases are predicted to adopt a reentrant topology

Three candidates – LpxM, LpxL and LpxP – are of particular interest. These enzymes belong to

a large family of lipid A biosynthesis acyltransferases: over 4,000 homologs of each sequence

were identified during covariance analyses. LpxM, LpxL and LpxP catalyze the transfer of

myristoyl, lauroyl, or palmitoleoyl groups, respectively, to Kdo_2-lipid IV to produce lipid A, the

lipidic component of outer membrane lipopolysaccharide found in most Gram-negative bacteria.

As such, these enzymes carry significant therapeutic relevance as antibiotic targets. Additionally,

the three enzymes belong to an extensive and diverse superfamily of lysophospholipid

acyltransferases, which comprises members from all three domains of life and also includes

several families of enzymes involved in triglyceride biosynthesis.[28-29] The biochemical function

of these enzymes in lipid A biosynthesis dictates that the soluble C-terminal domain of each

must localize to the cytoplasmic face of the inner membrane; accordingly, the predicted

transmembrane topology places the N-terminus in the periplasm. However, we noted that

multiple positively charged residues precede the hydrophobic domain in each sequence, making

localization of the N-terminus to the periplasm less likely. In addition, the three hydrophobic

domains contain polar (Gln, Thr) and multiple aromatic (Trp, Tyr) residues not typically found in

the middle of transmembrane helices,[6] as well as helix-breaking Pro and Gly residues. Finally,

covariance analyses of LpxM, LpxL and LpxP identified several instances of covariance between

residues in the hydrophobic and globular domains (Appendix), suggesting interactions between

the two. On the basis of these observations, we hypothesized that LpxM, LpxL and LpxP adopt

reentrant membrane topologies, rather than the predicted membrane-spanning ones.

A reentrant topology is confirmed for LpxM

Accordingly, we performed a SCAM analysis to determine the topology of the N-terminal

domain in a representative member of the lipid A biosynthesis acyltransferase family. LpxM was

chosen of the three acyltransferases as it has the fewest native cysteines, making it most

amenable to the SCAM method. LpxM has two native cysteines in the globular domain (C73

and C240), neither of which is highly conserved; thus, either one or both were substituted with

serine to create unique cysteine variants C73 and C240, and a "Cysless" variant. Unique

cysteines were introduced to the "Cysless" LpxM at non-conserved positions at the N-terminus

(S8, I11 and S17) and at an additional location on the globular domain (M89). SCAM analysis of "Cysless" LpxM and the six unique cysteine variants revealed that both the N-terminus and the globular domain of LpxM are located in the cytoplasm (**Figure 3-1**).
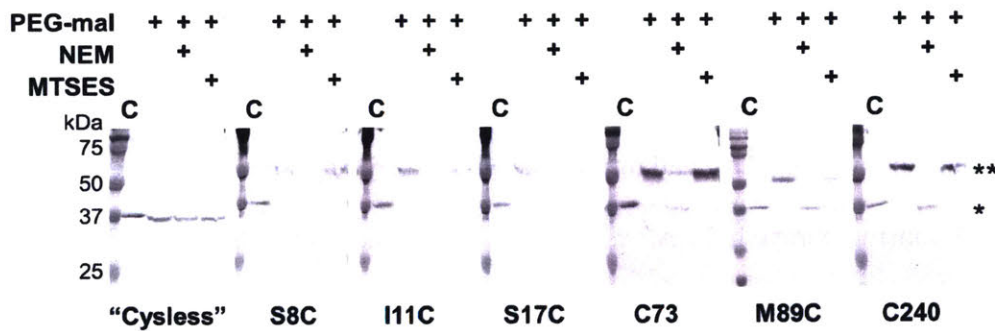
Figure 3-1 SCAM analysis of LpxM from *E. coli* indicates a reentrant membrane topology. Legend: * = native LpxM; ** = LpxM labeled with PEG-mal; C = control, no PEG-mal labeling. All SCAM experiments were performed in duplicate or more. Representative Western blots are shown.

Thus, SCAM analysis confirms a reentrant topology for LpxM. Notably, the only lipid A biosynthesis acyltransferase for which a structure has been reported to date is the LpxM homolog from *Acinetobacter baumannii*.[30] The N-terminus of the LpxM homolog from *A. baumannii* is similarly reported to be a predicted transmembrane domain; however, in the reported structure, this domain protrudes from the globular domain as a helix-break-helix (**Figure 3-2**), reminiscent of a reentrant domain.
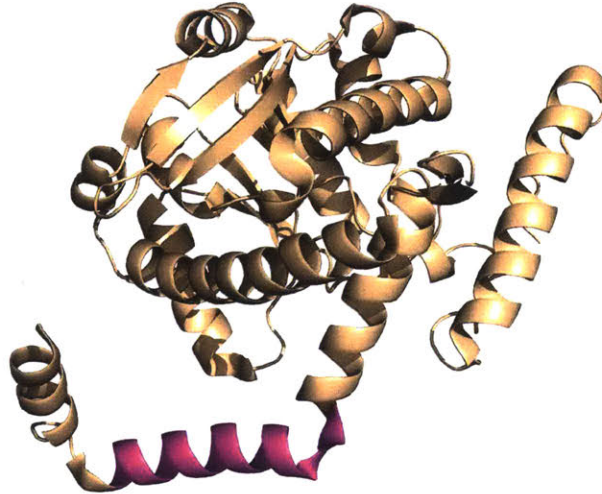
Figure 3-2 Structure of LpxM from *A. baumannii.*
The predicted transmembrane domain, as reported for the structure, is shown in pink.
PDB ID 5KN7.[30]

We noted that the predicted transmembrane region is less hydrophobic in the homolog

from *A. baumannii* than in those from *E. coli* and other Gram-negative pathogens (**Figure 3-3**),

making a membrane-spanning topology particularly unlikely. On the contrary, this could indicate

that the N-terminus of LpxM from *A. baumannii* forms a reentrant domain that penetrates the

membrane much more shallowly than the reentrant domains of LpxM homologs from *E. coli* and

other bacteria. Coloring of the N-terminal domain of LpxM from *A. baumannii* according to

hydrophobicity suggested that parts of the N-terminus may even form an amphipathic helix

Figure 3-4. Due to the low hydrophobicity of the N-terminus in this homolog, we were not able

to position this fold in the membrane using the PPM server.

Figure 3-3 Sequence alignment of LpxM homologs.

LpxM homologs from *A. baumannii, Pseudomonas aeruginosa, Klebsiella pneumoniae, E. coli* and *Salmonella enterica* are compared. Only the N-terminus is shown. The predicted transmembrane region, corresponding to residues 23-40 of LpxM from *E. coli*, is underlined with a black bar. Black dots indicate the location of unique cysteines introduced into the N-terminus of LpxM from *E. coli* for SCAM analysis.
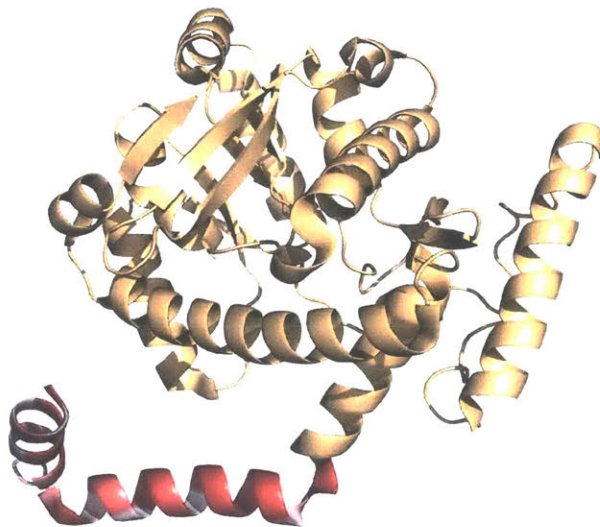
Figure 3-4 Coloring of the N-terminal domain of LpxM from *A. baumannii* by hydrophobicity. The N-terminus is colored by hydrophobicity from hydrophobic (*red*) *to* hydrophilic (*white*); the coloring pattern is suggestive of an amphipathic helix.

LpxL and LpxP are likely to also adopt reentrant topologies

Given the similarity between LpxM, LpxL and LpxP, we propose that the reentrant topology is conserved among the related lipid A biosynthesis acyltransferases. Peptide folding predictions using the QUARK algorithm[31-32] suggested that the N-termini of both LpxM and LpxP would adopt helix-break-helix structures with a conserved proline residue contributing to the break point (**Figure 3-5**).
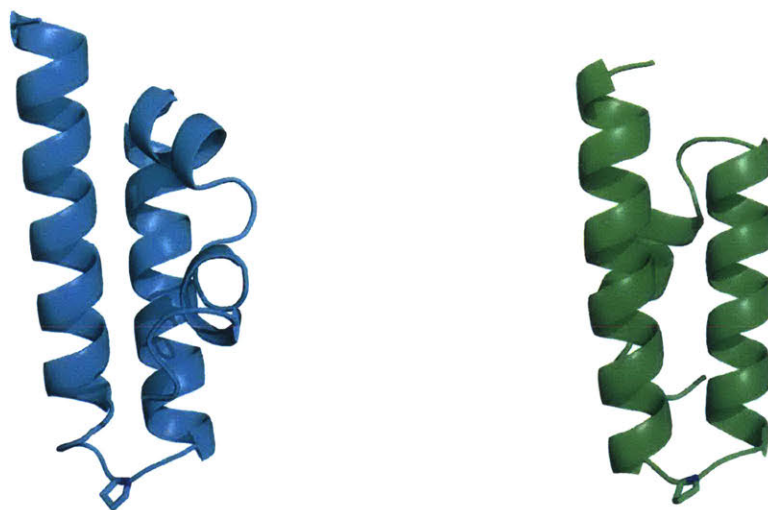
Figure 3-5 Predicted structures of the N-terminal domains of LpxM and LpxP. Peptides representing the N-terminal domains of LpxM (*left*) and LxpP (*right*) from *E. coli* were folded using the QUARK algorithm.[31-32] A conserved proline at the break in each helix is shown as sticks.

We noted that the predicted hydrophobic domain of LpxM is shorter (18 residues) than the predicted hydrophobic domains of LpxP and LpxL (each 23 residues, **Table 3-3**), suggesting that LpxP and LpxL might sit deeper in the membrane than LpxM. Accordingly, the peptide folds predicted by the QUARK algorithm for the N-terminal domains of LpxM and LpxP were

121

submitted to the PPM server, which calculates the rotational and translational positions of membrane proteins in the membrane plane.[33] As anticipated, the more hydrophobic, N-terminal domain of LpxP was found to sit deeper in the membrane than the less hydrophobic N-terminal domain of LpxM (**Figure 3-6**). This indicates that although all three Lipid A biosynthesis acyltransferases likely adopt similar reentrant topologies, subtle differences in the sequences of the respective N-terminal domains may dictate how deeply each fold embeds in the membrane. As the hydrophobic domains of LpxL and LpxP are very similar in sequence, they are expected to embed similarly in the membrane, while LpxM embeds more shallowly.
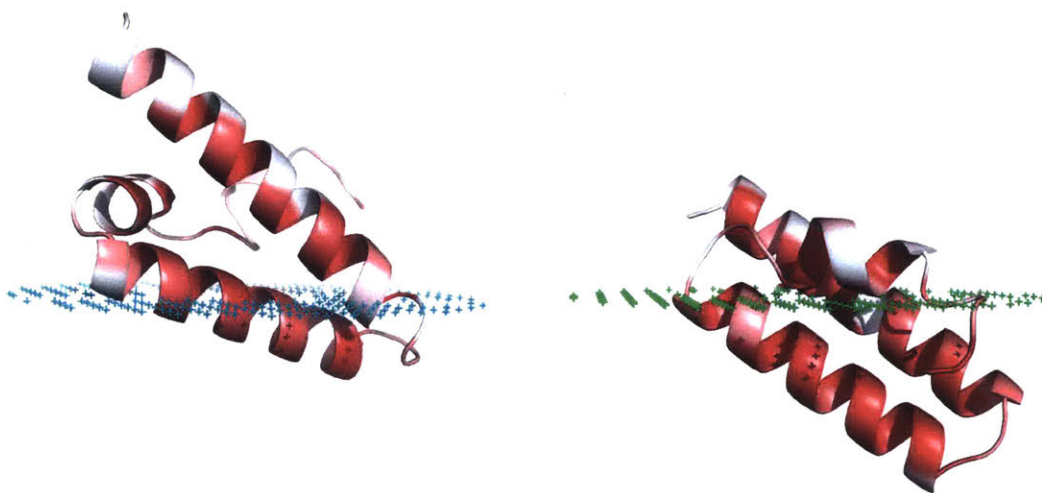
Figure 3-6 LpxP sits deeper in the membrane than LpxM.
The positions of peptides representing the N-terminal domains of LpxM (*left*) and LxpP (*right*) in the membrane were calculated using the PPM server.[33] Peptides are colored by hydrophobicity from hydrophobic (*red*) to hydrophilic (*white*). The membrane is represented as dots.

DISCUSSION

Identification of a novel family of monotopic membrane proteins

As PglC was previously predicted to adopt a bitopic topology, we investigated whether

knowledge of topology determination in PglC might facilitate the identification of additional

examples of membrane proteins that are predicted to be bitopic, but in fact adopt monotopic

topologies. The studies herein demonstrate that the principles of RMH formation identified in

PglC in Chapter 2 are broadly generalizable. We use these principles to identify a reentrant

topology in LpxM, a fatty acid transferase involved in the biosynthesis of lipid A. LpxM

represents a large family of enzymes, unrelated to PglC and the monotopic PGTs, previously

thought to be bitopic.

We predicted that this family of enzymes might be monotopic based on close

examination of the predicted transmembrane domain sequences, knowledge of the biological

function of the acyltransferases, and complementary covariance analysis. *In vivo* SCAM analysis

confirmed a reentrant topology for LpxM. Identification of a reentrant conformation for the

predicted transmembrane domain of LpxM is compatible with the previously reported structure

of LpxM from *A. baumannii*.[30] Finally, peptide folding and membrane insertion prediction tools

support the hypothesis that additional family members LpxL and LpxP adopt a similarly

reentrant fold that embeds deeper in the membrane than LpxM.

Proposed guidelines for identifying reentrant domains in monotopic membrane proteins

The presence of positively-charged residues preceding, and helix-disrupting residues throughout,

the hydrophobic domains of predicted bitopic proteins appears to bias towards a reentrant

topology. Such features can thus be taken as indicators of a reentrant topology, particularly if

such a topology is compatible with biological function, as in the case of both PglC and LpxM. In our parsing of the Membranome database, we looked for predicted transmembrane domains that contained polar residues and helix-breaking prolines and glycines reminiscent of the Ser-Pro motif in PglC, as well as large aromatics, reminiscent of the tryptophan found natively in some P24W PglC variants (the significance of the P24W substitution is discussed in detail in Chapter 2).

An abundance of available computational and bioinformatics tools also facilitates the identification of monotopic folds. Covariance analysis can be an invaluable tool for topology prediction, as it has the capacity to identify evolutionarily conserved interactions between residue pairs that strongly support a reentrant topology; indeed, covariance analysis identified interactions between the RMH and the globular domain of PglC before structural information from x-ray crystallography was available.[27] Covariance analysis for this study was performed using the GREMLIN server.[24] We note that, due to the large number of homologous sequences required for accurate covariance analysis, this tool is more amenable to bacterial proteins with homologs in many species for which genomic data is readily available. However, we anticipate that with the identification of additional examples of reentrant domains in membrane proteins from all domains of life, comprehensive prediction of reentrant topologies in eukaryotes as well as prokaryotes will become more accessible.

In addition, peptide folding prediction tools such as QUARK[31-32] or Peplook[34] (Chapter 2) can provide useful preliminary predictions of the conformations of hydrophobic domains. Finally, the development of tools such as the PPM server,[33] which positions both computationally and experimentally determined structures in a model lipid bilayer, are invaluable for defining the depth at which various monotopic folds are likely to embed in the

membrane. Importantly, all the computational and bioinformatics tools used in this study to perform covariance analysis, predict peptide folding and position folds in a model membrane are available online for general use. These resources are complementary to each other and to biochemical methodologies, such as the *in vivo* SCAM analysis, and can play an important role in the accurate prediction of membrane topology in proteins of interest.

CONCLUSIONS

The examples of LpxM, LpxL and LpxP underscore the importance of developing a deeper understanding of membrane topology determination, particularly with respect to single hydrophobic domains in monotopic and bitopic proteins. Structural information can be limited for membrane proteins, particularly monotopic membrane proteins, which are often reticent to classical experimental methods of structure determination. Complementary biochemical and computational approaches can be used both to determine the topology of membrane proteins when other structural information is not available and to validate existing structures.

Whereas it is currently standard to annotate stretches of hydrophobic residues as membrane-spanning helices, this study demonstrates that such helices are capable of adopting a greater diversity of topologies. In this chapter we outline guidelines for the identification of hydrophobic domains that adopt reentrant, rather than transmembrane, domains. We anticipate that these guidelines can serve as a starting point for the development of novel algorithms capable of predicting reentrant topologies among putative transmembrane domains in a variety of proteins of interest.

EXPERIMENTAL PROCEDURES

Identification of candidate reentrant helical domains from the Membranome database

The Membranome database of single-helix transmembrane proteins[23] was parsed manually for

candidates with a reentrant topology. Candidates were selected from the available list of 196

putative bitopic proteins from *E. coli*; only proteins longer than 100 residues were considered. A

preliminary list of candidates was composed of sequences in which the predicted transmembrane

domain was at the N-terminus (within the first 50 residues), was preceded by positively charged

residues, and contained polar or charged, large aromatic (Trp/Tyr) and/or helix-breaking

(Pro/Gly) residues. These sequences were then submitted to the GREMLIN webserver

(http://gremlin.bakerlab.org/)[24] for multiple sequence alignment and covariance analysis.

Notably, whereas a multiple sequence alignment was always returned, a covariance analysis was

only performed by the server if enough homologs were identified to allow for accurate analysis.

A final list of candidates, shown in **Table 3-3**, contained only those for which a) the N-terminal

positively charged residues and helix-disrupting residues identified previously were found to be

well-represented among homologs in the multiple sequence alignment, and b) for which

covariance analysis was able to identify interactions with >80% probability between residues at

the N-terminus or within the hydrophobic domain and residues in the globular domain

(Appendix 1).

LpxM variants and SCAM analysis

LpxM from *E. coli* K12 was codon-optimized, synthesized and cloned into the pET24a vector by

GenScript (Piscataway, NJ), using the NdeI and XhoI restriction sites to insert a C-terminal His$_6$-

tag. Serine substitutions at C73 and C240, to give a "Cysless" variant, were introduced using

QuikChange II Site-Directed Mutagenesis (Agilent Technologies, Santa Clara, CA) according to

the manufacturer's instructions. Unique cysteines at S8, I11, S17 and M89, for SCAM analysis,

were all similarly introduced by Quikchange. All primers used for mutagenesis are listed in

Table 3-4. SCAM analysis of LpxM variants was carried out with expression in *E. coli* followed

by labeling and Western blot analysis as described in Chapter 2 for the analysis for PglC-His$_6$.

Table 3-4 – Primers used for mutagenesis of LpxM	
C73S_F	5'-CAACCTGAGCCTGAGCTTCCCGGAACG-3'
C73S_R	5'-CGTTCCGGGAAGCTCAGGCTCAGGTTG-3'
C240S_F	5'-CTGATGAAGGTGAGCCGTGCGCGTGTG-3'
C240S_R	5'-CACACGCGCACGGCTCACCTTCATCAG-3'
S8C_F	5'-CCAAGAAAAACAACTGCGAATACATCCCGGAGTTC-3'
S8C_R	5'-GAACTCCGGGATGTATTCGCAGTTGTTTTTCTTGG-3'
I11C_F	5'-GAAACCAAGAAAAACAACAGCGAATACTGCCCGGAGTTC-3'
I11C_R	5'-GAACTCCGGGCAGTATTCGCTGTTGTTTTTCTTGGTTTC-3'
S17C_F	5'-GGAGTTCGACAAATGCTTTCGTCACCCGCG-3'
S17C_R	5'-CGCGGGTGACGAAAGCATTTGTCGAACTCC-3'
M89C_F	5'-GAACGTGAGGCGATTGTGGACGAGTGTTTTGCGACCG-3'
M89C_R	5'-CGGTCGCAAAACACTCGTCCACAATCGCCTCACGTTC-3'

Predicting peptide folding and membrane insertion using the QUARK and PPM tools

Sequences corresponding to the N-terminal domains of LpxM and LpxP were submitted to the

QUARK peptide fold prediction algorithm[31-32]

(https://zhanglab.ccmb.med.umich.edu/QUARK/). The sequences submitted were

METKKNNSEYIPEFDKSFRHPRYWGAWLGVAAMAGIALT-

PPKFRDPILARLGRFAGRLGKSSRR and MFPQCKFSREFLHPRYWLTWFGLGVLWLW-

VQLPYPVLCFLGTRIGAMARPFLKRR, respectively. The first model returned was taken in

each case. The coordinates for folded peptides were subsequently submitted to the PPM server[33]

(https://opm.phar.umich.edu/ppm_server).

REFERENCES

(1) Kyte, J.; Doolittle, R. F., A simple method for displaying the hydropathic character of a protein. *J Mol Biol* **1982,** *157* (1), 105-32.

(2) Eisenberg, D.; Weiss, R. M.; Terwilliger, T. C., The helical hydrophobic moment: a measure of the amphiphilicity of a helix. *Nature* **1982,** *299* (5881), 371-4.

(3) Claros, M. G.; von Heijne, G., TopPred II: an improved software for membrane protein structure predictions. *Comput Appl Biosci* **1994,** *10* (6), 685-6.

(4) Jones, D. T.; Taylor, W. R.; Thornton, J. M., A model recognition approach to the prediction of all-helical membrane protein structure and topology. *Biochemistry* **1994,** *33* (10), 3038-49.

(5) von Heijne, G., The distribution of positively charged residues in bacterial inner membrane proteins correlates with the trans-membrane topology. *EMBO J* **1986,** *5* (11), 3021-7.

(6) von Heijne, G., Membrane-protein topology. *Nat Rev Mol Cell Biol* **2006,** *7* (12), 909-18.

(7) von Heijne, G., Membrane protein structure prediction. Hydrophobicity analysis and the positive-inside rule. *J Mol Biol* **1992,** *225* (2), 487-94.

(8) Tusnady, G. E.; Simon, I., Principles governing amino acid composition of integral membrane proteins: application to topology prediction. *J Mol Biol* **1998,** *283* (2), 489-506.

(9) Sonnhammer, E. L.; von Heijne, G.; Krogh, A., A hidden Markov model for predicting transmembrane helices in protein sequences. *Proc Int Conf Intell Syst Mol Biol* **1998,** *6*, 175-82.

(10) Krogh, A.; Larsson, B.; von Heijne, G.; Sonnhammer, E. L., Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J Mol Biol* **2001,** *305* (3), 567-80.

(11) Moller, S.; Croning, M. D.; Apweiler, R., Evaluation of methods for the prediction of membrane spanning regions. *Bioinformatics* **2001,** *17* (7), 646-53.

(12) Viklund, H.; Elofsson, A., OCTOPUS: improving topology prediction by two-track ANN-based preference scores and an extended topological grammar. *Bioinformatics* **2008,** *24* (15), 1662-8.

(13) Tsirigos, K. D.; Peters, C.; Shu, N.; Kall, L.; Elofsson, A., The TOPCONS web server for consensus prediction of membrane protein topology and signal peptides. *Nucleic Acids Res* **2015,** *43* (W1), W401-7.

(14) Viklund, H.; Bernsel, A.; Skwark, M.; Elofsson, A., SPOCTOPUS: a combined predictor of signal peptides and membrane protein topology. *Bioinformatics* **2008,** *24* (24), 2928-9.

(15) Elazar, A.; Weinstein, J. J.; Prilusky, J.; Fleishman, S. J., Interplay between hydrophobicity and the positive-inside rule in determining membrane-protein topology. *Proc Natl Acad Sci U S A* **2016,** *113* (37), 10340-5.

(16) Aoki, S.; Thomas, A.; Decaffmeyer, M.; Brasseur, R.; Epand, R. M., The role of proline in the membrane re-entrant helix of caveolin-1. *J Biol Chem* **2010**, *285* (43), 33371-80.

(17) Okamoto, T.; Schlegel, A.; Scherer, P. E.; Lisanti, M. P., Caveolins, a family of scaffolding proteins for organizing "preassembled signaling complexes" at the plasma membrane. *J Biol Chem* **1998**, *273* (10), 5419-22.

(18) Lee, J.; Glover, K. J., The transmembrane domain of caveolin-1 exhibits a helix-break-helix structure. *Biochim Biophys Acta* **2012**, *1818* (5), 1158-64.

(19) Ray, L. C.; Das, D.; Entova, S.; Lukose, V.; Lynch, A. J.; Imperiali, B.; Allen, K. N., Membrane association of monotopic phosphoglycosyl transferase underpins function. *Nat Chem Biol* **2018**, *14*, 538-541.

(20) Yan, C.; Luo, J., An analysis of reentrant loops. *Protein J* **2010**, *29* (5), 350-4.

(21) Viklund, H.; Granseth, E.; Elofsson, A., Structural classification and prediction of reentrant regions in alpha-helical transmembrane proteins: application to complete genomes. *J Mol Biol* **2006**, *361* (3), 591-603.

(22) Almen, M. S.; Nordstrom, K. J.; Fredriksson, R.; Schioth, H. B., Mapping the human membrane proteome: a majority of the human membrane proteins can be classified according to function and evolutionary origin. *BMC Biol* **2009**, *7*, 50.

(23) Lomize, A. L.; Lomize, M. A.; Krolicki, S. R.; Pogozheva, I. D., Membranome: a database for proteome-wide analysis of single-pass membrane proteins. *Nucleic Acids Res* **2017**, *45* (D1), D250-D255.

(24) Balakrishnan, S.; Kamisetty, H.; Carbonell, J. G.; Lee, S. I.; Langmead, C. J., Learning generative models for protein fold families. *Proteins* **2011**, *79* (4), 1061-78.

(25) Ovchinnikov, S.; Kamisetty, H.; Baker, D., Robust and accurate prediction of residue-residue interactions across protein interfaces using evolutionary information. *Elife* **2014**, *3*, e02030.

(26) Ovchinnikov, S.; Park, H.; Varghese, N.; Huang, P. S.; Pavlopoulos, G. A.; Kim, D. E.; Kamisetty, H.; Kyrpides, N. C.; Baker, D., Protein structure determination using metagenome sequence data. *Science* **2017**, *355* (6322), 294-298.

(27) Lukose, V.; Luo, L. Q.; Kozakov, D.; Vajda, S.; Allen, K. N.; Imperiali, B., Conservation and covariance in small bacterial phosphoglycosyltransferases identify the functional catalytic core. *Biochemistry* **2015**, *54* (50), 7326-7334.

(28) Shi, Y.; Cheng, D., Beyond triglyceride synthesis: the dynamic functional roles of MGAT and DGAT enzymes in energy metabolism. *Am J Physiol Endocrinol Metab* **2009**, *297* (1), E10-8.

(29) Takeuchi, K.; Reue, K., Biochemistry, physiology, and genetics of GPAT, AGPAT, and lipin enzymes in triglyceride synthesis. *Am J Physiol Endocrinol Metab* **2009**, *296* (6), E1195-209.

(30) Dovala, D.; Rath, C. M.; Hu, Q.; Sawyer, W. S.; Shia, S.; Elling, R. A.; Knapp, M. S.; Metzger, L. E. t., Structure-guided enzymology of the lipid A acyltransferase LpxM reveals a dual activity mechanism. *Proc Natl Acad Sci U S A* **2016**, *113* (41), E6064-E6071.

(31) Xu, D.; Zhang, Y., Ab initio protein structure assembly using continuous structure fragments and optimized knowledge-based force field. *Proteins* **2012**, *80* (7), 1715-35.

(32) Xu, D.; Zhang, Y., Toward optimal fragment generations for ab initio protein structure assembly. *Proteins* **2013**, *81* (2), 229-39.

(33) Lomize, M. A.; Pogozheva, I. D.; Joo, H.; Mosberg, H. I.; Lomize, A. L., OPM database and PPM web server: resources for positioning of proteins in membranes. *Nucleic Acids Res* **2012**, *40*, D370-6.

(34) Thomas, A.; Deshayes, S.; Decaffmeyer, M.; Van Eyck, M. H.; Charloteaux, B.; Brasseur, R., Prediction of peptide structure: how far are we? *Proteins* **2006**, *65* (4), 889-97.

Chapter 4: Polyprenol phosphate geometry dictates key molecular interactions with conserved Pro24 in PglC

The work presented in this chapter was performed in collaboration with Dr. Ziqiang Guan at the Duke University Medical Center in Durham, North Carolina. Dr. Ziqiang Guan performed all of the NPHPLC/MS quantification of PrenP in lipidic extracts from *C. jejuni* and SMA-lipoparticles.

Much of the work described in this chapter is currently under review as part of a manuscript entitled:

Entova, S.; Guan, Z.; Imperiali, B., Polyprenol phosphate substrate geometry and a conserved proline in PglC dictate key molecular interactions in initiation of N-linked glycan assembly in *Campylobacter jejuni*.

INTRODUCTION

Glycoconjugates constitute a diverse and complex class of macromolecules that play a wide

variety of essential roles in both prokaryotic and eukaryotic biology. The assembly of many

complex glycans is associated with cellular membranes (e.g. the endoplasmic reticulum

membrane in eukaryotes and the cell membrane in prokaryotes). In such biosynthetic pathways,

a series of integral and peripheral membrane-bound glycosyltransferase enzymes catalyze the

step-wise addition of monosaccharides from nucleotide sugar donors to an amphiphilic

polyprenol phosphate (PrenP) acceptor in the membrane (**Figure 4-1**). PrenPs are linear, long-

chain products from the same terpene biosynthesis pathway responsible for the biosynthesis of

farnesyl- and geranylgeranyl diphosphate for post-translational protein prenylation as well as

cholesterol and other steroids. However, unlike cholesterol, which can make up as much as 30%

of the plasma membrane in eukaryotes,[1] polyprenol phosphates have been estimated to comprise

only ≤1% of bacterial and eukaryotic membranes.[2-4]

The precise structural features of the polyprenol phosphates used in glycan assembly vary

by species (**Figure 4-1**). Archaea and eukaryotes utilize α-saturated polyisoprenol phosphates,

called dolichols, which range in length from as few as eight isoprene units in some archaea to as

many as 20 in mammals.[5-8] Archaeal polyisoprenol phosphates can be additionally saturated at

other isoprene units.[9-10] Bacteria, on the other hand, generally feature α-unsaturated polyprenol

phosphates. The PrenP found in most bacteria investigated to date, including model organisms

Escherichia coli and *Staphylococcus aureus*, is the 11 isoprene unit-long undecaprenol

phosphate (UndP)[2, 11] however, a preference for shorter PrenPs has also been reported for a

handful of species.[12-14]

135
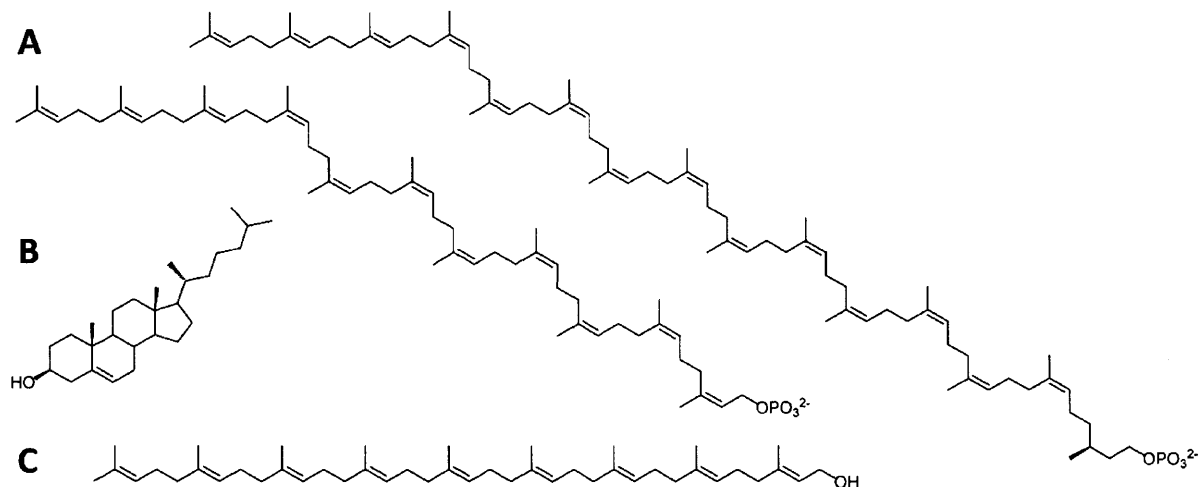
Figure 4-1 Structures of relevant terpenes.

(**A**) *Left:* Structure of bacterial undecaprenol phosphate (UndP). (Note: The stereochemistry of the plant-derived UndP used for *in vitro* assays described herein is 3-*trans* to 7-*cis* rather than 2-*trans* to 8-*cis* as for bacterial UndP, shown).[15] *Right:* Structure of C_{75} dolichol phosphate (DolP).

(**B**) Structure of cholesterol.

(**C**) Structure of the all-*trans* nonaprenol, solanesol.

Despite the apparent diversity in polyprenol phosphate structures across domains, both

the unique isoprenoid character and a preference for *cis* (*Z*) double-bond geometry in the

isoprene units proximal to the phosphate are strictly conserved in all glycan assembly pathways.

Although there exist linear polyprenols composed exclusively of *trans* (*E*) isoprene units, such as

the nine-isoprene solanesol abundant in tobacco and other solanaceous crops (**Figure 4-1**),[16]

these molecules do not serve as carriers in glycan assembly pathways. The structural

conservation of a series of *cis* configuration isoprenes suggests that the unique geometry of

PrenP chains may indicate a more purposed role in glycoconjugate biosynthesis, beyond that of a

hydrophobic membrane anchor on which glycan assembly can occur. A variety of roles in

136

mediating substrate-enzyme binding and multi-enzyme complex formation, as well as in

modulating the biophysical properties of the membrane, have been proposed.[17-22] As yet, it is

also poorly understood how a single pool of PrenP in bacterial membranes simultaneously

supports multiple glycoconjugate biosynthesis pathways including those leading to

lipopolysaccharide (LPS), peptidoglycan and teichoic acid. Herein, we describe studies designed

to provide additional insight into the role played by the structural features of PrenP in glycan

assembly in the protein glycosylation (pgl) pathway from the Gram-negative pathogenic

bacterium, *Campylobacter jejuni*.

Glycan biosynthesis in the *C. jejuni* pgl pathway is initiated by the integral membrane

protein PglC, a phosphoglycosyl transferase (PGT) that catalyzes the transfer of a C1-

phosphosugar from UDP-*N,N'*-diacetylbacillosamine (UDP-diNAcBac) to PrenP, which is

associated with the membrane.[23] The PrenPP-diNAcBac product is then further elaborated by

various downstream membrane-associated glycosyltransferases to yield a heptasaccharide that is

flipped across the membrane and ultimately transferred to a variety of protein acceptors by an

oligosaccharyl transferase.[24-26] The by-product of protein glycosylation is a PrenPP that is then

dephosphorylated to recycle PrenP.[27-28] The ability to synthesize and transfer complete N-linked

glycans has been linked to *C. jejuni* virulence;[26, 29] thus, a deeper understanding of the pgl

pathway is of direct significance to human health. Furthermore, the pgl pathway shares a

common logic with other diverse glycoconjugate assembly pathways in bacteria, including those

for LPS and peptidoglycan biosynthesis,[30-31] and also represents a more tractable functional

homolog for N-linked glycosylation in eukaryotes.[15, 32-33] This makes the pgl pathway an ideal

system for probing the role of polyprenol phosphates in glycan assembly across domains.

As the initiator of the pgl pathway and one of few pgl enzymes with a substantial hydrophobic domain, PglC in particular is likely to be involved in key interactions with PrenP in the membrane. Recently, we reported the structure of the PglC homolog from *Campylobacter concisus*, which shares 72% sequence identity with PglC from *C. jejuni*. PglC is a monotopic membrane protein that adopts a unique architecture in the membrane. An N-terminal hydrophobic domain makes a reentrant helix-break-helix that penetrates the membrane to a depth of 14 Å and associates with several amphipathic helices to anchor the protein into the cytoplasmic face of the inner membrane (**Figure 4-2**).[34] The enzyme active site, formed by one of the amphipathic helices and part of the soluble domain, is positioned at the membrane interface, allowing for efficient transfer of the phosphosugar from a soluble nucleotide sugar donor to the membrane-anchored PrenP. A putative PrenP-binding site, located adjacent to the active site, was proposed based on electron density attributed to a co-crystallizing PEG molecule (**Figure 4-2**).

14 Å
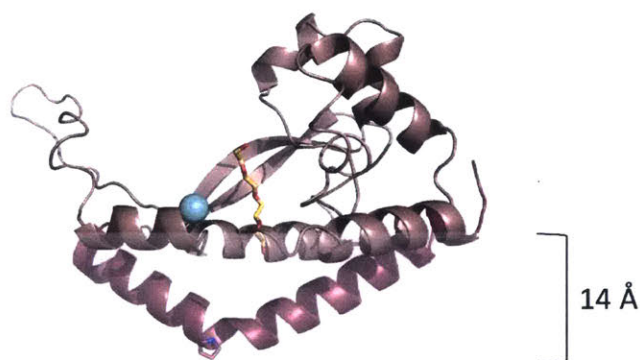
Figure 4-2 Structure of PglC.
The reported structure of PglC (PDB 5WL7).[34] The N-terminal reentrant helical domain is shown in dark pink. Pro24 is shown as sticks. The Mg^{2+} in the active site is shown in cyan. A co-crystallized PEG molecule marking the putative PrenP-binding site is shown in yellow. Pro24 in the membrane-resident domain is shown in sticks. The membrane is represented in gray.

PGTs can be divided into two superfamilies based on their distinct structures.[35] One superfamily comprises polytopic membrane proteins with 10- and 11-transmembrane helices (**Figure 1-9**), represented by MraY and WecA, respectively.[36] PglC is a representative member of a second superfamily of PGTs (**Figure 1-11**) and shows a monotopic architecture.[37] The PGTs in this superfamily share a PglC-like functional core with a single, reentrant hydrophobic domain. A conserved proline, Pro24, located at the break in the reentrant domain of PGTs in this superfamily has been shown to be important for PglC activity (**Figure 4-3**).[37-38]
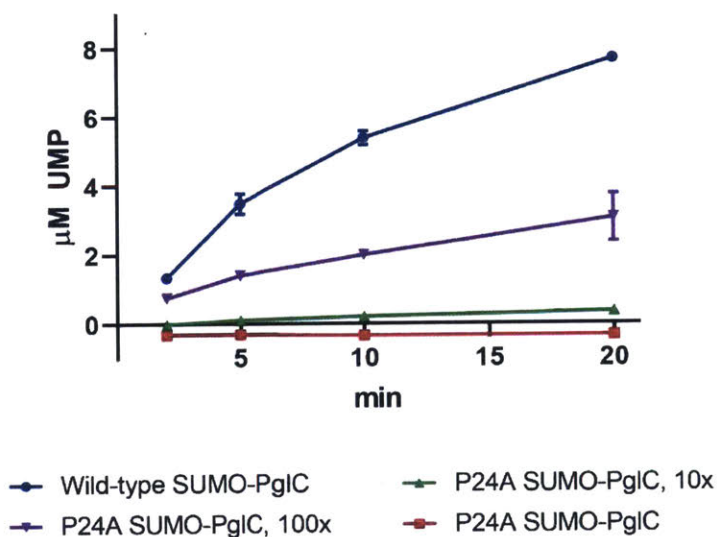
Figure 4-3 Importance of Pro24 for PglC activity.
Activity assays of wild-type and P24A SUMO-PglC variants using UMP-Glo to monitor UMP release. P24A SUMO-PglC does not achieve native levels of activity even when assayed at 100x concentration relative to wild-type SUMO-PglC. Error bars are given for mean ± SD, n = 2.

Based on the position of Pro24 within the only membrane-resident domain of PglC, and on the known conservation of prolines in polyisoprenol recognition sequences (PIRSs) in diverse proteins,[18] it has been hypothesized that Pro24 may play a key role in recognition and binding of PrenP substrates in the membrane, both in the context of PglC and in other members of the monotopic PGT superfamily.[35, 37] Herein, we apply a range of biochemical, biophysical and mass-spectrometry based methodologies to identify and quantify PrenP in *C. jejuni* membranes, and to investigate the substrate recognition and binding of PrenP to PglC, both in detergent-solubilized conditions and in a native-like model membrane environment. The significance of this work for the broader understanding of polyprenol phosphates in glycan assembly across domains is also discussed.

RESULTS

Establishing a growth curve of *C. jejuni*

A previous study of PrenP abundance in *E. coli* and *S. aureus* indicated higher levels of PrenP in bacteria during log phase growth than during stationary phase;[2] this is likely due to the fact that PrenP is required for the synthesis of many cell surface glycoconjugates that are essential for growth. Thus, it was desirable to establish a growth curve for *C. jejuni* to be able to quantify PrenP abundance in log-phase cultures. Determination of a growth curve for *C. jejuni* was made technically difficult by the fact that *C. jejuni* cultures prefer to grow while shaking under microaerophilic conditions (10% CO_2, 5% O_2, 85% N_2); cultures are generally grown in sealed containers under a microaerophilic gas mix. Thus, it was necessary to develop a method that

would allow culture density in a sealed flask to be measured at distinct time points without introducing atmospheric oxygen.

The method used to establish a growth curve is depicted in **Figure 4-4**. Cultures were grown in a 1L Erlenmeyer flask fitted with a rubber septum. The septum was pierced with a long needle attached to a Leur-lock stopcock, such that the culture could be closed off from the surrounding atmosphere as desired. Following inoculation, the flask headspace was filled with a microaerophilic gas mix using the attached needle, and the stopcock was closed. At time points during culture growth, the stopcock could be opened briefly to allow culture aliquots to be removed through the attached needle, with minimal disruption to the microaerophilic environment inside. A growth curve for *C. jejuni* was thus established and log-phase was determined to occur at OD600 ~0.15-35.
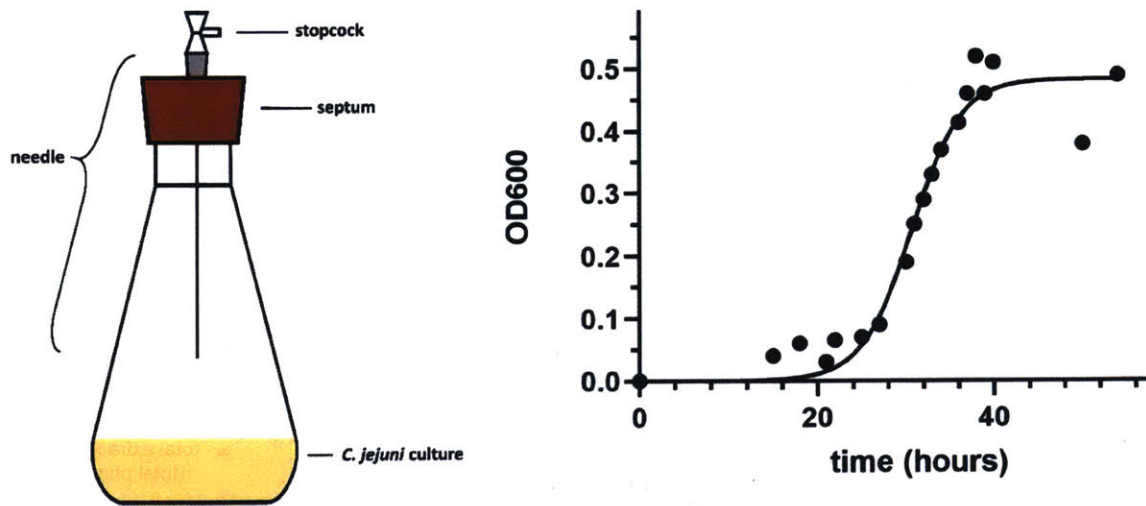
Figure 4-4 Determination of a growth curve for *C. jejuni*.
A modified culture flask set-up (*left*) allowed sampling of the culture for density measurements with minimal disruption to the microaerophilic environment inside. A growth curve (*right*) was established over five independent cultures of *C. jejuni*.

141

Quantification of PrenP abundance in *C. jejuni*

Quantification of PrenP abundance in *C. jejuni* was performed by NPLC/MS analysis of total lipids extracted from log-phase *C. jejuni* cultures. Lipid extraction was performed using a modified Bligh-Dyer method.[39] High-resolution MS was performed in the negative ion mode, and showed that the predominant PrenPs in *C. jejuni* are decaprenol phosphate (DecP) and UndP, whose [M-H]⁻ ions were detected at *m/z* 777.6 and 845.6, respectively. The total PrenP in extracts was quantified by spiking in an internal UndP standard. The total pool of PrenP in Gram-negative bacteria is distributed between free PrenP and PrenPP-glycoconjugate biosynthetic intermediates, including PrenPP-linked glycans and PrenPP. A preliminary analysis of extracted lipids identified PrenP but no pathway intermediates, suggesting that these moieties might only be present in very low abundance, or they are not extractable by the Bligh-Dyer method. For quantification of the entire PrenP pool, lipid extracts in subsequent analyses were first subjected to a mild alkaline hydrolysis, a step shown previously to release PrenP from PrenPP and PrenPP-linked glycans.[40] The abundance of PrenP in lipid extracts is reported as the percentage of PrenP relative to the total lipids which were quantified by microdetermination of total phosphate (**Figure 4-5**).
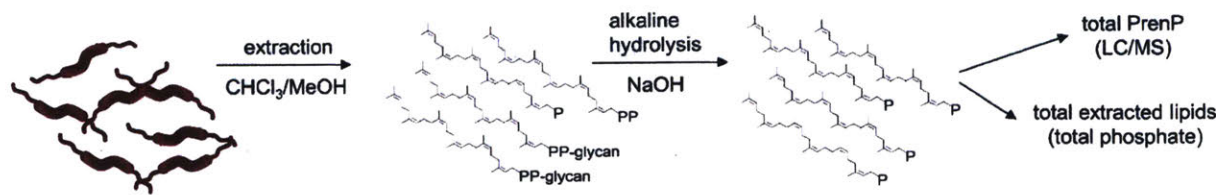
Figure 4-5 Work flow for quantification of PrenP *C. jejuni*.
Schematic showing work flow for extraction, alkaline hydrolysis and quantification of PrenP in *C. jejuni* membranes.

As UndP is the most common PrenP identified in bacteria studied to date, previous reports have ascribed UndP the role of glycan carrier in *C. jejuni* as well.[24, 26] Indeed, some UndP was identified in the lipidic fraction. However, the majority of PrenP identified in the lipid extracts was actually the shorter 10 isoprene-long PrenP homolog, DecP (**Figure 4-6**). The total abundance of PrenP (DecP + UndP) in *C. jejuni* lipid extracts was found to be ~0.1%, with DecP accounting for over 80% of the total PrenP.
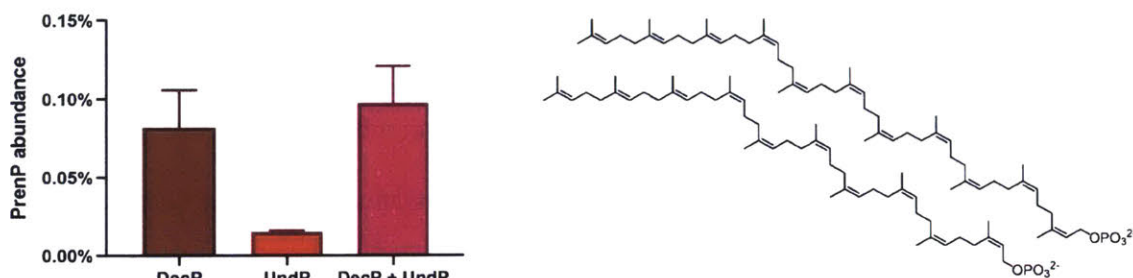
Figure 4-6 PrenP abundance in *C. jejuni.*
Left: Total abundance of DecP and UndP as a percentage of the total lipidic extract from *C. jejuni*. Error bars are given for mean ± SD, n = 3. *Right:* Structures of UndP (top) and DecP (bottom).

Enrichment of PrenP in the local membrane around PglC

Quantification of PrenP levels in native extracted lipids provides valuable insight into the

identity and average abundance of these species in the bulk *C. jejuni* membrane, but it does not

describe the abundance of PrenP in the membrane local to PglC and other PrenP-binding

enzymes. For example, it is possible that specific interactions with such enzymes may result in

PrenP enrichment in the local membrane environment. In particular, Pro24 in PglC[37] and

similarly conserved prolines in polyisoprenol-recognition sequences in various enzymes were

previously proposed to contribute to polyprenol phosphate binding.[18]

Localization of PrenP to PglC in the membrane was investigated using the recently-

introduced SMALP technology,[41-43] which uses a styrene-maleic acid (SMA) polymer to

solubilize membrane proteins and lipids directly from cell membranes to make SMA-

lipoparticles (SMALPs). The use of SMA solubilization obviates the need for detergent or other

non-native surfactants, and importantly, solubilizes proteins together with lipids from their native

membrane environment. For this reason, extraction and quantification of lipids from SMALPs

has been used to provide insight into the lipid composition of the local membrane environment

surrounding membrane proteins.[44-45] To this end, PglC from *C. jejuni* with a C-terminal His$_6$ tag

was heterologously overexpressed in *E. coli* and solubilized into SMALPs; particles containing

PglC-His$_6$ were then isolated by Ni-NTA affinity chromatography (**Figure 4-7** and **Figure 4-8**).

Lipids extracted both from the Ni-NTA-purified PglC-SMALPs and from SMA-solubilized

membranes not enriched on Ni-NTA for PglC-His$_6$ ("background") were assessed qualitatively

by TLC, and determined to be congruent with commercially-available *E. coli* lipid extract

(**Figure 4-8**). PrenP and PrenPP-linked intermediates in the extracted lipids were subjected to

mild alkaline hydrolysis and quantified by LC/MS against a pure UndP standard as described above; notably, in this case only UndP was quantified, as this is the only PrenP found in *E. coli*.
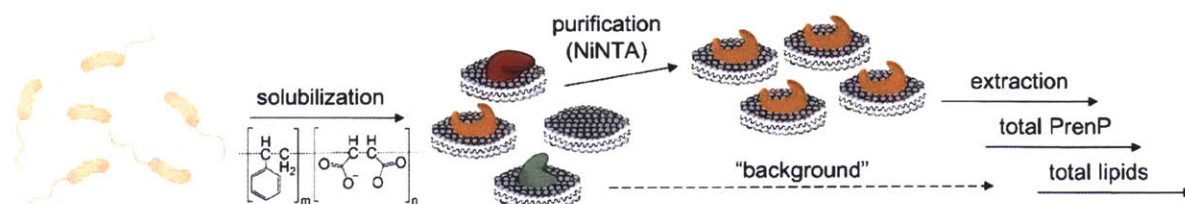
Figure 4-7 Work flow for quantification of PrenP enrichment around SMA-solubilized PglC. Schematic showing work-flow for solubilization and purification of SMA-solubilized PglC-His$_6$ and extraction and quantification of the surrounding lipids.
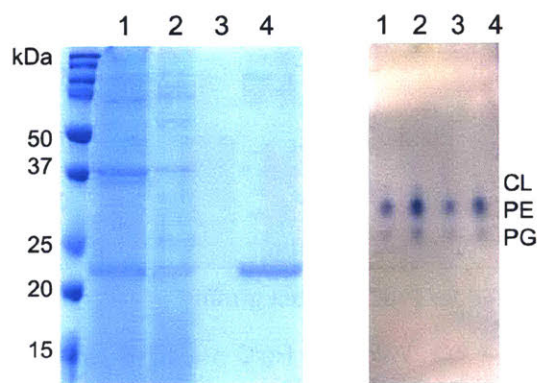
Figure 4-8 Ni-NTA purification and lipid extraction of SMA-lipoparticles.
Left: Coomassie-stained SDS-PAGE gel showing purification of SMA-lipoparticles with PglC-His$_6$, purified by Ni-NTA affinity chromatography. Lanes: 1 = load, 2 = flow-through, 3 = wash, 4 = elution. *Right:* Thin-layer chromatography analysis of extracted lipids. Lanes: 1, 2 = *E. coli* lipid commercial standard. Lanes 3, 4 = extracted lipids from *C. jejuni*. CL = cardiolipin, PE = phosphotidylethanolamine, PG = phosphotidylglycerol.

The UndP abundance in background lipid extracts was ~0.075%, similar to the bulk abundance of PrenP determined in native lipid extracts from *C. jejuni* membrane (**Figure 4-9**). However, UndP was found to be 1.91±0.35-fold more abundant in PglC-SMALPs, relative to background, indicating an enrichment of UndP around PglC in the local membrane environment. To probe the role of conserved PglC Pro24 in the observed localization of UndP, the same experimental protocol was applied to a P24A PglC-His$_6$ variant. In this case, UndP was not found to be significantly more abundant in PglC-SMALPs relative to background. Thus, the enrichment of UndP around PglC is largely abrogated by the mutation of Pro24 to Ala, supporting a role for the conserved proline in PrenP binding in the native membrane environment.
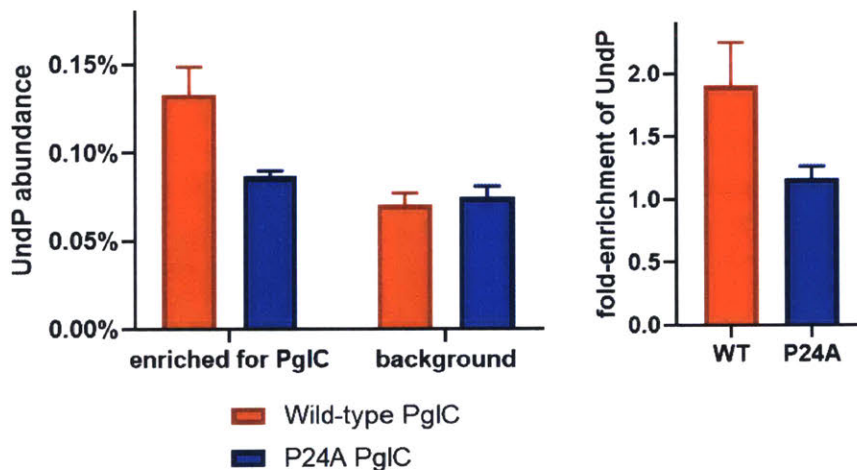
Figure 4-9 PrenP enrichment around PglC in SMA-lipoparticles.
Left: Quantification of UndP in lipid extracts from PglC- and background SMA-lipoparticles. * *p* = 0.02; n.s. = not significant. *Right:* Fold-enrichment of UndP in PglC SMA-lipoparticles over background. Error bars are given for mean ± SD, n = 3.

146

The role of Pro24 in recognition and binding of PrenP substrate

The putative role played by Pro24 in binding PrenP was investigated further *in vitro* using wild-type and P24A SUMO-PglC variants. Incorporation of an N-terminal SUMO-tag was previously shown to significantly improve purification of PglC,[34, 37] and SUMO-PglC has been shown to adopt a native membrane topology [38] and be catalytically active.[37, 46] The interactions between SUMO-PglC variants and the UndP substrate were assessed both by circular dichroism (CD) spectroscopy and by a thermal denaturation assay. CD studies revealed that wild-type and P24A SUMO-PglC adopt very similar, fairly α-helical secondary structures (**Figure 4-10**), and showed cooperative unfolding upon heating with similar T_ms (**Figure 4-11**). The two variants demonstrated comparable increases in thermal stability in the presence of 300 μM of UndP substrate (**Figure 4-11**), indicating secondary structure stabilization by UndP.
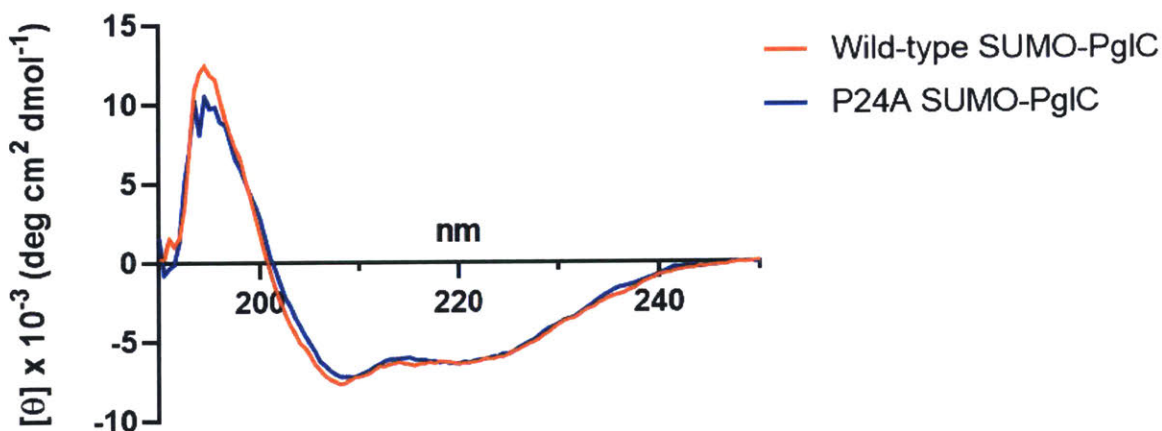
Figure 4-10 CD spectroscopy scans of wild-type and P24A SUMO-PglC.
The secondary structure of wild-type and P24A SUMO-PglC appeared to adopt very similar by circular dichroism. Mean residue ellipticity is reported.
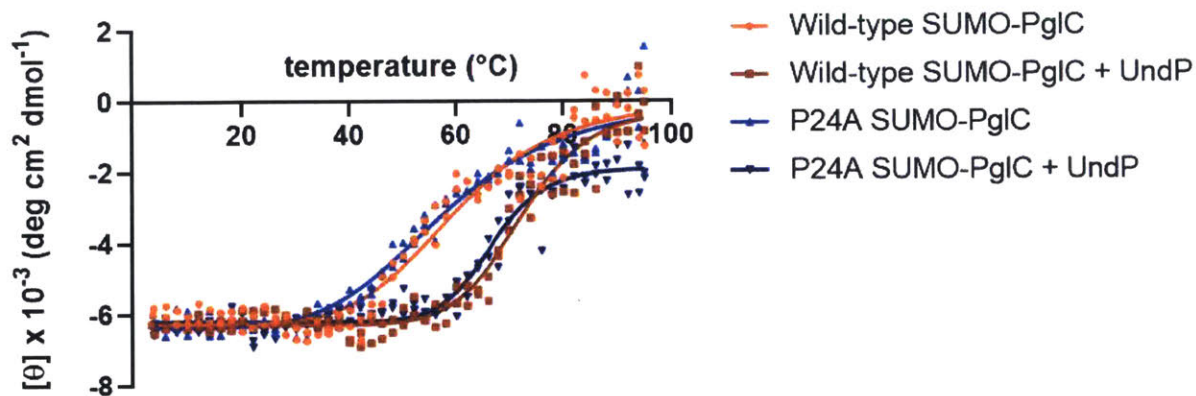
Figure 4-11 CD spectroscopy thermal melts of wild-type and P24A SUMO-PglC. The secondary structure of wild-type and P24A SUMO-PglC show similar, cooperative melting, and both are stabilized by 300 μM UndP. $T_{m,WT} = 59.7\pm1.5$ °C; $T_{m,P24A} = 57.4\pm1.5$ °C; $T_{m,WT+UndP} = 72.2\pm0.9$ °C; $T_{m,P24A+UndP} = 66.7\pm0.7$ °C. Mean residue ellipticity at 222 nm is reported, n = 2.

It has been noted that unfolding of α-helical membrane proteins is often driven through loss of tertiary contacts rather than secondary structure;[47-48] thus, the effect of UndP on PglC tertiary structure stability was also investigated, using a thermal denaturation assay reported previously to assess the thermal stability of PglC[38] and the membrane protein, dolichylphosphate mannose synthase.[49] In this assay, thermal stability was measured as a function of resistance to denaturation and precipitation upon heating: following heating, precipitated protein was removed by centrifugation and the remaining soluble fraction was quantified by gel densitometry. Both wild-type and P24A SUMO-PglC exhibited complete, cooperative in a similar temperature range (**Figure 4-13**). However, the two variants were affected differently by the presence of UndP. Wild-type SUMO-PglC continued to show cooperative denaturation in the presence of 300 μM UndP, but experienced a significant (> 20 °C) increase in thermal stability. By contrast, although P24A SUMO-PglC showed apparent stabilization at higher temperatures in 300 μM UndP, this

variant no longer exhibited cooperative denaturation and remained partially soluble even at

elevated temperatures.
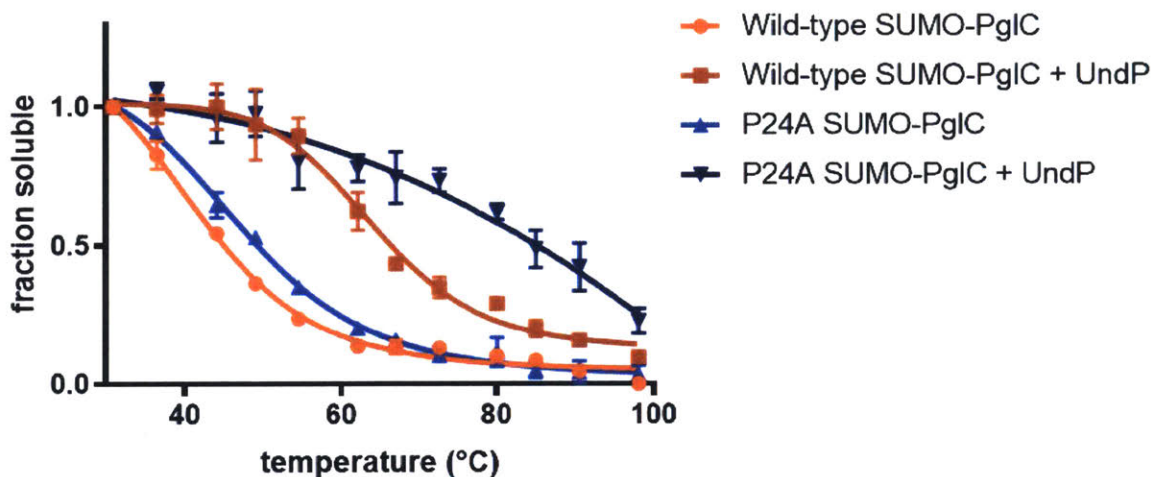
Figure 4-12 Thermal denaturation assays of wild-type and P24A SUMO-PglC with UndP. Thermal denaturation assays of wild-type and P24A SUMO-PglC variants with and without UndP. $T_{m,WT} = 41.9\pm1.3\ °C$; $T_{m,P24A} = 47.4\pm1.1\ °C$; $T_{m,WT+UndP} = 64.2\pm1.4\ °C$; $T_{m,P24A+UndP}$ not determined. Error bars are given for mean $\pm$ SEM, n = 4.

This atypical thermal denaturation behavior suggested that P24A SUMO-PglC was not

being stabilized in a native conformation. Indeed, a very similar behavior was observed for both

wild-type and P24A SUMO-PglC in the presence the non-substrate all-*trans* nonaprenol,

solanesol phosphate (SolP) (**Figure 4-13**). Again, although wild-type and P24A SUMO-PglC

both appeared more stable in the presence of 300 μM SolP, both variants no longer exhibited

cooperative denaturation, and remained partially soluble at high temperatures, suggesting that

interaction with SolP resulted in stabilization of a non-native configuration.

149
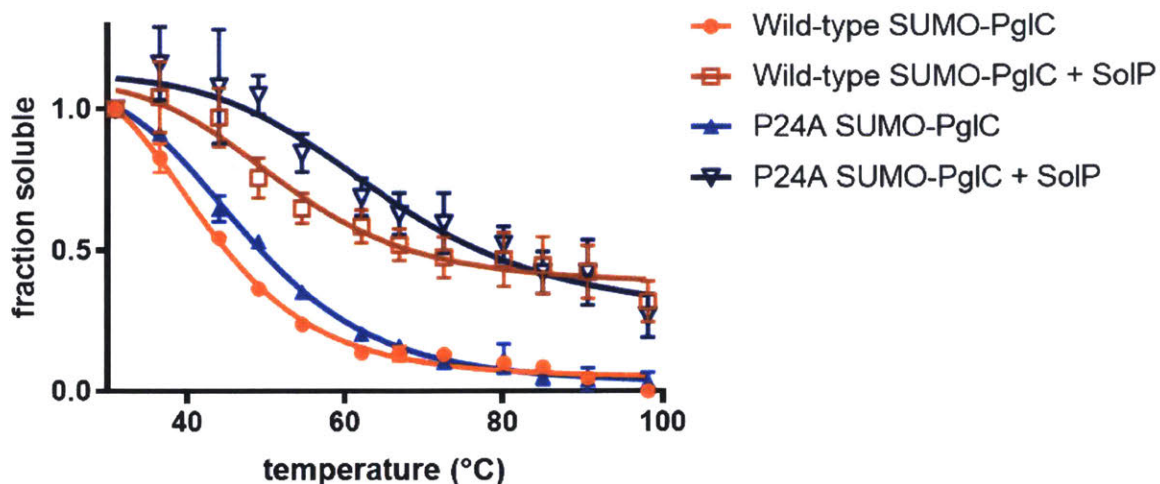
Figure 4-13 Thermal denaturation assays of wild-type and P24A SUMO-PglC with SolP. Thermal denaturation assays of wild-type and P24A SUMO-PglC variants with and without SolP. $T_{m,WT} = 41.9 \pm 1.3$ °C; $T_{m,P24A} = 47.4 \pm 1.1$ °C; $T_{m,WT+UndP}$ and $T_{m,P24A+UndP}$ not determined. Error bars are given for mean ± SEM, n = 4.

Towards probing UndP localization to PglC *in vitro* using Nanodiscs

As a complementary approach, we sought to determine whether UndP localization to PglC could be recreated *in vitro* in Nanodiscs composed from purified SUMO-PglC, UndP and mix of synthetic lipids. Nanodiscs are small self-assembling structures that effectively mimic the hydrophobic environment of a lipid membrane.[50-51] Each particle is composed of a planar phospholipid bilayer, the hydrophobic circumference of which is protected by a membrane scaffolding protein (MSP). Unlike SMA-solubilized lipoparticles which capture the native lipid environment, a variety of different lipids and membrane proteins can be easily incorporated into the Nanodisc bilayer, making this a very flexible system for the study of membrane proteins. The details and advantages of Nanodisc-based methods for characterization of membrane proteins are discussed in greater depth in Chapter 5.

To promote solubilization of no more than one target protein per Nanodisc, assembly reactions typically include an excess of lipid and MSP to target protein;[51] following assembly, Nanodiscs carrying the target protein are separated from empty Nanodiscs by affinity chromatography. Recently, Wagner et al. reported the use of covalently circularized MSP constructs for enhanced membrane protein solubilization and stability in Nanodiscs.[52] SUMO-PglC was previously successfully reconstituted into Nanodiscs using the MSP1E3D1 scaffolding protein, which yields Nanodiscs ~13 nm in diameter.[51, 53] We compared incorporation of SUMO-PglC into Nanodiscs when using the MSP1E3D1 scaffold and the corresponding covalently circularized construct, NW11. Incorporation of SUMO-PglC (which contains an N-terminal His_6 tag) into Nanodiscs was assessed by gel densitometry of fractions from Ni-NTA purification of SUMO-PglC Nanodiscs (**Figure 4-14**). Percent incorporation was defined by the amount of SUMO-PglC observable in the elution fraction (corresponding to properly formed SUMO-PglC Nanodiscs), relative to the total amount of SUMO-PglC used in Nanodisc assembly. Approximately 25-30% of total SUMO-PglC was found to incorporate into Nanodiscs; the remaining 70-75% could be seen in flow-through and wash fractions (**Figure 4-14A**), and may correspond to unincorporated or precipitated SUMO-PglC. As reported for other targets,[52] the circularized NW11 scaffold protein did result in somewhat enhanced solubilization of SUMO-PglC into Nanodiscs (**Figure 4-14B**); thus, all subsequent experiments were performed using this MSP.
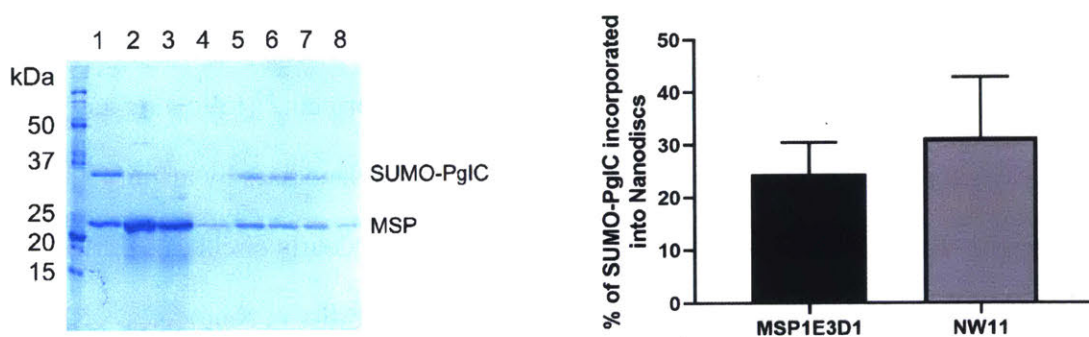
Figure 4-14 Comparison of MSP1E3D1 and NW11 for SUMO-PglC Nanodisc assembly. *Left:* Incorporation of SUMO-PglC into Nanodiscs was assessed by gel densitometry of Ni-NTA purification fractions. Lanes: 1 = load, 2 = flow-through, 3-4 = wash, 5-8 = elution. *Right:* Percent of total SUMO-PglC determined to be in the elution fraction. Error bars represent mean ± SEM, n = 4.

To determine whether localization of UndP to SUMO-PglC could be observed in Nanodiscs, Nanodisc assemblies were carried out in the presence of SUMO-PglC, the NW11 scaffolding protein, a synthetic mix of lipids mimicking a bacterial membrane, and increasing concentrations of UndP. The concentrations of UndP to be used were chosen such that the highest concentration (2% of total lipids) was expected to result in more than 98% of the Nanodiscs formed receiving at least one molecule of UndP, while the lowest concentration (0.016% of total lipids) was expected to result in less than 5% of the Nanodiscs formed receiving any UndP. At an intermediate concentration (0.3% of total lipids), ~50% of the Nanodiscs formed were expected to receive at least one molecule of UndP. At the highest concentration of UndP it was expected that UndP would be able to "saturate" the Nanodisc population, such that UndP would be distributed evenly throughout the population. When provided at significantly lower concentrations wherein very few Nanodiscs would be expected to receive any UndP, UndP was expected to partition preferentially into Nanodiscs that contained SUMO-PglC. As above,

Nanodiscs containing SUMO-PglC were isolated by Ni-NTA affinity chromatograph. Small amounts of isotopically labeled Und-[^{33}P]-P (supplemented with unlabeled UndP) were used so that the distribution of UndP in Nanodisc assemblies could be assessed by scintillation counting of Ni-NTA purification fractions (**Figure 4-15**).
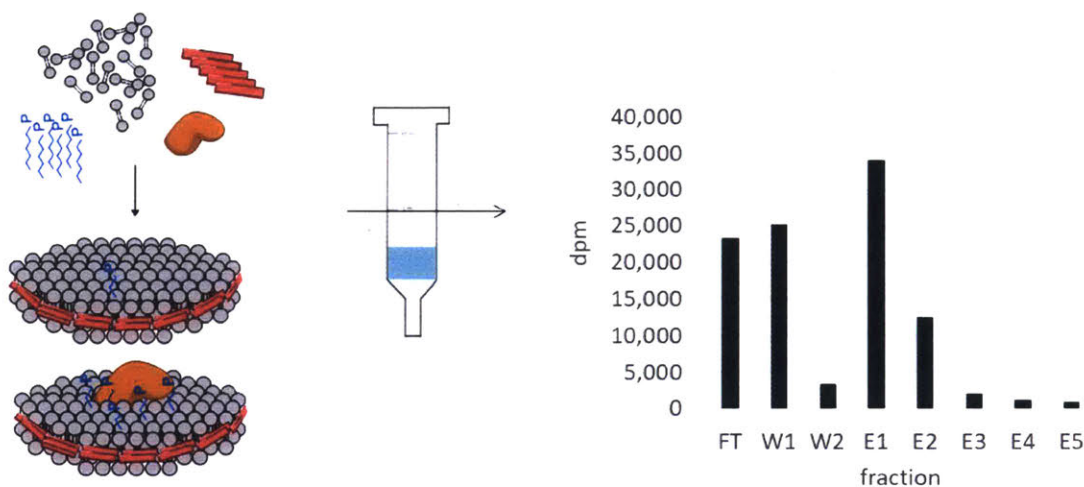
Figure 4-15 Schematic of Nanodisc methodology for assessing Und-[^{33}P]-P localization to PglC. Nanodiscs were assembled from NW11 circularized MSP (red), SUMO-PglC (orange), a synthetic lipid mix (gray) and a mix of unlabeled UndP and Und-[^{33}P]-P (blue). Nanodiscs containing SUMO-PglC were isolated using Ni-NTA affinity purification. The Und-[^{33}P]-P content of all fractions was measured using scintillation counting (FT = flow-through, W = wash, E = elution).

The fraction of counts associated with the elution fraction (corresponding to SUMO-PglC Nanodiscs) relative to the sum of all purification fractions is shown in **Figure 4-16** for Nanodisc preparations containing 0.016%, 0.3% and 2% UndP. The fraction of UndP associated with the elution fraction was higher when Nanodiscs were assembled with scarce (0.016%) UndP than when Nanodiscs were assembled with saturating (2%) UndP ($p = 0.043$). These results suggested that UndP-PglC binding might be resulting in localization of UndP to SUMO-PglC in the *in vitro* Nanodisc model. Thus, the same experiments were repeated with the P24A SUMO-PglC variant, which was shown above to be unable to bind UndP.
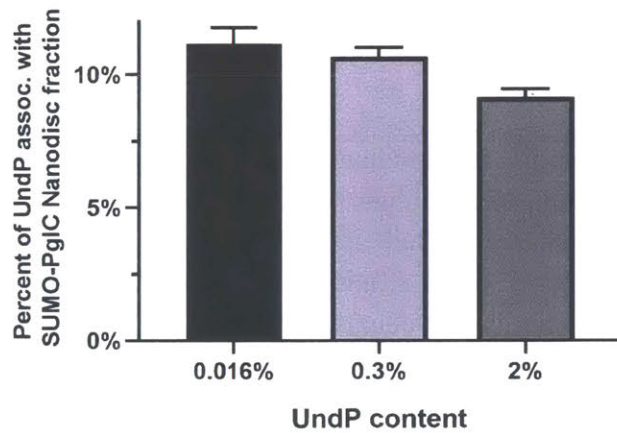
Figure 4-16 Assessing Und-[^{33}P]-P localization to wild-type SUMO-PglC in Nanodiscs. The percentage of scintillation counts associated with the elution fraction (containing SUMO-PglC Nanodiscs) was evaluated in Nanodiscs made with 0.016, 0.3% or 2% UndP. Error bars are given for mean ± SEM, n = 3, *p*-value between 0.016% and 2% UndP = 0.043 (Student's test).

The fraction of counts associated with the elution fraction for P24A SUMO-PglC Nanodiscs is shown in **Figure 4-17**. In the case of P24A SUMO-PglC, the overall amount of UndP associated with elution in all Nanodisc assemblies was generally lower than in assemblies with wild-type SUMO-PglC, suggesting a decrease in efficiency of Nanodisc assembly in the P24A SUMO-PglC variant. Unexpectedly, the same trend was observed as for wild-type SUMO-PglC Nanodiscs, with more UndP associating with the elution fraction for assemblies with low (0.016%) UndP than with high (2%) UndP ($p = 0.094$). Such a trend in the presence of both wild-type and P24A SUMO-PglC variants suggests that the observed enrichment is not mediated by UndP-PglC binding. There was more variability in experiments with P24A SUMO-PglC than with wild-type SUMO-PglC, which likely explains the higher p value.
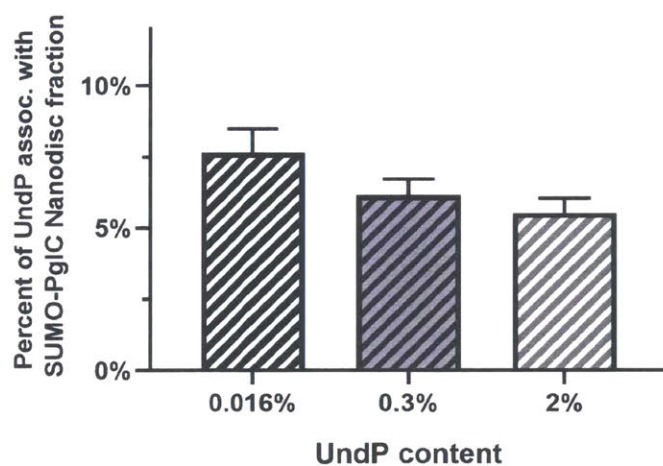
Figure 4-17 Assessing Und-[^{33}P]-P localization to P24A SUMO-PglC in Nanodiscs. The percentage of scintillation counts associated with the elution fraction for P24A SUMO-PglC Nanodiscs was evaluated in Nanodiscs made with 0.016, 0.3% or 2% UndP. Error bars are given for mean ± SEM, n = 3, p-value between 0.016% and 2% UndP = 0.094 (Student's test).

The ability of PrenP to modulate the lipid membrane depends on isoprene geometry

NMR and molecular modeling studies of polyprenol phosphates in the membrane previously suggested that these amphiphilic molecules could modulate the fluidity of the membrane and induce changes in membrane structure.[18, 20-22] We investigated this effect further using pyrene, an aromatic membrane probe that embeds in the lipid bilayer proximal to the headgroup region.[54] The characteristic fluorescence spectra of pyrene change with the polarity of the surrounding environment, with an increase in the ratio between two vibrionic peak intensities, I_1/I_3, corresponding to an increase in polarity. Pyrene fluorescence has been used to characterize membrane properties as a function of membrane composition, for example, as a result of lipid acyl chain saturation in bacterial membranes[55] and cholesterol depletion in hippocampal membranes.[56]

The ratio between the first (372 nm) and third (384 nm) emission peaks of pyrene was measured in liposomes composed of 3:1 POPE:POPG and increasing amounts of UndP (**Figure 4-18**). The I_1/I_3 ratio in such liposomes was found to increase with UndP concentration, indicating that UndP caused an increase in polarity in the local membrane environment. The evident trend was approximately linear up to 10% UndP, but an effect was observable even at 0.5% UndP, the lowest concentration. Pyrene fluorescence was also measured in POPE/POPG liposomes with increasing amounts of the all-*trans* SolP; in this case, no change in membrane polarity was observable, even at 10% SolP, the highest concentration tested.
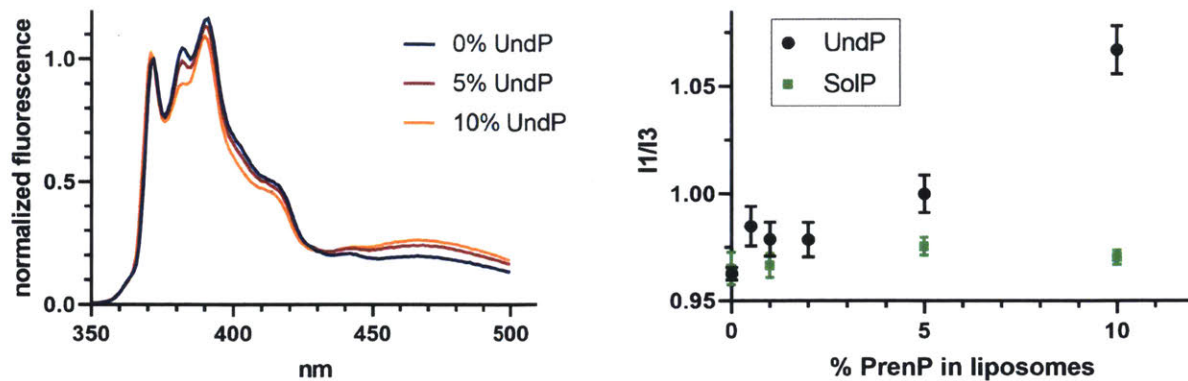
Figure 4-18 Pyrene fluorescence reflects changes in membrane environment.
Left: Fluorescence spectra of pyrene in liposomes composed of POPE/POPG and increasing concentrations of UndP. Representative data is shown normalized to fluorescence at 372 nm.
Right: The ratio of between the first (372 nm) and third (382 nm) vibrionic peaks of pyrene increases with increasing UndP concentration but is not affected by the presence of SolP. Error bars are given for mean ± SEM, n = 4-8.

DISCUSSION

PrenP abundance in *C. jejuni* is consistent with that of *E. coli* and eukaryotes

Previous work to quantify PrenP levels in bacteria has focused on model bacteria such as *E. coli*. One such study determined levels of ~350 ng UndP per gram of dry cell weight in *E. coli*; we estimate that this corresponds to ~0.3% of the total lipid pool.[57] Thus, the current quantification of ~0.1% PrenP abundance in *C. jejuni* approximately agrees with what has been determined for *E. coli*. Differences in methodology and interspecies variability likely account for deviations in the numbers. The total dolichol phosphate content of eukaryotic membranes has similarly been estimated to be ~0.1%,[3-4] suggesting that the abundance of polyprenol phosphate in membranes may be similar across domains of life.

Although the PrenP utilized for glycoconjugate assembly in *C. jejuni* was previously presumed to be UndP, the use of alternate PrenP lipid carriers is not unfounded: DecP is the preferred PrenP of *Mycobacterium tuberculosis*, and even shorter PrenP carriers are observed in other bacteria.[12-14] A previous study of PrenP substrate specificity in PglC and other enzymes from the pgl pathway demonstrated that while a series of *cis* double bonds and unsaturation at the α-isoprene unit of PrenP play important roles in substrate recognition, the length of the chain (ca. 8-11 isoprene units) plays a more minor role.[15]

Enrichment of PrenP in the local membrane environment

Studies to quantify PrenP abundance in *C. jejuni* and other bacteria are an important step towards understanding the role of PrenP in glycan assembly. However, several lines of evidence suggest that PrenP is not uniformly distributed throughout the membrane, but instead enriched in the local membrane environment of enzymes involved in glycoconjugate assembly.

PrenP is utilized in the biosynthetic pathways leading to many different bacterial glycoconjugates; thus, multiple pathways must compete for a single pool of PrenP in the membrane. [58-59] Early studies of bacterial glycoconjugate biosynthesis pathways suggested that these pathways function in large multi-enzyme complexes that assemble completed glycoconjugates without releasing pathway intermediates into the membrane.[60-62] Such a strategy capitalizes on the limited PrenP available and facilitates efficient flux of PrenP through the biosynthetic pathway.

Prior studies have suggested that enzymes in the *C. jejuni* pgl pathway similarly associate into a macromolecular complex.[53] Significant levels of PrenPP-glycoconjugate intermediates in were not observed in this analysis of membrane extracts, consistent with a model of glycoconjugate assembly in which enzymes in close proximity work in concert to minimize the release of intermediates. In such a model, the local concentration of PrenP substrate is expected to be higher in the membrane surrounding the glycoconjugate biosynthesis machinery than in the bulk membrane. We hypothesize that a role in the enrichment of PrenP would be especially important with PglC and PGTs in general, as these enzymes are typically responsible for initiating glycoconjugate assembly. Members of the monotopic PGT superfamily in particular also include a membrane-resident domain with a conserved proline previously predicted to be a key player in PrenP substrate binding. Using SMA copolymer to solubilize PglC together with the local lipid constituents, we quantified the abundance of PrenP in the membrane surrounding PglC relative to the rest of the solubilized membrane (**Figure 4-9**). For wild-type PglC heterologously expressed in *E. coli* membranes, we determined an almost two-fold enrichment of UndP relative to background. This finding supports a model of glycan assembly in which PglC "recruits" PrenP to areas of the membrane where glycan biosynthesis is occurring, thereby

supporting flux through the pathway. We note that in this *E. coli* model, the *C. jejuni*-specific

phosphosugar substrate for PglC, UDP-*N,N'*-diacetylbacillosamine, as well as the other enzymes

of the pgl pathway, are not present, allowing us to study enrichment of PrenP around PglC in

isolation from catalysis and pathway flux. In a native C. jejuni system with the entire pgl

pathway intact and participating in glycan assembly, we might anticipate an even greater

enrichment of PrenP in the local membrane environment.

Pro24 plays a key role in PrenP "recruitment" by PglC

Whereas UndP was found to be almost two-fold enriched around SMA-solubilized wild-type

PglC relative to background, the lipids surrounding a P24A PglC variant did not show significant

enrichment in UndP (**Figure 4-9**), strongly supporting a critical role for Pro24 in PrenP binding.

Thermal denaturation assays were used to further investigate substrate binding *in vitro*. Assays

performed with SUMO-PglC revealed a significant increase in T_m from 41.9±1.3 °C in 10%

DMSO to 64.2±1.4 °C in 300 μM UndP + 10% DMSO. In contrast, a P24A SUMO-PglC variant

appeared to be stabilized only via non-specific interactions with UndP, as evidenced by a lack of

cooperativity in denaturation. Thus, a critical role for Pro24 in binding PrenP substrate is further

confirmed *in vitro*.

Importantly, although Pro24 is part of a conserved two-residue motif that was previously

found to drive folding of the reentrant membrane domain of PglC, the P24A PglC variant was

found to adopt a native-like reentrant topology.[38] In addition, both CD spectroscopy and thermal

denaturation assays of wild-type and P24A SUMO-PglC in the absence of UndP indicate that the

two variants have a similar thermal stability (**Error! Reference source not found.**). Thus,

although minor changes in structure between the wild-type and P24A PglC variants cannot be

ruled out, the available data suggest that loss of UndP binding in P24A PglC reflects an inability

of the variant to bind PrenP substrate, rather than a global misfolding event.

Increased concentrations of PrenP modulate the local membrane environment

The exact consequences of PrenP enrichment in the membrane are unclear, although many

possible modes of action have been suggested. It has been proposed that PrenP may serve as a

structural scaffold for the organization of complexes with downstream glycosyltransferases[18, 63]

In addition, it has been found that PrenP alters the biophysical properties of the local membrane,

increasing fluidity and modifying membrane structure;[18, 20-22] such local changes may also be

important for promoting multi-enzyme complex formation or for increasing substrate flux

through the pathway.[17] An analogous role in organizing lipids and proteins into membrane

microdomains has been proposed for cholesterol.[1, 64]

Our studies using the membrane polarity probe, pyrene, confirm earlier reports that UndP

modulates the biophysical properties of the membrane (**Figure 4-18**). Whereas prior studies were

carried out with high levels (5-10%) of PrenP, the trend observed in the pyrene fluorescence

studies persists at much lower, and more physiologically relevant, UndP concentrations. As little

as 0.5% UndP in liposomes is found to increase the polarity of the local membrane environment,

an effect which is attributed to increased water penetration into the membrane due to loss of

membrane order surrounding UndP.[56] Our findings support the proposed model wherein

increased local concentrations of PrenP modulate the biophysical properties of the membrane to

facilitate multi-enzyme complex formation and support substrate flux through the pathway.

Specificity for *cis* isoprene geometry

The pyrene fluorescence studies show that polarity in the membrane environment does not increase in the presence of up to 10% of the all-*trans* SolP, indicating that double-bond geometry plays a key role in the reported ability of PrenP to disrupt lipid membranes. NMR and molecular modeling studies of polyprenol phosphates have suggested that the *cis*-isoprene units of UndP and dolichol phosphate chains create a coiled structure within the membrane, while the *trans*-isoprene units form a non-coiled "tail";[18] it may be that this *cis*-isoprene coiled structure of UndP is more disruptive to local lipid packing than the all-*trans* linear structure of SolP.

In vitro thermal stability studies additionally provide new insights into the specificity of PglC for the *cis*-isoprene geometry of PrenP substrates. Thermal denaturation assays of wild-type and P24A SUMO-PglC in the presence of 300 μM UndP revealed two very different behaviors (**Figure 4-12**). Wild-type SUMO-PglC exhibited complete cooperative denaturation in the presence of UndP, but was significantly more stabilized, corresponding to a ΔT_m of 22.3±1.9 °C, indicating stabilization due to substrate binding. By contrast, P24A SUMO-PglC in the presence of UndP appeared more stable, but no longer exhibited cooperative denaturation, and remained partially soluble even at elevated temperatures.

The thermal stability studies suggest that P24A SUMO-PglC may not be stabilized by binding of UndP in the active site; instead, it may be stabilized in a non-native conformation by interaction with UndP. Due to the amphiphilic nature of PrenPs, it is possible that such molecules are able to have a detergent-like solubilizing effect in solution, stabilizing the hydrophobic portion of the PglC fold even as tertiary structure is lost during thermal denaturation. Indeed, CD spectroscopy studies of wild-type and P24A SUMO-PglC indicated that the secondary structures of both variants were significantly stabilized by the presence of

UndP (**Figure 4-11**). Importantly, unfolding of α-helical membrane proteins has been reported to occur through loss of tertiary, rather than secondary, structure.[47-48] Thus, while UndP stabilized the secondary structure of both wild-type and P24A SUMO-PglC, and contributed to increasing the solubility of both, only the wild-type PglC tertiary structure appeared to be stabilized in a native, UndP-binding conformation. Such a model agrees well with the observation that the P24A SUMO-PglC variant is inactive (**Figure 4-3**).

Non-cooperative and incomplete denaturation was also observed for both wild-type and P24A SUMO-PglC in the presence of 300 μM SolP (**Figure 4-13**), indicating that both variants are similarly unable to bind the all-*trans* PrenP. These findings are corroborated by previous reports of the importance of the geometry of the internal *cis* double-bond of PrenP for PglC activity.[15] Thus, only interaction between wild-type SUMO-PglC and the substrate UndP resulted in stabilization of a native conformation of PglC. These findings further support the significance of Pro24 as a key mediator of PrenP recognition and binding, and suggest an additional role in discerning substrate PrenP from phospholipid acyl tails and non-substrate isoprenoids in the membrane on the basis of double-bond geometry.

Structural models predict that peptides containing polyisoprenol-recognition sequences bind to the coiled, *cis*-isoprene region of PrenPs.[18, 63] In the reported structure of PglC, a co-crystallized PEG molecule reveals the putative PrenP-binding site, which is near the active site and directly above Pro24 (**Figure 4-2**). Thus, we can envision a model of PrenP binding wherein Pro24, buried within the membrane, associates with the unique coiled structure of the PrenP substrate and facilitates positioning of the phosphate group of PrenP at the active site, which is located at the membrane interface, for catalysis. Notably, unlike members of the polytopic PGT superfamily, which have 10- or 11-transmembrane helices which may interact with PrenP,

members of the monotopic PGT superfamily rely on a single reentrant helix, carrying the conserved Pro24, to facilitate PrenP binding.

A molecular basis for substrate specificity mediated by Pro24

The ability of Pro24 to distinguish PrenP substrate is likely due to the unique structure of the pyrrolidine side chain of proline. Proline is known to break α-helices by perturbing hydrogen bonding in the peptide backbone.[65] It has been suggested that the change in peptide-backbone hydrogen bonding around conserved prolines in polyisoprenol-recognition sequences facilitates interactions between PrenP and PrenP-binding proteins.[63] Indeed, recent structural and molecular dynamics analyses of PglC revealed an extensive and dynamic network of modified hydrogen bonds in the peptide backbone surrounding Pro24 in the hydrophobic domain of PglC.[34, 38]

A bioinformatics analysis of the monotopic PGT superfamily found that among members that do not carry the conserved Pro24, this residue is most commonly substituted with tryptophan.[37] A P24W PglC variant was found to be less active than wild-type PglC but significantly more active than P24A PglC, indicating that it was able to bind PrenP substrate. As discussed in Chapter 2, P24W substitutions are commonly accompanied by S23W substitutions, and that both P24W and S23W/P24W variants are expected to adopt native, reentrant topologies. In such homologs, it is possible that the Trp side chain, which contains an indole ring and can act as a hydrogen bond donor to disrupt the peptide backbone, is able to partially substitute for the molecular interactions that would otherwise be facilitated by Pro24.

Probing UndP localization to PglC *in vitro* using Nanodiscs

Attempts to observe UndP localization to SUMO-PglC in *in vitro* Nanodisc experiments proved

challenging. While initial experiments with wild-type SUMO-PglC appeared to suggest that

UndP became enriched in Nanodiscs carrying SUMO-PglC under conditions of limited UndP,

the same general trend was also observed in Nanodiscs carrying the P24A SUMO-PglC control

variant, indicating that the observed enrichment was not mediated by UndP-PglC binding. In

addition, the poorer efficiency of Nanodisc assembly and increased variability with the P24A

SUMO-PglC variant ultimately made any trends in the data very difficult to interpret.

An alternative explanation for the observed trend may be that Nanodisc formation does

not proceed with the same efficiency at all concentrations of UndP, as it was assumed to be.

Although the Nanodisc assembly reaction is considered to be fairly robust, minor variability on

the basis of lipid composition cannot be ruled out. As the pyrene fluorescence studies herein

show, even small amounts of UndP can modulate the biophysical properties of the lipid bilayer.

If the presence of 2% UndP during Nanodisc assembly decreases the efficiency of Nanodisc

formation, the total incorporation of SUMO-PglC into Nanodiscs formed is expected to be lower,

and subsequently less UndP would be observed in the elution fraction. In this case, the observed

trend would persist regardless of UndP-PglC binding. Notably, such minor changes would be

unlikely to impact most applications for which Nanodisc technology can be applied; the ability to

discern this particular change between Nanodiscs with 0.016% and 2% UndP was due to the use

of high-sensitivity scintillation counting. Thus, although unsuitable for the particular

experimental application described herein, Nanodisc methodologies remain an important and

versatile tool for the study of membrane proteins.

CONCLUSIONS

The elusive role of PrenP in glycan assembly by the pgl pathway is explored with respect to both its effect on the local membrane environment and interaction with PglC, a monotopic PGT that acts in initiation of N-glycan assembly. The *cis*-double bond geometry of PrenP is found to be essential both to interactions with the local membrane and with PglC. We also confirm a role for the conserved Pro24 in PglC in binding PrenP substrate, and propose an additional role in driving specificity for the unique *cis*-isoprene geometry of the PrenP substrate. Our findings build upon models, proposed previously, of PrenP as a modulator of the local membrane environment and facilitator of multi-enzyme complex formation for glycan assembly. Furthermore, as the *cis*-isoprene geometry is conserved in polyprenol phosphates used for glycoconjugate biosynthesis in all domains, we predict that many of the findings herein will be broadly applicable to a diversity of prokaryotic and eukaryotic glycan biosynthesis pathways of interest.

EXPERIMENTAL PROCEDURES

C. jejuni growth

C. jejuni strain 81-176 from a glycerol stock was streaked on selective blood agar plates (BD,

Franklin Lakes, NJ; cat. 221727) at 37 °C under microaerophilic conditions (10% CO_2, 5% O_2,

85% N_2). After 20-24 hr of growth cells were resuspended in a small volume of Mueller-Hinton

broth. A 500 µL aliquot of resuspension at an OD600 of 0.5 was used to inoculate a 0.5 L culture

of Mueller-Hinton broth + 10 µg/mL trimethoprim (selection antibiotic) in a 1 L Erlenmeyer

flask. Cultures were grown at 37 °C under microaerophilic conditions (see above) with shaking

at 200 rpm. In preliminary experiments, cultures were sampled periodically to determine a

growth curve under these conditions; log-phase growth was found to occur at an OD600 of 0.15-

0.35. In subsequent experiments, cultures were harvested at an OD600 within this range and

stored at -80 °C until analysis.

Total lipid extraction

Lipids were extracted using a modification of the protocol by Bligh and Dyer.[39] Cells from 250

mL of C. jejuni culture were washed once with Tris-buffered saline (25 mM Tris, 150 mM NaCl,

pH 7.5) and resuspended in 2 mL of water acidified to a pH of 1-2 by the addition of

concentrated HCl (a low pH facilitated the partition of UndP into the lower organic phase during

extraction). Next, 4 mL of 2:1 (v/v) $CHCl_3$:MeOH were added to the cell suspension and

samples were vortexed for 15-30 sec and centrifuged at 1,000 x g for 5-10 min. The organic

phase was removed and the aqueous phase was extracted three additional times with 1-2 mL

$CHCl_3$. All organic phases were pooled, concentrated under N_2 and washed with an additional 3

mL of water acidified with HCl to a pH of 1-2. Final lipid extracts were dried completely under

N_2 and stored at -20 °C until analysis.

Quantification and alkaline hydrolysis of total lipid extracts

Lipid extracts were quantified by a modification of the ammonium molybdate method for

microdetermination of total phosphate described previously[66] against a phosphorus standard

solution (Sigma-Aldrich, St. Louis, MO).

Prior to LC/MS quantification of total PrenP, lipid extracts were subjected to mild

alkaline hydrolysis[40] to release PrenP from PrenPP and PrenPP-linked glycans. Dry lipids (~500

nmol) were resuspended in 4 mL of a single-phase Bligh-Dyer mix containing 0.2 M NaOH

(1:2:0.8 (v/v) $CHCl_3$:MeOH:0.95M NaOH) and incubated at 60 °C for 1 hr with periodic

vortexing. The addition of 1 mL $CHCl_3$ and 0.8 mL 1 M HCl neutralized the solution and

converted the solvent mix to two phases. The organic phase was removed and the aqueous phase

was extracted three additional times with 1 mL $CHCl_3$. All organic phases were pooled,

concentrated under N_2 and washed with an additional 2 mL of water acidified with HCl to a pH

of 1-2. Total hydrolyzed lipids were divided equally into five aliquots (~100 nmol each) and

supplemented with 0-100 pmol of pure UndP standard (see below). Samples were dried under N_2

and stored at -20 °C until analysis.

Purification of an UndP standard for LC/MS

Synthesis of UndP by phosphorylation of plant-derived UndOH was described previously.[53] This

method yields a mixture of C_{50}-C_{60} PrenP (predominantly UndP, C_{55}) which was used for

activity and thermal denaturation assays. For use as a standard for PrenP quantification, this mix

was purified further by ion-pair reverse-phase HPLC to give only the UndP.[67] Separation of C_{50}, C_{55} and C_{60} PrenP was successfully performed on a YMC-Pack ODS-A RP-18 column (4.6 × 250 mm, 5 μm, 120 Å) under isocratic elution in 76:19:5:2 (v/v) EtOH:CH_3CN:hexanes:ion-pair reagent (ion pair-reagent prepared as 8 g 85% H_3PO_4 + 45 mL 40% aq. tetrabutyl ammonium hydroxide). The fraction containing pure UndP (C_{55}) was quantified by the ammonium molybdate method described above, dried under N_2 and stored at -20 °C until analysis.

Quantification of total polyprenol-P by LC/MS

Each dried lipid extract was resuspended in 0.5 mL 2:1 (v/v) $CHCl_3$:MeOH and 10 μL of sample was injected for normal-phase LC/MS analysis. Normal-phase LC was performed on an Agilent 1200 Quaternary LC system equipped with an Ascentis Silica HPLC column, 5 μm, 25 cm x 2.1 mm (Sigma-Aldrich). Mobile phase A consisted of $CHCl_3$:MeOH:aq. NH_4OH (800:195:5, v/v); mobile phase B consisted of $CHCl_3$:MeOH:aq. NH_4OH (600:340:50:5, v/v); mobile phase C consisted of $CHCl_3$:MeOH:aq. NH_4OH (450:450:95:5, v/v). The elution program consisted of the following: 100% mobile phase A was held isocratically for 2 min, then linearly increased to 100% B over 14 min and held at 100% B for 11 min. The LC gradient was then changed to 100% C over 3 min and held at 100% C for 3 min, and finally returned to 100% A over 0.5 min and held at 100% A for 5 min. The LC eluent (with a total flow rate of 300 μl/min) was introduced into the ESI source of a high resolution TripleTOF5600 mass spectrometer (Sciex, Framingham, MA). Instrumental settings for negative ion ESI and MS/MS analysis of lipid species were as follows: IS= -4500 V; CUR= 20 psi; GSI= 20 psi; DP= -55 V; and FP= -150 V. The MS/MS analysis used nitrogen as the collision gas. Data analysis was performed using Analyst TF1.5 software (Sciex).

PrenP abundance is reported as total PrenP (determined by LC/MS) as a percentage of total extracted lipids (determined by total phosphate quantification, described above). Data were plotted using Graphpad Prism 8 (GraphPad Software).

PglC variants and expression

PglC from *C. jejuni* strain 11168 was cloned into the pET24a vector to insert a C-terminal His$_6$-tag[38] or into the pE-SUMO vector.[37] P24A variants of both constructs were generated using QuikChange II Site-Directed Mutagenesis (Agilent Technologies, Santa Clara, CA) as described previously [38].

PglC-His$_6$ and SUMO-PglC variants were expressed in BL21-CodonPlus (DE3)-RIL *E. coli* cells (Agilent Technologies) using the Studier auto-induction method.[68] Overnight cultures grown in 3 mL MDG media (0.5% (w/v) glucose, 0.25% (w/v) sodium aspartate, 2 mM MgSO$_4$, 25 mM Na$_2$HPO$_4$, 25 mM KH$_2$PO$_4$, 50 mM NH$_4$Cl, 5 mM Na$_2$SO$_4$ and 0.2x trace metal mix (from 1000x stock, Teknova, Hollister, CA)) with kanamycin and chloramphenicol (30 μg/mL each) were used to inoculate 0.5 L auto-induction media (1% (w/v) tryptone, 0.5% (w/v) yeast extract, 0.5% (v/v) glycerol, 0.05% (w/v) glucose, 0.2% (w/v) α-D-lactose, 2 mM MgSO$_4$, 25 mM Na$_2$HPO$_4$, 25 mM KH$_2$PO$_4$, 50 mM NH$_4$Cl, and 5 mM Na$_2$SO$_4$, 0.2x trace metal mix) containing kanamycin (90 μg/mL) and chloramphenicol (30 μg/mL). Cells were grown for 4 hr at 37 °C followed by an additional 16-18 hr at 16 °C and then harvested at 3700 x *g* for 30 min. Cells were stored at -80 °C until purification.

SMA solubilization and Ni-NTA purification of PglC-His$_6$ variants

All purification steps were carried out at 4 °C. Cells from 0.5 L of autoinduction culture were resuspended in 20 mL lysis buffer (50 mM Tris, 150 mM NaCl, pH 7.5) and lysed by French Press at 25 psi. Lysate was supplemented with 0.5 mg/mL lysozyme (Research Products International, Mount Prospect, IL), 1:1000 dilution of EDTA-free protease inhibitor cocktail (EMD Millipore, Burlington, MA) and 1 unit/mL DNase I (New England Biolabs, Ipswich, MA) and incubated on ice for 15 min, then centrifuged at 9,000 x g for 45 min. The resulting supernatant was further centrifuged at 140,000 x g for 65 min to yield the cell envelope fraction (CEF), which was homogenized into 20-30 mL solubilization buffer (50 mM Tris, 150 mM NaCl, pH 8.0) to a final protein concentration of 50 mg/mL, as assessed by absorbance at 280 nm.

Homogenized CEF was mixed 1:1 with a 5% (w/v) solution of XIRAN SL40005 styrene-maleic anhydride copolymer (Polyscope, Geleen, Netherlands) in solubilization buffer (50 mM Tris, 150 mM NaCl, adjusted to a final pH 8.0) and tumbled on a rotating mixer for 3 hr at room temperature. Solubilized lysate was centrifuged at 100,000 x g for 45 min to remove any insoluble material. Approximately 1/10 of the clarified supernatant volume was retained as a "background" sample, while the remaining supernatant was incubated with 5 mL Ni-NTA resin (Thermo Fisher Scientific, Waltham, MA) pre-treated with equilibration buffer (50 mM Tris, 150 mM NaCl, 10 mM imidazole, pH 8.0) overnight at 4 °C on a rotating mixer. Resin was washed with 10 column volumes of wash buffer (50 mM Tris, 150 mM NaCl, 20 mM imidazole, pH 8.0). SMA lipoparticles carrying PglC-His$_6$ were eluted in 2 column volumes of elution buffer (50 mM Tris, 150 mM NaCl, 500 mM imidazole, pH 8.0). The purity of eluted particles was assessed by SDS-PAGE with Coomassie staining.

Total lipid extraction from SMA lipoparticles

The extraction protocol for SMA lipoparticles was adapted from Dorr et al.[44] A 30 kDa MWCO centrifugal filter was used to exchange eluted SMA lipoparticles into solubilization buffer (50 mM Tris, 150 mM NaCl, pH 8.0) and to concentrate to ~1 mL. For extraction, 1 mL Ni-NTA enriched or "background" lipoparticles were added to 3.4 mL 10:23:1 (v/v) $CHCl_3$:MeOH:1 M Tris pH 8.0 to make a single phase and vortexed intermittently for 30 min at room temperature. Phase separation was induced by the addition of 0.5 mL 0.1 M Tris, pH 8.0 and 0.5 mL $CHCl_3$; samples were vortexed for 15-30 sec and centrifuged at 1,000 x g for 5-10 min. The organic phase was removed and the aqueous phase was extracted three additional times with 1 mL $CHCl_3$. All organic phases were pooled, concentrated under N_2 and washed with 1 mL ion-switch buffer (50 mM Tris, 100 mM NaCl, 100 mM EDTA, pH 8.2). Approximate lipid composition and concentration were estimated by TLC against a commercial *E. coli* polar lipid extract (Avanti Polar Lipids, Inc., Alabaster, AL), mobile phase 65:25:4 (v/v) $CHCl_3$:MeOH:water, visualization with molybdenum blue stain. Final lipid extracts were dried completely under N_2 and stored at -20 °C until analysis.

Purification and activity assays of SUMO-PglC

SUMO-PglC variants were purified by Ni-NTA affinity as described in detail previously.[38] Activity assays were performed using the UMP/CMP-Glo assay (Promega, Madison, WI) on a 10 μL scale, as described previously.[38] Data were plotted using Graphpad Prism 8 (GraphPad Software).

Circular dichroism

Circular dichroism was performed on a JASCO Model J-1500 Circular Dichroism Spectrometer. Spectral scans were performed with 6 μM purified SUMO-PglC in phosphate buffer (50 mM Na_2HPO_4, 100 mM NaCl, 0.03% DDM, pH 7.5). Reads were taken at 4 °C from 190-250 nm at 0.5 nm intervals. Thermal melts were performed with 10 μM purified SUMO-PglC in assay buffer (50 mM HEPES, 100 mM NaCl, 5 mM $MgCl_2$, 0.1% reduced Triton X-100, pH 7.5). Samples were heated from 4-95°C at a rate of 1 °C per minute. Reads were taken at 222 nm at 2 °C intervals.

Thermal denaturation assay

Thermal denaturation assays were based on a protocol described previously.[49] Aliquots of 3 μM purified SUMO-PglC in assay buffer (50 mM HEPES, 100 mM NaCl, 0.1% Triton X-100, 5 mM $MgCl_2$, pH 7.5) were supplemented with 300 μM UndP or SolP (synthesis and purification described previously)[15] from a 10x stock in DMSO, or with 10% DMSO (vehicle control), and heated for 10 min at 30–99 °C in a thermocycler (MJ Mini Thermal Cycler; BioRad, Hercules, CA). Precipitate was immediately removed by centrifugation at 16,000 x g for 10 min at 4°C. The resulting supernatant, containing protein that remained soluble, was analyzed by SDS-PAGE with Coomassie staining and quantified by gel densitometry using a Molecular Imager Gel Doc XR+ System with Image Lab software (BioRad). Data were fitted using Graphpad Prism 8 (GraphPad Software).

Pyrene fluorescence in liposomes

Lipids (3:1 (mol/mol) POPE:POPG, Avanti Lipids, Inc.) from chloroform stocks, supplemented with UndP or SolP as necessary, were dried first under N_2 and then overnight under vacuum.

Lipids were hydrated in PBS at 42 °C for at least two hours, with intermittent vortexing. The final lipid concentration was 50 μM. Pyrene was added to a final concentration of 0.5 μM from a 100x stock in ethanol and allowed to incubate with liposomes for at least 30 min at room temperature.

Fluorescence was measured on a FluoroMax-P (Horiba, Kyoto, Japan), with excitation at 335 nm and emission at 350-500 nm with 4 nm bandwidth in both. Although background fluorescence was low, a blank liposome sample (from which pyrene was omitted) was read and subtracted from each pyrene-containing sample.

Purification of UndOH

Crude UndOH was extracted from staghorn suman (*Rhus typhina*) leaves as described previously.[53] The extract was purified by RP-HPLC on a YMC-Pack ODS-A RP-18 column (20 × 250 mm, S-5 μm, 120 Å) with the following solvent mixes: A = 60:40:5 (v/v) MeOH:iPrOH:water, B = 60:40 (v/v) hexanes:iPrOH, and a gradient of 0% B for 20 min followed by 10-20% B over 30 min. Three product peaks were isolated as oils corresponding to C_{50}, C_{55}, and C_{60} polyprenol (yield = 7%, 17% and 8%, respectively). Polyprenols were dried under rotary evaporation and stored at -20 °C until use.

Chemoenzymatic synthesis of Und-[^{33}P]-P

Diacylglycerol kinase from *Streptococcus mutans* was expressed and purified as a cell envelope fraction as described previously.[69] Chemoenzymatic synthesis of Und-[^{33}P]-P followed the protocol described previously:[53] 55 nmol UndOH, 55 nmol ATP, ~50 pmol γ-[^{33}P]-ATP (3,000 Ci/mmol) and 25 uL *S. mutans* diacylglycerol kinase CEF were reacted in a buffer containing 3% DMSO, 1% Triton X-100, 30 mM Tris pH 8.0, 50 mM $MgCl_2$ reaction for 1 h at room

temperature. The reaction was quenched into 2 mL 2:1 (v/v) $CHCl_3$:MeOH and washed with 3 x 2 mL Pure Solvent Upper Phase (for 500 mL combine: 235 ml of H_2O, 1.83 g of KCl, 240 ml MeOH, 15 ml $CHCl_3$).

The resulting organic phase is dried down under N_2, resuspended in a small amount of 7:1 $CHCl_3$:MeOH and loaded into a Pasteur pipette silica column (45 mm silica height) pre-wetted with the same solvent mix. The column was run with 4 column volumes 7:1 (v/v) $CHCl_3$:MeOH to remove unreacted UndOH and Triton-X followed by 4 column volumes methanol to elute the UndP product; a small amount of lipids from the CEF also eluted in this fraction. As MeOH was found to dissolve some silica binder, the methanol phase was dried down under N_2, resuspended in $CHCl_3$ and filtered through a cotton plug. The final filtrate was concentrated under N_2 quantified by scintillation counting using an LS 6500 Scintillation Counter (Beckman Coulter, Brea, CA), and aliquoted. The final yield of Und-[^{33}P]-P after purification was 5-10%.

Nanodisc assembly and Ni-NTA purification

The pET28a-MSP1E3D1 construct was obtained from the Gerhard Wagner lab and is also available to the community (Addgene, Watertown, MA). Purification of MSP1E3D1 scaffold protein proceeded as described in Chapter 5 for MSP1D1ΔH5. Lipid chloroform stocks were quantified by a modification of the ammonium molybdate method for microdetermination of total phosphate described previously[66] against a phosphorus standard solution (Sigma-Aldrich). Lipids from chloroform stocks were measured into Eppendorf tubes and dried first under N_2 and then under high vacuum overnight.

Nanodiscs with SUMO-PglC and MSP1E3D1 were assembled in Eppendorf tubes on a

300 μL scale using 50 μM MSP1E3D1, 5 μM SUMO-PglC, 3.5 mM lipids (71:24:5 (mol/mol)

POPE:POPG:cardiolipin, Avanti Polar Lipids, Inc.) and 25 mM sodium cholate in Nanodisc

buffer (50 mM HEPES, 150 mM NaCl, pH 7.5). Assembly reactions were rotated at 4 °C for 20

min. Meanwhile, adsorbent BioBeads SM-2 (BioRad) were prepared by washing with 4 x 1

volume of MeOH followed by 4 x 10 volumes of water. BioBeads (300 μL) were added to the

Nanodisc assembly reaction and the Nanodisc-BioBead mix was rotated at 4 °C overnight.

NW11 circularized MSP was purchased from Now Scientific (Boston, MA). Nanodiscs

with SUMO-PglC and NW11 were assembled on a 300 μL scale using 50 μM NW11, 5 μM

SUMO-PglC, 6 mM lipids (71:24:5 (mol/mol) POPE:POPG:cardiolipin) and 25 mM sodium

cholate in 50 mM HEPES, 150 mM NaCl, pH 7.5. For experiments with isotopically labeled

UndP, 0.96 mM Und-[^{33}P]-P (for 0.016% UndP), 0.96 μM Und-[^{33}P]-P plus 17 μM unlabeled

UndP (for a total of 0.3% UndP), or 0.96 μM Und-[^{33}P]-P plus 119 μM unlabeled UndP (for a

total of 2% UndP), were also included in the lipids. Assembly proceeded similarly to Nanodiscs

with MSP1E3D1, but as described previously,[52] components were incubated on ice for 30-60

min, instead of rotating for 20 min, prior to the addition of BioBeads.

The following day, Nanodiscs were removed from the Eppendorf tube by piercing the

bottom of the tube with a small-gauge needle and applying centrifugal force. The BioBeads

remaining in the tube were washed with an additional 200 μL of Nanodisc buffer. The eluted

Nanodiscs and wash were combined and incubated with 200 μL Ni-NTA resin equilibrated with

Nanodisc buffer for 1 hr at 4 °C. The resin was washed with 2 x 2 mL cold Wash I buffer

(Nanodisc buffer + 20 mM imidazole) followed by 2 x 2 mL cold Wash II buffer (Nanodisc

buffer + 45 mM imidazole). SUMO-PglC Nanodiscs were eluted in 3 x 300 μL elution buffer

(Nanodisc buffer + 500 mM imidazole) followed by 2 x 300 μL 1 M imidazole. For gel

densitometry analysis, aliquots (10-100 μL) of Ni-NTA purification fractions were analyzed by

SDS-PAGE with Coomassie staining and quantified by gel densitometry using a Molecular

Imager Gel Doc XR+ System with Image Lab software (BioRad).

Liquid scintillation counting

Aliquots (5-500 μL) of Ni-NTA purification fractions were added to glass scintillation vials and

treated with 250 μL SOLVABLE (PerkinElmer, Waltham, MA) for 30-60 min at 50 °C with

intermittent vortexing. Samples were allowed to cool to room temperature. Five mL of Opti-

Fluor (PerkinElmer) was added to each vial, with vortexing. Samples were incubated at room

temperature for 30 min before being subjected to scintillation counting using an LS 6500

Scintillation Counter (Beckman Coulter).

REFERENCES

(1) Krause, M. R.; Regen, S. L., The structural role of cholesterol in cell membranes: from condensed bilayers to lipid rafts. *Acc Chem Res* **2014,** *47* (12), 3512-21.

(2) Barreteau, H.; Magnet, S.; El Ghachi, M.; Touze, T.; Arthur, M.; Mengin-Lecreulx, D.; Blanot, D., Quantitative high-performance liquid chromatography analysis of the pool levels of undecaprenyl phosphate and its derivatives in bacterial membranes. *J Chromatogr B Analyt Technol Biomed Life Sci* **2009,** *877* (3), 213-20.

(3) Kaiden, A.; Krag, S. S., Dolichol metabolism in Chinese hamster ovary cells. *Biochem Cell Biol* **1992,** *70* (6), 385-9.

(4) Krag, S. S., The importance of being dolichol. *Biochem Biophys Res Commun* **1998,** *243* (1), 1-5.

(5) Lechner, J.; Wieland, F.; Sumper, M., Biosynthesis of sulfated saccharides N-glycosidically linked to the protein via glucose. Purification and identification of sulfated dolichyl monophosphoryl tetrasaccharides from halobacteria. *J Biol Chem* **1985,** *260* (2), 860-6.

(6) Adair, W. L., Jr.; Cafmeyer, N., Characterization of the Saccharomyces cerevisiae cis-prenyltransferase required for dolichyl phosphate biosynthesis. *Arch Biochem Biophys* **1987,** *259* (2), 589-96.

(7) Eggens, I.; Chojnacki, T.; Kenne, L.; Dallner, G., Separation, quantitation and distribution of dolichol and dolichyl phosphate in rat and human tissues. *Biochim Biophys Acta* **1983,** *751* (3), 355-68.

(8) Guan, Z.; Meyer, B. H.; Albers, S. V.; Eichler, J., The thermoacidophilic archaeon Sulfolobus acidocaldarius contains an unusually short, highly reduced dolichyl phosphate. *Biochim Biophys Acta* **2011,** *1811* (10), 607-16.

(9) Guan, Z.; Naparstek, S.; Kaminski, L.; Konrad, Z.; Eichler, J., Distinct glycan-charged phosphodolichol carriers are required for the assembly of the pentasaccharide N-linked to the Haloferax volcanii S-layer glycoprotein. *Mol Microbiol* **2010,** *78* (5), 1294-303.

(10) Kuntz, C.; Sonnenbichler, J.; Sonnenbichler, I.; Sumper, M.; Zeitler, R., Isolation and characterization of dolichol-linked oligosaccharides from Haloferax volcanii. *Glycobiology* **1997,** *7* (7), 897-904.

(11) Higashi, Y.; Strominger, J. L.; Sweeley, C. C., Structure of a lipid intermediate in cell wall peptidoglycan synthesis: a derivative of a C55 isoprenoid alcohol. *Proc Natl Acad Sci U S A* **1967,** *57* (6), 1878-84.

(12) Wolucka, B. A.; de Hoffmann, E., Isolation and characterization of the major form of polyprenyl-phospho-mannose from Mycobacterium smegmatis. *Glycobiology* **1998,** *8* (10), 955-62.

(13) Kaur, D.; Brennan, P. J.; Crick, D. C., Decaprenyl diphosphate synthesis in Mycobacterium tuberculosis. *J Bacteriol* **2004,** *186* (22), 7564-70.

(14) Ishii, K.; Sagami, H.; Ogura, K., A novel prenyltransferase from Paracoccus denitrificans. *Biochem J* **1986,** *233* (3), 773-7.

(15) Chen, M. M.; Weerapana, E.; Ciepichal, E.; Stupak, J.; Reid, C. W.; Swiezewska, E.; Imperiali, B., Polyisoprenol specificity in the Campylobacter jejuni N-linked glycosylation pathway. *Biochemistry* **2007,** *46* (50), 14342-8.

(16) Yan, N.; Liu, Y.; Zhang, H.; Du, Y.; Liu, X.; Zhang, Z., Solanesol biosynthesis in plants. *Molecules* **2017,** *22* (4).

(17) Hartley, M. D.; Imperiali, B., At the membrane frontier: a prospectus on the remarkable evolutionary conservation of polyprenols and polyprenyl-phosphates. *Arch Biochem Biophys* **2012,** *517* (2), 83-97.

(18) Zhou, G. P.; Troy, F. A., 2nd, Characterization by NMR and molecular modeling of the binding of polyisoprenols and polyisoprenyl recognition sequence peptides: 3D structure of the complexes reveals sites of specific interactions. *Glycobiology* **2003,** *13* (2), 51-71.

(19) Kern, N. R.; Lee, H. S.; Wu, E. L.; Park, S.; Vanommeslaeghe, K.; MacKerell, A. D., Jr.; Klauda, J. B.; Jo, S.; Im, W., Lipid-linked oligosaccharides in membranes sample conformations that facilitate binding to oligosaccharyltransferase. *Biophys J* **2014,** *107* (8), 1885-1895.

(20) Valtersson, C.; van Duyn, G.; Verkleij, A. J.; Chojnacki, T.; de Kruijff, B.; Dallner, G., The influence of dolichol, dolichol esters, and dolichyl phosphate on phospholipid polymorphism and fluidity in model membranes. *J Biol Chem* **1985,** *260* (5), 2742-51.

(21) Wang, X.; Mansourian, A. R.; Quinn, P. J., The effect of dolichol on the structure and phase behaviour of phospholipid model membranes. *Mol Membr Biol* **2008,** *25* (6-7), 547-56.

(22) Janas, T.; Chojnacki, T.; Swiezewska, E.; Janas, T., The effect of undecaprenol on bilayer lipid membranes. *Acta Biochim Pol* **1994,** *41* (3), 351-8.

(23) Glover, K. J.; Weerapana, E.; Chen, M. M.; Imperiali, B., Direct biochemical evidence for the utilization of UDP-bacillosamine by PglC, an essential glycosyl-1-phosphate transferase in the Campylobacter jejuni N-linked glycosylation pathway. *Biochemistry* **2006,** *45* (16), 5343-50.

(24) Glover, K. J.; Weerapana, E.; Imperiali, B., In vitro assembly of the undecaprenylpyrophosphate-linked heptasaccharide for prokaryotic N-linked glycosylation. *Proc Natl Acad Sci U S A* **2005,** *102* (40), 14255-9.

(25) Scott, N. E.; Parker, B. L.; Connolly, A. M.; Paulech, J.; Edwards, A. V.; Crossett, B.; Falconer, L.; Kolarich, D.; Djordjevic, S. P.; Hojrup, P.; Packer, N. H.; Larsen, M. R.; Cordwell, S. J., Simultaneous glycan-peptide characterization using hydrophilic interaction chromatography and parallel fragmentation by CID, higher energy collisional dissociation, and electron transfer dissociation MS applied to the N-linked glycoproteome of Campylobacter jejuni. *Mol Cell Proteomics* **2011,** *10* (2), M000031-MCP201.

(26) Kelly, J.; Jarrell, H.; Millar, L.; Tessier, L.; Fiori, L. M.; Lau, P. C.; Allan, B.; Szymanski, C. M., Biosynthesis of the N-linked glycan in Campylobacter jejuni and addition onto protein through block transfer. *J Bacteriol* **2006,** *188* (7), 2427-34.

(27) Lu, Y. H.; Guan, Z.; Zhao, J.; Raetz, C. R., Three phosphatidylglycerol-phosphate phosphatases in the inner membrane of Escherichia coli. *J Biol Chem* **2011,** *286* (7), 5506-18.

(28) El Ghachi, M.; Derbise, A.; Bouhss, A.; Mengin-Lecreulx, D., Identification of multiple genes encoding membrane proteins with undecaprenyl pyrophosphate phosphatase (UppP) activity in Escherichia coli. *J Biol Chem* **2005,** *280* (19), 18689-95.

(29) Karlyshev, A. V.; Everest, P.; Linton, D.; Cawthraw, S.; Newell, D. G.; Wren, B. W., The Campylobacter jejuni general glycosylation system is important for attachment to human epithelial cells and in the colonization of chicks. *Microbiology* **2004,** *150* (Pt 6), 1957-64.

(30) Weerapana, E.; Imperiali, B., Asparagine-linked protein glycosylation: from eukaryotic to prokaryotic systems. *Glycobiology* **2006,** *16* (6), 91R-101R.

(31) Tytgat, H. L.; Lebeer, S., The sweet tooth of bacteria: common themes in bacterial glycoconjugates. *Microbiol Mol Biol Rev* **2014,** *78* (3), 372-417.

(32) Glover, K. J.; Weerapana, E.; Numao, S.; Imperiali, B., Chemoenzymatic synthesis of glycopeptides with PglB, a bacterial oligosaccharyl transferase from Campylobacter jejuni. *Chem Biol* **2005,** *12* (12), 1311-5.

(33) Lizak, C.; Gerber, S.; Numao, S.; Aebi, M.; Locher, K. P., X-ray structure of a bacterial oligosaccharyltransferase. *Nature* **2011,** *474* (7351), 350-5.

(34) Ray, L. C.; Das, D.; Entova, S.; Lukose, V.; Lynch, A. J.; Imperiali, B.; Allen, K. N., Membrane association of monotopic phosphoglycosyl transferase underpins function. *Nat Chem Biol* **2018,** *14* (6), 538-541.

(35) Lukose, V.; Walvoort, M. T. C.; Imperiali, B., Bacterial phosphoglycosyl transferases: initiators of glycan biosynthesis at the membrane interface. *Glycobiology* **2017,** *27* (9), 820-833.

(36) Anderson, M. S.; Eveland, S. S.; Price, N. P., Conserved cytoplasmic motifs that distinguish sub-groups of the polyprenol phosphate:N-acetylhexosamine-1-phosphate transferase family. *FEMS Microbiol Lett* **2000,** *191* (2), 169-75.

(37) Lukose, V.; Luo, L.; Kozakov, D.; Vajda, S.; Allen, K. N.; Imperiali, B., Conservation and Covariance in Small Bacterial Phosphoglycosyltransferases Identify the Functional Catalytic Core. *Biochemistry* **2015,** *54* (50), 7326-34.

(38) Entova, S.; Billod, J. M.; Swiecicki, J. M.; Martin-Santamaria, S.; Imperiali, B., Insights into the key determinants of membrane protein topology enable the identification of new monotopic folds. *Elife* **2018,** *7*.

(39) Bligh, E. G.; Dyer, W. J., A rapid method of total lipid extraction and purification. *Can J Biochem Physiol* **1959,** *37* (8), 911-7.

180

(40) Wang, X.; Ribeiro, A. A.; Guan, Z.; Raetz, C. R., Identification of undecaprenyl phosphate-beta-D-galactosamine in Francisella novicida and its function in lipid A modification. *Biochemistry* **2009,** *48* (6), 1162-72.

(41) Knowles, T. J.; Finka, R.; Smith, C.; Lin, Y. P.; Dafforn, T.; Overduin, M., Membrane proteins solubilized intact in lipid containing nanoparticles bounded by styrene maleic acid copolymer. *J Am Chem Soc* **2009,** *131* (22), 7484-5.

(42) Long, A. R.; O'Brien, C. C.; Malhotra, K.; Schwall, C. T.; Albert, A. D.; Watts, A.; Alder, N. N., A detergent-free strategy for the reconstitution of active enzyme complexes from native biological membranes into nanoscale discs. *BMC Biotechnol* **2013,** *13*, 41.

(43) Dorr, J. M.; Scheidelaar, S.; Koorengevel, M. C.; Dominguez, J. J.; Schafer, M.; van Walree, C. A.; Killian, J. A., The styrene-maleic acid copolymer: a versatile tool in membrane research. *Eur Biophys J* **2016,** *45* (1), 3-21.

(44) Dorr, J. M.; Koorengevel, M. C.; Schafer, M.; Prokofyev, A. V.; Scheidelaar, S.; van der Cruijsen, E. A.; Dafforn, T. R.; Baldus, M.; Killian, J. A., Detergent-free isolation, characterization, and functional reconstitution of a tetrameric K+ channel: the power of native nanodiscs. *Proc Natl Acad Sci U S A* **2014,** *111* (52), 18607-12.

(45) Prabudiansyah, I.; Kusters, I.; Caforio, A.; Driessen, A. J., Characterization of the annular lipid shell of the Sec translocon. *Biochim Biophys Acta* **2015,** *1848* (10 Pt A), 2050-6.

(46) Das, D.; Kuzmic, P.; Imperiali, B., Analysis of a dual domain phosphoglycosyl transferase reveals a ping-pong mechanism with a covalent enzyme intermediate. *Proc Natl Acad Sci U S A* **2017,** *114* (27), 7019-7024.

(47) Grinberg, A. V.; Gevondyan, N. M.; Grinberg, N. V.; Grinberg, V. Y., The thermal unfolding and domain structure of Na+/K+-exchanging ATPase. A scanning calorimetry study. *Eur J Biochem* **2001,** *268* (19), 5027-36.

(48) Vogel, R.; Siebert, F., Conformation and stability of alpha-helical membrane proteins. 2. Influence of pH and salts on stability and unfolding of rhodopsin. *Biochemistry* **2002,** *41* (11), 3536-45.

(49) Gandini, R.; Reichenbach, T.; Tan, T. C.; Divne, C., Structural basis for dolichylphosphate mannose biosynthesis. *Nat Commun* **2017,** *8* (1), 120.

(50) Denisov, I. G.; Grinkova, Y. V.; Lazarides, A. A.; Sligar, S. G., Directed self-assembly of monodisperse phospholipid bilayer Nanodiscs with controlled size. *J Am Chem Soc* **2004,** *126* (11), 3477-87.

(51) Nath, A.; Atkins, W. M.; Sligar, S. G., Applications of phospholipid bilayer nanodiscs in the study of membranes and membrane proteins. *Biochemistry* **2007,** *46* (8), 2059-69.

(52) Nasr, M. L.; Baptista, D.; Strauss, M.; Sun, Z. J.; Grigoriu, S.; Huser, S.; Pluckthun, A.; Hagn, F.; Walz, T.; Hogle, J. M.; Wagner, G., Covalently circularized nanodiscs for studying membrane proteins and viral entry. *Nat Methods* **2017,** *14* (1), 49-52.

(53) Hartley, M. D.; Schneggenburger, P. E.; Imperiali, B., Lipid bilayer nanodisc platform for investigating polyprenol-dependent enzyme interactions and activities. *Proc Natl Acad Sci U S A* **2013,** *110* (52), 20863-70.

(54) Hoff, B.; Strandberg, E.; Ulrich, A. S.; Tieleman, D. P.; Posten, C., 2H-NMR study and molecular dynamics simulation of the location, alignment, and mobility of pyrene in POPC bilayers. *Biophys J* **2005,** *88* (3), 1818-27.

(55) Aricha, B.; Fishov, I.; Cohen, Z.; Sikron, N.; Pesakhov, S.; Khozin-Goldberg, I.; Dagan, R.; Porat, N., Differences in membrane fluidity and fatty acid composition between phenotypic variants of Streptococcus pneumoniae. *J Bacteriol* **2004,** *186* (14), 4638-44.

(56) Saxena, R.; Shrivastava, S.; Chattopadhyay, A., Exploring the organization and dynamics of hippocampal membranes utilizing pyrene fluorescence. *J Phys Chem B* **2008,** *112* (38), 12134-8.

(57) Neidhardt, F. C., *Physiology of the Bacterial Cell: A Molecular Approach.* Sinauer Associates Inc.: 1990.

(58) Jorgenson, M. A.; Kannan, S.; Laubacher, M. E.; Young, K. D., Dead-end intermediates in the enterobacterial common antigen pathway induce morphological defects in Escherichia coli by competing for undecaprenyl phosphate. *Mol Microbiol* **2016,** *100* (1), 1-14.

(59) Jorgenson, M. A.; Young, K. D., Interrupting biosynthesis of O antigen or the lipopolysaccharide core produces morphological defects in Escherichia coli by sequestering undecaprenyl phosphate. *J Bacteriol* **2016,** *198* (22), 3070-3079.

(60) Leaver, J.; Hancock, I. C.; Baddiley, J., Fractionation studies of the enzyme complex involved in teichoic acid synthesis. *J Bacteriol* **1981,** *146* (3), 847-52.

(61) Anderson, R. G.; Hussey, H.; Baddiley, J., The mechanism of wall synthesis in bacteria. The organization of enzymes and isoprenoid phosphates in the membrane. *Biochem J* **1972,** *127* (1), 11-25.

(62) Guan, Z.; Breazeale, S. D.; Raetz, C. R., Extraction and identification by mass spectrometry of undecaprenyl diphosphate-MurNAc-pentapeptide-GlcNAc from Escherichia coli. *Anal Biochem* **2005,** *345* (2), 336-9.

(63) Zhou, G. P.; Troy, F. A., 2nd, NMR study of the preferred membrane orientation of polyisoprenols (dolichol) and the impact of their complex with polyisoprenyl recognition sequence peptides on membrane structure. *Glycobiology* **2005,** *15* (4), 347-59.

(64) Sezgin, E.; Levental, I.; Mayor, S.; Eggeling, C., The mystery of membrane organization: composition, regulation and roles of lipid rafts. *Nat Rev Mol Cell Biol* **2017,** *18* (6), 361-374.

(65) Piela, L.; Nemethy, G.; Scheraga, H. A., Proline-induced constraints in alpha-helices. *Biopolymers* **1987,** *26* (9), 1587-600.

(66) Chen, P. S. T., T.Y.; Warner, H, Microetermination of phosphorus. *Anal Chem* **1956,** *28,* 1756-1758.

(67) Kozlov, V. V.; Danilov, L. L., Reversed-phase ion-pair high-performance liquid chromatography assay of polyprenyl diphosphate oligomer homologues. *J Sep Sci* **2016**, *39* (3), 525-7.

(68) Studier, F. W., Protein production by auto-induction in high density shaking cultures. *Protein Expr Purif* **2005**, *41* (1), 207-34.

(69) Hartley, M. D.; Larkin, A.; Imperiali, B., Chemoenzymatic synthesis of polyprenyl phosphates. *Bioorg Med Chem* **2008**, *16* (9), 5149-56.

Chapter 5: Cell-free expression of PglC into model membrane Nanodiscs

INTRODUCTION

Cell-free expression, first described over 50 years ago as a method for studying mRNA translation,[1] has since become a valuable tool for the study of a wide variety of cellular components, including membrane proteins. Over-expression of membrane proteins *in vivo* for biochemical and structural studies is often hindered by the hydrophobic nature of these proteins, which can lead to aggregation, cell toxicity, poor solubility and low yield. Cell-free expression, by contrast, is particularly well suited for the expression of membrane proteins. By isolating the transcription and translation processes *in vitro*, cell-free expression circumvents all problems associated with cell toxicity and membrane transport, and also allows the introduction of detergents and lipids during expression to stabilize the protein product and reduce issues of aggregation and insolubility.

A typical *in vitro* expression reaction requires a cell extract that typically contains ribosomes, aminoacyl tRNA synthases, translation factors, residual lipids, and other important components of transcription and translation. This extract is supplemented with several additional components necessary for transcription and translation, including (but not limited to): a DNA template, T7 RNA polymerase, NTPs, tRNAs, amino acids, and an energy-regeneration system composed of two enzyme-substrate pairs: pyruvate kinase - phosphoenolpyruvate, and acetate kinase - acetylphosphate. Cell-free protein expression is commonly performed using cell extracts from *Escherichia coli,* yeast, wheat germ, insect cells or rabbit reticulocytes.[2] The study presented herein was performed using cell extract from *E. coli*, which is among the simplest and most cost-effective extracts to produce and suitable for the expression of diverse bacterial proteins.

Whereas early attempts at cell-free expression used a batch format, many current protocols utilize a "continuous exchange" methodology.[3-4] In continuous exchange cell-free expression, transcription and translation occur in a reaction chamber separated by a semipermeable membrane from a second, "feed" chamber containing additional nucleotides, amino acids, and energy sources. Influx of small molecules from the "feed" chamber, and export of inhibitory byproducts such as pyrophosphate, significantly improves the protein expression yield in the reaction chamber. Membrane proteins expressed by cell-free methods are expected to precipitate unless a solubilizing medium is provided. Cell-free expression of various membrane proteins into detergent micelles, liposomes, and lipid Nanodiscs has been described.[4-7]

Lipid Nanodiscs are small (8-20 nm in diameter) self-assembling structures that effectively mimic the hydrophobic environment of a lipid bilayer.[8] Each particle is composed of a planar phospholipid bilayer, the hydrophobic circumference of which is protected by a membrane scaffolding protein (MSP). Importantly, unlike liposomes and detergent micelles, Nanodiscs produce particles that are monodisperse and composed of true planar lipid bilayers. The original scaffolding protein described by the Sligar group,[9] MSP1D1, was derived from human apolipoprotein A-I, an important factor in fat efflux and HDL cholesterol transport in the human body. MSP1D1 is composed of a series of α-helices separated by proline kinks which allow the MSP to wrap around the circumference of the Nanodisc particle. The diameter of the Nanodisc can thus be modulated by augmenting or removing α-helices from the MSP.

Nanodiscs are self-assembling in solution and can be formed from MSP and lipids alone ("empty" Nanodiscs) or with a membrane protein, solubilizing the later in the Nanodisc lipid bilayer core. A variety of different lipids and membrane proteins can be easily incorporated into the Nanodisc bilayer, making Nanodiscs a flexible system for the study of membrane proteins.

This technology has been used previously to reconstitute and characterize a variety of detergent-purified membrane proteins, including several biologically relevant CYPs and GPCRs.[9-10] Additionally, Nanodiscs were previously used to investigate functional interactions between PglC and PglA from the N-linked protein glycosylation (pgl) pathway C. jejuni.[11] More recently, Nanodiscs have been applied to the co-translational solubilization of membrane proteins expressed in cell-free expression systems.[6-7] Translation directly into lipid Nanodiscs obviates the need to expose the membrane proteins to denaturants, and can result in a more native membrane-inserted fold.

The aim of this study was to develop an in-house continuous flow cell-free expression system for expression of PglC into lipid Nanodiscs, with the ultimate goal of applying such a methodology towards diverse biochemical and biophysical studies of PglC and the rest of the pgl pathway.

RESULTS

Preliminary cell-free expression of PglC using a commercially available kit

Initial small-scale trials of cell-free expression of PglC were performed using the commercially available Expressway Cell-Free E. coli Expression System (Invitrogen, Carlsbad, CA). Expression of the provided control construct, His_6-CALML3 (calmodulin-like protein 3), as well as His_6-PglC and PglC-His_6 variants was tested following the manufacturer's protocol. Expression of all three constructs was confirmed by Western blot (**Figure 5-1**); all three constructs were well-expressed.

189
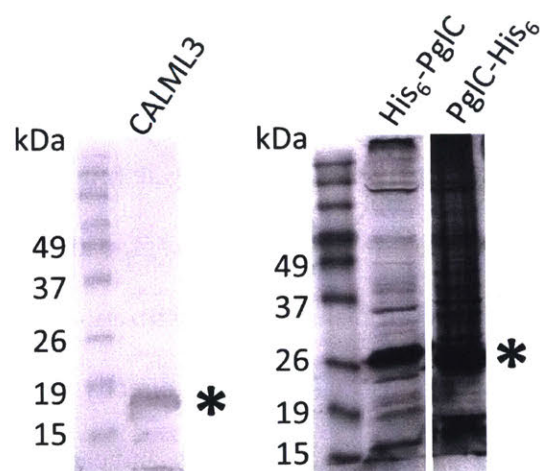
Figure 5-1 Expression of His$_6$-CALM3, His$_6$-PglC and PglC-His$_6$ using the Expressway system. Expression of all three constructs were confirmed via Western blot against the His$_6$-tag. Asterisks indicate expected molecular weights as follows: His$_6$-CALM3, 17 kDa; His$_6$-PglC, 27 kDa; PglC-His$_6$, 24 kDa.

As PglC is an integral membrane protein, translated PglC is not expected to remain soluble unless the cell-free expression reaction is supplemented with detergent or a membrane mimetic. Thus, a detergent screen was performed to identify a suitable detergent that would solubilize newly-translated PglC without interfering with the expression reaction. The following commonly used detergents were tested: n-dodecyl β-D-maltoside (DDM), Triton X-100 (TX-100), n-octyl β-D-glycoside (βOG), lauryl maltose neopentyl glycol (LMNG), and digitonin (dig). Each detergent was provided in excess of its critical micelle concentration (CMC). Following cell-free expression of PglC, expression reactions were clarified by centrifugation and the soluble and insoluble fractions were analyzed by Western blot (**Figure 5-2**). Many detergents (DDM, TX-100, LMNG) had little effect on the expression and solubility of PglC, while one (βOG) hindered expression. Digitonin was the only detergent tested that significantly increased the solubility of the expressed PglC.
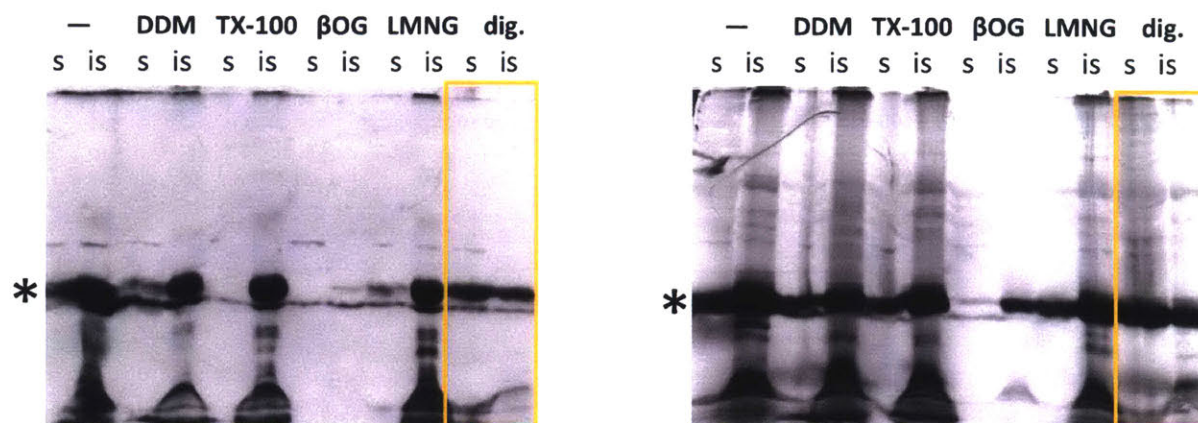
Figure 5-2 Detergent screen.

Five detergents were tested for solubilization of expressed His_6-PglC (*left*) and PglC-His_6 (*right*) variants via Western blot against the His_6-tag. Asterisk indicates expected molecular weight of PglC; s = soluble fraction, is = insoluble fraction. Detergent abbreviations as in main text; "–" = no detergent. Digitonin, outlined in yellow, solubilized expressed PglC most effectively.

PglC expression into lipid Nanodiscs

Next, cell-free expression of PglC into lipid Nanodiscs was tested. Expression directly into a membrane mimetic such as Nanodiscs is advantageous because it obviates the need to expose membrane proteins to detergent, and thus likely captures a more native structure and membrane topology of the target protein. Nanodiscs for cell-free expression were prepared using the MSP1D1ΔH5 scaffolding protein, reported to yield Nanodiscs 8 nm in diameter,[12] and a 3:1 mix of DMPC/DMPG lipids. Preparations were monodisperse and of the expected radius, as assessed by size-exclusion chromatography and dynamic light scattering (**Figure 5-3**).
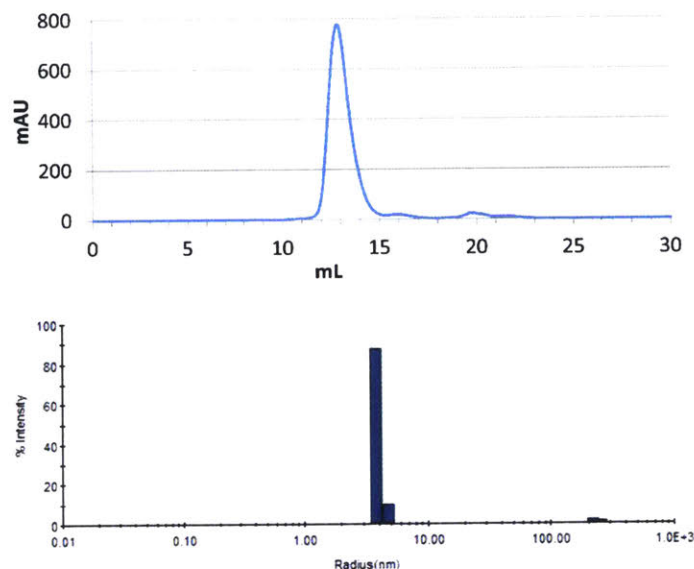
Figure 5-3 Preparation of lipid Nanodiscs for cell-free expression.
Analysis of prepared by size-exclusion chromatography (*top*) and dynamic light scattering
(*bottom*) indicated a monodisperse population with a diameter of ~8 nm.

When supplemented into PglC cell-free expression reaction mixtures, the prepared

Nanodiscs were able to solubilize expressed PglC (**Figure 5-4**). The efficiency of PglC

solubilization increased with the concentration of Nanodiscs provided: when Nanodiscs were

provided at 12.5 µM, PglC was solubilized at approximately the same efficiency as in the

presence of digitonin detergent, while in 25 µM Nanodiscs, solubilization was markedly

improved.
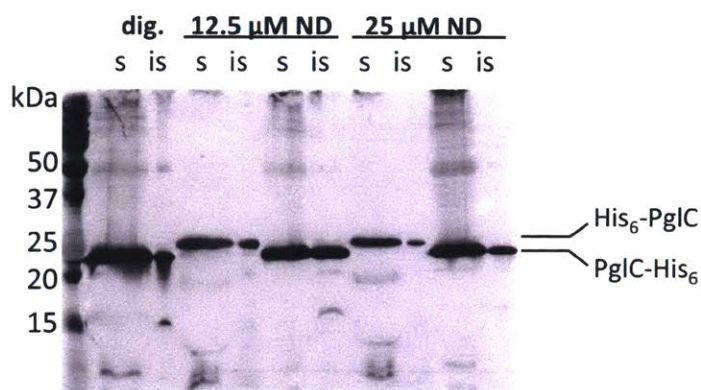
Figure 5-4 Cell-free expression of PglC into Nanodiscs.
Nanodiscs were provided to the cell-free reaction at 12.5 or 25 μM. Solubilization of both His_6-PglC and PglC-His_6 constructs was tested via Western blot against the His_6-tag. s = soluble fraction, is = insoluble fraction.

Nanodiscs carrying solubilized PglC were purified by Ni-NTA affinity chromatography and subjected to a radioactivity-based activity assay. Both the His_6-PglC and PglC-His_6 constructs, following cell-free expression and solubilization into Nanodiscs, were found to have comparable activity to a SUMO-PglC control expressed in *E. coli* and detergent-purified by standard biochemical methods (**Figure 5-5**). Notably, as the assay conditions contain Nanodisc-disrupting detergent, the activity measured does not reflect the activity of PglC when embedded in a Nanodisc, but rather reports on the proper folding of PglC when expressed and solubilized by these two very different methodologies.
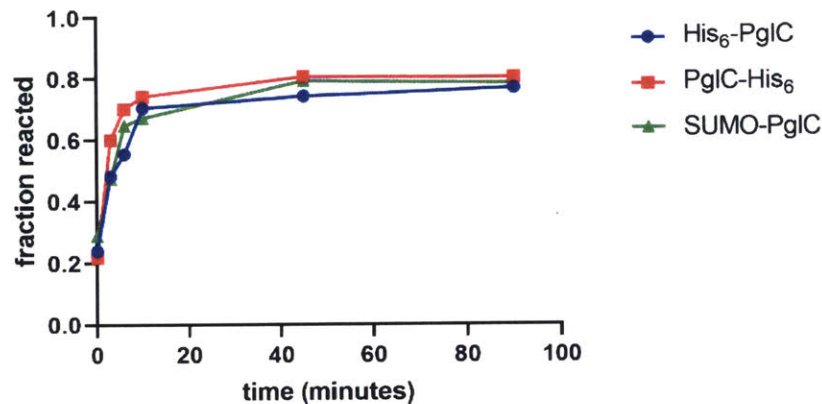
Figure 5-5 PglC that has been cell-free expressed into Nanodiscs is active. PglC-Nanodiscs purified from cell-free expression reactions performed in 25 μM Nanodiscs were found to contain active PglC.

Purification of T7 polymerase for large-scale cell-free expression

Once PglC constructs were shown to be active when expressed directly into Nanodiscs, production of in-house recombinant T7 RNA polymerase and *E. coli* cell extract, to allow for more cost-effective and large-scale expression, was undertaken. Notably, whereas the Expressway kit (Invitrogen) utilizes a batch method, the in-house cell-free expression components described herein allowed for continuous exchange cell-free expression, which often results in better yields.

Recombinant T7 polymerase was expressed from a commercially-available plasmid and purified by anion exchange as described previously[4-5] (**Figure 5-6**). Fractions containing T7 polymerase were pooled and assayed for transcriptase activity, which was quantified relative to a commercially available T7 RNA polymerase of known activity. An activity of 900 U/μL was

194

determined for the purified T7 polymerase. Previous reports indicated that purified T7

polymerase would be only ~50% pure;[4] based on the known specific activity of T7 polymerase

and the determined activity and concentration of our purified protein, we estimated the purified

T7 polymerase to be 36% pure.
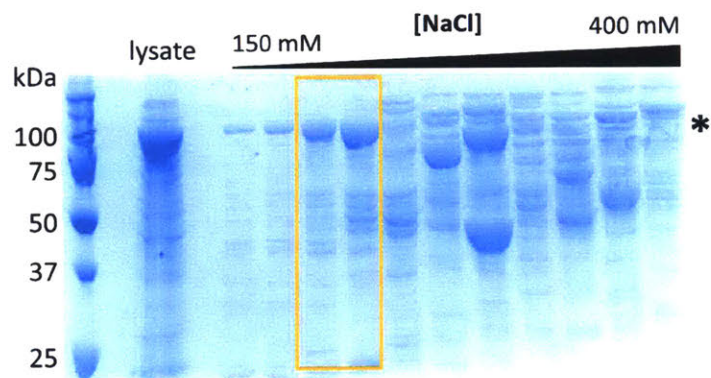
Figure 5-6 Purification of recombinant T7 RNA polymerase.
T7 RNA polymerase was purified by anion exchange according to previously published
protocols.[4] Fractions were analyzed by SDS-PAGE. Asterisk indicates the expected molecular
weight of T7 polymerase (99 kDa). Fractions found to contain T7 polymerase, such as those
outlined in yellow, were pooled.

Isolation of *E. coli* cell-free extract

Cell-free extract from BL21-CodonPlus (DE3)-RIL *E. coli* cells (Agilent Technologies, Santa Clara, CA) was isolated according to described methods.[4-5] Activity of the *E. coli* extract in conjunction with the purified T7 polymerase was confirmed using expression of the control soluble protein, CALML3. An optimization of Mg^{2+} and K^+ concentrations during cell-free expression was also performed using CALML3 (**Figure 5-7**). It was determined that 15 mM Mg^{2+} and 310 mM K^+ gave optimal activity.
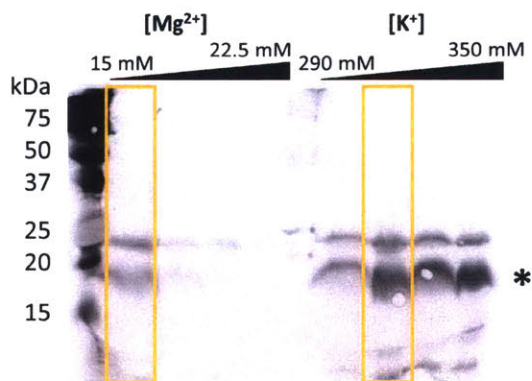
Figure 5-7 Cation optimization for cell-free expression using extract isolated from *E. coli*. *E. coli* extract isolation and optimization of Mg^{2+} and K^+ concentrations was performed according to previously published protocols[4] and analyzed by Western blot against the His_6-tag Asterisk indicates the expected molecular weight of CALML3 (16 kDa). Conditions giving maximum expression activity, outlined in yellow, were chosen.

Optimization of PglC expression into lipid Nanodiscs

The isolated T7 RNA polymerase and *E. coli* extract were next used for cell-free expression of His_6-PglC into lipid Nanodiscs. Preliminary trials of PglC expression into Nanodiscs had been performed previously using the commercially available Expressway kit (**Figure 5-4**). However, it was desirable to both confirm expression using our in-house system, and to optimize

conditions wherein sufficient Nanodiscs were available during expression to allow insertion of a single His$_6$-PglC per Nanodisc. Singular insertion was desirable for many potential down-stream applications, such as single molecule fluorescence and structure determination methods.

Thus, a titration of Nanodisc concentrations during cell-free expression of His$_6$-PglC was performed (**Figure 5-8**). PglC-Nanodiscs purified by Ni-NTA from such reactions were quantified by SDS-PAGE gel densitometry to determine the amounts of PglC and MSP1D1ΔH5 present: as each Nanodisc is known to be bound by two molecules of MSP1D1ΔH5,[12] this method allowed us to calculate how many molecules of His$_6$-PglC, on average, were solubilized per Nanodisc. It was found reproducibly that 60 µM of Nanodiscs were necessary in the cell-free expression reaction to ensure that a single molecule of His$_6$-PglC was solubilized per Nanodisc. Interestingly, the recovered PglC-Nanodiscs were found to account for only ~0.5% of the total Nanodiscs provided to cell-free expression reaction, indicating that most of the Nanodiscs do not receive a molecule of expressed His$_6$-PglC.
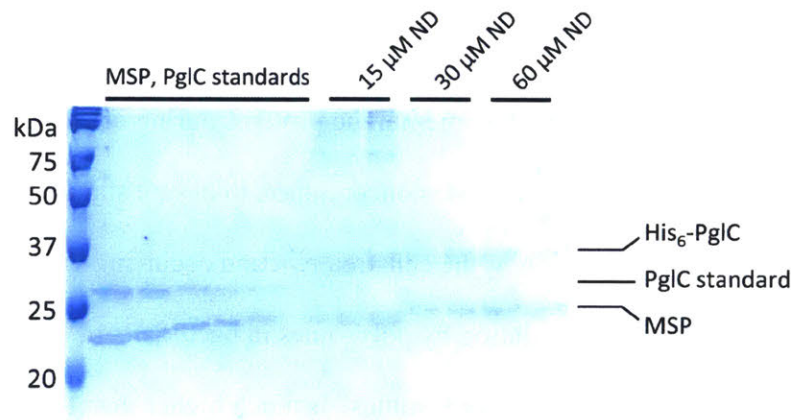
Figure 5-8 Nanodisc titration into cell-free expression of His$_6$-PglC.
PglC and MSP1D1ΔH5 concentrations in PglC-Nanodiscs were quantified using gel densitometry relative to known standards. An example is shown here. Note: the PglC standard used had been cleaved previously to remove the His$_6$ tag, and thus is of slightly lower molecular weight than the His$_6$-PglC produced by cell-free expression.

DISCUSSION

We have found that cell-free expression can be used to both generate detergent-solubilized PglC and to express PglC directly into lipid Nanodiscs. Expression is confirmed both with the commercially available Expressway kit and with T7 RNA polymerase and *E. coli* cell extract isolated in-house.

Cell-free expression of PglC directly into Nanodiscs

Cell-free expression of PglC directly into lipid Nanodiscs is a particularly useful tool, as it allows PglC to be incorporated directly into a model membrane environment without precipitation or the need for detergent solubilization, and is thus more likely to provide a faithful representation of the native PglC fold. We find that a relatively high concentration of lipid Nanodiscs, 60 μM, must be available during cell-free expression to ensure solubilization of a single molecule of PglC per Nanodisc. Unexpectedly, very few of these Nanodiscs (~0.5%) are found to receive expressed PglC; most of the Nanodiscs provided to such a reaction do not solubilize a molecule of PglC. Nevertheless, similar concentrations of Nanodiscs (10-100 μM) have been reported in other studies using Nanodiscs to solubilize membrane proteins during cell-free expression.[6-7] We hypothesize that the need for a high Nanodisc concentration to ensure singular insertion of PglC may be attributed to protein translation in the cell-free reaction occurring via multiple ribosomes along a single mRNA, similar to translation by polysomes in bacteria.[13] In this case, the local concentration of PglC surrounding such "polysomes" is much higher than the average concentration in the reaction, and a relatively high concentration of Nanodiscs is necessary to solubilize PglC into individual Nanodiscs.

Reaction yield

The yield for PglC solubilized into Nanodiscs was determined to be ~10 μg PglC per mL of reaction mix. However, significantly higher yields of one or more mg of protein per mL of reaction mix have been reported for cell-free expression of diverse membrane proteins.[5, 14-15] Although it is possible that some loss in yield is due to precipitation of PglC, the high concentration of available Nanodiscs during expression makes this unlikely. In addition, we note that unlike previous reports,[4-5] we were not able to grow the cultures used for isolating extract in a fermenter; thus, the volume of cells harvested was significantly lower than that reported elsewhere, and the extract we isolated is likely to be more dilute. The yield of cell-free expressed PglC could almost certainly be improved by systematic optimization of the various reaction components. While some optimization with regard to Mg^{2+} and K^+ concentration is described herein, this is only one of many parameters; optimization of many additional expression parameters, including, but not limited to, DNA template construction and concertation, protein tags, and reaction container set-up, have all been reported to lead to significant increases in yield.[6, 14-15]

CONCLUSIONS

The initial set-up of an in-house cell-free expression system for co-translational solubilization of PglC into Nanodiscs has been described. The use of such a system enables many applications of interest to the study of PglC both in isolation and in complex with downstream enzymes of the pgl pathway. For example, cell-free expression is well suited for use with methods for non-canonical amino acid incorporation as a strategy for site-specific protein labeling[16-18] and subsequent single-molecule fluorescence studies. This methodology could be applied to the labeling and co-solubilization of multiple enzymes of the pgl pathway into Nanodiscs to support the study of multi-enzyme complex formation in the membrane. Nanodisc-solubilized membrane proteins are also amenable to structure-determination studies, particularly using cryo-EM.[19-20]

EXPERIMENTAL PROCEDURES

Cloning of PglC into expression vectors

PglC was cloned into the pEXP5-NT/TOPO and pEXP5-CT/TOPO vectors, provided with the Expressway system (Invitrogen) to add an N- or C-terminal His$_6$-tags, respectively. PglC from *Campylobacter concisus*, a more tractable homolog that was later also used for x-ray crystallography structure determination, was used. The primers used for cloning are listed in **Table 5-1**.

Table 5-1 – Primers used for TOPO cloning of PglC	
PglC_Cc_TOPO F	ATGTATAGAAATTTTTTAAAGAGAGTGATCGACATTTTGGGAGCTTTAT
PglC_Cc_TOPO_CT_R	GTTTTTGCCATTAAATTTCTCCGTCGTCGCCTGCCC
PglC_Cc_TOPO_NT_R	TTAGTTTTTGCCATTAAATTTCTCCGTCGTCGCCTGCCC

Note: TOPO cloning makes use of a topoisomerase enzyme, attached covalently to the linear vector DNA, and the 3'-A overhang added to the insert by Taq polymerase during PCR amplification; thus, cloning proceeds without the need for restriction enzyme digest or DNA ligase. The manufacturer's recommended protocol was followed for cloning.

Cell-free expression optimization using the Expressway system

Cell-free expression using the Expressway system was performed following the manufacturer's instructions for a 50 μL scale reaction. Reactions proceeded for 30 min at 30 °C prior to the addition of 50 μL feeding buffer, and for an additional 4-6 hr at 30 °C after.

For Western blot analysis, 5 μL of reaction mix was precipitated in acetone and resuspended in SDS-PAGE loading dye. Proteins were separated by SDS-PAGE on 12% acrylamide gels and transferred to nitrocellulose membranes for Western blot analysis against the His$_6$-tag. Blots were blocked with 3% bovine serum albumin in TBST (10 mM Tris, 150 mM

NaCl, 0.05% Tween-20, pH 7.5) for 1 hr. The primary antibody, mouse anti-His (LifeTein, cat. # LT0426) was applied in a 1:3,000 dilution for 1 hr. The secondary antibody, goat anti-mouse AP (Invitrogen, cat. # 31328) was applied in a 1:10,000 dilution for 1 hr. Immunoblots were developed using 1-Step NBT/BCIP Substrate Solution (ThermoFisher Scientific, Waltham, MA, cat. # 34042) and imaged using a Molecular Imager Gel Doc XR+ System (BioRad, Hercules, CA).

For detergent screening, expression reactions were performed at one-half the scale described above. Detergents from stock solutions were added during reaction set-up to the following final concentrations: DDM, 0.2%; TX-100, 0.3%; βOG, 1%; LMNG, 0.02%; digitonin, 0.5%. After completion of the cell-free expression reaction as described above, reactions were centrifuged at 14,000 x g for 10 minutes. The supernatant ("soluble fraction") was removed; the pellet ("insoluble fraction") was washed once with buffer (50 mM HEPES, 150 mM NaCl, pH 7.5). Western blot analysis was performed as above.

In-house T7 polymerase expression and purification

T7 RNA polymerase expression and purification was performed following previously published protocols.[4-5] The plasmid pAR1219, carrying T7 RNA polymerase under an IPTG-inducible promoter, was purchased from Sigma-Aldrich. T7 polymerase was expressed in BL21-CodonPlus (DE3)-RIL *E. coli* cells (Agilent Technologies) in LB media supplemented with 50 μg/mL carbenicillin. Cultures were grown with shaking at 37 °C to an OD600 of 0.6-0.8 before induction with 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG, Chem-Impex, Wood Dale, IL). Following induction for an additional 5 hr at 37 °C, cells were harvested at 3,700 x g for 30 min and kept at -80 °C until purification.

Cells from 4 L of culture were resuspended in 120 mL of T7 buffer (30 mM Tris, 50 mM

NaCl, 10 mM EDTA, 5% glycerol, 10 mM β-mercaptoethanol, pH 8.0) and lysed by sonication

(Vibra-Cell, 50% amplitude, 1 sec ON – 1 sec OFF, 2 X 1 min; Sonics, Newtown, CT). Lysate

was centrifuged at 22,000 x g for 30 min at 4 °C. The supernatant (~120 mL) was supplemented

with 30 mL of 10% streptomycin sulfate, added dropwise with gentle stirring at 4 °C, for a final

concentration of 2% streptomycin sulfate. Precipitated DNA was removed by centrifugation at

30,000 x g for 30 min at 4 °C. The supernatant was filtered through a 0.45 μm filter and loaded on

a 20 mL Q HiTrap anion exchange column (GE Healthcare, Chicago, IL) pre-equilibrated with T7

buffer. The column was washed with one column volume of T7 buffer. T7 polymerase was eluted

using a gradient from 50-500 mM NaCl in T7 buffer over 15 column volumes. Fractions were

analyzed by SDS-PAGE (**Figure 5-6**); T7 polymerase was found to elute at ~200-240 mM NaCl.

Fractions containing T7 polymerase were pooled and dialyzed against 10 mM Tris, 10 mM NaCl,

5% glycerol, 1 mM EDTA, 1 mM DTT, pH 8.0 at 4 °C. The dialyzed sample was supplemented

to a final glycerol concentration of 50% and aliquoted for storage at -80 °C.

The concentration of the protein was determined to be 2.1 mg/mL by BCA assay

(ThermoFisher Scientific). Based on this protein concentration, the measured activity of 900 U/μL

(described below), and the known specific activity (1,200,000 U/mg), the purity of T7 RNA

polymerase was estimated to be (900 U/μL / 1,200,000 U/mg) / 2.1 mg/mL = 36%.

T7 polymerase activity assay

The purified T7 RNA polymerase was assayed against a commercially available polymerase

(New England Biolabs, Ipswich, MA) with a reported activity of 50 U/μL. Dilutions of purified

or commercial T7 polymerase were assayed at 37 °C in transcription buffer (40 mM Tris, 12 mM

MgCl$_2$, 5 mM DTT, 1 mM spermidine, pH 8.0) with 40-50 ng of a linearized vector template

containing a T7 promotor (i.e. pETAcrA vector linearized with XhoI restriction enzyme), 20 mM

each of ATP, CTP, GTP and UTP, and 0.1 U/μL RiboLock RNAse inhibitor (ThermoFisher

Scientific) in a final volume of 25-30 μL. Transcription was quenched after 2 hr by the addition

of 2 mM EDTA. Transcription reaction products were visualized on agarose gels with ethidium

bromide staining and quantified by gel densitometry using a Molecular Imager Gel Doc XR+

System with Image Lab software (BioRad) and compared between dilutions of purified and

commercial T7 polymerase; in this manner, the purified T7 polymerase was determined to have

an activity of ~900 U/μL.

In-house isolation of _E. coli_ cell-free extract

Isolation of cell-free extract from BL21-CodonPlus (DE3)-RIL _E. coli_ cells (Agilent

Technologies) was performed following previously published protocols.[4-5] Note: previously

described protocols recommended culture growth in a 10-L fermenter, but as one was not readily

available, cultures were grown in Fernbach flasks instead, and the purification protocol scaled

accordingly as much as possible. Cells were grown in 6 x 1 L cultures in Fernbach flasks in 2 x

YPTG broth (for 6 L, combine: 17.9 g KH$_2$PO$_4$ and 54.8 K$_2$HPO$_4$ in 2 L water, autoclaved; 60 g

yeast extract, 30 g NaCl and 96 g tryptone in 3 L water, autoclaved; and 119 glucose in 1 L

water, filter-sterilized). Cultures were grown at 37 °C with shaking and harvested during log

phase (OD600 1.3-1.6), harvested at 3,700 x _g_ for 30 min and kept at -80 °C until purification.

Cell were resuspended in 35 mL S30-B buffer (10 mM Tris-acetate, 14 mM Mg^{2+} acetate,

0.6 mM KCl, 1 mM DTT, 0.1 mM PMFS, pH 8.2) and lysed using a microfluidizer at 22,000 psi.

Lysate was centrifuged at 30,000 x _g_ for 30 min at 4 °C. The upper 2/3 of the supernatant was

transferred to a new centrifuge tube and centrifuged a second time. The upper 2/3 of the

supernatant after the second centrifugation step was collected. A 5 M NaCl stock was used to

adjust the extract to a final concentration of 400 mM NaCl. The extract was incubated in a 42 °C

waterbath for 45 min, then dialyzed twice at 4 °C against S30-C buffer (10 mM Tris-acetate, 14

mM Mg^{2+} acetate, 0.6 mM K^+ acetate, 0.5 mM DTT, pH 8.2). The dialyzed extract was

centrifuged at 30,000 x g for 30 min at 4 °C. The upper 2/3 of the supernatant was collected and

concentrated ~1.5-fold using a 10 kDa MWCO centrifugal filter, aliquoted and flash-frozen for

storage at -80 °C.

Cell-free expression using prepared cell-free extract

Cell-free expression was performed following previously published protocols.[4-5] All reaction

components for a sample reaction are shown in (**Figure 5-9**). Digitonin detergent or lipid

Nanodiscs were included in the reaction as necessary. The reaction components were combined

and added to a small dialysis bag, which was placed in a tissue culture dish with a feed buffer

containing extra nucleotides, amino acids and other small molecule components. A schematic

representation of the reaction set-up is shown in (**Figure 5-10**). The reaction proceeded with

gentle shaking for 6 hr at room temperature. A needle and syringe were used to pierce the

dialysis bag and remove the reaction and any expressed protein.

Master mix	uL added	stock	final
NaN₃ (%)	10.7	10	0.1
PEG 8000 (%)	107.0	40	2.0
KOAc (mM)	91.4	4000	170.8
Mg(OAc)₂ (mM)	21.6	1000	10.1
HEPES (mM)	89.2	2.4	0.1
protease inhibitor (-fold)	42.8	50	1.0
Folinic acid (mg/mL)	21.4	10	0.1
DTT (mM)	8.6	500	2.0
NTP-mix (-fold)	28.5	75	1.0
PEP (mM)	42.8	1000	20.0
AcP (mM)	42.8	1000	20.0
AA-mix (mM)	267.5	4	0.5
RCWMDE-mix (mM)	128.1	16.7	1.0
Final Volume:	902.4		

Reaction			
Master mix	59.2		
Pyruvate kinase (mg/mL)	0.6	10	0.04
Plasmid template (mg/mL)	4.2	0.5	0.015
RiboLock (U/uL)	1.1	40	0.3
T7RNAP (U/uL)	0.9	900	6
tRNA (mg/mL)	1.8	40	0.5
E. coli extract (x)	49.0	1	0.35
Milli-Q water	23.3		
Final Volume:	140		

Feed			
Master mix	842		
AA-mix	250	4	1
S30-C buffer (x)	14	50	0.35
DTT (mM)	0.7	500	0.175
Milli-Q water	894		
Final Volume:	2,000		

Figure 5-9 Reaction components for cell-free expression using isolated *E. coli* extract. A sample reaction using 50 μL cell-free extract is shown. NTP-mix = 90 mM ATP, 60 mM GTP, CTP, UTP; PEP = phospho(enol)pyruvate, AcP = acetylphosphate; AA-mix = 4 mM of each of the 20 amino acids; RCWMDE-mix = 16.7 mM each Arg, Cys, Trp, Met, Asp, Glu. RiboLock = RiboLock RNAse inhibitor (ThermoFisher Scientific), tRNA = *E. coli* tRNA (Roche, Penzberg, Germany).
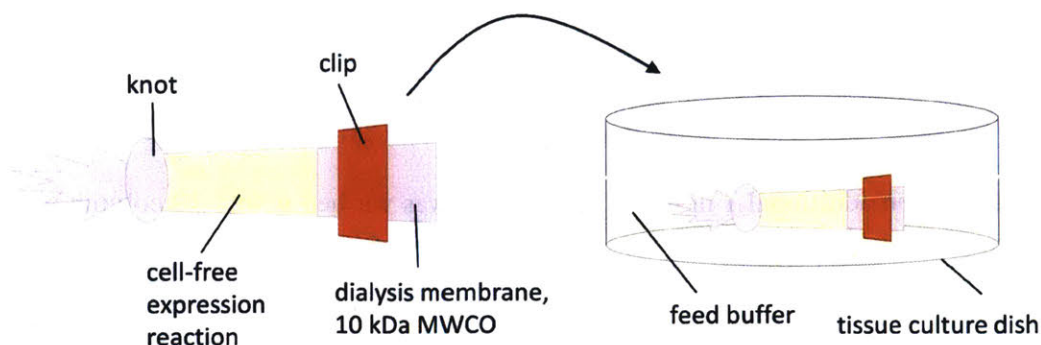
Figure 5-10 Reaction set-up for cell-free expression.
Cell-free expression occurred in a dialysis bag placed in a tissue culture dish with feed buffer.

Purification of MSP1D1ΔH5

The pET28a-MSP1D1ΔH5 construct was obtained from the Gerhard Wagner lab and is also available to the community (Addgene, Watertown, MA). MSP1D1ΔH5 was expressed in BL21-CodonPlus (DE3)-RIL *E. coli* cells (Agilent Technologies). Overnight cultures were used to inoculate 0.5 L Terrific Broth medium (Research Products International, Mount Prospect, IL) containing kanamycin and chloramphenicol (30 µg/mL each). Cells were grown at 37 °C to an OD of 0.9-1.2 before induction with 1 mM IPTG. Following induction for an additional 3-4 hr at 37 °C, cells were harvested at 3,700 x *g* for 30 min and kept at -80 °C until purification.

Cells were resuspended in 10 mL / 1 g cell pellet of lysis buffer (50 mM HEPES, 500 mM NaCl, 1% Triton X-100, pH 8.0), supplemented with 0.5 mg/mL lysozyme (Research Products International), 1:1000 dilution of EDTA-free protease inhibitor cocktail (EMD Millipore, Burlington, MA) and 1 unit/mL DNase I (New England Biolabs). Cells were lysed by sonication (Vibra-Cell, 50% amplitude, 1 sec ON – 1 sec OFF, 2 X 1 min; Sonics). Lysate was centrifuged at 30,000 x *g* for 45 min at 4 °C.

The supernatant was incubated with Ni-NTA resin (4 mL per L original culture volume; ThermoFisher Scientific) pre-treated with lysis buffer, for 1 hr at 4 °C. Flow-through from this incubation was applied to a second batch of Ni-NTA resin (2 mL per L original culture volume) and incubated for an additional 1 hr at 4 °C. The resin was washed with 5-10 column volumes of lysis buffer, followed by 5-10 column volumes of cholate buffer (50 mM HEPES, 500 mM NaCl, 50 mM sodium cholate, pH 8.0), 5-10 column volumes of non-detergent buffer (50 mM HEPES, 500 mM NaCl, pH 8.0), and 5-10 column volumes of wash buffer (50 mM HEPES, 500 mM NaCl, 20 mM imidazole, pH 8.0). Protein was eluted in 3 column volumes of elution buffer (50 mM HEPES, 500 mM NaCl, 500 mM imidazole, pH 8.0). The eluate was supplemented with TEV protease (75-100 nmol protease per 1 L original culture volume) and dialyzed against 25 mM Tris, 50 mM NaCl, 5 mM dithiotreitol, pH 8.0 for 4-5 hr at room temperature. Fresh dialysis buffer was provided and dialysis was repeated overnight at 4 °C.

The following day, dialyzed MSP1D1ΔH5 was supplemented with 0.5 M Tris, pH 8.0 and 5 M NaCl to increase the final NaCl concentration to 500 mM. Protein was incubated with Ni-NTA (4 mL per L original culture volume) pre-treated with detergent-free buffer (50 mM Tris, 500 mM NaCl, pH 8.0), for 1 hr at 4 °C. The column was washed with 2-3 column volumes of the same buffer. Optionally, TEV protease and impurities were eluted in 2-3 column volumes of elution buffer for quality-control purposes. The flow-through and washes, containing TEV-cleaved MSP1D1ΔH5, were concentrated using a 10 kDa MWCO centrifugal filter and dialyzed against Nanodisc buffer (50 mM HEPES, 150 mM NaCl, pH 7.5), overnight at 4 °C. Dialyzed protein was aliquoted and flash frozen for storage at -80 °C.

Nanodisc formation

The following protocol was used for formation of "empty" (not containing a solubilized membrane protein) Nanodiscs. Lipid chloroform stocks were quantified by a modification of the ammonium molybdate method for microdetermination of total phosphate described previously[21] against a phosphate standard solution (Sigma-Aldrich). Lipids from chloroform stocks were measured into Eppendorf tubes and dried first under N_2 and then under high vacuum overnight.

Nanodiscs assembly was typically performed on a 300-500 µL scale with 10-100 µM MSP1D1ΔH5 and 3:1 (mol/mol) DMPC:DMPG lipids (Avanti Polar Lipids, Alabaster, AL). The optimal MSP:lipid molar ratio was determined empirically to be 1:60. Nanodiscs were assembled in Eppendorf tubes in Nanodisc buffer (50 mM HEPES, 150 mM NaCl, pH 7.5) supplemented with 25 mM sodium cholate. Assembly reactions were rotated at room temperature for 20 min. Meanwhile, adsorbent BioBeads SM-2 (Bio-Rad) were prepared by washing with 4 x 1 volume of MeOH followed by 4 x 10 volumes of water. BioBeads (50-100 µL) were added to the Nanodisc assembly reaction and the Nanodisc-BioBead mix was rotated at room temperature for at least 4 hr, or overnight. Nanodiscs were removed from the Eppendorf tube by piercing the bottom of the tube with a small-gauge needle and applying centrifugal force.

Ni-NTA purification of PglC-Nanodiscs from cell-free expression

Following completion of cell-free expression, the reaction was incubated with Ni-NTA resin (one-eighth- to one-quarter the reaction volume), pre-equilibrated with Nanodisc buffer, for 1 hr at room temperature. The resin was washed with 10 column volumes wash buffer (Nanodisc buffer + 40 mM imidazole). PglC-Nanodiscs were eluted in 4 column volumes of elution buffer (Nanodisc buffer + 500 mM imidazole). For gel densitometry analysis, aliquots (10-20 µL) of

purified PglC-Nanodiscs and standards of MSP1D1ΔH5 and PglC were analyzed by SDS-PAGE

with Coomassie staining and quantified by gel densitometry using a Molecular Imager Gel Doc

XR+ System with Image Lab software (BioRad).

Radioactivity-based PglC activity assay

The radioactivity extraction-based assay for PglC activity has been described in detail

previously.[22] Assays contained 20 µM Und-P, 3% DMSO, 0.2% Triton X-100, 30 mM Tris (pH

8.0), 10 µM [³H]UDP-diNAcBac (5.4 mCi/mmol), and 1 nM SUMO-PglC or PglC-Nanodiscs in

a final volume of 100 µL. After initiation of the reaction with [³H]UDP-diNAcBac, aliquots (20

µL) were taken at 3, 6, 10, 45 and 90 min time points, quenched into 2 mL 2:1 (v/v)

$CHCl_3$:MeOH and washed with 3 x 400 µL of Pure Solvent Upper Phase (for 500 mL combine:

235 ml of H_2O, 1.83 g of KCl, 240 ml MeOH, 15 ml $CHCl_3$). The resulting aqueous layers were

combined with 5 mL of EcoLite liquid scintillation cocktail (MP Biomedicals, Santa Ana, MA).

Organic layers were combined with 5 mL of OptiFluor (PerkinElmer, Waltham, MA). Both

layers were analyzed on an LS 6500 Scintillation Counter (Beckman Coulter, Brea, CA).

REFERENCES

(1) Nirenberg, M. W.; Matthaei, J. H., The dependence of cell-free protein synthesis in E. coli upon naturally occurring or synthetic polyribonucleotides. *Proc Natl Acad Sci U S A* **1961**, *47*, 1588-602.

(2) Carlson, E. D.; Gan, R.; Hodgman, C. E.; Jewett, M. C., Cell-free protein synthesis: applications come of age. *Biotechnol Adv* **2012**, *30* (5), 1185-94.

(3) Spirin, A. S.; Baranov, V. I.; Ryabova, L. A.; Ovodov, S. Y.; Alakhov, Y. B., A continuous cell-free translation system capable of producing polypeptides in high yield. *Science* **1988**, *242* (4882), 1162-4.

(4) Schwarz, D.; Junge, F.; Durst, F.; Frolich, N.; Schneider, B.; Reckel, S.; Sobhanifar, S.; Dotsch, V.; Bernhard, F., Preparative scale expression of membrane proteins in Escherichia coli-based continuous exchange cell-free systems. *Nat Protoc* **2007**, *2* (11), 2945-57.

(5) Schneider, B.; Junge, F.; Shirokov, V. A.; Durst, F.; Schwarz, D.; Dotsch, V.; Bernhard, F., Membrane protein expression in cell-free systems. *Methods Mol Biol* **2010**, *601*, 165-86.

(6) Roos, C.; Kai, L.; Haberstock, S.; Proverbio, D.; Ghoshdastider, U.; Ma, Y.; Filipek, S.; Wang, X.; Dotsch, V.; Bernhard, F., High-level cell-free production of membrane proteins with nanodiscs. *Methods Mol Biol* **2014**, *1118*, 109-30.

(7) Henrich, E.; Peetz, O.; Hein, C.; Laguerre, A.; Hoffmann, B.; Hoffmann, J.; Dotsch, V.; Bernhard, F.; Morgner, N., Analyzing native membrane protein assembly in nanodiscs by combined non-covalent mass spectrometry and synthetic biology. *Elife* **2017**, *6*.

(8) Denisov, I. G.; Grinkova, Y. V.; Lazarides, A. A.; Sligar, S. G., Directed self-assembly of monodisperse phospholipid bilayer Nanodiscs with controlled size. *J Am Chem Soc* **2004**, *126* (11), 3477-87.

(9) Nath, A.; Atkins, W. M.; Sligar, S. G., Applications of phospholipid bilayer nanodiscs in the study of membranes and membrane proteins. *Biochemistry* **2007**, *46* (8), 2059-69.

(10) D'Antona, A. M.; Xie, G.; Sligar, S. G.; Oprian, D. D., Assembly of an activated rhodopsin-transducin complex in nanoscale lipid bilayers. *Biochemistry* **2014**, *53* (1), 127-34.

(11) Hartley, M. D.; Schneggenburger, P. E.; Imperiali, B., Lipid bilayer nanodisc platform for investigating polyprenol-dependent enzyme interactions and activities. *Proc Natl Acad Sci U S A* **2013**, *110* (52), 20863-70.

(12) Hagn, F.; Etzkorn, M.; Raschle, T.; Wagner, G., Optimized phospholipid bilayer nanodiscs facilitate high-resolution structure determination of membrane proteins. *J Am Chem Soc* **2013**, *135* (5), 1919-25.

(13) Miller, O. L., Jr.; Hamkalo, B. A.; Thomas, C. A., Jr., Visualization of bacterial genes in action. *Science* **1970**, *169* (3943), 392-5.

(14) Haberstock, S.; Roos, C.; Hoevels, Y.; Dotsch, V.; Schnapp, G.; Pautsch, A.; Bernhard, F., A systematic approach to increase the efficiency of membrane protein production in cell-free expression systems. *Protein Expr Purif* **2012,** *82* (2), 308-16.

(15) Klammt, C.; Lohr, F.; Schafer, B.; Haase, W.; Dotsch, V.; Ruterjans, H.; Glaubitz, C.; Bernhard, F., High level cell-free expression and specific labeling of integral membrane proteins. *Eur J Biochem* **2004,** *271* (3), 568-80.

(16) Hong, S. H.; Kwon, Y. C.; Jewett, M. C., Non-standard amino acid incorporation into proteins using Escherichia coli cell-free protein synthesis. *Front Chem* **2014,** *2*, 34.

(17) Ozawa, K.; Loscha, K. V.; Kuppan, K. V.; Loh, C. T.; Dixon, N. E.; Otting, G., High-yield cell-free protein synthesis for site-specific incorporation of unnatural amino acids at two sites. *Biochem Biophys Res Commun* **2012,** *418* (4), 652-6.

(18) Chemla, Y.; Ozer, E.; Schlesinger, O.; Noireaux, V.; Alfonta, L., Genetically expanded cell-free protein synthesis using endogenous pyrrolysyl orthogonal translation system. *Biotechnol Bioeng* **2015,** *112* (8), 1663-72.

(19) Wild, R.; Kowal, J.; Eyring, J.; Ngwa, E. M.; Aebi, M.; Locher, K. P., Structure of the yeast oligosaccharyltransferase complex gives insight into eukaryotic N-glycosylation. *Science* **2018,** *359* (6375), 545-550.

(20) Gao, Y.; Cao, E.; Julius, D.; Cheng, Y., TRPV1 structures in nanodiscs reveal mechanisms of ligand and lipid action. *Nature* **2016,** *534* (7607), 347-51.

(21) Chen, P. S. T., T.Y.; Warner, H, Microetermination of phosphorus. *Anal Chem* **1956,** *28*, 1756-1758.

(22) Lukose, V.; Luo, L. Q.; Kozakov, D.; Vajda, S.; Allen, K. N.; Imperiali, B., Conservation and covariance in small bacterial phosphoglycosyltransferases identify the functional catalytic core. *Biochemistry* **2015,** *54* (50), 7326-7334.

Appendix

Covariance analysis of HofN, LpxL, LpxM, LpxP and YihG suggested likely contacts between the hydrophobic and globular domains, indicating a reentrant topology. Key interactions are highlighted in red (>90% probability) and orange (80–90% probability). Only contacts involving the first ~50 residues are shown.

Analysis was performed using the GREMLIN web-server (http://gremlin.bakerlab.org/).

Column labels:
i, j = the ith and jth sequence positions
i_id, j_id = the residues at the ith and jth positions in the submitted sequence
r_sco = the predicted score for interaction between the ith and jth positions
s_sco = the scaled score for interaction between the ith and jth positions
prob = probability interaction between the ith and jth positions

HofN							LpxL						
i	j	i_id	j_id	r_sco	s_sco	prob	i	j	i_id	j_id	r_sco	s_sco	prob
4	7	4_P	7_F	0.1074	1.004	0.926	10	13	10_A	13_H	0.1674	0.971	0.908
4	8	4_P	8_L	0.0885	0.828	0.788	11	16	11_L	16_Y	0.1606	0.931	0.883
5	95	5_I	95_W	0.0953	0.892	0.852	11	14	11_L	14_P	0.14	0.812	0.769
5	88	5_I	88_Q	0.0851	0.796	0.749	11	15	11_L	15_R	0.1268	0.736	0.664
6	10	6_N	10_W	0.1105	1.033	0.938	11	305	11_L	305_R	0.0983	0.57	0.387
6	9	6_N	9_P	0.0997	0.932	0.883	12	16	12_L	16_Y	0.1452	0.842	0.803
7	92	7_F	92_T	0.0833	0.779	0.726	12	17	12_L	17_W	0.1068	0.619	0.469
9	126	9_P	126_G	0.0882	0.825	0.784	13	16	13_H	16_Y	0.0998	0.579	0.402
10	14	10_W	14_R	0.1282	1.199	0.979	14	17	14_P	17_W	0.219	1.27	0.987
10	15	10_W	15_R	0.1064	0.995	0.921	14	20	14_P	20_W	0.1079	0.625	0.479
11	21	11_R	21_F	0.1304	1.22	0.982	17	20	17_W	20_W	0.1719	0.997	0.922
11	17	11_R	17_A	0.0997	0.933	0.884	18	158	18_L	158_N	0.1082	0.627	0.483
14	18	14_R	18_F	0.1379	1.29	0.989	19	23	19_T	23_I	0.1176	0.682	0.576
15	21	15_R	21_F	0.0845	0.791	0.742	19	160	19_T	160_L	0.0977	0.566	0.38
15	57	15_R	57_A	0.0777	0.727	0.65	20	23	20_W	23_I	0.1177	0.683	0.578
17	20	17_A	20_R	0.1303	1.219	0.982	20	24	20_W	24_G	0.1085	0.629	0.486
17	47	17_A	47_E	0.0861	0.805	0.76	22	26	22_G	26_L	0.0983	0.57	0.387
17	87	17_A	87_R	0.0803	0.751	0.686	23	164	23_I	164_L	0.195	1.131	0.967
18	22	18_F	22_W	0.0878	0.821	0.78	23	27	23_I	27_W	0.1678	0.973	0.909
18	112	18_F	112_L	0.0786	0.735	0.662	24	27	24_G	27_W	0.1057	0.613	0.459
20	39	20_R	39_I	0.0819	0.766	0.708	24	28	24_G	28_L	0.1052	0.61	0.454
22	26	22_W	26_F	0.097	0.908	0.865	26	94	26_L	94_G	0.164	0.951	0.896

23	26	23_L	26_F	0.0917	0.858	0.82	27	31	27_W	31_Q	0.2555	1.482	0.997
24	36	24_L	36_I	0.0845	0.791	0.742	27	30	27_W	30_V	0.1301	0.754	0.691
27	68	27_V	68_T	0.0824	0.771	0.715	28	32	28_L	32_L	0.1069	0.62	0.471
28	**63**	**28_A**	**63_R**	**0.1394**	**1.304**	**0.99**	29	91	29_V	91_M	0.0996	0.578	0.4
28	31	28_A	31_L	0.1198	1.121	0.965	30	98	30_V	98_F	0.1157	0.671	0.558
28	**83**	**28_A**	**83_S**	**0.0979**	**0.916**	**0.872**	30	33	30_V	33_P	0.0985	0.571	0.388
28	77	28_A	77_Q	0.0839	0.784	0.733	31	36	31_Q	36_V	0.1523	0.883	0.844
30	**110**	**30_L**	**110_A**	**0.1074**	**1.004**	**0.926**	32	40	32_L	40_L	0.1509	0.875	0.837
30	37	30_L	37_T	0.1063	0.994	0.921	32	36	32_L	36_V	0.1246	0.722	0.642
30	38	30_L	38_L	0.1002	0.937	0.887	32	35	32_L	35_P	0.1218	0.706	0.616
30	**56**	**30_L**	**56_Q**	**0.0999**	**0.935**	**0.886**	33	36	33_P	36_V	0.364	2.111	1
31	34	31_L	34_V	0.1324	1.239	0.984	34	38	34_Y	38_Y	0.1669	0.968	0.907
31	35	31_L	35_G	0.103	0.964	0.904	35	39	35_P	39_R	0.2243	1.301	0.99
32	36	32_L	36_I	0.094	0.88	0.841	35	38	35_P	38_Y	0.1568	0.91	0.867
34	**55**	**34_V**	**55_L**	**0.1353**	**1.265**	**0.987**	36	39	36_V	39_R	0.1615	0.937	0.887
35	**48**	**35_G**	**48_A**	**0.1017**	**0.952**	**0.897**	**37**	**91**	**37_I**	**91_M**	**0.2148**	**1.246**	**0.985**
35	90	35_G	90_Q	0.0894	0.836	0.797	**37**	**95**	**37_I**	**95_M**	**0.1525**	**0.885**	**0.846**
35	164	35_G	164_Q	0.0825	0.772	0.716	**38**	**88**	**38_Y**	**88_M**	**0.176**	**1.021**	**0.933**
36	40	36_I	40_L	0.1532	1.433	0.996	38	42	38_Y	42_C	0.1249	0.724	0.645
36	**170**	**36_I**	**170_T**	**0.0912**	**0.853**	**0.815**	38	41	38_Y	41_G	0.1112	0.645	0.514
38	41	38_L	41_R	0.1137	1.064	0.949	39	43	39_R	43_G	0.2507	1.454	0.996
38	42	38_L	42_L	0.107	1.001	0.924	39	42	39_R	42_C	0.2302	1.335	0.992
38	46	38_L	46_A	0.0953	0.891	0.851	40	43	40_L	43_G	0.1348	0.782	0.73
39	43	39_I	43_T	0.112	1.048	0.944	40	91	40_L	91_M	0.1011	0.586	0.413
39	42	39_I	42_L	0.0907	0.849	0.811	41	87	41_G	87_G	0.3817	2.214	1
39	137	39_I	137_L	0.0849	0.794	0.746	41	45	41_G	45_G	0.1611	0.934	0.885
40	44	40_L	44_G	0.1382	1.293	0.989	42	84	42_C	84_E	0.2721	1.578	0.998
42	45	42_L	45_S	0.1062	0.993	0.92	42	88	42_C	88_M	0.2451	1.422	0.995
42	143	42_L	143_Q	0.0903	0.845	0.807	42	46	42_C	46_K	0.0998	0.579	0.402
43	102	43_T	102_L	0.0849	0.794	0.746	43	47	43_G	47_L	0.2181	1.265	0.987
44	72	44_G	72_L	0.0807	0.755	0.692	43	46	43_G	46_K	0.1517	0.88	0.841
45	53	45_S	53_V	0.0878	0.821	0.78	43	84	43_G	84_E	0.099	0.574	0.393
45	59	45_S	59_Q	0.0875	0.819	0.777	44	87	44_L	87_G	0.1311	0.76	0.699
45	83	45_S	83_S	0.086	0.804	0.759	45	87	45_G	87_G	0.1286	0.746	0.679
45	54	45_S	54_L	0.0858	0.802	0.756	45	49	45_G	49_L	0.125	0.725	0.647
46	114	46_A	114_T	0.0829	0.775	0.721	45	50	45_G	50_R	0.1145	0.664	0.546
46	62	46_A	62_A	0.0793	0.742	0.673	45	84	45_G	84_E	0.1121	0.65	0.522
47	50	47_E	50_I	0.0804	0.752	0.688	46	84	46_K	84_E	0.4006	2.324	1
48	51	48_A	51_D	0.1102	1.031	0.937	46	80	46_K	80_V	0.1674	0.971	0.908
49	53	49_R	53_V	0.0818	0.765	0.706	46	49	46_K	49_L	0.1133	0.657	0.534
50	55	50_I	55_L	0.1507	1.409	0.995	46	50	46_K	50_R	0.1072	0.622	0.474
51	55	51_D	55_L	0.1057	0.989	0.918	46	77	46_K	77_K	0.0979	0.568	0.384

i	j	i_id	j_id	r_sco	s_sco	prob	i	j	i_id	j_id	r_sco	s_sco	prob
51	56	51_D	56_Q	0.0939	0.878	0.839	47	51	47_L	51_F	0.1632	0.947	0.894
53	57	53_V	57_A	0.1309	1.224	0.983	49	80	49_L	80_V	0.2394	1.388	0.994
53	56	53_V	56_Q	0.1069	1	0.924	49	56	49_L	56_A	0.2066	1.198	0.979
54	57	54_L	57_A	0.0904	0.846	0.808	49	79	49_L	79_V	0.1846	1.07	0.951
54	58	54_L	58_E	0.0892	0.835	0.796	49	76	49_L	76_R	0.1368	0.793	0.745
54	92	54_L	92_T	0.0846	0.791	0.742	49	60	49_L	60_H	0.1092	0.633	0.493
55	83	55_L	83_S	0.1055	0.987	0.917	50	80	50_R	80_V	0.1581	0.917	0.872
55	58	55_L	58_E	0.0873	0.817	0.775	50	76	50_R	76_R	0.1346	0.781	0.729

LpxM							LpxP						
i	j	i_id	j_id	r_sco	s_sco	prob	i	j	i_id	j_id	r_sco	s_sco	prob
17	20	17_S	20_H	0.1649	0.97	0.908	10	13	10_E	13_H	0.1724	0.991	0.919
18	23	18_F	23_Y	0.1635	0.961	0.902	11	16	11_F	16_Y	0.1662	0.955	0.899
18	21	18_F	21_P	0.1467	0.862	0.824	11	14	11_F	14_P	0.1335	0.767	0.709
18	22	18_F	22_R	0.1211	0.712	0.626	11	15	11_F	15_R	0.1272	0.731	0.656
18	24	18_F	24_W	0.1161	0.683	0.578	11	304	11_F	304_L	0.1047	0.602	0.44
19	23	19_R	23_Y	0.1467	0.862	0.824	12	16	12_L	16_Y	0.1509	0.867	0.829
19	24	19_R	24_W	0.1271	0.747	0.68	12	17	12_L	17_W	0.1071	0.615	0.462
19	27	19_R	27_W	0.1017	0.598	0.433	13	16	13_H	16_Y	0.0988	0.568	0.384
21	24	21_P	24_W	0.2372	1.394	0.995	14	17	14_P	17_W	0.218	1.253	0.986
21	27	21_P	27_W	0.1155	0.679	0.571	14	20	14_P	20_W	0.1036	0.595	0.428
21	25	21_P	25_G	0.1154	0.678	0.57	17	20	17_W	20_W	0.1737	0.999	0.923
22	27	22_R	27_W	0.1023	0.602	0.44	18	158	18_L	158_N	0.1061	0.61	0.454
23	27	23_Y	27_W	0.117	0.688	0.587	19	23	19_T	23_L	0.1188	0.683	0.578
24	27	24_W	27_W	0.1697	0.997	0.922	19	164	19_T	164_V	0.1003	0.576	0.397
27	31	27_W	31_A	0.0958	0.563	0.375	20	23	20_W	23_L	0.1134	0.651	0.524
28	31	28_L	31_A	0.0996	0.586	0.413	20	24	20_W	24_G	0.1061	0.61	0.454
29	33	29_G	33_M	0.0988	0.581	0.405	21	24	21_F	24_G	0.0977	0.561	0.372
30	**172**	**30_V**	**172_V**	**0.1951**	**1.147**	**0.971**	22	26	22_G	26_L	0.0982	0.565	0.379
30	34	30_V	34_A	0.1568	0.922	0.876	**23**	**164**	**23_L**	**164_V**	**0.1999**	**1.149**	**0.971**
31	34	31_A	34_A	0.1007	0.592	0.423	23	27	23_L	27_W	0.1661	0.955	0.899
33	**101**	**33_M**	**101_A**	**0.1594**	**0.937**	**0.887**	24	28	24_G	28_L	0.1051	0.604	0.443
34	38	34_A	38_L	0.2537	1.491	0.997	24	27	24_G	27_W	0.1009	0.58	0.403
34	37	34_A	37_A	0.1296	0.762	0.702	**26**	**94**	**26_L**	**94_G**	**0.1645**	**0.945**	**0.892**
35	39	35_G	39_T	0.1047	0.616	0.464	27	31	27_W	31_Q	0.2615	1.503	0.997
36	98	36_I	98_A	0.0964	0.567	0.382	27	30	27_W	30_V	0.1337	0.769	0.712
37	105	37_A	105_I	0.1094	0.643	0.51	28	32	28_L	32_L	0.1129	0.649	0.52
37	40	37_A	40_P	0.097	0.57	0.387	29	91	29_W	91_V	0.1019	0.586	0.413
38	43	38_L	43_F	0.1528	0.898	0.857	30	98	30_V	98_F	0.1148	0.66	0.539
38	67	38_L	67_L	0.0996	0.586	0.413	30	33	30_V	33_P	0.1017	0.585	0.412
39	47	39_T	47_I	0.1577	0.927	0.88	31	36	31_Q	36_V	0.15	0.862	0.824

39	43	39_T	43_F	0.1213	0.713	0.627	32	40	32_L	40_L	0.1586	0.912	0.869
39	42	39_T	42_K	0.115	0.676	0.566	32	36	32_L	36_V	0.1311	0.754	0.691
40	43	40_P	43_F	0.3679	2.163	1	32	35	32_L	35_P	0.1222	0.702	0.61
41	45	41_P	45_D	0.1636	0.962	0.903	33	36	33_P	36_V	0.3682	2.116	1
41	102	41_P	102_E	0.0959	0.564	0.377	34	38	34_Y	38_C	0.1629	0.936	0.886
42	46	42_K	46_P	0.2039	1.198	0.979	35	39	35_P	39_F	0.2087	1.2	0.979
42	45	42_K	45_D	0.1536	0.903	0.861	35	38	35_P	38_C	0.158	0.908	0.865
43	46	43_F	46_P	0.1604	0.943	0.891	36	39	36_V	39_F	0.1561	0.897	0.856
44	98	44_R	98_A	0.2152	1.265	0.987	**37**	**91**	**37_L**	**91_V**	**0.2127**	**1.223**	**0.982**
44	102	44_R	102_E	0.1443	0.848	0.81	**37**	**95**	**37_L**	**95_M**	**0.1508**	**0.867**	**0.829**
45	95	45_D	95_Q	0.1823	1.072	0.952	**38**	**88**	**38_C**	**88_M**	**0.1789**	**1.028**	**0.936**
45	49	45_D	49_A	0.1207	0.71	0.623	38	42	38_C	42_T	0.125	0.718	0.635
45	48	45_D	48_L	0.1048	0.616	0.464	38	41	38_C	41_G	0.1082	0.622	0.474
46	50	46_P	50_R	0.2387	1.403	0.995	39	43	39_F	43_R	0.2502	1.438	0.996
46	49	46_P	49_A	0.2217	1.303	0.99	39	42	39_F	42_T	0.2268	1.303	0.99
47	50	47_I	50_R	0.1326	0.78	0.727	40	43	40_L	43_R	0.1275	0.733	0.659
47	98	47_I	98_A	0.1008	0.593	0.425	**41**	**87**	**41_G**	**87_G**	**0.3799**	**2.184**	**1**
48	94	48_L	94_P	0.3788	2.227	1	41	45	41_G	45_G	0.1604	0.922	0.876
48	52	48_L	52_G	0.1619	0.952	0.897	42	84	42_T	84_R	0.2764	1.589	0.999
49	91	49_A	91_A	0.2714	1.596	0.999	42	88	42_T	88_M	0.2433	1.399	0.995
49	95	49_A	95_Q	0.2345	1.379	0.994	42	46	42_T	46_A	0.099	0.569	0.385
49	53	49_A	53_R	0.0958	0.563	0.375	43	47	43_R	47_M	0.2061	1.185	0.977
50	54	50_R	54_F	0.2036	1.197	0.979	43	46	43_R	46_A	0.1532	0.88	0.841
50	53	50_R	53_R	0.1465	0.861	0.823	43	84	43_R	84_R	0.1111	0.639	0.503
50	91	50_R	91_A	0.1013	0.596	0.43	44	87	44_I	87_G	0.1324	0.761	0.701
51	94	51_L	94_P	0.132	0.776	0.722	45	87	45_G	87_G	0.1286	0.739	0.668
52	56	52_G	56_G	0.1264	0.743	0.674	45	49	45_G	49_R	0.1281	0.737	0.665
52	94	52_G	94_P	0.1258	0.74	0.67	45	50	45_G	50_P	0.1135	0.652	0.526
52	91	52_G	91_A	0.1158	0.681	0.575	45	84	45_G	84_R	0.1094	0.629	0.486
53	91	53_R	91_A	0.3902	2.294	1	46	84	46_A	84_R	0.399	2.293	1
53	87	53_R	87_D	0.1705	1.002	0.925	46	80	46_A	80_A	0.1798	1.034	0.939
53	56	53_R	56_G	0.1126	0.662	0.543	46	49	46_A	49_R	0.1198	0.689	0.588
54	58	54_F	58_L	0.1805	1.061	0.948	46	77	46_A	77_K	0.1005	0.578	0.4
56	87	56_G	87_D	0.232	1.364	0.993	46	50	46_A	50_P	0.098	0.563	0.375
56	63	56_G	63_R	0.2143	1.26	0.986	47	51	47_M	51_F	0.1645	0.946	0.893
56	86	56_G	86_V	0.183	1.076	0.953	49	80	49_R	80_A	0.2346	1.348	0.992
56	83	56_G	83_E	0.1392	0.818	0.776	49	56	49_R	56_E	0.2121	1.219	0.982
56	67	56_G	67_L	0.1181	0.694	0.596	49	79	49_R	79_I	0.188	1.081	0.955
56	90	56_G	90_F	0.1021	0.6	0.437	49	76	49_R	76_E	0.1414	0.812	0.769
57	87	57_R	87_D	0.1647	0.968	0.907	49	60	49_R	60_R	0.1136	0.653	0.527
57	83	57_R	83_E	0.1397	0.821	0.78	50	80	50_P	80_A	0.1611	0.926	0.879
59	62	59_G	62_S	0.1274	0.749	0.683	50	76	50_P	76_E	0.1396	0.802	0.756

YihG						
i	j	i_id	j_id	r_sco	s_sco	prob
12	15	12_R	15_A	0.0597	0.9	0.609
12	16	12_R	16_A	0.0525	0.791	0.465
13	16	13_I	16_A	0.0735	1.109	0.827
13	188	13_I	188_F	0.049	0.738	0.396
13	51	13_I	51_V	0.0478	0.72	0.373
14	21	14_L	21_L	0.0693	1.045	0.772
14	18	14_L	18_T	0.0691	1.042	0.769
14	80	14_L	80_H	0.0541	0.816	0.498
15	137	15_A	137_L	0.0587	0.884	0.589
15	20	15_A	20_L	0.052	0.785	0.457
16	185	16_A	185_I	0.0589	0.888	0.594
16	155	16_A	155_A	0.0486	0.733	0.389
17	**42**	**17_I**	**42_I**	**0.0817**	**1.232**	**0.902**
17	20	17_I	20_L	0.0662	0.999	0.726
17	21	17_I	21_L	0.0583	0.878	0.581
17	55	17_I	55_K	0.055	0.829	0.516
17	27	17_I	27_I	0.0527	0.794	0.469
17	28	17_I	28_L	0.0524	0.789	0.462
17	59	17_I	59_F	0.0523	0.788	0.461
17	197	17_I	197_E	0.048	0.724	0.378
18	228	18_T	228_F	0.0608	0.917	0.631
19	**247**	**19_L**	**247_F**	**0.126**	**1.9**	**0.997**
19	**251**	**19_L**	**251_S**	**0.0734**	**1.106**	**0.824**
19	41	19_L	41_G	0.0549	0.827	0.513
19	22	19_L	22_S	0.0529	0.798	0.474
20	23	20_L	23_I	0.0974	1.469	0.97
20	286	20_L	286_Q	0.0498	0.751	0.412
20	38	20_L	38_I	0.0484	0.73	0.385
20	45	20_L	45_L	0.0479	0.722	0.375
21	46	21_L	46_L	0.0515	0.777	0.446
21	229	21_L	229_D	0.0484	0.73	0.385
22	27	22_S	27_I	0.0743	1.12	0.835
22	74	22_S	74_L	0.0569	0.857	0.553
22	25	22_S	25_L	0.0555	0.837	0.527
22	117	22_S	117_R	0.052	0.785	0.457
23	159	23_I	159_R	0.0529	0.798	0.474
23	38	23_I	38_I	0.0478	0.72	0.373
24	31	24_V	31_I	0.0573	0.864	0.562
24	291	24_V	291_L	0.0489	0.737	0.394

25	140	25_L	140_A	0.0714	1.076	0.8
26	74	26_T	74_L	0.0791	1.193	0.882
26	30	26_T	30_T	0.0693	1.044	0.771
27	260	27_I	260_H	0.0511	0.77	0.437
29	71	29_V	71_L	0.0581	0.876	0.578
31	35	31_I	35_V	0.105	1.583	0.984
31	46	31_I	46_L	0.05	0.754	0.416
33	66	33_C	66_C	0.0854	1.288	0.926
33	63	33_C	63_M	0.0619	0.933	0.65
33	40	33_C	40_A	0.057	0.86	0.557
33	70	33_C	70_G	0.051	0.77	0.437
33	43	33_C	43_V	0.0495	0.747	0.407
34	38	34_S	38_I	0.0786	1.185	0.878
34	37	34_S	37_I	0.0478	0.72	0.373
35	38	35_V	38_I	0.0526	0.793	0.468
35	154	35_V	154_R	0.0516	0.778	0.448
37	41	37_I	41_G	0.0585	0.883	0.587
37	63	37_I	63_M	0.0498	0.751	0.412
39	51	39_I	51_V	0.052	0.784	0.456
40	59	40_A	59_F	0.069	1.04	0.767
40	162	40_A	162_E	0.0568	0.857	0.553
40	233	40_A	233_N	0.0561	0.845	0.537
40	56	40_A	56_V	0.0485	0.732	0.388
42	47	42_I	47_L	0.0498	0.751	0.412
43	52	43_V	52_I	0.0626	0.944	0.664
44	48	44_K	48_P	0.0486	0.733	0.389
44	57	44_K	57_S	0.0479	0.722	0.375
44	252	44_K	252_G	0.0478	0.721	0.374
46	301	46_L	301_S	0.0495	0.747	0.407
47	261	47_L	261_V	0.0596	0.899	0.608
47	52	47_L	52_I	0.049	0.739	0.397
49	52	49_V	52_I	0.0967	1.459	0.969
50	54	50_P	54_R	0.0884	1.333	0.941
50	180	50_P	180_L	0.0619	0.933	0.65