DISS. ETH NO. 25304

Structural basis of glycan transfer by bacterial oligosaccharyltransferase PglB

A thesis submitted to attain the degree of

DOCTOR OF SCIENCES of ETH ZURICH
(Dr. sc. ETH Zurich)

presented by

MAJA HANNA NAPIÓRKOWSKA

M.Sc. Medical Biotechnology, University of Warsaw

born on 18.04.1989
citizen of Poland

accepted on the recommendation of

Prof. Dr. Kaspar Locher
Prof. Dr. Markus Aebi
Prof. Dr. Jean-Louis Reymond

2018

Table of Contents

Abbreviations

Å	Angström
A_{600}	Absorbance measured at 600nm
ATP	Adenosine triphosphate
ADP	Adenosine diphosphate
Bac	Bacillosamine
Bicine	N,N-Bis(2-hydorxyethyl)glycine
bp	Base pair(s)
BSA	Bovine serum albumin
CDR	Complementarity determining region
CMC	Critical micelle concentration
Cymal5	5-Cyclohexyl-1-Pentyl-β-D-Maltoside
Cymal6	6-Cyclohexyl-1-Hexyl-β-D-Maltoside
Dab	2,4-diaminobutanoic acid
DDM	*n*-Dodecyl-β-D-Maltoside
DM	*n*-Decyl-β-D-Maltopyranoside
DMF	*N,N*-dimethylformamide
DNA	Deoxyribonucleic acid
DTT	Dithiothreitol
EDTA	Ethylenediaminetetraacetic acid
EL	External loop
ELISA	Enzyme-linked immunosorbent assay
EM	Electron microscopy
ER	Endoplasmic reticulum
Fab	Fragment, antigen binding
FR	Framework region
Fv	Variable fragment
GalNAc	*N*-acetyl-D-galactosamine
Glc	glucose
GlcNAc	*N*-acetyl-D-glucosamine
GT	Glycosyltransferase
His-Tag	Histidine affinity tag
HRP	horseradish peroxidase
IgG	Immunoglobulin G
IPTG	Isopropyl-β-D-1-thiogalactopyranoside
k_{cat}	turnover number
K_d	dissociation constant
kDa	kilodalton
K_M	Michaelis-Menten constant
LB	Lysogeny Broth
LDAO	Lauryl-DimethylAmine-*N*-Oxide
LMNG	Lauryl-Maltose-Neopentyl-Glycol
LLO	Lipid-linked oligosaccharide
Man	Mannose
MES	2-(*N*-morpholino)ethanesulfonic acid
MST	Microscale- Thermophoresis
MW	Molecular weight
Nb	Nanobody

NDM	*n*-Nonyl-β-D-Maltopyranoside
NG	*n*-Nonyl-β-D-Glucopyranoside
NiNTA	Nickel-nitrilotriacetic acid
NMR	Nuclear magnetic resonance
OD	Optical density
OG	*n*-Octyl-β-D-Glucopyranoside
OGNG	Octyl-Glucose-Neopentyl-Glycol
OST	Oligosaccharyltransferase
PAGE	Polyacrylamide gel electrophoresis
PDB	Protein Data Bank
PBS	Phosphate-buffered saline
PEG	Polyethylene glycol
PC	Phosphocholine
PCR	Polymerase chain reaction
PP	Pyrophosphate
RNA	Ribonucleic acid
RT	Room temperature
RTC	Ribosome-translocon complex
SDS	Sodium dodecyl sulfate
SEC	Size-exclusion chromatography
SLS	Swiss Light Source
TB	Teriffic broth
TBE	Tris/Borate/EDTA
TM	Transmembrane
Tris	Tris(hydroxymethyl)aminomethane
Ts	Thermostability
UDM	*n*-Undecyl-β-D-Maltopyranoside
VH	Variable domain of heavy chain
VHH	Variable domain of heavy-chain antibodies
VL	Variable domain of light chain
WT	Wild type
YT	Yeast/tryptone

Summary

N-linked protein glycosylation is one of the most abundant post-translational modifications and results in the covalent attachment of an oligosaccharide onto asparagine residues of proteins. While it is essential in eukaryotes, being implicated in many biological processes such as protein folding, trafficking and self-non-self interactions, it was also found to be important in prokaryotes. In the bacterial foodborne pathogen *Campylobacter jejuni*, *N*-glycans were proven to facilitate host adhesion and colonization. The key enzyme of the *N*-glycosylation pathway is oligosaccharyltransferase (OST) that catalyzes the *en bloc* transfer of the glycan from lipid linked oligosaccharide (LLO) onto acceptor asparagines located within the conserved sequon N-X-S/T. Despite some differences among bacteria, archaea and eukaryotes, the glycosylation mechanism is thought to be conserved. Single-subunit OST from *C. jejuni*, the PglB protein, is a well-studied, *bona fide* model system that allows the reaction mechanism of the OST-catalyzed process to be studied. The first structure of OST, peptide-bound PglB from *C. lari*, was published in 2011. This structure, combined with functional studies, revealed the mechanism of sequon recognition. However, owing to the absence of structural insight, the interactions of OST with LLO and the transfer mechanism remain poorly understood. A higher resolution structure would be required to identify the residues and interactions involved in the activation of the amide nitrogen of the acceptor asparagine.

In this thesis, I combined structural and functional approaches to study the mechanism of glycan transfer catalyzed by *C. lari* PglB. In chapters 2 and 3 I describe two attempts to facilitate PglB crystallization. The first approach was to generate conformational nanobodies to improve lattice contacts by increasing the polar surface area for protein-protein contacts and restrict the flexibility of mobile parts of the PglB, such as extracellular loop 5 (EL5). Despite an exceptionally good immune response from the alpaca in the heavy-chain class, the binders did not improve protein crystallization. Nevertheless, six nanobodies were thermostabilizing and may therefore be useful for future cryo-EM and NMR studies. The second approach was to engineer PglB to stabilize it by inserting an additional glycosylation sequon (named the Inseq constructs). Binding of the inserted sequon to the peptide binding site stabilizes the protein and increases the local concentration of the acceptor peptide during crystallization trials. The best Inseq construct, Inseq 4, showed much higher stability and crystallization efforts resulted in crystals that diffracted to 4.5 Å. This may provide a basis to obtain higher resolution structures of binary and ternary complexes of PglB in the future.

In order to get insight into the interactions between OST, LLO, peptide and the divalent metal ion, and ultimately understand the molecular mechanism of glycan transfer, I pursued functional and structural studies of PglB ternary complexes. Two different approaches to trap ternary complexes are described in detail in chapters 4 and 5. The main challenge in trapping a ternary-complex intermediate of OST was to prevent catalysis without compromising the binding affinity of the peptide and LLO substrates. In this thesis I screened different inhibitory and reactive peptide and LLO analogs, and identified the best candidates for crystallization experiments. I subsequently co-crystallized and determined the structure of two important intermediates of the glycosylation reaction, the first ternary complexes of any OST. By combining this structural insight with chemo-enzymatic approaches and further functional studies, I could propose a glycosylation mechanism. This work not only gives insight into structural changes that may be required for the function of PglB, but for all OSTs, as the glycosylation mechanism is likely to be broadly conserved in all domains of life. Although the results give detailed mechanistic insight, they cannot clarify the exact mechanism of activation of the asparagine carboxamide group. They do however provide a basis for studying ternary complexes of OST and give a promising starting point for computational approaches aimed at exploring the glycan transfer step in molecular detail.

Zusammenfassung

N-Glykosylierung ist eine der am weitesten verbreiteten post-transnationalen Modifikationen und resultiert in der kovalenten Bindung eines Oligosaccharids an einer Asparagin-Seitenkette von Proteinen. Während dieser Prozess in Eukaryoten essenziell ist, da er in vielen biologischen Prozessen, wie der Proteinfaltung, dem Proteintransport sowie der Unterscheidung zwischen körpereigenen und körperfremden Proteinen involviert ist, wurde dieser Prozess auch für Prokaryoten als wichtig befunden. Es wurde gezeigt, dass in dem bakteriellen Lebensmittelpathogen *Campylobacter jejuni*, N-Glykane die Adhäsion am Wirt, sowie dessen Kolonisierung vereinfachen. Die Schlüsselstelle in dem enzymatischen Stoffwechselweg der N-Glykosylierung ist die Oligosaccharyltransferase (OST), welche den en bloc Transfer des Glykans vom lipidgebundenen Oligosaccharid (LLO) auf die Akzeptor-Asparagin-Seitenkette, welche sich im konservierten Sequon N-X-S/T befindet, katalysiert. Obwohl es einige Unterschiede zwischen Bakterien, Archaea, und Eukaryoten gibt, geht man davon aus, dass der Mechanismus der N-Glykosylierung konserviert ist. Die nur aus einer Untereinheit bestehende *C. jejuni* OST PglB stellt ein weitgehend erforschtes, angemessenes Modellsystem dar, welches es erlaubt den OST-katalysierten Reaktionsmechanismus zu studieren. Im Jahr 2011 wurde die erste OST-Struktur in Form von Peptidgebundener PglB von *C. lari*, publiziert. In Kombination mit funktionellen Studien, liess sich aus dieser Struktur der Mechanismus der Sequon-Erkennung ableiten. Aufgrund der fehlenden strukturellen Information, konnte jedoch wenig über die Interaktion der OST mit dem LLO sowie über die Transfer-Reaktion ausgesagt werden. Es bedurfte einer höher aufgelösten Struktur um die Seitenketten sowie die Interaktionen zu identifizieren, welche in der Aktivierung des Amid-Stickstoffs des Akzeptor-Asparagins involviert sind.

In dieser Dissertation habe ich funktionelle sowie strukturelle Ansätze kombiniert um den Mechanismus des von *C. lari* PglB katalysierten Glykan-Transfers zu studieren. In Kapiteln zwei und drei beschreibe ich zwei Ansätze zur Erleichterung der Kristallisation vom PglB. Der erste Ansatz bestand in der Erzeugung konformativer Einzeldomänenantikörper zur Verbesserung von Kristallgitterkontakten durch eine Vergrösserung der polaren Oberfläche für Protein-Protein Interaktionen, sowie der Restriktion der Flexibilität der mobilen Teile von PglB, wie zum Beispiel der Extrazellulären Schlinge 5 (EL5). Trotz der aussergewöhnlich starken Immunantwort des Alpakas für Antikörper in der Klasse der schweren Ketten, konnten die Binder die Protein Kristallisation nicht verbessern. Nichtsdestotrotz wurden sechs der Einzeldomänenkörper für thermostabilisierend befunden. Diese könnten sich in der Zukunft für cryo-EM und NMR Studien als nützlich erweisen. Der zweite Ansatz war die Konstruktion eines PglB Proteins, welches durch den Einbau eines zusätzlichen Glykosylierungssequons (als Inseq Konstrukte benannt) stabilisiert werden sollte. Durch das Binden des eingebauten Sequons an die Peptidbindungsstelle wird das Protein stabilisiert. Des Weiteren wird die lokale Konzentration an Akzeptorpeptiden in den Kristallisationsversuchen erhöht. Für das beste Inseq Konstrukt (Inseq 4) konnte eine viel höhere Stabilität aufgezeigt werden. Kristallisationsversuche mit diesem Protein führten zu Kristallen, welche bis auf 4.5 Å streuten. Dies stellt die Basis für zukünftige, höher auflösende Strukturen von Binär- und Ternärkomplexen von PglB dar.

Um Einblicke in die Interaktionen zwischen OST, LLO, Peptid und dem bivalenten Metal Ion zu gewinnen, sowie letztendlich den molekularen Mechanismus des Glykan Transfers zu verstehen, habe ich funktionelle und strukturelle Studien an Ternärkomplexen von PglB durchgeführt. In Kapiteln vier und fünf werden zwei weitere Ansätze zum Einfangen der Ternärkomplexe beschrieben. Beim Einfangen eines Ternärkomplex-Zwischenprodukts der

OST bestand die grösste Herausforderung darin die Katalyse zu verhindern, ohne dabei die Bindungsaffinintät der Peptid- und LLO-Substrate zu beeinträchtigen. In dieser Dissertation habe ich verschiedene inhibierende und reaktive Peptid- und LLO-Analoga überprüft und dabei die besten Kandidaten für Kristallisationsexperimente identifiziert. Ich habe zwei wichtige Zwischenprodukte der Glykosylierungsreaktion co-kristalisiert und deren Struktur aufgeklärt. Durch die Kombination dieses strukturellen Einblicks mit chemo-emzymatischen Ansätzen, sowie weiteren Studien konnte ich einen Glykosyslieringsmechanismus vorschlagen. In dieser Arbeit präsentiere ich die ersten Ternärkomplexstrukturen, nicht nur von PglB, sondern auch überhaupt von einer OST. Obschon diese Resultate beispiellose mechanistische Einblicke gewähren, können sie den exakten Mechanismus der Aktivierung der Asparagin Carboxamid Gruppe nicht erklären. Allerdings stellen sie einen vielversprechenden Ausgangspunkt für rechnergestützte Ansätze zur Untersuchung des Glykantransferschritts auf molekularer Ebene dar. Ebenfalls gewährt diese Arbeit Einsicht in die strukturellen Veränderungen, welche für die Funktion aller OST benötigt werden könnten, da der Glykosylierungsmechanismus höchstwahrscheinlich in allen Domänen des Lebens konserviert ist.

Chapter 1: Introduction

N-linked protein glycosylation

N-linked protein glycosylation is one of the most abundant post-translational modifications and results in the covalent attachment of an oligosaccharide onto asparagine residues of proteins. Whereas *N*-glycosylation is essential in eukaryotes, facilitating many biological processes such as protein folding, quality control, trafficking, cellular signalling and host–pathogen interactions, it is also important in bacteria and archaea. In *Campylobacter jejuni*, a bacterial food pathogen, *N*-glycans improve host cell- adhesion and colonization, indicating their role in virulence[1,2]. Although there are differences, the *N*-glycosylation pathways in all three domains of life are thought to be homologous and follow a conserved mechanism[3]. In all three pathways the glycan is firstly assembled on a lipid carrier by various glycosyltransferases and then is transferred *en bloc* to an asparagine side chain[4,5]. Only asparagine residues located within the consensus sequence N-X-S/T can be glycosylated[6,7].

The structure of the lipid-linked oligosaccharide

The lipid-linked oligosaccharide (LLO) is assembled at the membrane of the endoplasmic reticulum (ER) in eukaryotes and at the plasma membrane in prokaryotes[5,8]. It is composed of a glycan moiety attached to a phosphorylated lipid carrier. Lipid anchors are phosphorylated polyisoprenoids of different lengths and saturation. Whereas in eukaryotes and archaea glycans are assembled onto dolichyl-pyrophosphate (and dolichyl-phosphate in archaea) lipid carriers, in bacteria the LLO is built on an undecaprenyl-pyrophosphate[4,9,10] (Fig. 1). The presence of a saturated bond in dolichol increases the rotational mobility of the oligosaccharide moiety of the eukaryotic LLO, which may play an important role in glycan elongation in the ER lumen[4].

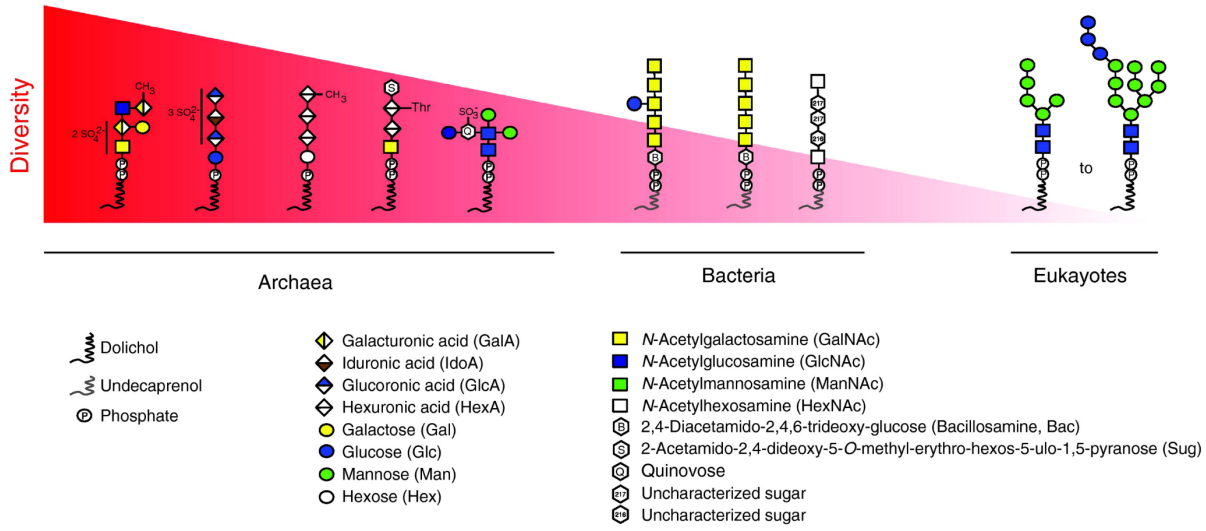


Figure 1. Diversity of the lipid-linked oligosaccharides in in all domains of life. Prokaryotes synthesize a vast array of lipid-linked oligosaccharides, whereas eukaryotes synthesize more conserved structures. Different sugar residues are represented by geometric symbols. Taken from Schwarz and Aebi, 2011 (ref. 5).

Dolichyl-phosphate is also the acceptor in the synthesis of sugar donors: dolichyl-phosphate-mannose (Dol-P-Man) and dolichyl-phosphate-glucose (Dol-P-Glc), the substrates for

glycosyltransferases that elongate LLO in the ER lumen. The availability of the lipid represents a key factor in the assembly of the lipid-linked oligosaccharide because it is the substrate of a number of enzymes in the glycosylation pathway[5,9]. Phosphorylated lipids serve as good leaving groups of nucleophilic substitution reactions and after glycan transfer they can be recycled for further rounds of biosynthesis[5,10,11].

The glycan attached to the lipid-pyrophosphate carrier differs in all domains of life. Archaea show the biggest diversity in their glycan precursors (Fig. 1). They use a vast array of sugar building blocks and the transferred glycans display a broad structural variety, which may be explained by the diverse habitats of the archaeal species that produce them[5,8,12]. For example *Halobacterium salinarum* can synthesize two different oligosaccharides, with either hexose or *N*-acetylhexosamine as the reducing-end sugar, and dolichyl-pyrophosphate or dolichyl-phosphate as the lipid carrier[13]. On the other hand, *N*-glycans in bacteria display more limited diversity with *N*-glycosylation being found in a small number of species known as the epsilon subdivision of proteobacteria, which includes *Campylobacter*, *Helicobacter* and *Wolinella* genera[5,8]. In addition, a subset of *Deltaproteobacteria*, including *Desulfovibrio* species, also possess the *N*-glycan pathway[14]. The bacterial food pathogen *Campylobacter jejuni* assembles a heptasaccharide (Glucose-Galactosamine$_5$-2,4-di-*N*-acetyl-Bacillosamine-) on a undercarenyl-pyrophosphate (Fig. 2), whereas *Helicobacter pullorum* synthesizes a linear pentasaccharide containing uncharacterized hexosamines attached to undecaprenyl-pyrophosphate[5,15]. In higher eukaryotes, in contrast, the glycans transferred to asparagine residues have a conserved structure of Glc_3-Man_9-$GlcNAc_2$ and the diversity of glycans attached to proteins is due to their consecutive modification in the ER and Golgi[5]. The *N*-glycan diversity of extracellular proteins found in all domains of life is therefore a result of convergent evolution and is probably connected to the function of their glycans.

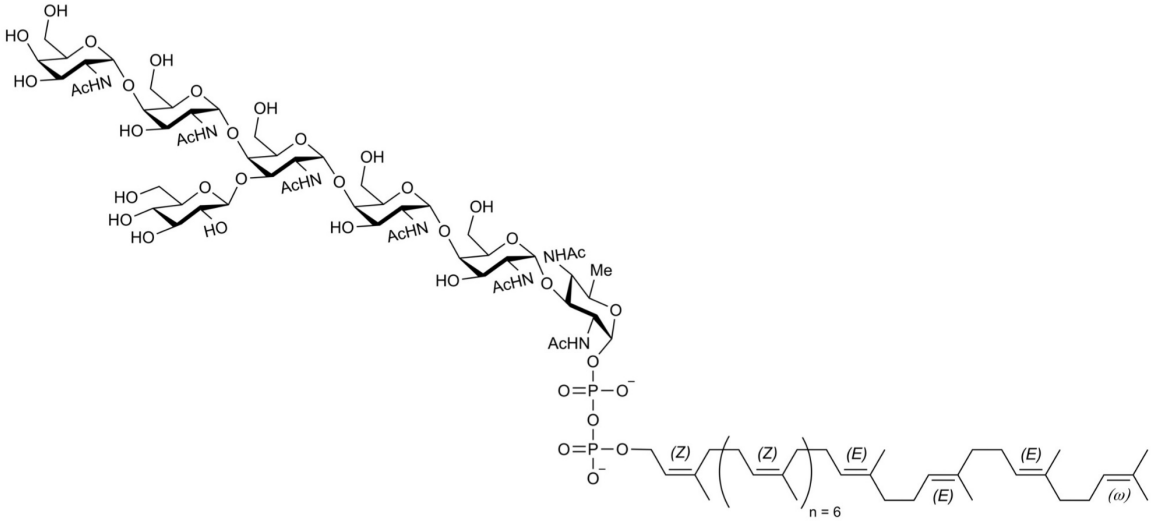


Figure 2. Structure of full-length, wild type *C. jejuni* LLO.

The assembly of the lipid-linked oligosaccharide

The assembly of the LLO is a complex process where various integral-membrane or membrane-associated glycosyltransferases are involved in the stepwise addition of monosaccharides onto a lipid anchor. Studies on the Alg- and Pgl- pathways, from *Saccharomyces cerevisiae* and *Campylobacter jejuni* respectively, gave insight into LLO biosynthesis[9,16,17] (Fig. 3 and 4). The eukaryotic Alg- pathway is initiated at the cytosolic side of the ER where GlcNAc-phosphate is transferred to dolichyl-phosphate by the Alg7 protein (GlcNAc-1-phosphotransferase), resulting in dolichyl-pyrophosphate-GlcNAc (Dol-PP-GlcNAc)[18,19] (Fig. 3). After generating Dol-PP-GlcNAc a second GlcNAc and five mannose moieties are added by cytosolic

glycosyltransferases (Alg13/14, Alg1, Alg2, Alg11), resulting in Dol-PP-GlcNAc$_2$Man$_5$[9,20] (Fig. 3). All these cytosolic glycosyltransferases use nucleotide-activated sugars, UDP-GlcNAc and GDP-Man, as substrates[9,20]. The LLO is subsequently flipped across the membrane to the ER lumen. Translocation is probably facilitated by the Rft1 protein, however it's role has not been confirmed by any *in vitro* studies[21]. In the ER lumen the heptasaccharide LLO is further elongated by luminal glycosyltransferases (Alg3, Alg9, Alg12, Alg6, Alg8 and Alg10) to finally form Dol-PP-GlcNAc$_2$Man$_9$Glc$_3$ (Fig. 3). The luminal glycosyltransferases use lipid-phosphate activated sugar donors (Dol-P-Man and Dol-P-Glc)[4,5,9]. Once the LLO glycan is fully assembled it is transferred to acceptor proteins by oligosaccharyltransferase (OST). Alg proteins are either membrane-integral or membrane-associated enzymes with hydrophobic surfaces, making them challenging targets for structural studies. The human paralog of Alg7, DPAGT1, is so far the only Alg protein whose structure is known. The X-ray structure of DPAGT1 in complex with substrate and tunicamycin, a bactericidal nucleoside analog, was recently published revealing the substrate binding mode and suggesting the mechanism of GlcNAc-P transfer[18].

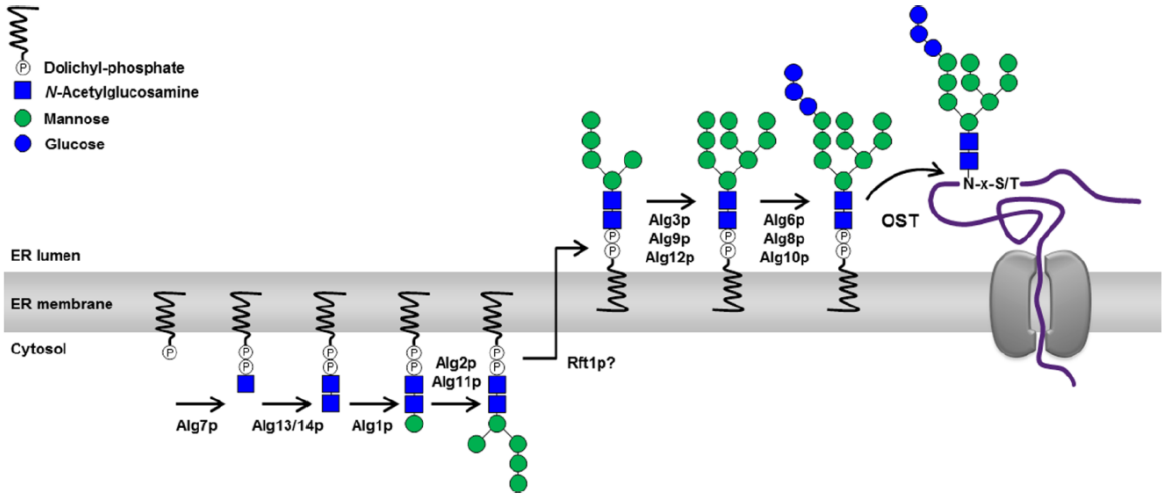


Figure 3. Synthesis of the *N*-linked oligosaccharide and its transfer to a polypeptide chain in *Saccharomyces cerevisiae* (Alg pathway). The synthesis of the LLO takes place in the cytoplasm and then the putative flippase Rft1 translocates Dol-PP-GlcNAc2Man5 to the ER lumen where it is further elongated. When it is fully assembled, the OST transfers the glycan onto polypeptides. Different sugar residues are represented by geometric symbols. Taken from Lizak, 2011 (ref. 17).

The *N*-linked glycosylation pathway in *C. jejuni* is similar to the first half of the dolichol pathway in *S. cerevisiae* because all of the enzymes assemble the LLO on the cytoplasmic side of the membrane using nucleotide-activated sugars (Fig. 3 and 4). In the Pgl pathway three enzymes, a dehydratase (PglF), an aminotransferase (PglE), and an acetyltransferase (PglD) are responsible for the synthesis of UDP-2,4-di-*N*-acetyl-Bacillosamine (UDP-diNAcBac), the UDP activated reducing-end sugar of bacterial LLO (Fig. 4). This unusual sugar is a derivative of the amino sugar bacillosamine, which was originally discovered in *Bacillus licheniformis* and found in bacterial glycans including the lipopolysaccharide of *Vibrio cholerae* and the pilin of *Neisseria meningitides*[8,22,23]. Bac is transferred to the lipid-carrier by the PglC protein, and then elongated by two GalNAc moieties by PglA and PglJ (yielding α-1,3 and α-1,4 linkages, respectively)[24,25]. The next three α-1,4 GalNAc units are added by the processive glycosyltransferase PglH and the last step of LLO assembly is catalysed by PglI, which transfers a glucose moiety, leading to a branched β-1,3-linkage with the third GalNAc unit[26,27]. Once the glycan is fully synthesized in the cytoplasm, it is translocated by the ATP Binding Cassette (ABC) transporter PglK to the periplasm, where it is directly used by the

oligosaccharyltransferase PglB[28] (Fig. 4). The entire *pgl* locus has been inserted into *Escherichia coli* and proven to be responsible for facilitating *N*-linked glycosylation[7,29].

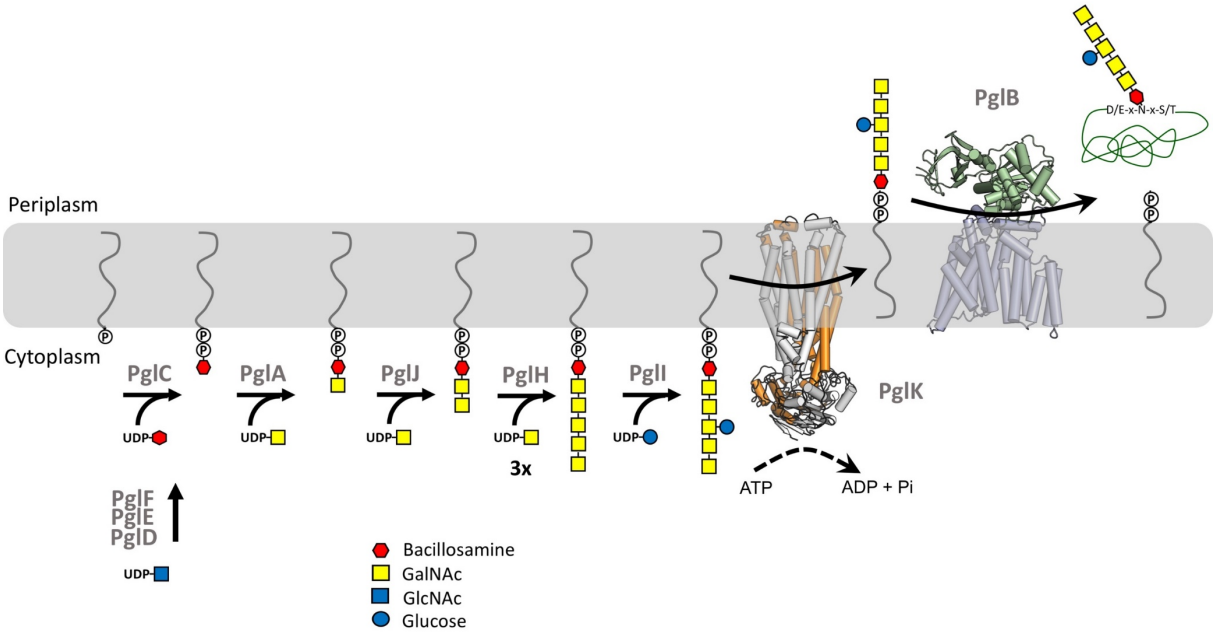


Figure 4. Synthesis of the *N*-linked oligosaccharide and its transfer to a polypeptide chain in *Campylobacter jejuni* (Pgl pathway). Biosynthesis of the LLO occurs on the cytoplasmic side of the plasma membrane. The fully assembled glycan is translocated to the periplasm, where it is used by oligosaccharyltransferase (PglB) to glycosylate folded proteins. Different sugar residues are represented by geometric symbols. Redrawn from Perez *et al.*, 2013 (ref. 28).

The role of each of the Pgl enzymes in *N*-linked glycosylation was validated by *in vitro* experiments where the wild type heptasaccharide LLO was chemo-enzymatically synthesized[15,30]. The X-ray structures of PglH and PglK from *C. jejuni* have been determined[26,28]. The structure of PglH, combined with molecular dynamics studies, revealed its glycan counting mechanism. PglH has an amphipathic helix that plays a dual role: It associates with the membrane and, through interactions between its three positively charged residues and the pyrophosphate group of the LLO substrate, it ensures that a maximum of three GalNAc units are added[26]. The crystal structures of PglK captured the transporter in different conformations: two apo inward-facing conformations and an outward-occluded conformation bound to ADP. Based on structural studies, and *in vitro* and *in vivo* assays, the authors proposed a flipping mechanism, where the hydrophilic pyrophosphate-oligosaccharide group of LLO interacts with conserved arginine residues and enters the translocation cavity of PglK, leaving the lipidic tail exposed to the lipid bilayer[28].

Oligosaccharyltransferase is a key enzyme of the *N*-linked protein glycosylation pathway
A key step in the *N*-glycosylation pathway is the *en bloc* transfer of the oligosaccharide onto the acceptor asparagine, resulting in the formation of an *N*-glycosidic bond between the amide nitrogen of the acceptor asparagine and the C1 carbon of the reducing-end sugar of a lipid-linked oligosaccharide. The reaction occurs with an inversion of the configuration at the substituted C1 carbon, resulting in a β–*N*-glycosidic linkage. The glycosylation reaction is catalyzed by an integral-membrane enzyme called oligosaccharyltransferase (OST). OST is a member of the C family of glycosyltransferases, which consists of membrane proteins that are located in the endoplasmic reticulum or the plasma membrane and are predicted to contain 8–13 transmembrane helices[11,31]. OST of higher eukaryotes is a multiprotein complex with STT3

as the catalytic subunit. In contrast, archaea, kinetoplastids and some bacteria contain single-subunit OST enzymes that are homologous to STT3[32–34].

The first structure of STT3, peptide-bound PglB of *C. lari*, was published in 2011 by Lizak and colleagues[35]. This structure revealed the architecture of the enzyme. The same fold was found in the following structures of its archaeal homolog, AglB from *Archaeoglobus fulgidus*[36,37]. The X-ray structure of PglB revealed that OST contains two domains: a transmembrane (TM) domain and a periplasmic domain. The TM domain is composed of 13 TM helices and two extracellular loops, EL1 and EL5[35] (Fig. 5). Whereas EL1 mainly has a structural role, EL5 was found to contribute to both peptide binding and catalysis[35,38,39]. The C-terminal part of EL5 plays an important role in peptide binding and coordination of the catalytic divalent metal ion. It was fully ordered in the peptide bound structure of PglB, where it pinned the bound sequon against the periplasmic domain and provided an essential residue (E319) to the catalytic site[35]. The N-terminal part of EL5 was disordered in the peptide-bound structure, however *in vitro* studies showed that it contains an important sequence motif (Tyr plug) with a conserved tyrosine residue (Tyr293) whose aromatic side chain is essential for catalysis[39]. Because enzymatic cleavage of EL5 has been shown to result in only a slight reduction in activity, the two functions of EL5 are independent of each other[39]. The peptide-bound structure showed that PglB forms two large cavities above the membrane surface, located at opposite sides of the protein (Fig. 5). An acceptor peptide binds to the cavity on the left-hand side, whereas the cavity on the right forms the catalytic site and probably binds the LLO. The two cavities are connected by a tunnel, through which the side chain of the acceptor asparagine reaches the catalytic site[35].

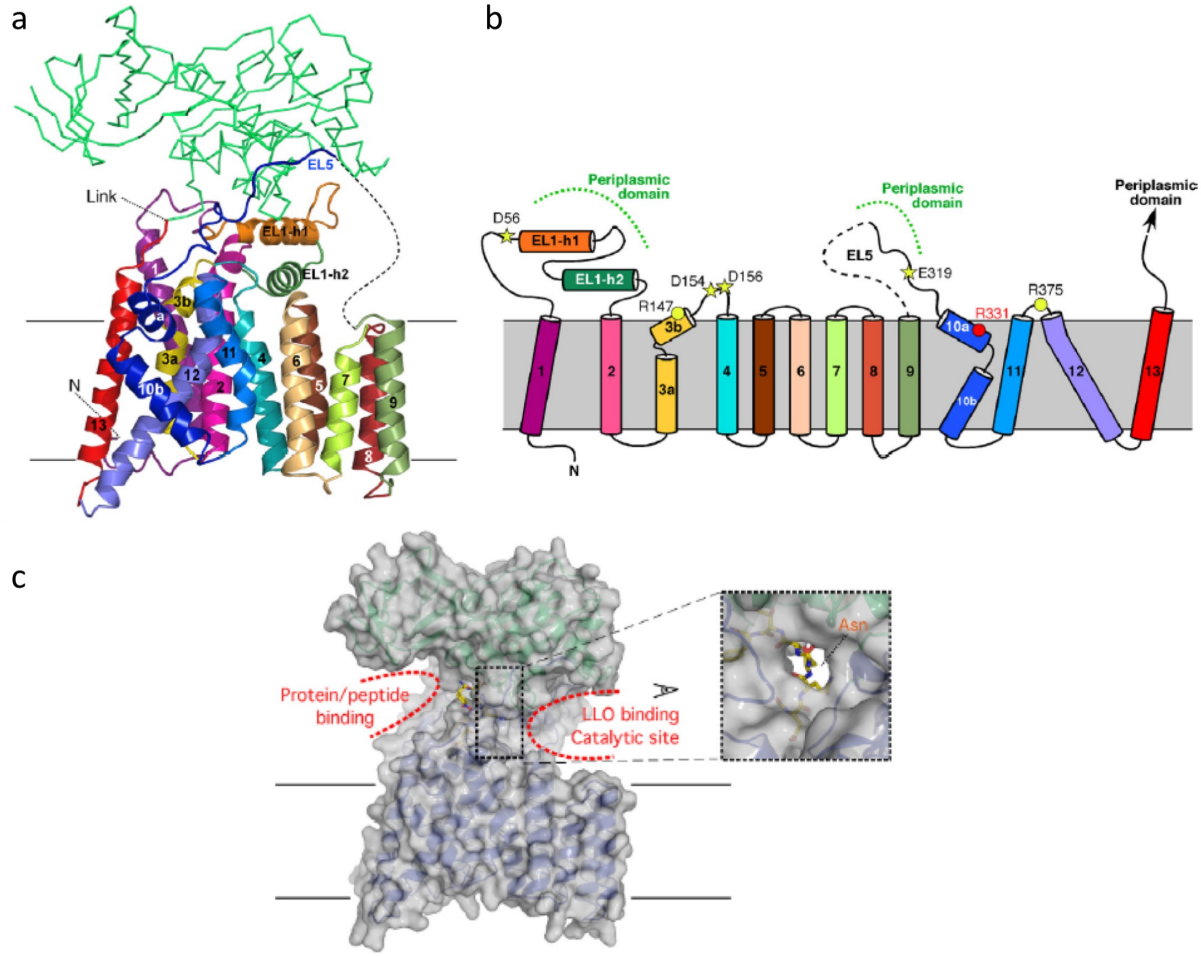


Figure 5. X-ray structure of PglB from *Campylobater jejuni*. (**a**) Ribbon diagram with TM helices numbered. The periplasmic domain is shown as a green backbone trace. (**b**) Topology of the

transmembrane domain. Conserved residues forming the active site of PglB are indicated by yellow stars (residues involved in M^{2+} coordination) and spheres and are labelled. TM helices are colored as in a. (**c**) Surface representation in semi-transparent grey. Two cavities connected by a tunnel are positioned at opposite sides of PglB. The cavities are indicated by dashed red lines. Taken from Lizak *et al*., 2011 (ref. 35).

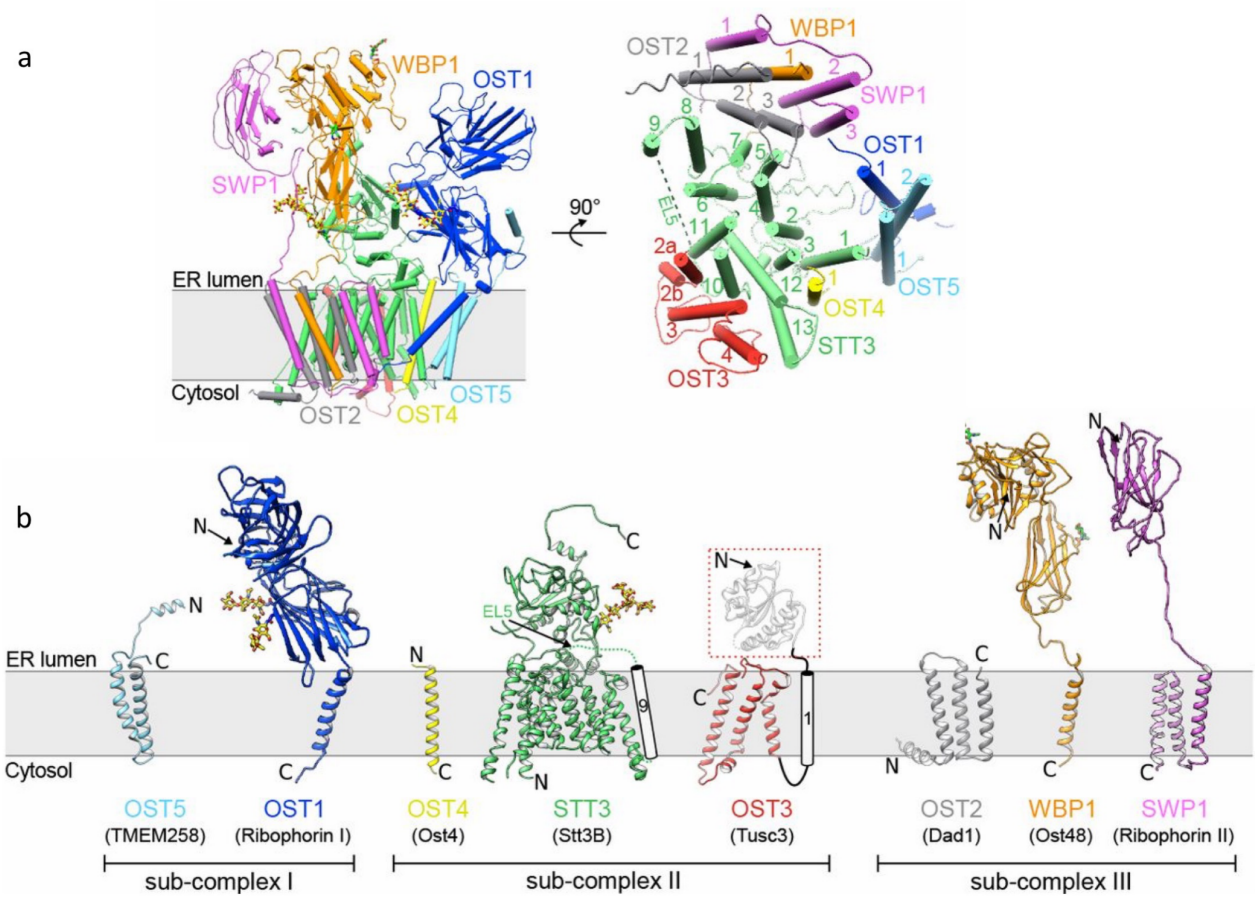


Figure 6. cryo-EM structure of the yeast OST complex. (**a**) Cartoon representation of the yeast OST structure, with a side view of OST on the left and a view from the cytoplasm on the right. Ordered glycans are shown as yellow sticks. (**b**) Cartoon representation of the single subunit structures. EL5 and TM9 in STT3 as well as TM1 of OST3 were disordered and are shown schematically. Taken from Wild *et al*., 2018 (ref. 44).

The octameric OST complex has been studied most comprehensively in *S. cerevisiae*. It consists of eight subunits that are all membrane- integral or membrane-anchored proteins[32,33]. Five of the subunits were found to be essential for yeast viability: STT3, OST1, OST2, WBP1 and SWP1[40,41]. The absence of either the OST3, OST4, OST5 or OST6 subunits results in general hypoglycosylation of *N*-glycoproteins[32,42]. The yeast OST complex exists in two isoforms: either containing OST3 or OST6, which are products of paralogous genes. OST3 and OST6 are responsible for the different polypeptide specificity of the two isoforms[43]. The high-resolution cryo-EM structures of yeast OST, combined with biochemical data, showed that OST assembles into three subcompelexes: Sub-complex 1 is composed of OST1 and OST5, sub-complex 2 of STT3, OST3, and OST4, and sub-complex 3 of OST2, WBP1, and SWP1[32,44,45] (Fig. 6). Interactions between the membrane-embedded interfaces of the sub-complexes are mediated by phospholipids. All five OST soluble domains are in the ER lumen. In the structures 4 out of 5

luminal domains and 26 out of 28 transmembrane helices were visible[44,45]. The catalytic subunit, STT3, has a very similar fold to the ssOSTs with 13 TM helices and a large soluble domain[35,36] (Fig. 6). In the structure of apo yeast OST EL5 was disordered, which agreed with the previously proposed hypothesis that EL5 only becomes fully ordered when bound to both substrates[39,46]. TM9, connected to EL5, was also poorly resolved and is probably flexible in the absence of bound substrate. In the structure of yeast OST, five of seven *N*-glycans (at N539 of STT3, N336 and N400 of OST1, N60 and N332 of WBP1) could be built. Interestingly, the best-ordered *N*-glycan is attached to the conserved STT3 residue N539 and it interacts with WBP1 and SWP1, possibly contributing to LLO binding[44].

The cryo-EM structure of the mammalian OST complex has a very similar arrangement to yeast OST and contains proteins homologous to the yeast OST components: DAD1 (OST2), TUSC3 and MAGT1 (OST3 and OST6), OST48 (WBP1), Ribophorin I (OST1), Ribophorin II (SWP1), STT3A and STT3B (STT3)[42]. OST of higher eukaryotes exists in two isoforms: STT3A and STT3B- containing OST. The STT3B complex (homologous to the yeast enzyme) contains several STTB-specific subunits: oxido-reductases TUSC3 (Tumor suppressor candidate 3) and MAGT1 (Magnesium transporter protein 1). The OST complex with STT3A associates with the translocon, and the proteins DC2 and KCP2 (Keratinocyte-associated protein 2), found only in STT3A complexes, mediate the interaction between the oligosaccharyltransferase and the ER translocon[47,48]. Cryo-EM reconstructions of the mammalian STT3A complex revealed the position and orientation of the OST catalytic site in relation to the ribosome-translocon complex (RTC) (Fig. 7).

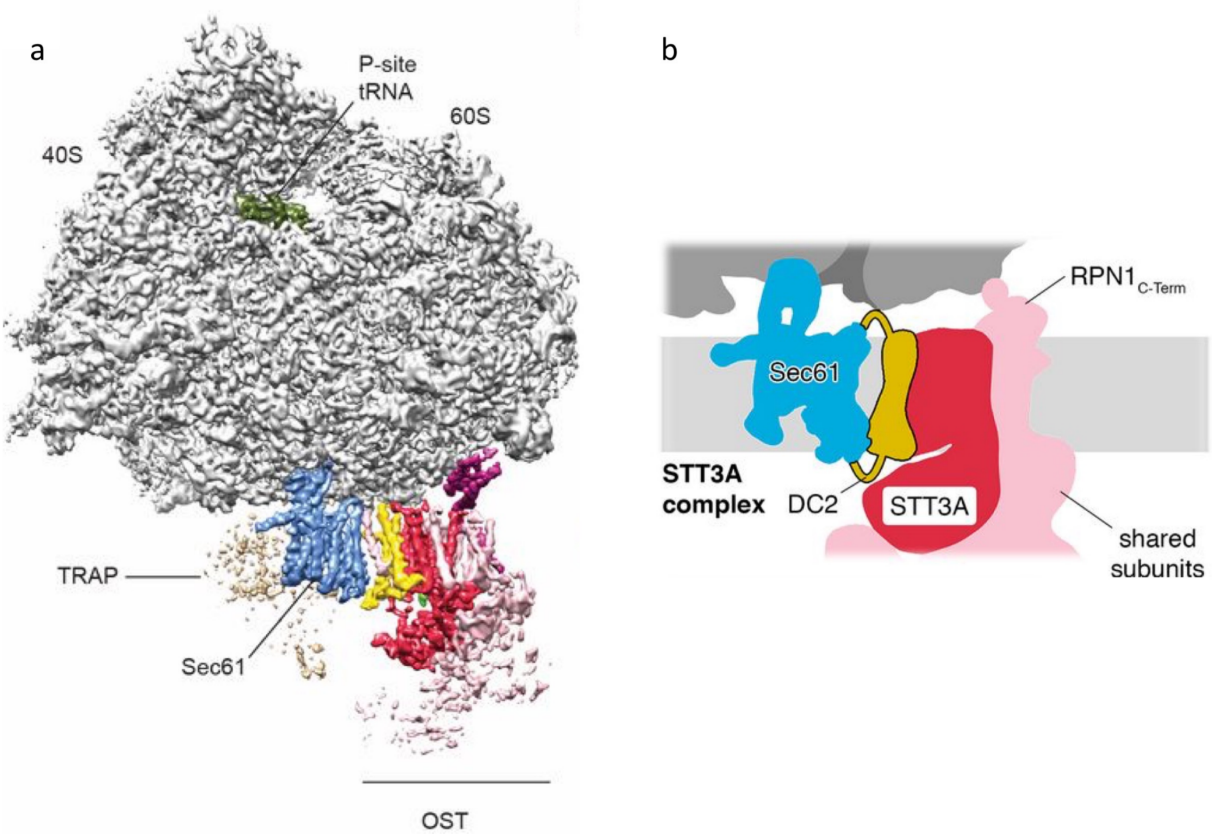


Figure 7. Cryo-EM structure of ribosome-bound mammalian OST showing the location of STT3A, RPN1 and DC2. (**a**) Cryo-EM map of the active solubilized ribosome-translocation complex (RTC). The translocon (Sec61) is shown in blue, STT3A in red and DC2 in yellow. (**b**) Schematic

representation of the molecular basis for STT3 paralog specificity in the RTC. The DC2 and RPN1 subunits tie the STT3A complex into the RTC. Taken from Braunger *et al*., 2018 (ref. 47).

The mammalian OST structure indicates the path for a nascent glycosylation substrate and shows how translocon- associated OST scans the nascent polypeptide chain for glycosylation sequons. It also gives new insight into STT3A specificity in associating with the RTC. There are two interfaces between RTC and STT3A-OST: One between the ribosome and the cytosolic RPN1 domain and one between DC2 and the translocon (Sec61) (Fig. 7). These recent cryo-EM structures of yeast and mammalian OST complexes, and the X-ray structures of ssOSTs, confirmed that, despite low sequence similarity, the fold of the catalytic subunit is conserved in all OSTs (Fig. 8)[35,36,44,45,47].

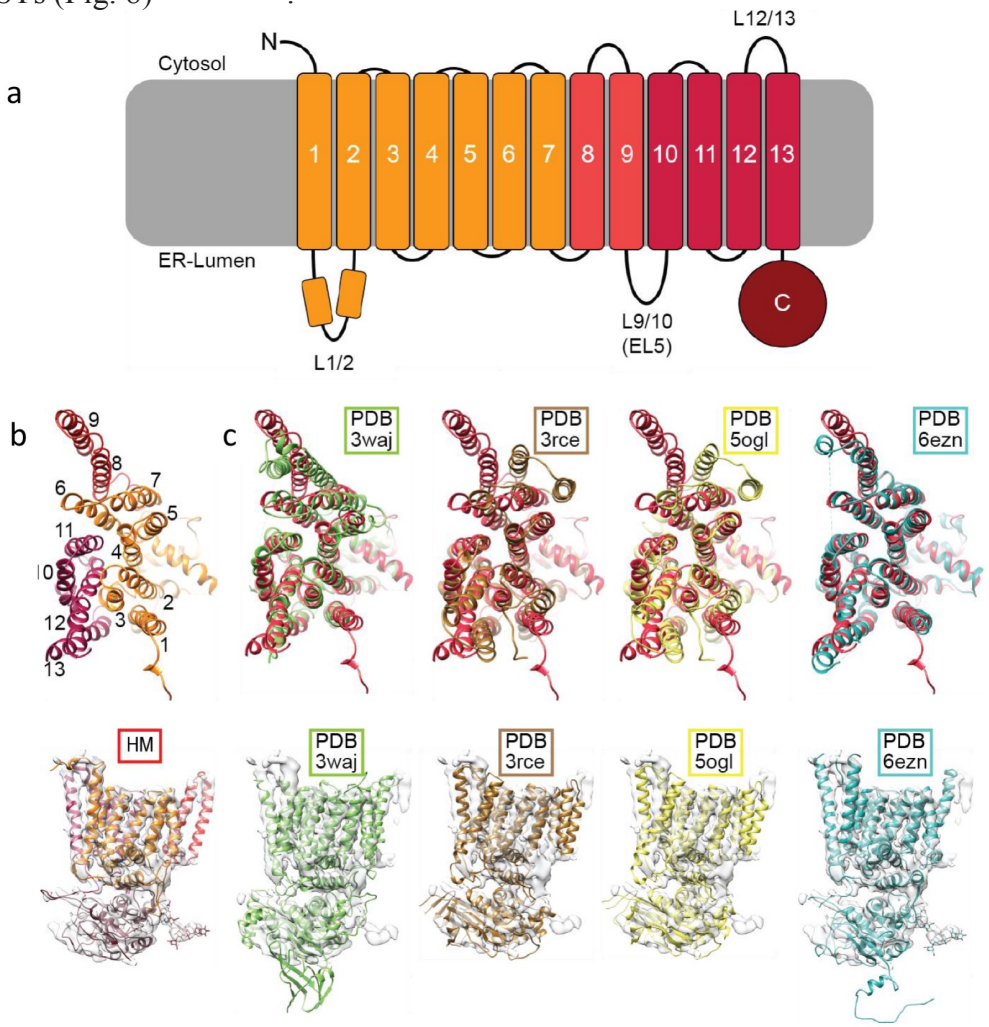


Figure 8. Homology model of mammalian STT3A. (**a**) Membrane topology of STT3A. (**b**) Homology model of mammalian STT3A as seen from the cytosol. Order of TM segments is indicated by Arabic numbers (**c**) Superposition of the STT3A homology model (red) with X-ray structures of archaeal (green, PDB 3WAJ), bacterial (brown, PDB 3RCE; yellow, PDB 5OGL) and cryo-EM structure of yeast (blue, PDB 6EZN), and the fit of those models into cryo EM density of mammalian OST. All homologs show a very similar arrangement of TM helices. Taken from Braunger *et al*., 2018 (ref. 47).

OST as a member of GT-C superfamily of glycosylatransferases

Glycosyltransferases (GTs) are defined as enzymes that use an activated donor sugar substrate. They are mainly activated in the nucleoside diphosphate or lipid-phosphate/ pyrophosphate form. Glycosyltransferases utilizing nucleotide-sugars are called Leloir enzymes[11,31]. The Leloir GTs have been thoroughly studied and two general folds were defined for these glycosyltransferases: GT-A and GT-B[11,31]. The GTA fold is an open twisted β-sheet surrounded by α-helices on both sides and was first observed for the SpsA enzyme from *Bacillus subtilis*[49]. The GT-B fold consists of two similar Rossmann subdomains and was described for the first time in the β-glucosyltransferase from bacteriophage T4[50]. More recently a third glycosyltransferase fold, GT-C, was identified. According to the Carbohydrate Active enZYme (CAZy) database the GT-C superfamily consists 19 GT families with OST (STT3 family) as one of them. The GT-C superfamily consists of membrane proteins that are located in the endoplasmic reticulum or the plasma membrane and contain 8–13 predicted transmembrane helices. The large hydrophobic transmembrane domain of GT-C members is thought to be essential for their glycosylation activity[11,31]. They also use lipid-phosphate/pyrophosphate activated sugars. Peptide-bound PglB was not only the first structure of an STT3 protein but also the first structure of GT-C family member[35].

Acceptor sequon specificity, binding and recognition

Statistical analysis of the eukaryotic glycoproteome, as well as biochemical studies, revealed that the N-X-S/T sequence, where X cannot be proline, is the minimum sequence required for *N*-linked protein glycosylation[51]. Statistical analyses have shown that sequences with threonine at the +2 position are three times more common than those containing serine[52,53]. Further *in vitro* studies revealed that peptides containing N-X-T sites are much better substrates than those with N-X-S[54–56]. Intriguingly, proteins containing N-X-C, N-X-V and N-G-X sites were also found to be glycosylated in the mouse glycoproteome[57]. Only sequons containing L-serine or L-threonine, and not their D-stereoisomers, were recognized as substrates, indicating that the binding of the hydroxyl group is stereospecific[58].

The structure of peptide-bound PglB provided more insight into sequon binding and recognition[35]. The acceptor peptide DQNATF is tightly bound to the surface of PglB at the interface of the periplasmic and transmembrane domains where it forms a loop that turns 180° (Fig. 9a). Adopting this conformation would not be possible with a proline at the +1 position of the sequon. The peptide-bound structure also explained the specificity for a β- hydroxyl group at the +2 position[35]. The strictly conserved WWD motif, located in the periplasmic domain of PglB, forms three hydrogen bonds with the β- hydroxyl group of threonine (Fig. 9b). These hydrogen bonding interactions contribute significantly to peptide binding and they can be formed primarily with amino acids containing a beta hydroxyl group (T or S) or a thiol group of cysteine[38]. The structure, combined with comprehensive functional studies, revealed why the threonine- rather than the serine- containing acceptor peptides are better substrates for OST. The γ- methyl group of the threonine, which is not present in serine, forms van der Waals contacts with I572. Mutation of I572 to alanine reduced both peptide binding and glycosylation activity[38] (Fig. 9b). Compared to eukaryotic OSTs, bacteria have an extended consensus sequon (D/E-X-N-X-S/T) with an acidic amino acid at the -2 position. The peptide-bound structure showed that the R331 of PglB makes a salt bridge to the negatively charged aspartate of the acceptor peptide resulting in stronger PglB-peptide interactions. The R331A mutant showed significantly decreased peptide binding affinity and therefore glycosylation activity[38].

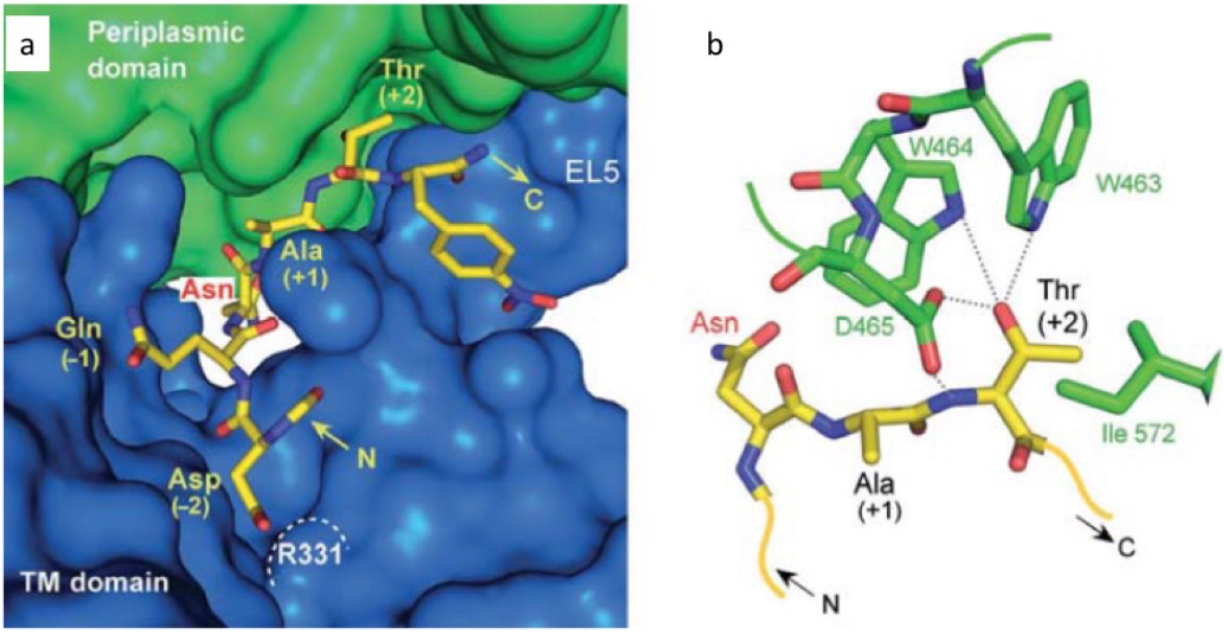


Figure 9. Sequon binding and recognition. (**a**) Interactions between the acceptor peptide DQNATF and PglB. Transmembrane and periplasmic domains of PglB are colored in blue and green, respectively. The peptide is shown as yellow sticks. (**b**) Schematic of the interactions between threonine at the -2 position and conserved residues. The WWD motif forms three hydrogen bonds with the β- hydroxyl group and I572 provides van der Waals interactions to the γ-methyl group. Taken from Lizak *et al*., 2011 (ref. 35).

Catalytic site of PglB

The catalytic site of PglB is located in a cavity at the boundary between the periplasmic and transmembrane domains (Fig. 5c)[35]. In the catalytic site the bound divalent metal ion is coordinated by the conserved acidic residues D56, D154 and E319 (in *C. lari*) (Fig. 10). Residue D154, as well as D156, belong to the D-X-D motif that is conserved in divalent metal-dependent glycosyltransferases, where the aspartate residues are involved in metal- and/or ribose-binding[11,35]. Residues D56 and E319, on the other hand, are involved in binding of both a divalent metal ion and an acceptor peptide (Fig. 10). Mutation of these residues to alanine decreased *in vivo* and *in vitro* glycosylation activity and also severely reduced *in vitro* affinity of the acceptor peptide[38]. However, when they were mutated to glutamate and aspartate, respectively, they retained activity, although at a decreased level. The glycosylation activity of these two mutants was also dependent on the LLO concentration, indicating that the divalent cation is further involved in coordinating the pyrophosphate of the LLO molecule[59]. Therefore the divalent metal ion has a dual role: It possibly stabilizes the lipid-pyrophosphate leaving group during the glycosylation reaction and is contributes to peptide binding[38,60]. Different divalent metal ions were found to be cofactors for PglB and although Mn^{2+} and Mg^{2+} showed the highest turnover rates *in vitro*, PglB was also active in the presence of Ca^{2+} and Zn^{2+}, which is unusual for metal-dependent enzymes[60–62]. Two other essential residues that are present in the catalytic site, Y468 and R375, were proposed to be involved in the catalytic activity of OST as well as LLO binding. Whereas mutation of R375 to lysine did not significantly affect glycosylation activity of PglB, it's mutation to alanine abolished catalytic activity, suggesting that the positively charged side chain of this arginine interacts with the pyrophosphate moiety of LLO[35,38]. Y468 on the other hand belongs to the extended WWD motif (^{463}WWDYGY) and it is thought to interact with the C-2 substituent of the reducing-end sugar[35].

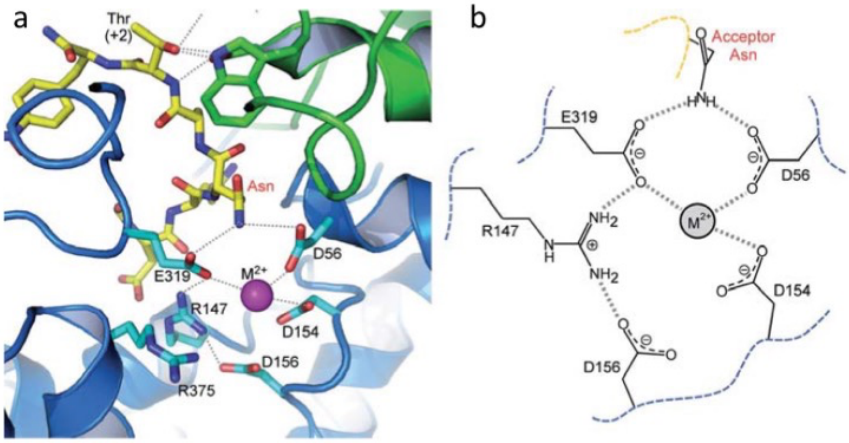


Figure 10. Catalytic site of PglB. (**a**) Catalytic site of PglB. The periplasmic and transmembrane domains are colored green and blue respectively. Residues forming catalytic site are shown as light blue sticks. The acceptor peptide DQNATF is shown as yellow sticks. (**b**) Schematic representation of the catalytic site of PglB. The interactions between residues and the metal ion (hydrogen bonds or interactions with divalent ion) are indicated by dashed lines. Taken from Lizak *et al*., 2011 (ref. 35).

Catalytic mechanism of glycan transfer

OST is an inverting glycosyltransferase, because the catalysed transfer of the glycan to an acceptor asparagine happens with the inversion of configuration at the anomeric centre (C1 of the reducing-end sugar). The mechanism of the glycan transfer catalysed by OST, especially the activation of the carboxamide of acceptor asparagine, has been debated with two main mechanisms being proposed. According to the first one the reaction occurs via a direct S_N2-like displacement reaction that proceeds through a single nucleophilic substitution step, facilitated by the activated carboxamide group of the acceptor asparagine (Fig. 11). The reaction is believed to have an oxocarbenium ion-like intermediate[63]. The second mechanism involves an S_N1-like reaction, where the lipid-pyrophosphate group of LLO is initially released, resulting in generation of an oxocarbenium ion, followed by nucleophilic attack of the activated carboxamide group of the acceptor asparagine (Fig. 11). It has also been proposed that the formed oxocarbenium ion might be further stabilized by the *N*-acetyl group at the C-2 position of the reducing-end sugar[3,63] (Fig. 11). The importance of the *N*-acetyl group at the C-2 position of the reducing-end sugar of LLO for OST activity was suggested by functional studies on bacterial and eukaryotic homologs[3,34,64]. Experiments with synthetic LLO analogs of yeast OST showed that minor modifications to this group, such as substitution of hydrogens with fluorines, abolished glycosylation activity but not LLO binding, suggesting that it is important during glycan transfer[64]. PglB can process lipid-linked oligo- and polysaccharides without an *N*-acetyl group at the C-2 position albeit with much lower efficiency[65,66]. In both mechanisms, the carboxamide group has to be activated. In principle, this could be done by a base catalyst triggering deprotonation of the acceptor asparagine side chain. The potential base was not identified, however. The second hypothesis, supported by the peptide-bound structure and functional studies, is based on twisting of the C-N bond of the acceptor asparagine carboxamide group. This twisting is a result of two strong hydrogen bonds between the amide nitrogen and the side chains of residues D56 and E319 and breaks conjugation of the carboxamide group[35,60] (Fig. 12). Transition states at the catalytic site, identified by *in silico* experiments, further support this twisting amide hypothesis[67].

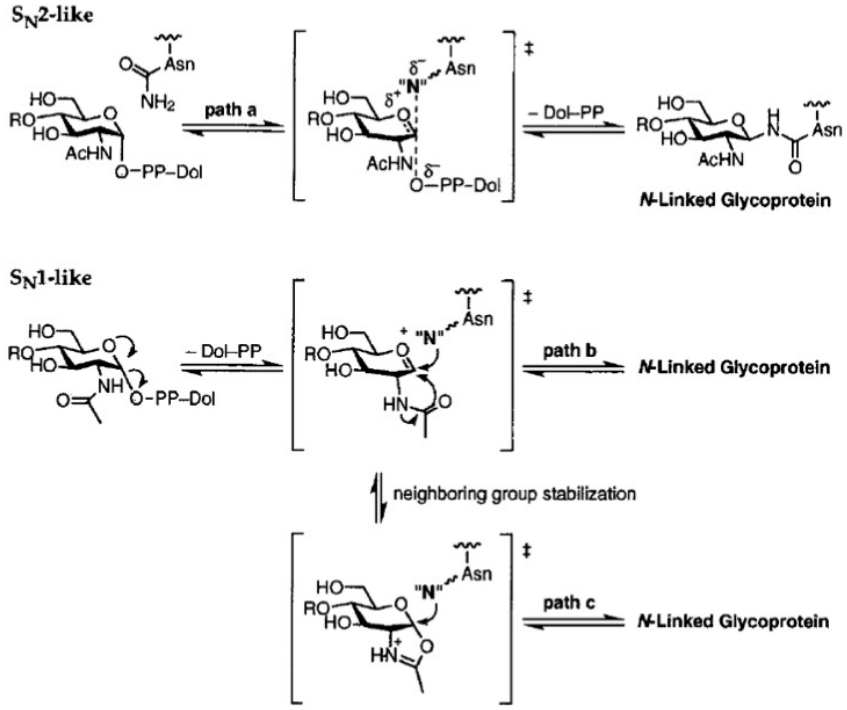


Figure 11. Proposed mechanism of glycan transfer catalysed by OST. The glycan transfer might occur either via S_N2-like or S_N1-like displacement. Taken from Imperiali and Tai, 2005 (ref. 63).

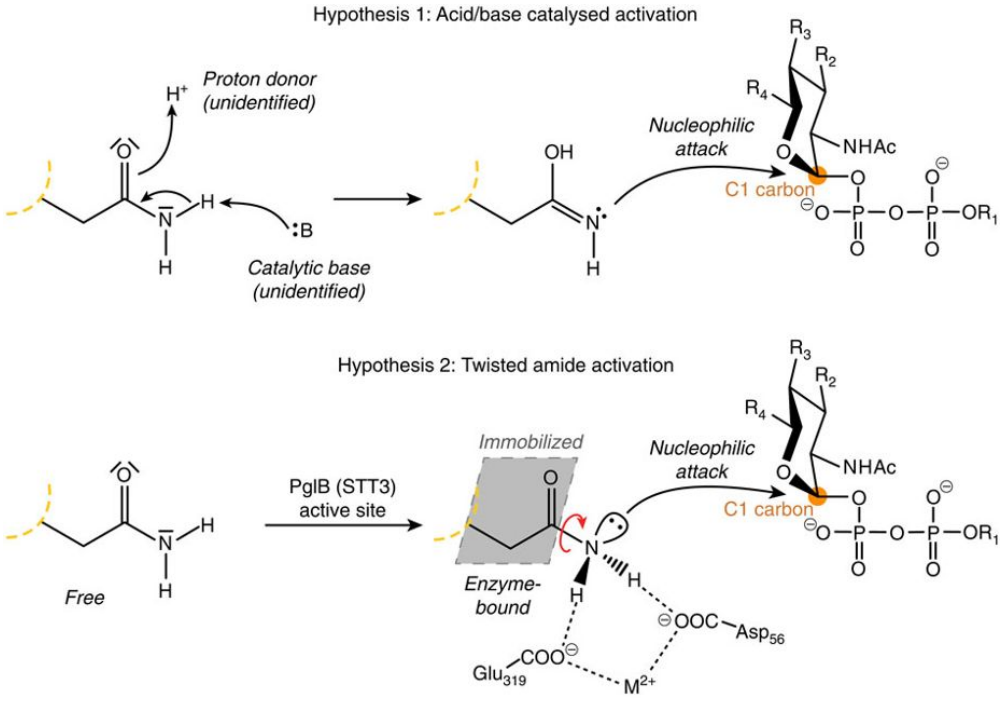


Figure 12. Mechanistic hypotheses for catalytic activation of the amide nitrogen. The carboxyamide group is thought to either be activated by a catalytic base (by deprotonation) or by forming two strong hydrogen bonds with D56 and E319 that break the conjugation of the carboxyamide group. Taken from Lizak *et al.*, 2013 (ref. 60).

Open questions and aims of this thesis

The previously reported structures of single-subunit OSTs, peptide-bound *C. lari* PglB and apo *A. fulgidus* AglB, revealed the fold of this class of enzymes and, combined with biochemical studies, gave insight into the mechanism of sequon recognition. However, how the peptide, LLO and divalent metal ion interact, how the amide nitrogen of the acceptor asparagine is activated, and the exact mechanism of glycan transfer are still poorly understood. Higher resolution structures are required to identify residues and interactions involved in the activation of the amide nitrogen of the acceptor asparagine, i.e. reliably measure the lengths of the potential hydrogen bonds formed with catalytic residues and to detect the presence of the water molecules in the catalytic site. In addition ternary complex structures of PglB bound to peptide, LLO and the divalent metal ion are required to understand how the substrates interact to facilitate glycan transfer. In this thesis I combined structural and functional approaches to study the mechanism of glycan transfer catalyzed by PglB, a ssOST from *C. lari*. PglB is an excellent model system to study the glycosylation reaction as the fold of the catalytic subunit is conserved from the ssOSTs in bacteria and archaea to the multi-subunit complexes in yeast and mammals. Studies on PglB therefore allow us to investigate the OST-catalyzed process without the complexity of additional subunits.

PglB is a membrane protein with 13 transmembrane helices, a large periplasmic domain and mobile extracellular loops, making it a challenging crystallization target. Indeed the production of high-quality crystals that would allow high-resolution structural determination of PglB remains a major hurdle. In the pursuit of higher resolution structures, I adopted two strategies. The first strategy, described in chapter 2, was to generate conformational nanobodies to facilitate PglB crystallization. Nanobodies bound to PglB could potentially increase the polar surface area for protein-protein contacts, improving lattice contacts, and restricting the flexibility of mobile parts of PglB, namely EL5. In addition, a thermostabilizing nanobody would allow us to rescreen purification and crystallization conditions. The second approach, described in chapter 3, was to engineer PglB to stabilize it by inserting a glycosylaton sequon into the periplasmic domain or the region linking TM13 and periplasmic domain. Binding of the inserted sequon would likely stabilize the protein and increase the local concentration of the acceptor peptide during crystallization trials. Although both approaches yielded promising results that may be used for future applications, neither yielded crystals that diffracted to better than 4 Å resolution.

In order to get insight into the interactions between OST, LLO, peptide and the divalent metal ion, and ultimately understand the molecular mechanism of glycan transfer, I pursued functional and structural studies of PglB ternary complexes. This work is described in chapters 4 and 5. The studies presented are a result of close collaborative work between the Locher, Reymond and Aebi laboratories. The main challenge in trapping ternary-complex intermediates of OST is to prevent catalysis without compromising the binding affinity of the peptide and LLO substrates. Co-crystallization of the ternary complex is also very challenging due to the highly hydrophobic nature of wild-type LLO, which requires high concentrations of a detergent for solubilization. To overcome these problems we used synthetic water-soluble analogs of LLO. I therefore screened different inhibitory and reactive peptide and LLO analogs, and identified the best candidates for crystallization experiments. I co-crystallized and determined the structure of two important intermediates of the glycosylation reaction. By combining this structural insight with chemo-enzymatic approaches and further functional studies, I could propose a glycosylation mechanism for PglB that may also be applicable to OSTs from different kingdoms of life.

References

1. Szymanski, C. M., Burr, D. H. & Guerry, P. Campylobacter Protein Glycosylation Affects Host Cell Interactions Campylobacter Protein Glycosylation Affects Host Cell Interactions. *Infect. Immun.* **70,** 2242–2244 (2002).
2. Valguarnera, E., Kinsella, R. L. & Feldman, M. F. Sugar and Spice Make Bacteria Not Nice: Protein Glycosylation and Its Influence in Pathogenesis. *J. Mol. Biol.* **428,** 3206–3220 (2016).
3. Wacker, M. *et al.* Substrate specificity of bacterial oligosaccharyltransferase suggests a common transfer mechanism for the bacterial and eukaryotic systems. *Proc. Natl. Acad. Sci. U. S. A.* **103,** 7088–93 (2006).
4. Dell, A., Galadari, A., Sastre, F. & Hitchen, P. Similarities and differences in the glycosylation mechanisms in prokaryotes and eukaryotes. *Int. J. Microbiol.* **2010,** (2010).
5. Schwarz, F. & Aebi, M. Mechanisms and principles of N-linked protein glycosylation. *Curr. Opin. Struct. Biol.* **21,** 576–582 (2011).
6. Nita-Lazar, M., Wacker, M., Schegg, B., Amber, S. & Aebi, M. The N-X-S/T consensus sequence is required but not sufficient for bacterial N-linked protein glycosylation. *Glycobiology* **15,** 361–367 (2005).
7. Wacker, M. N-Linked Glycosylation in Campylobacter jejuni and Its Functional Transfer into E. coli. *Science* **298,** 1790–1793 (2002).
8. Larkin, A. & Imperiali, B. The expanding horizons of asparagine-linked glycosylation. *Biochemistry* **50,** 4411–4426 (2011).
9. Burda, P. & Aebi, M. The dolichol pathway of N-linked glycosylation. *Biochim. Biophys. Acta - Gen. Subj.* **1426,** 239–257 (1999).
10. Calo, D., Kaminski, L. & Eichler, J. Protein glycosylation in Archaea: Sweet and extreme. *Glycobiology* **20,** 1065–1076 (2010).
11. Lairson, L. L., Henrissat, B., Davies, G. J. & Withers, S. G. Glycosyltransferases: Structures, Functions, and Mechanisms. *Annu. Rev. Biochem.* **77,** 521–555 (2008).
12. Eichler, J. Extreme sweetness: protein glycosylation in archaea. *Nat. Rev. Microbiol.* **11,** 151–156 (2013).
13. Lechner, J. & Wieland, F. Structure and Biosynthesis of P : Rokaryotic. *Annu. Rev. Biochem.* 173–194 (1989).
14. Ielmini, M. V. & Feldman, M. F. Desulfovibrio desulfuricans PglB homolog possesses oligosaccharyltransferase activity with relaxed glycan specificity and distinct protein acceptor sequence requirements. *Glycobiology* **21,** 734–742 (2011).
15. 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.* **102,** 14255–9 (2005).
16. 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.* **32,** 1022–1030 (1999).
17. Lizak, C. A. N-linked protein glycosylation in Escherichia coli. A system for basic nad applied research. (2011).
18. Yoo, J. *et al.* GlcNAc-1-P-transferase-tunicamycin complex structure reveals basis for inhibition of N-glycosylation. *Nat. Struct. Mol. Biol.* **25,** 217–224 (2018).
19. Heifetz, A. & Elbein, A. Solubilization and properties of mannose and N-acetylglucosamine transferases involved in formation of polyprenyl-sugar intermediates. *J. Biol. Chem.* **252,** 3057–3063 (1977).
20. Ramírez, A. S. *et al.* Chemo-enzymatic synthesis of lipid-linked

GlcNAc2Man5oligosaccharides using recombinant Alg1, Alg2 and Alg11 proteins. *Glycobiology* **27,** 726–733 (2017).
21. Helenius, J. *et al.* Translocation of lipid-linked oligosaccharides across the ER membrane requires Rft1 protein. *Nature* **415,** 447–450 (2002).
22. Sharon, N. Celebrating the golden anniversary of the discovery of bacillosamine, the diamino sugar of a Bacillus. *Glycobiology* **17,** 1150–1155 (2007).
23. Nothaft, H. & Szymanski, C. M. Bacterial protein N-glycosylation: New perspectives and applications. *J. Biol. Chem.* **288,** 6912–6920 (2013).
24. 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* **45,** 5343–5350 (2006).
25. Chen, M. M. *et al.* Polyisoprenol specificity in the Campylobacter jejuni N-linked glycosylation pathway. *Biochemistry* **46,** 14342–8 (2007).
26. Ramírez, A. S. *et al.* Structural basis of the molecular ruler mechanism of a bacterial glycosyltransferase. *Nat. Commun.* **9,** (2018).
27. Troutman, J. M. & Imperiali, B. Campylobacter jejuni PglH is a single active site processive polymerase that utilizes product inhibition to limit sequential glycosyl transfer reactions. *Biochemistry* **48,** 2807–2816 (2009).
28. Perez, C. *et al.* Structure and mechanism of an active lipid-linked oligosaccharide flippase. *Nature* **524,** 433–438 (2015).
29. Linton, D. *et al.* Functional analysis of the Campylobacter jejuni N-linked protein glycosylation pathway. *Mol. Microbiol.* **55,** 1695–1703 (2005).
30. Weerapana, E., Glover, K. J., Chen, M. M. & Imperiali, B. Investigating bacterial N-linked glycosylation: synthesis and glycosyl acceptor activity of the undecaprenyl pyrophosphate-linked bacillosamine. *J. Am. Chem. Soc.* **127,** 13766–7 (2005).
31. Liu, J. & Mushegian, A. Three monophyletic superfamilies account for the majority of the known glycosyltransferases. *Protein Sci.* **12,** 1418–31 (2003).
32. Kelleher, D. J. & Gilmore, R. An evolving view of the eukaryotic oligosaccharyltransferase. *Glycobiology* **16,** 47R–62R (2005).
33. Nasab, F. P., Schulz, B. L., Gamarro, F., Parodi, A. J. & Aebi, M. All in one: Leishmania major STT3 proteins substitute for the whole oligosaccharyltransferase complex in Saccharomyces cerevisiae. *Mol. Biol. Cell* **19,** 3758–68 (2008).
34. Ramírez, A. S. *et al.* Characterization of the single-subunit oligosaccharyltransferase STT3A from Trypanosoma brucei using synthetic peptides and lipid-linked oligosaccharide analogs. *Glycobiology* **27,** 525–535 (2017).
35. Lizak, C., Gerber, S., Numao, S., Aebi, M. & Locher, K. P. X-ray structure of a bacterial oligosaccharyltransferase. *Nature* **474,** 350–355 (2011).
36. Matsumoto, S. *et al.* Crystal structures of an archaeal oligosaccharyltransferase provide insights into the catalytic cycle of N-linked protein glycosylation. *Proc. Natl. Acad. Sci. U. S. A.* **110,** 17868–73 (2013).
37. Matsumoto, S., Taguchi, Y., Shimada, A., Igura, M. & Kohda, D. Tethering an N-Glycosylation Sequon-Containing Peptide Creates a Catalytically Competent Oligosaccharyltransferase Complex. *Biochemistry* **56,** 602–611 (2017).
38. Gerber, S. *et al.* Mechanism of bacterial oligosaccharyltransferase: In vitro quantification of sequon binding and catalysis. *J. Biol. Chem.* **288,** 8849–8861 (2013).
39. Lizak, C. *et al.* A catalytically essential motif in external loop 5 of the bacterial oligosaccharyltransferase PglB. *J. Biol. Chem.* **289,** 735–746 (2014).
40. te Heesen, S., Janetzky, B., Lehle, L. & Aebi, M. The yeast WBP1 is essential for oligosaccharyl transferase activity in vivo and in vitro. *EMBO J.* **11,** 2071–2075

(1992).
41. te Heesen, S., Knauer, R., Lehle, L. & Aebi, M. Yeast Wbp1p and Swp1p form a protein complex essential for oligosaccharyl transferase activity. *EMBO J.* **12,** 279–284 (1993).
42. Aebi, M. N-linked protein glycosylation in the ER. *Biochim. Biophys. Acta - Mol. Cell Res.* **1833,** 2430–2437 (2013).
43. Schwarz, M., Knauer, R. & Lehle, L. Yeast oligosaccharyltransferase consists of two functionally distinct sub-complexes, specified by either the Ost3p or Ost6p subunit. *FEBS Lett.* **579,** 6564–6568 (2005).
44. Wild, R. *et al.* Structure of the yeast oligosaccharyltransferase complex gives insight into eukaryotic N-glycosylation. *Science.* **359,** 545–550 (2018).
45. Bai, L., Wang, T., Zhao, G., Kovach, A. & Li, H. The atomic structure of a eukaryotic oligosaccharyltransferase complex. *Nature* **555,** 328–333 (2018).
46. Napiórkowska, M. *et al.* Molecular basis of lipid-linked oligosaccharide recognition and processing by bacterial oligosaccharyltransferase. *Nat. Struct. Mol. Biol.* **24,** 1100–1106 (2017).
47. Braunger, K. *et al.* Structural basis for coupling protein transport and N-glycosylation at the mammalian endoplasmic reticulum. *Science.* **360,** 215–219 (2018).
48. Shrimal, S., Cherepanova, N. A. & Gilmore, R. DC2 and KCP2 mediate the interaction between the oligosaccharyltransferase and the ER translocon. *J. Cell Biol.* **216,** 3625–3638 (2017).
49. Charnock, S. J. & Davies, G. J. Structure of the nucleotide-diphospho-sugar transferase, SpsA from Bacillus subtilis, in native and nucleotide-complexed forms. *Biochemistry* **38,** 6380–6385 (1999).
50. Vrielink, A., Rüger, W., Driessen, H. P. & Freemont, P. S. Crystal structure of the DNA modifying enzyme beta-glucosyltransferase in the presence and absence of the substrate uridine diphosphoglucose. *EMBO J.* **13,** 3413–22 (1994).
51. Marshall, R. D. Glycoproteins. *Annu. Rev. Biochem.* **41,** 673–702 (1972).
52. Gavel, Y. & Heyne, G. Von. Sequence differences between glycosylated and non-glycosylated Asn-X-Thr / Ser acceptor sites : implications for protein engineering. **3,** 433–442 (1990).
53. Petrescu, A. J., Milac, A. L., Petrescu, S. M., Dwek, R. A. & Wormald, M. R. Statistical analysis of the protein environment of N-glycosylation sites: Implications for occupancy, structure, and folding. *Glycobiology* **14,** 103–114 (2004).
54. Bause, E. Model studies on N-glycosylation of proteins. *Biochem. Soc. Trans.* **12,** (1984).
55. Bause, E. & Legler, G. The role of the hydroxy amino acid in the triplet sequence Asn-Xaa-Thr(Ser) for the N-glycosylation step during glycoprotein biosynthesis. *Biochem. J.* **195,** 639–644 (1981).
56. Kasturi, L., Chen, H. & Shakin-Eshleman, S. H. Regulation of N-linked core glycosylation: use of a site-directed mutagenesis approach to identify Asn-Xaa-Ser/Thr sequons that are poor oligosaccharide acceptors. *Biochem. J.* **323 (Pt 2,** 415–9 (1997).
57. Zielinska, D. F., Gnad, F., Wiśniewski, J. R. & Mann, M. Precision mapping of an in vivo N-glycoproteome reveals rigid topological and sequence constraints. *Cell* **141,** 897–907 (2010).
58. Breuer, W., Klein, R. A., Hardt, B., Bartoschek, A. & Bause, E. Oligosaccharyltransferase is highly speci ¢ c for the hydroxy amino acid in. *FEBS Lett.* **501,** 106–110 (2001).
59. Jaffee, M. B. & Imperiali, B. Exploiting topological constraints to reveal buried sequence motifs in the membrane-bound N-linked oligosaccharyl transferases.

Biochemistry **50,** 7557–67 (2011).
60. Lizak, C. *et al.* Unexpected reactivity and mechanism of carboxamide activation in bacterial N-linked protein glycosylation. *Nat. Commun.* **4,** 2627 (2013).
61. Bock, C. W., Katz, A. K., Markham, G. D. & Glusker, J. P. Manganese as a replacement for magnesium and zinc: Functional comparison of the divalent ions. *J. Am. Chem. Soc.* **121,** 7360–7372 (1999).
62. Glusker, J. P. Structural aspects of metal liganding to functionla groups in proteins. *Adv. Protein Chem.* **42,** 1–76 (1991).
63. Imperiali, B. & Tai, V. W. Chemistry and Biochemistry of Asparagine-Linked Protein Glycosylation. 281–303 (2005).
64. Tai, V. W. F. & Imperiali, B. Substrate specificity of the glycosyl donor for oligosaccharyl transferase. *J. Org. Chem.* **66,** 6217–6228 (2001).
65. Ihssen, J. *et al.* Increased efficiency of Campylobacter jejuni N-oligosaccharyltransferase PglB by structure-guided engineering. *Open Biol.* **5,** 140227 (2015).
66. Garcia-Quintanilla, F., Iwashkiw, J. A., Price, N. L., Stratilo, C. & Feldman, M. F. Production of a recombinant vaccine candidate against Burkholderia pseudomallei exploiting the bacterial N-glycosylation machinery. *Front. Microbiol.* **5,** 1–10 (2014).
67. Pedebos, C., Arantes, P. R., Giesel, G. M. & Verli, H. *In silico* Investigation of the PglB Active Site Reveals Transient Catalytic States and Octahedral Metal Ion Coordination. *Glycobiology* **25,** 1183–1195 (2015).

Chapter 2: Screening for conformational nanobodies against oligosaccharyltransferase PglB

Abstract

Large hydrophobic domains make membrane proteins challenging crystallization targets. Protein- specific binders, such as nanobodies, can not only increase the soluble surface of the protein, and therefore improve lattice contacts, but can also stabilize flexible regions, locking them in specific conformations. Such chaperons could be useful tools to improve crystallization of the oligosaccharyltransferase PglB, and therefore help to obtain high-resolution structures in the presence and absence of substrates. This chapter describes the full process of generating and screening for conformational nanobodies against PglB. We identified 29 binders, 6 of which thermostabilized PglB. Although none of the nanobodies identified improved PglB crystallization, the thermostabilizing nanobodies could be used in the future for NMR or cryo-EM studies.

Introduction

To identify residues and interactions involved in the activation of the amido group of the acceptor asparagine, high-resolution X-ray crystal structures are required. The production of diffraction-quality crystals remains the major bottleneck in macromolecular X-ray crystallography. Only well- ordered crystals with strong crystal contacts are suitable for high resolution X-ray analysis. Inherent flexibility of protein domains and/or inadequate surface interactions between protein molecules may result in low quality crystals.
Conformation-specific binders, such as nanobodies or Fab fragments, can facilitate protein crystallization. PglB has a large hydrophobic transmembrane domain and a flexible loop (EL5) which limits the formation of well-diffracting crystals. Using binders during crystallization could help increase the polar surface area for protein-protein contacts and restrict the flexibility of EL5.
Nanobodies, which naturally occur in Camelids, are small (15 kDa), antibody-derived single domain proteins that contain the unique structural and functional properties of the antibody heavy chain[1]. Nanobodies can be used as chaperones for crystallizing membrane proteins[2,3], stabilizing transient conformational states[2] or intrinsically disordered proteins[4]. They have been shown to selectively recognize and stabilize proteins in native conformations[5,6].
In this chapter, the screening of conformation specific nanobodies for PglB and PglB-Nb cocrystallization experiments are described.

Structural features of heavy chain antibodies and nanobodies
The classical immunoglobulin-γ (IgG) antibodies are assembled from two identical heavy (H)-chain and two identical light (L)-chain polypeptides[7] (Fig. 1). The L chain contains two domains, whereas the H chain folds into four domains. The sequences of the N-terminal domain of the H and L polypeptide chains varies between antibodies. These variable domains are designated as VH and VL, respectively. The corresponding VH-VL domains constitute the variable fragment (Fv) that recognizes the antigen[7,8]. Sequence variability within V domains is localized in three hypervariable regions called complementarity-determining regions (CDRs), that form a surface complementary to the surface of the epitope. CDRs are surrounded by more

conserved framework regions (FR)[1,7] (Fig. 2). The remaining H and L sequences (abbreviated as CH and CL, respectively) are more conserved. The two last CH regions are important for recruitment of immune cells (e.g., macrophages and natural killer cells) or for effector functions (e.g., complement activation)[8].

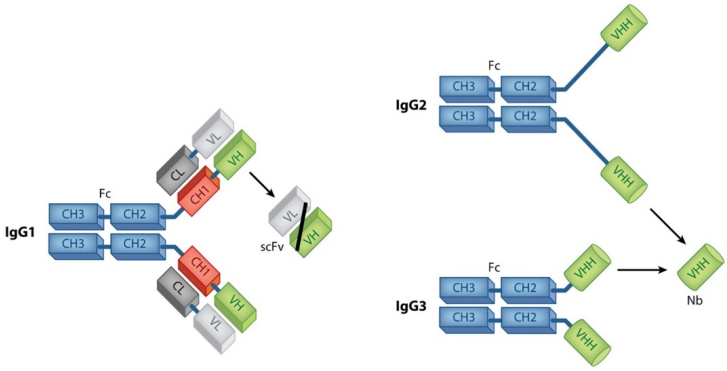


Figure 1. Schematic representation of conventional and heavy-chain only antibodies. Conventional antibodies (IgG1) contain two light (L) chains (the VL and CL domains) and two heavy (H) chains (composed of the VH, CH1, hinge, and CH2 and CH3 domains). The two types of homodimeric heavy-chain antibodies, IgG2 and IgG3, comprise only heavy chains, where each contains a VHH, hinge, and CH2 and CH3 domains. The hinge of the IgG2 fraction is longer than that of the IgG3 type. Abbreviations: CH, constant domain of immunoglobulin heavy chain; Fab, antigen-binding fragment; Fc, crystallizable fragment; scFv, single-chain variable fragment. The figure was taken from Muyldermans, S. *et al.* (2013)[8].

In sera of Camelidae, in addition to the conventional heterotetrameric antibodies, special IgG antibodies known as heavy-chain antibodies were found[1]. The two constant domains (CH2–CH3) are highly homologous to the Fc domains of classical antibodies. Heavy-chain antibodies lack CH1 and the domain from the L chain polypeptide (Fig. 1)[9], but their variable domain and the hinge are intact. This is substantially different to pathogenic human/mouse heavy chain antibodies, where various parts of the VH, CH1, and hinge are deleted[10].

The shorter distance between the two paratopes within one heavy chain antibody might compromise their capacity to cross-link antigens. However, it has been speculated that the long IgG2 hinge containing Pro-Gln repeats might form an extended spacer that structurally replaces the CH1 region[1,11]. At the N-terminal region of heavy chain antibodies, the H chain of the homodimeric protein contains a dedicated variable domain (VHH) which serves to associate with an antigen. The VHH, also called a nanobody, is the structural and functional equivalent of the Fab fragment (antigen-binding fragment) of conventional antibodies[8] (Fig. 1).

The structural organization of the VHH and VH is similar, however there are a few notable differences that explain the autonomous antigen-binding capacity of the VHH's. Firstly, because the VHH acts independently (without VL, as seen in conventional antibodies), the antigen is recognized by only three instead of six loops. Possibly, to provide a sufficiently large antigen-binding surface, the first and the third hypervariable loops (CDR1 and CDR3) are longer in a VHH than in a VH of a conventional antibody. In Camelidae the CDR3 loops of VHHs are almost twice as long as those in VHs[10,12].

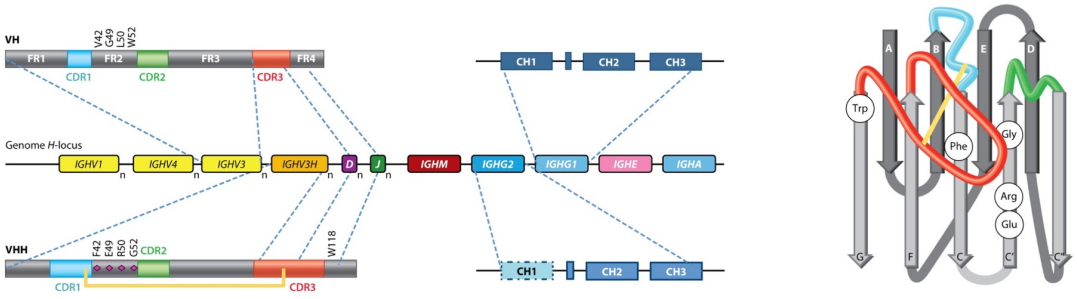


Figure 2. Schematic representation of the *H* locus in the genome of camelids. To form the VH domain (top left), during B-cell lymphopoiesis, an IGHV3 gene rearranges to one D and one J element (dotted lines) to form the VH domain. CDR 3 is formed by the IGHV3-D-J junction. To assemble the conventional antibody (IgG1), this rearranged gene is first expressed from an upstream promoter as part of an IGHM gene and after a class switch, it is coexpressed with the dedicated IGHG1 genes (with a CH1, hinge, and CH2 and CH3 exons; top left). In other B cells, the IGHV3H is rearranged to one D and one J element from the same locus (dotted lines) to form a VHH domain (below the genome H locus) and is eventually coexpressed with a an HCAb-dedicated IGHG2 (or IGHG3) gene (below the genome H locus). The CH1 exon region is eliminated during mRNA splicing. The VH and VHH domains are schematically shown with their framework (FR) and CDRs. A folded VHH domain is schematically shown on the right. The key amino acids of a VH that interact with a VL and that are substituted in a VHH are shown. The CDRs are color coded as in the aligned panel above. The figure was taken from Muyldermans, S. *et al.* (2013)[8].

Finally, most of the dromedary VHHs contain four Cys, of which two are conserved between VH and VHHs. The additional two Cys are located in the CDR1 and CDR3 and form an interloop disulfide[8,13,14] (Fig. 2). The generation of the non- canonical disulfide bonds restricts the flexibility of the loops in absence of the antigen and increases the stability of the domain[8,14]. From crystal structures of VHH we know that these Cys form a disulfide bond that possibly assists in shaping the loop structure[14,15].
Another striking distinction between VH and VHH occurs in FR2, where highly conserved hydrophobic amino acids (Val42, Gly49, Leu50, Trp52), that in conventional antibodies participate in the VH/VL interaction[10,16], are replaced in a VHH by smaller and/or hydrophilic amino acids (mostly Phe42, Glu49, Arg50, Gly52)[17] (Fig. 2). These amino acid substitutions in FR2 of the nanobody prevent any VL interactions and increases its solubility[8].
The use of nanobodies as protein-specific binders has several significant advantages. In addition to their small size, ease of production and handling, VHHs are more stable and robust under stringent conditions (resist chemical and thermal denaturation) than conventional antibody fragments[18]. Moreover, they are often found to have a more compact architecture and unusual epitope specificities, particularly an improved ability to bind active site pockets to inhibit enzymes [8,19,20] or mediate the inhibition of transporter proteins[21,22].

Experimental Procedures

Materials

All materials were obtained from AppliChem or Sigma unless stated otherwise. Restriction enzymes were purchased from New England Biolabs and Fermentas.

Sample preparation for the immunization of alpacas

Cloning, expression and purification of a cross-linked PglB double cysteine mutant F308C/ K528C

The gene encoding the *C. lari* PglB protein carrying the K2E, C17A, C30A, F308C, C350S, K528C) (Pos 3, Fig. 3a) mutations was cloned into a modified pBAD (Invitrogen) expression plasmid with a C-terminal decahistidine affinity tag fused to PglB [23]. A C-terminal 3C protease cleavage site was introduced at the EcoRI restriction site using oligo duplex cloning (Fig 3b). Expression and purification of the resulting PglB double cysteine mutant was done following a previously described protocol[23]. Briefly, PglB was overexpressed in *Escherichia coli* BL21-Gold cells (DE3) (Strategene) at 37 °C in five-liter flasks using Terrific Broth medium supplemented with 1 % glycerol (w/v). The cells were induced at 37 °C for 4 h at an A600 of 3.0 by adding 0.1 % arabinose (w/v). Cells were harvested by centrifugation and cell pellets were stored at −80 °C. All following steps were carried out at 4 °C. Cells were resuspended in 25 mM Tris-HCl, pH 8.0, 250 mM NaCl, 10 mM β-mercaptoethanol and disrupted in a M-110-l microfluidizer (Microfluidics) at 15,000 p.s.i. chamber pressure. Membranes were pelleted by centrifugation at 100,000 g for 30 min and solubilized in 25 mM Tris-HCl, pH 8.0, 250 mM NaCl, 10 % glycerol (v/v), 10 mM β-mercaptoethanol and 1 % N-dodecyl-β-D-maltopyranoside (w/v) (DDM, Anatrace) for 1.5 h. PglB was purified on a nickel-nitrilotriacetic acid (Ni-NTA) Superflow affinity column (Qiagen). The Ni-NTA bound protein was washed with 20 column volumes of 25 mM Tris-HCl, pH 8.0, 250 mM NaCl, 60 mM imidazole-HCl pH 8.0, 10 % glycerol (v/v), 0.016 % DDM buffer, followed by 5 column volumes of imidazole-free wash and a 5 column volumes of 25 mM Tris-HCl, pH 8.0, 250 mM NaCl, 10 % glycerol (v/v), 0.016 % DDM supplemented with 1 mM $CuCl_2$.

After 15-min incubation at 4 °C, $CuCl_2$ was washed out, and the protein was eluted in 25 mM Tris-HCl, pH 8.0, 250 mM NaCl, 200 mM imidazole-HCl, pH 8.0, 10 % glycerol (v/v), 0.016 % DDM buffer and desalted in 25 mM Tris-HCl, pH 8.0, 250 mM NaCl, 10 % glycerol (v/v), 0.016 % DDM. The histidine tag was cleaved overnight at 4 °C with 3C protease in a 1:20 w/w ratio. The reaction mix was supplemented with 20 mM imidazole-HCl pH 8.0 and cleaved decahistidine tag and 3C protease were removed by rebinding to NiNTA. Tagless crosslinked PglB was desalted into 10 mM MES-NaOH, pH 6.5, 100 mM NaCl, 3% glycerol (v/v), 0.5 mM EDTA and 0.016% DDM, concentrated to 5 mg mL^{-1} in an Amicon Ultra-15 concentrator (Milipore) with a molecular mass cutoff of 100 kDa and purified by size exclusion chromatography (Superdex 200 10/300, GE Healthcare) in desalting buffer.

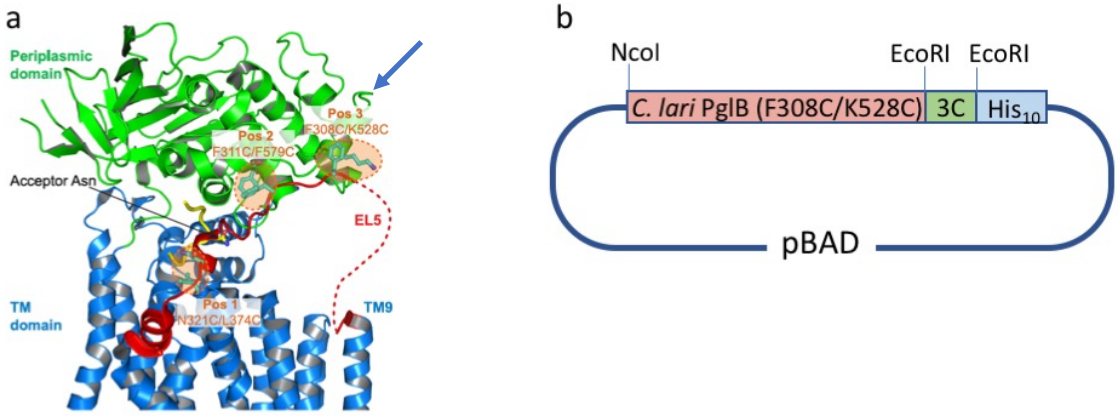


Figure 3. Sample preparation for alpaca immunization. (**a**) Schematic representation of *C. lari* PglB (PDB code 3RCE) showing the localization of the engineered disulfide cross-link at position 3 (marked with an arrow). The transmembrane domain is shown in blue, the periplasmic domain in green, EL5 in

red and bound acceptor peptide in yellow. Adapted from Lizak *et al.* (2014)[23]. (**b**) PglB expression construct. The 3C protease recognition sequence was introduced at the EcoRI restriction site using oligo duplex cloning.

Reconstitution of the PglB double cysteine mutant F308C/ K528C into proteoliposomes
An *E. coli* polar lipids: phosphocholine (ECPL:PC) (Avanti Polar Lipids) stock solution was prepared at a 3:1 (w/w) ratio as described previously[24]. The lipid mixture was finally resuspended at 20 mg mL^{-1} in a buffer containing 10mM MES pH 6.5 and 100mM NaCl, aliquoted and stored at -80 °C. The ECPL:PC mixture was extruded 13 times through a 400 nm filter and destabilized with 0.3 % (v/v) Triton X-100. Detergent-purified PglB was concentrated to 0.8 mgL^{-1} and then mixed with ECPL:PC at a 9:1 (w/w) lipid: protein ratio for 1h at room temperature. Detergent was removed with Bio-Beads (40 mg L^{-1} of mixture) in 4 steps. The proteoliposomes were spun at 100,000g, resuspended in 10 mM MES pH 6.5, 100 mM NaCl at a final lipid concentration of 20 mg mL^{-1}, and the reconstitution efficiency was determined by the protein assay of Schaffner and Weissman[25].

Generation of nanobodies against the cross-linked double cysteine mutant F308C/ K528C of PglB

Preparation of VHH library
An alpaca (Spot) was immunized with a mixture of PglB and two other target proteins following the protocols established previously[6,26]. 200 μg of the cross-linked PglB double cysteine mutant F308C/ K528C, reconstituted into proteoliposomes, was injected into the alpaca four times at fortnightly intervals and the immune response was monitored by ELISA (Saša Štefanić, Institute of Parasitology, University of Zurich). 10 days after the last immunization, 50 ml of blood was collected for subsequent peripheral blood lymphocyte (PBLs) isolation using Histopaque-1077 (Sigma) and ACCUSPIN tubes (Sigma) according to the manufacturer's instrucutions. The cells were aliquoted (~3×10^7 PBLs in each tube). The extraction of the total RNA was performed from one PBL aliquot and the rest was flash frozen in liquid nitrogen and stored at -80 °C. The extraction of the total RNA was performed using an RNeasy Mini Kit (Qiagen) according to manufacturer's instructions and the yield was calculated in a spectrophotometer at wavelengths of 260 nm and 280 nm. First strand cDNA synthesis was performed using the SuperScript®III reverse transcriptase system (ThermoFisher Scientific) and CALL002 primer: 5'- GGTACGTGCTGTTGAACTGTTCC-3'[6]. PCR amplification of the variable domains of all immunoglobulin heavy chains (VHs and VHHs) from the cDNA was done using KAPA Taq DNA polymerase (Sigma), the forward primer 5'- CTGARCTKGGTGGTCCTGGCTGC-3' (where R=A/G and K=G/T)[26,27], that anneals in the leader fragment and reverse primer mix (1:1 molar ratio) of 5'- ATGGAGAGGACGTCCTTGGGT -3' and 5'- TTCGGNGGGAAGAYRAAGAC -3' (where R= A/G, Y=C/T and N=A/T/C/G), that anneals in CH2 fragment[26] (Fig. 4). The PCR products were separated by electrophoresis on a 1% (w/v) TBE agarose gel and the correct band corresponding to VHH (600 bp) was cut out and gel- extracted using the QiAquick (Qiagen) extraction kit according to manufacturer's instructions. The variable domain of the heavy chain only antibody (VHH) was further amplified and the FX cloning sites[28] were added via second nested-PCR using KAPA Taq DNA polymerase (Sigma), 20ng of purified DNA from 1st PCR and primers: forward primer 5' - ATATATGCTCTTCGGCACAGNKGCARYTSGTRGAGTCTGGGGG-3' and reverse primer 5'- ATATATGCTCTTCCACTGGAGACGGTGACCSGGGTCCCYTKGCY-3', that anneal at framework 1 and framework 4, respectively [26,27]. The VHH fragments were then cloned by fragment exchange cloning[28] into the *Pml*I-linearized pDX phagemid vector using

SapI (New England Biolabs), a class IIS restriction enzyme, and T4 ligase (New England Biolabs).

The ligation mixtures carrying the VHH DNA library were purified using the Qiaquick PCR purification kit (Qiagen). 10 aliquots of 25 µl of electrocompetent TG1 cells were then transformed with 50-100 ng of the purified ligation mixture. Glycerol stocks of the recovered TG1 cells (VHH library) were stored at -80 ºC. The size of the library was determined by preparing serial dilutions.

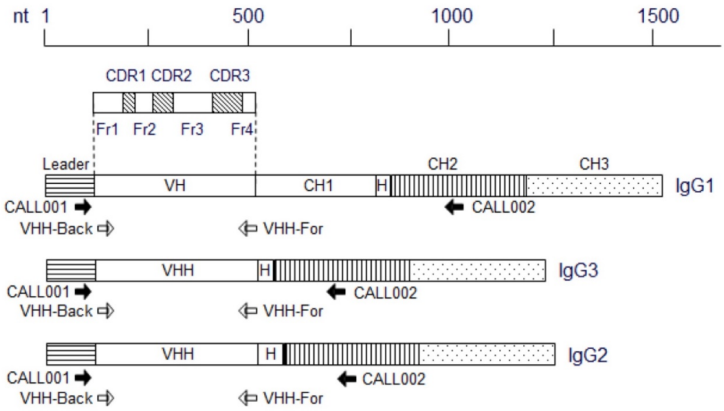


Figure 4. Strategies to amplify the nanobody repertoire by PCR from cDNA. For each heavy chain antibody (IgG1, IgG2 and IgG3) the locus of hybridization of the primers on the cDNA is shown. Primers CALL001 and CAL002 were used to amplify the variable domains of all camelid immunoglobulin heavy chains from cDNA, while the primers VHH-Back and VHH-For were used to amplify the nanobody repertoire via nested PCR. The figure was adapted from Pardon *et al*. (2014)[6].

Sample preparation for phage display- PglB biotinylation

In order to obtain an Avi-tagged PglB construct, the EcoRI restriction site between the 3C and decahistidine tag in the PglB double cysteine mutant F308C/K528C-pBAD was removed by Quikchange PCR. Subsequently, the Avi tag was cloned by oligoduplex annealing into the EcoRI-linearized PglB double cysteine mutant F308C/ K528C-pBAD plasmid between the C-terminal end of PglB and the 3C cleavage site (Fig. 5). The construct was expressed and the membranes were prepared and solubilized as described above. PglB was purified on a nickel-nitrilotriacetic acid (Ni-NTA) Superflow affinity column (Qiagen). The Ni-NTA bound protein was washed with 20 column volumes of 25 mM Tris-HCl pH 8.0, 250 mM NaCl, 60mM imidazole-HCl pH 8.0, 10 % glycerol (v/v), 0.016 % DDM buffer and the protein was eluted in 25 mM Tris-HCl pH 8.0, 250 mM NaCl, 200 mM imidazole-HCl pH 8.0, 10 % glycerol (v/v), 0.016 % DDM buffer and desalted in 50 mM bicine pH 8.0, 100 mM NaCl, 0.016 % DDM. Biotinylation was performed according to manufacturer's instructions (Avidity, protocol for the BirA enzyme). The buffer was supplemented with 10 mM ATP, 10 mM magnesium acetate, 20 µM biotin (Sigma), and BirA and 3C protease were added to 5 µM PglB in a 1:2 (w/w) and 1:20 (w/w) ratio, respectively. The reaction was incubated overnight at 4 °C. The reaction mix was supplemented with 20 mM imidazole pH 8.0 and the decahistidine tag, BirA and 3C protease were removed by rebinding to NiNTA. Tagless biotinylated PglB was concentrated and purified by size exclusion chromatography (Superdex 200 10/300, GE Healthcare) in 25 mM Tris-HCl, pH 8.0, 250 mM NaCl, 5 % glycerol (v/v), 0.016 % DDM. Biotinylation efficiency was determined by binding to Streptavidin resin (Sigma).

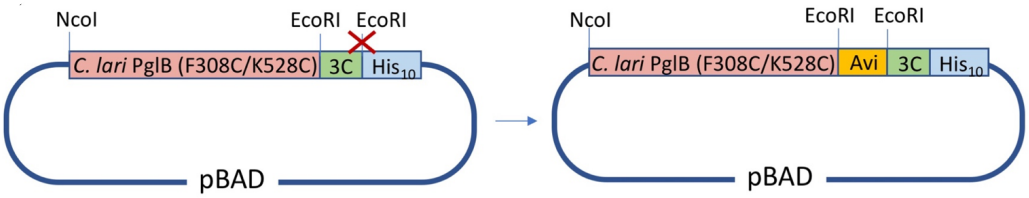


Figure 5. Avi-tagged PglB expression construct. The Avi tag sequence was introduced at the EcoRI restriction site using oligo duplex cloning.

Phage display: antigen presentation and phage selection by panning

Before each round of panning, the nanobody-displaying phage particles were rescued and amplified by adding helper phage according to the standard rescue protocol[6]. To estimate the number of infective recombinant phage (size of the phage library), the rescued phage was titrated by preparing serial tenfold dilutions and infecting fresh TG1 cells at OD_{600}=0.7.

22 µg of biotinylated PglB, buffer and 16 µg Stt3 protein from *T. brucei*, used as negative controls, were immobilized on a preblocked NeutrAvdin coated plate (Thermo Fisher Scientific) overnight at 4 °C. The unbound protein was washed out and the plate was blocked with PBS buffer containing 2 % milk and 0.016 % DDM for 2 h at RT. After removing blocking solution and 5 washing steps the plate was incubated for 30 min at RT with 10^{11} phages in PBS buffer with 1 % milk and 0.016 % DDM. After 15 washing steps the elution of bound phages in each selection well was performed with 0.25mg mL^{-1} trypsin (Sigma). The eluted phages were directly transferred into microcentrifuge tubes prefilled with 5 µl of a 5mg mL^{-1} AEBSF (Sigma) solution to inhibit protease activity. 50 µl of eluted phages from each well were recovered by infecting 350 µL of TG1 cells for 30 min at 37 °C. 4.5 mL of 2xYT medium supplemented with 100 µg mL^{-1} ampicillin and 2 % of glucose (w/v) was then added and the cells were grown overnight at 37 °C. The selected sublibraries were stored as glycerol stocks at -80 °C. To estimate the number of infective recombinant phage that eluted from each well, the remaining eluted phage were titrated as described above. Two rounds of panning were performed in total, although one round of selection was sufficient to detect enrichment and identify the binders. After the first round of panning 100 individual clones from the enriched PglB sub-library were sent for DNA sequencing (Microsynth).

Expression and purification of nanobodies

Nanobodies were overexpressed in *E. coli* WK6 cells. Cells were grown at 37 °C in five-liter flasks using Terrific Broth medium supplemented with 1% glucose (w/v) and 1mM $MgCl_2$ to OD600=0.7. The cells were induced at 25 °C overnight by adding 1mM IPTG. Cells were harvested by centrifugation and cell pellets from 1 liter of culture were resuspended in 15 ml of ice- cold TES buffer composed of 0.2 M Tris-HCl, pH 8.0, 0.5 M sucrose, 0.5 mM EDTA, and incubated on a shaking platform for 1 h at 4 °C. The cell suspension was diluted with double the volume of TES/4 buffer and incubated for another hour at 4 °C. The suspension was centrifuged for 30 min at 10,000g and 4 °C and the supernatant with periplasmic extract was recovered. Nanobodies were purified on a NiNTA Superflow affinity column (Qiagen) and desalted with PD10 columns in 20 mM Tris-HCl, pH 8.0, 250 mM NaCl, 10% glycerol (v/v).

Pull down assay

Sequence-optimized PglB was overexpressed as described before[29]. Protein was desalted in a buffer containing 25 mM Tris-HCl, pH 8.0, 250 mM NaCl, 10 % glycerol (v/v) and 0.016 % DDM. The histidine tag was cleaved overnight at 4 °C with 3C protease in a 1:20 w/w ratio. The reaction mix was supplemented with 20 mM imidazole-HCl pH 8.0 and the cleaved

decahistidine tag and 3C protease were removed by rebinding to NiNTA. Tagless PglB was desalted again into 25 mM Tris-HCl, pH 8.0, 250 mM NaCl, 10 % glycerol (v/v) and 0.016 % DDM. Tagless PglB (0.2 mg mL^{-1}) was incubated with a 2-fold molar excess of histidine-tagged nanobody at 4 °C for 30 min in presence of 25mM imidazole-HCl, pH 8.0. The mixture was then incubated with equilibrated NiNTA at 4 °C for 1 h. The Ni-NTA bound protein was washed with 25 CV of 25 mM Tris-HCl, pH 8.0, 40 mM imidazole-HCl, pH 8.0, 250 mM NaCl, 10 % glycerol (v/v) and 0.016 % DDM. The bound protein was eluted with 500mM imidazole pH 8.0 and all collected fractions were analyzed by SDS-PAGE.

Gel filtration binding assay

Purified sequence-optimized PglB was desalted into 25 mM Tris-HCl, 8.0, 150 mM NaCl, 3 % glycerol (v/v) and 0.016 % DDM and incubated with a 2-fold molar excess of nanobody for 1 h at 4 °C. The mixture was purified by size exclusion chromatography using a Superdex 200 10/30, GE Healthcare column equilibrated with 25 mM Tris-HCl, 8.0, 150 mM NaCl, 3 % glycerol (v/v) and 0.016 % DDM. The main peak fractions were collected and analyzed by SDS-PAGE.

***In vitro* glycosylation assay**

In vitro glycosylation assays were performed as described previously[30,31], using a labeled acceptor peptide 5-carboxyfluorescein-GS-DQNATF-NH2 (10 μM) as an acceptor substrate and wild-type LLO (50 μM) or farnesyl/nerylneryl-PP-GlcNAc (500 μM) as a donor substrate. Reaction mixtures with wild-type LLO and nerylneryl-PP-GlcNAc contained 1 nM or 100 nM PglB, respectively. Prior to the reaction PglB was preincubated with a 2-fold molar excess of nanobody. For determination of the turnover rate, a total of six samples were taken at different time intervals so that the reaction was in the linear range. Data was fitted by linear regression in GraphPad PRISM 7.0.

Size-exclusion chromatography thermostability (SEC-Ts) assay

Sequence-optimized PglB was overexpressed and purified as described before. Protein was desalted into 10 mM MES-NaOH, pH 6.5, 100 mM NaCl, 3% glycerol, 0.016% DDM. The size exclusion chromatography thermostability (SEC-Ts) of sequenced-optimized PglB was calculated to be 35 °C. In order to screen for thermostabilising nanobodies, 5 μM of PglB was preincubated with a 2-fold molar excess of nanobody at 4 °C and then mixtures were incubated at 35 °C. For thermostabilizing nanobodies samples were incubated at different temperatures (4, 20, 30, 35, 40, 45, 50, 55, 60 and 70 °C) for 10 min. The samples were cooled on ice and centrifuged for 30 min at 120, 000 g, and the supernatants were analyzed by size-exclusion chromatography using a TSK-gel G3000SWXL column equilibrated with 10 mM MES-NaOH, pH 6.5, 100 mM NaCl, 3 % glycerol, 0.016 % DDM. For each temperature, the height of the main peak was plotted against temperature. The data was then fitted using the 'sigmoidal dose-response (variable slope)' function in Graph Pad Prism 7.0 and the thermostability (Ts) of each sample was calculated.

MST nanobody-binding affinity assay

Nanobody 64 and Nanobody 13 (carrying 5 and 6 lysine residues, respectively) were labeled using the Nanotemper labeling kit RED-NHS according to manufacturer's protocol. The NT-647-NHS reactive dye contains NHS-ester chemistry, which reacts efficiently with primary amines of proteins (lysine residues) to form highly stable dye-protein-conjugates. Shortly, 20 μM Nb was incubated with a 4-fold molar excess of the NT-647 dye for 30 min at RT, then desalted into 10mM MES-NaOH, 100mM NaCl. The labeling efficiency was calculated by measuring the absorbance at 280nm for the protein and 650 nm for the dye. Sequence-optimized

PglB was purified as described previously and desalted into 10 mM MES-NaOH, pH 6.5, 100 mM NaCl, 3% glycerol, 0.016% DDM. For Microscale-Thermophoresis (MST) experiments PglB was concentrated to a final concentration of 30 µM and fifteen PglB titration series (1:1) were prepared. Each PglB dilution was incubated with a constant concentration of the fluorescently-labeled nanobody (100 nM and 200 nM for Nb64 and Nb13, respectively). The MST measurements were run at 10 °C. K_d data were fitted with the formula according to law of mass action:

$$f(c) = u + \frac{b-u}{2} \times (F + c + K_d - \sqrt{(F + c + K_d)^2 - 4F \times c}),$$

where *f(c)* is the normalized fluorescence signal dependent on the concentration of the unlabeled binding partner *c*, *F* is the concentration of the fluorescent binding partner, u and b are the normalized fluorescence signals of unbound and bound states, respectively, and K_d is the dissociation constant.

Crystallization of PglB-Nb complexes

Sequence-optimized PglB was overexpressed and purified as described before. Protein was desalted into 10 mM MES-NaOH, pH 6.5, 100 mM NaCl, 3 % glycerol, 0.016 % DDM. For crystallization experiments, desalted protein was concentrated to 6 mg mL−1 in an Amicon Ultra-15 concentrator (Milipore) with a molecular mass cutoff of 100 kDa and purified by size exclusion chromatography (Superdex 200 10/300, GE Healthcare) in desalting buffer. Collected fractions were pooled and concentrated to 12 mg mL−1. All nanobodies identified as PglB binders were concentrated to 500 µM. PglB was incubated with 0.75 mM DQNATF-p{NO_2}, 2 mM $MnCl_2$, 2 mM nerylneryl-PPC-GlcNAc and a 1.2-fold excess of Nb for 30 min at 4 °C. Crystallization experiments were performed at 20 °C with home-made 96-well screens using 1:1 and 2:1 protein to well solution ratios. The crystals were reproduced in 24-well screens. The crystals appeared in conditions containing low-molecular weight PEG as precipitants and those PEGs were used for cryoprotection. The crystals were screen at the Swiss Light Source at PSI, Villigen.

Results

Sample preparation for alpaca immunization

For immunization of alpacas we used an engineered PglB construct from *C. lari* with three naturally occurring cysteines (C17A, C30A, C350S) removed and with 2 new cysteines (F308C, K528C) introduced[23](Fig. 3a). The PglB mutant was designed to restrict the motion of EL5 (Gly^{282}-Gln^{330}) by formation of a disulfide bond between cysteine residues[23]. By using this construct for alpaca immunization, we wanted to increase the chances of generating nanobodies directed against the flexible part of PglB. The cross-linked mutants eluted as monodisperse peaks on size-exclusion chromatography (Fig. 6a). Cross-linking resulted in a size shift after SDS-PAGE, and cross-linking efficiencies were calculated from band intensities of Coomassie-stained gels (Fig. 6b) (determined to be 95%). The final concentration of cross-linked PglB in proteoliposomes was estimated as 1.1 mg/ml. We chose proteoliposome-reconstituted over detergent-solubilized protein as PglB is likely to be more stable in proteoliposomes, enabling an immune response against folded protein.

We confirmed that glycosylation activity of the double cysteine mutant is only slightly impaired by formation of the disulphide bond, which agrees with experiments described in the literature[23]. The cross-linked double cysteine mutant, in oxidizing conditions, showed a glycosylation activity of 0.88 ±0.01 peptide/s with WT LLO and 0.29 ±0.05 peptide/min with

farnesyl-PP-GlcNAc, which is only ~1.5-fold slower when compared with the activity of the same sample in reducing conditions. Therefore, although the motion of EL5 of cross-linked PglB is restricted, glycosylation activity is maintained.

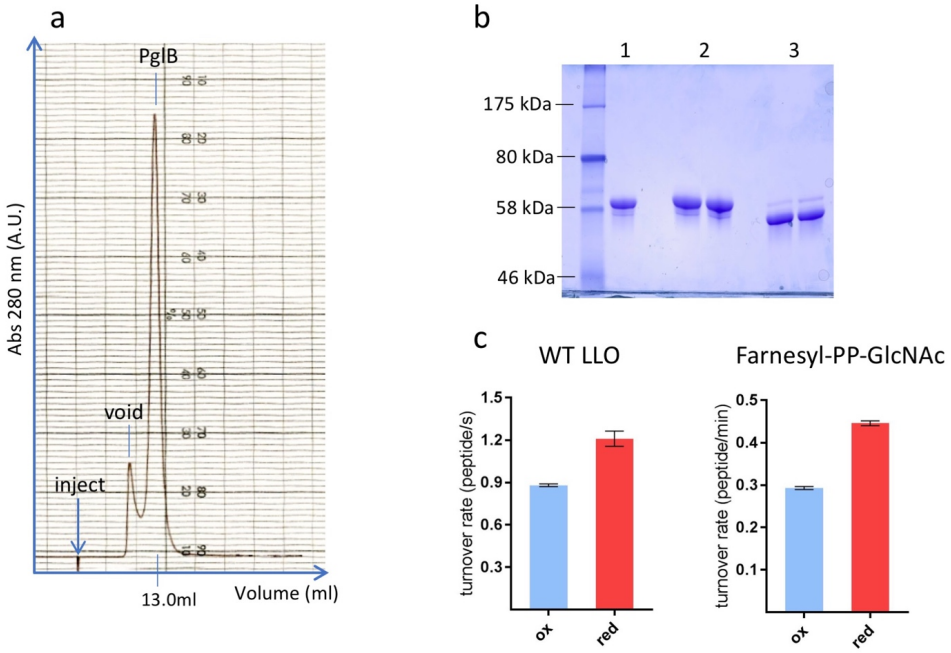


Figure 6. Sample preparation for alpaca immunization. (**a**) Size exclusion chromatogram of cross-linked PglB. (**b**) SDS-PAGE analysis of cross-linked PglB, where lane 1 corresponds to the PglB sample after preparative SEC, 2 to proteoliposome- reconstituted PglB in oxidizing conditions and 3 to proteoliposome- reconstituted PglB in reducing conditions (in the presence of 80 mM β-mercaptoethanol). (**c**) Glycosylation activity of the PglB F308C K528C double mutant using 50 μM wild-type LLO (WT, left) and 500 μM synthetic farnesyl-PP-GlcNAc (right). Bars labeled "ox" (oxidized) depict turnover rates of disulfide cross-linked samples, whereas bars labeled "red" (reduced) depict turnover rates of samples containing 80 mM β-mercaptoethanol, breaking the disulfide bond. Data in c represent three independent cell cultures (error bars indicate s.d., n = 3).

Selection of PglB-specific nanobodies.

The immune response of the alpacas was monitored at two weeks intervals by ELISA (Saša Štefanić, Institute of Parasitology, University of Zurich) (Fig. 7). The alpaca's immune serum was checked for response in three IgG classes: IgG1, IgG2 and IgG3, where IgG2 and IgG3 are the heavy-chain antibodies. According to the literature the molar ratio of IgG1:IgG2:IgG3 in alpaca serum should average 50:30:20[26]. The IgG1:IgG2:IgG3 ratio in the serum of the alpaca that was immunized with the cross-linked PglB sample was estimated at 31:24:45, which is an exceptionally good response in the heavy-chain class (Fig. 7).

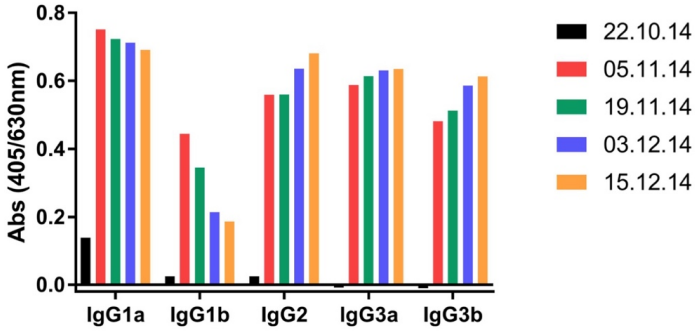


Figure 7. Alpaca's immune response analyzed by ELISA. The alpaca's immune serum was checked for response in three IgG classes: IgG1, IgG2 and IgG3 in two weeks intervals, where 22.10.14 was the day of the first injection.

The nanobody-cDNA encoding library was prepared from lymphocytes isolated from the alpaca's blood. The variable domains of all immunoglobulin heavy chains were PCR amplified resulting in two PCR products of 900 bp and 600 bp for conventional antibodies (VH) and heavy chain only antibodies (VHH), respectively (Fig. 8a). The nested-PCR of the variable domain of the heavy chain only antibody (VHH) was successfully completed (Fig. 8b). The final immune library consisted of 8×10^7 transformants per mL.

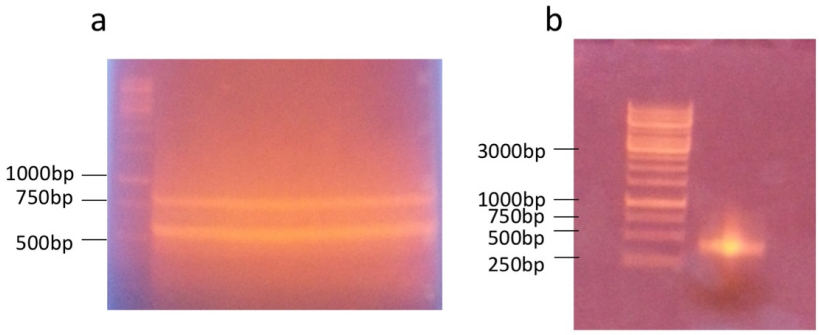


Figure 8. Agarose- gel DNA electrophoresis of amplified variable domains of heavy chains. (**a**) The gel showing the product of the first PCR with DNA repertoire of both conventional antibodies VH (band around 900 bp) and heavy chain only antibodies VHH (band around 600 bp). (**b**) The analysis of the second PCR product showing only one band of the DNA repertoire of the variable domains of heavy chain only antibodies.

The library was panned against PglB using phage display. For the panning experiments the C-terminally Avi-tagged cysteine double mutant of PglB was successfully purified and biotinylated (Fig. 9). The biotinylated PglB variant was stable and showed a monodisperse peak on size exclusion chromatography 24 h after purification (Fig. 9a).

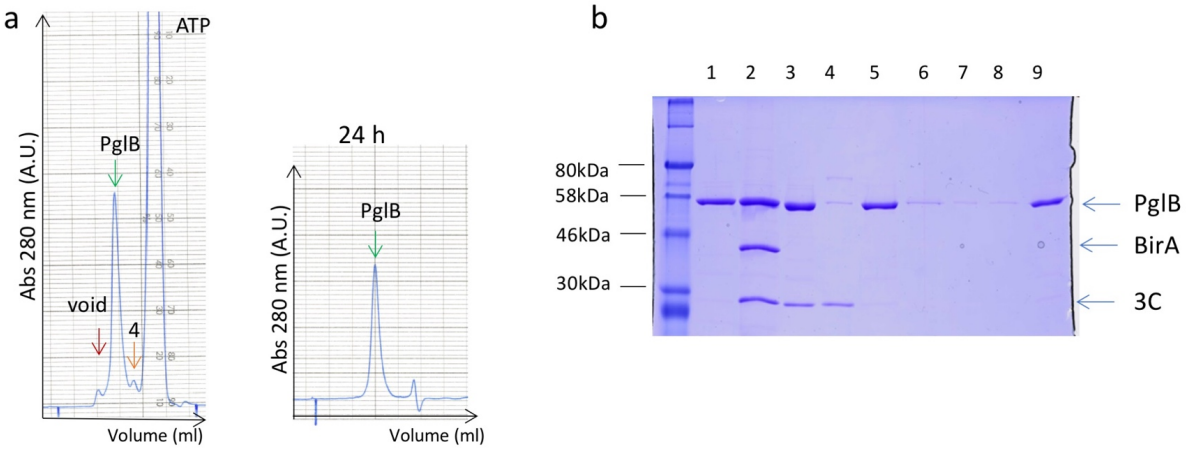


Figure 9. SEC and SDS-PAGE analysis of biotinylated Avi-tagged PglB. (**a**) size exclusion chromatogram of PglB after overnight biotinylation and cleavage of the histidine tag with 3C (left). The stability of the biotinylated Avi-tagged PglB was checked after incubation for 24 h at 4 ºC (right). (**b**) SDS-PAGE analysis of the biotinylated Avi- tagged PglB. Lane 1: sample after desalt, lane 2: sample after overnight incubation with 3C protease and BirA, lane 3: sample after reversed NiNTA; lane 4: the

second peak after preparative SEC (marked as 4 in a.), lane 5: sample from the main peak after preparative SEC, incubated with streptavidin resin, lane 6: sample from the flowthrough on the streptavidin column, lane 7 and 8: samples from the subsequent washing steps, lane 9: sample eluted from the streptavidin column.

Only one round of panning was necessary to detect specific enrichment over the control samples: 100-fold enrichment over both controls (Stt3A protein from *T. brucei* and desalting buffer) (Fig. 10). After the second round of panning an enrichment of 1,000 and 10,000-fold over Stt3A and buffer sample, respectively, was detected (Fig. 10).

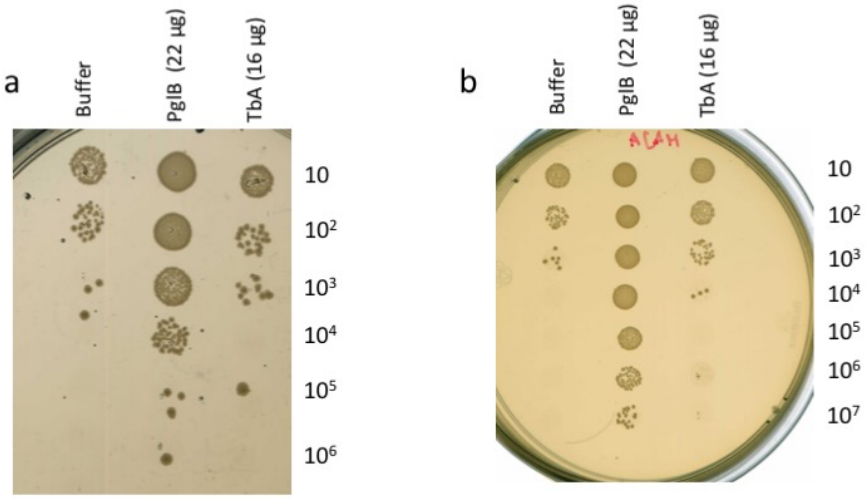


Figure 10. Enrichment analysis from 2 rounds of phage display. (**a**) After the first round of panning a 100-fold enrichment was detected over the control samples (Stt3A protein and desalting buffer). (**b**) The second round of panning showed 1,000 and 10,000-fold enrichment over Stt3A and the buffer sample, respectively.

100 clones were chosen and sequenced from the first round of phage display. From 91 different nanobody sequences we could identify only three distinct families. The representative nanobodies of each family and the unique outliers (34 different nanobodies (Fig. 11)), were expressed, purified and analyzed using a pull-down assay (Fig. 12).

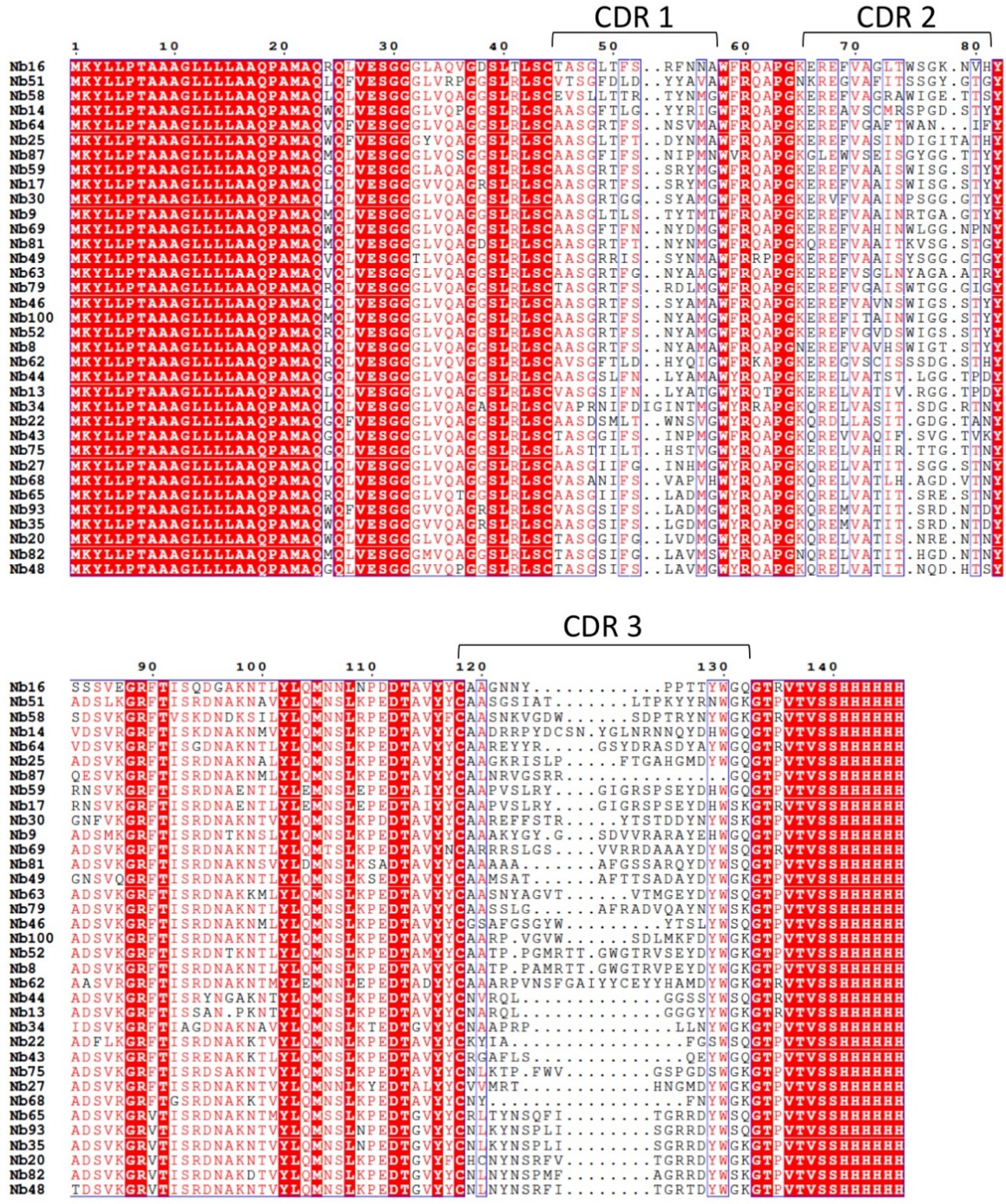


Figure 11. Sequence alignment of selected nanobodies. Amino acid sequence alignment of the nanobodies selected for the binding assay. The CDRs are marked with brackets. The ESPrit server was used to generate the alignment[32].

"Pull-down assays" with immobilized nanobodies and tagless PglB identified 29 PglB-specific nanobodies: 9, 13, 14, 16, 17, 20, 22, 25, 27, 30, 35, 43, 48, 49, 51, 52, 59, 63, 64, 65, 68, 69, 75, 79, 81, 82, 87, 93, 100. The results of the pull -down assay for representative nanobodies are shown in the Fig.12a. All nanobodies were confirmed to co-elute with PglB during size-exclusion chromatography, indicating that PglB forms tight complexes with the selected nanobodies (representative results are shown in Fig. 12b).

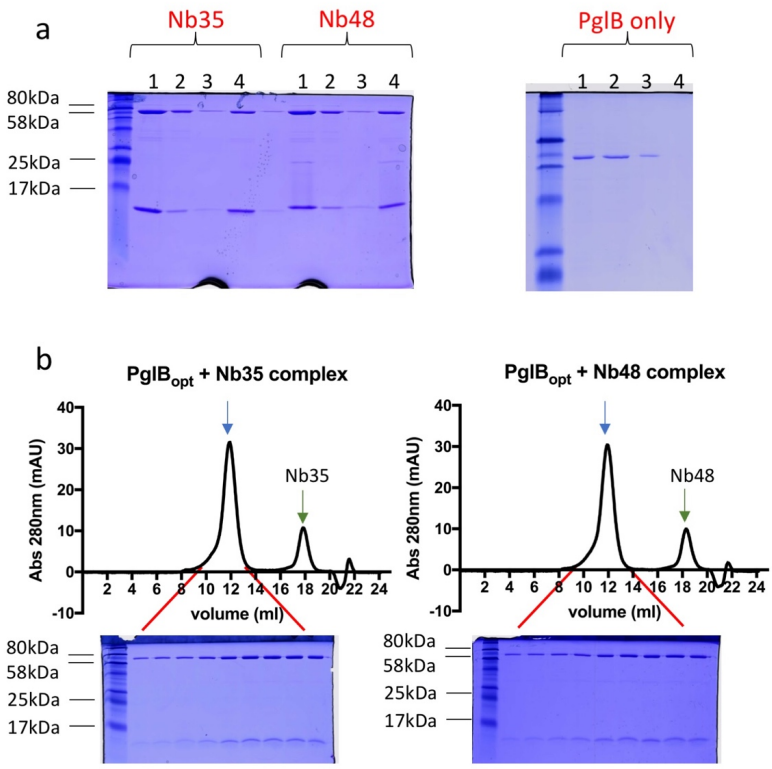


Figure 12. Binding assays of the representative nanobodies Nb35 and Nb48. (**a**) SDS-PAGE analysis of the pull- down assays for Nb35 and Nb48 and 'control' sample without any nanobody present. Lane 1: sample incubated with NiNTA resin, lane 2: collected flowthrough, that did not bind to the resin, lane 3: washing step, lane 4: sample eluted from the column. (**b**) SEC profiles of PglB - nanobody complexes (top) and SDS-PAGE gels of the main peak fractions of PglB- nanobody complexes (bottom).

PglB glycosylation activity in the presence of nanobodies

The effect of PglB-specific nanobodies on PglB glycosylation activity was analyzed by an *in vitro* glycosylation assay with the synthetic LLO. No significant inhibitory effect was observed. The results are summarized in Fig. 13.

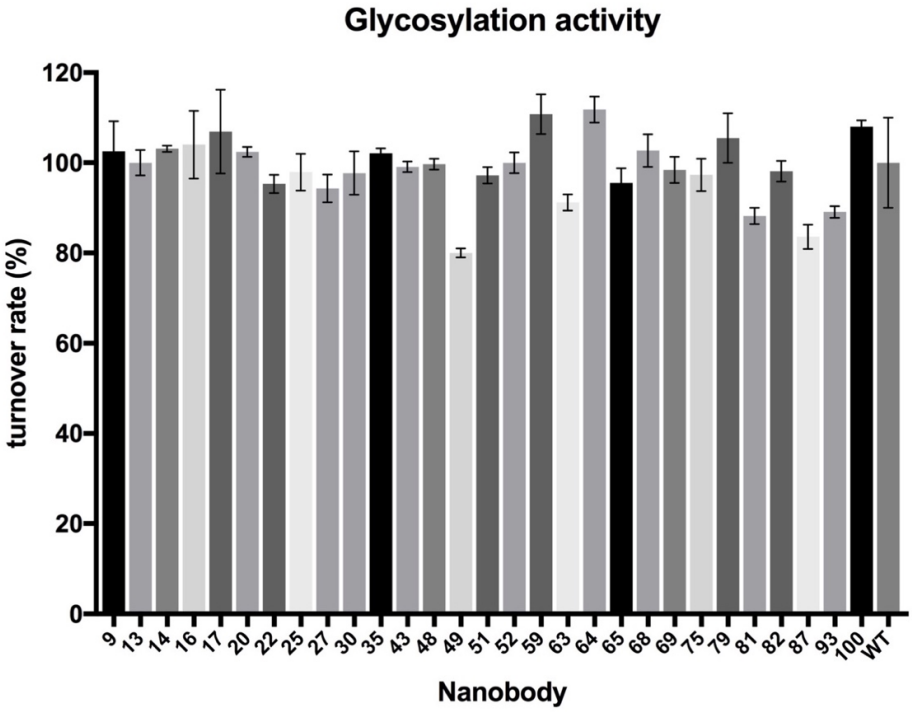


Figure 13. *In vitro* glycosylation activity of PglB in the presence of PglB-specific nanobodies. Normalized initial turnover rates of PglB in presence of selected nanobodies using 500 μM synthetic nerylneryl- PP-GlcNAc as a donor substrate. Data represent three technical replicates (error bars indicate s.d., n ≥ 3).

Thermostabilizing effect of PglB-specific nanobodies

For all PglB-specific nanobodies that had a thermostabilizing effect, thermostability (Ts) values were determined (Fig. 14a, Table 1). Nanobodies 17, 59 and 93 stabilized PglB by 5 ºC. While nanobodies 17 and 59 have very similar CDR sequences, nanobody 93 is completely different (Fig. 11). Nanobodies 43 and 75 stabilised PglB by about 10°C, while the most thermostabilizing nanobody was Nb64, which stabilized PglB by 13°C (Fig. 14b, Table 1).

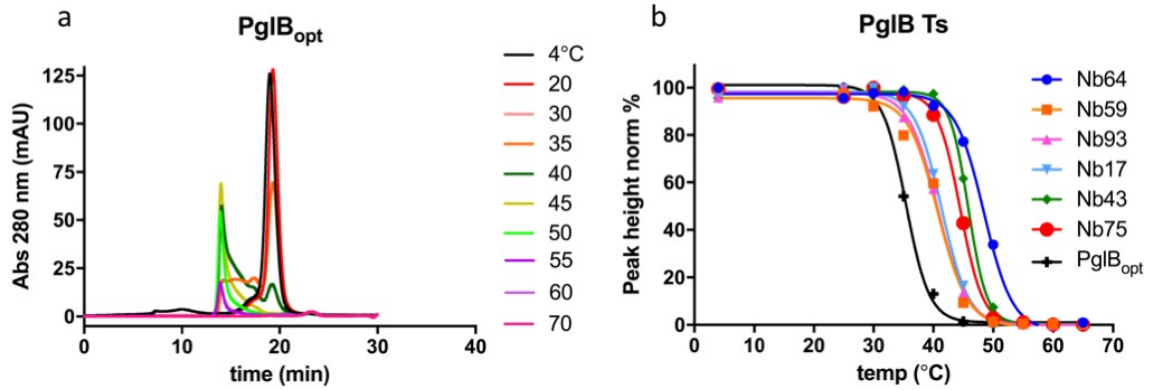


Figure 14. Thermostabilizing effect of the PglB-specific nanobodies. (**a**) SEC profiles of sequence-optimized PglB incubated at different temperatures. The height of the main peak was measured and plotted against the temperature to generate the $PglB_{opt}$ SEC-Ts curve in b. (**b**) The experiment in (a) was repeated for complexes of PglB with stabilizing nanobodies (17, 43, 59, 64, 75 and 93). The respective Ts values calculated are summarized in Table 1.

Sample	Ts (°C)
PglB	35.4 ± 0.2
+Nb17	41.4 ± 0.2
+Nb43	45.8 ± 0.1
+Nb59	40.7 ± 0.5
+Nb64	48.4 ± 0.3
+Nb75	44.5 ± 0.1
+Nb93	40.7 ± 0.2

Table 1. Thermostability values of PglB-Nb complexes

In order to check the affinity of nanobody-PglB complexes, two of the thermostabilizing nanobodies (Nb13 and Nb64) were fluorescently labeled and the PglB-Nb complexes were analysed using MST. The labeling efficiency was estimated to be 85%. MST parameters were optimized for each nanobody separately. Both K_d values were in the nanomolar-range (Fig. 15).

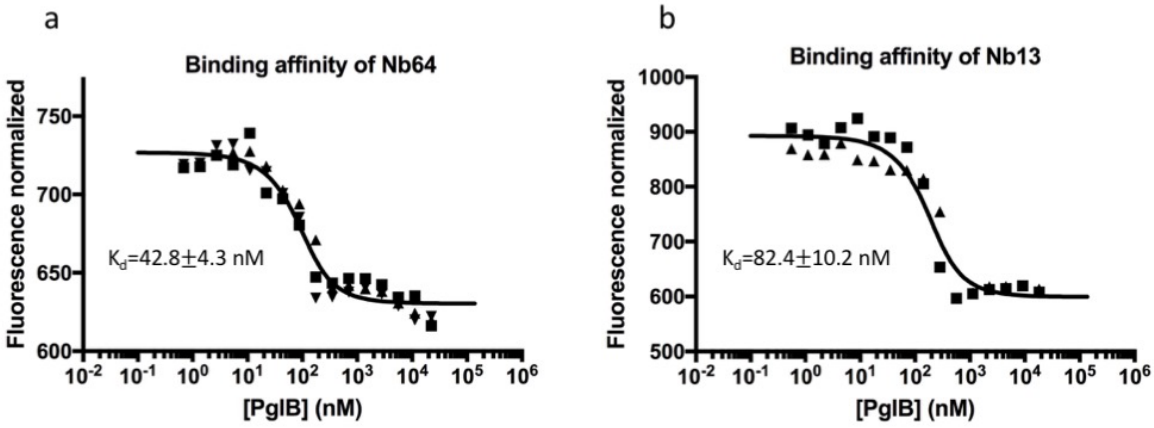


Figure 15. MST binding assay to determine the K_d of selected nanobodies. (a) PglB incubated with fluorescently labeled Nb64. (**b**) PglB incubated with fluorescently labeled Nb13. Data represent three technical replicates.

Crystallization of PglB-nanobody complexes

All PglB-Nb complexes were tested using homemade screens. The complexes were set up in the presence of both substrates- an acceptor peptide and an inhibitory LLO. Crystals appeared in conditions with PglB and three nanobodies: 27, 30 and 48 (Fig. 16). Interestingly, none of these nanobodies had a thermostabilizing effect on PglB. The crystals were reproduced, harvested and screened at the SLS, Villigen, Switzerland. Crystals of the ternary complex, in drops containing nanobody 27, diffracted to 3.3 Å, but did not have a nanobody molecule bound to PglB. Two other conditions gave crystals of different morphology to the ternary complex alone, indicating that they may have nanobody bound, but the diffraction data only went to 9-10 Å.

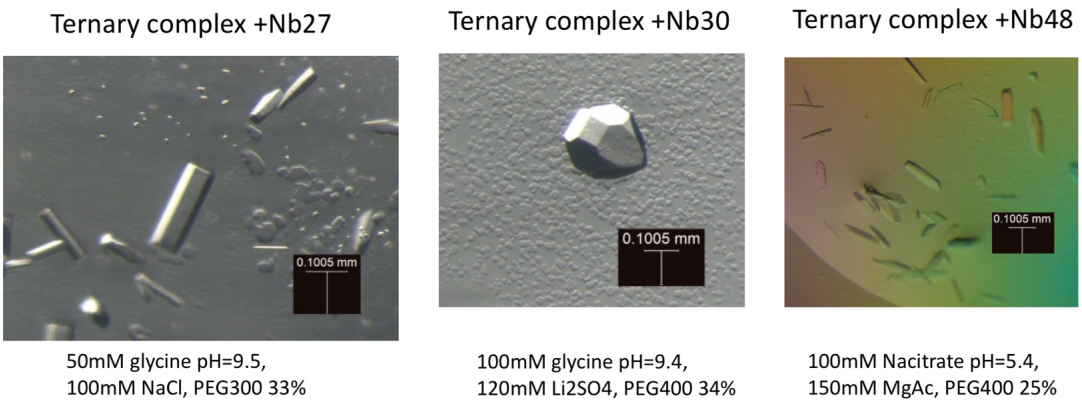


Figure 16. Crystallization of PglB- nanobody complexes. Crystals of the ternary complex grown in wells containing PglB-specific nanobodies. The conditions of each buffer are indicated below the pictures.

Conclusions

In summary, an engineered cross-linked mutant of PglB was expressed, purified and reconstituted into proteoliposomes for injection into an alpaca. The sample gave an exceptionally good immune response in the heavy-chain class. After panning the nanobody library against PglB using phage display, we were able to identify 91 different nanobody sequences. Pull-down assays identified 29 PglB-specific nanobodies, all forming tight complexes with PglB as indicated by co-elution after size exclusion chromatography. None of the nanobodies had a significant impact on the glycosylation activity of PglB. Five nanobodies thermostabilised PglB and Nb64, which increased the stability of PglB by 13 °C, was shown to bind PglB with a nanomolar affinity. Crystallization trials were set up with all nanobodies by forming complexes with PglB, an acceptor peptide and an inhibitory LLO. Crystals appeared in the presence of three nanobodies, 27, 30 and 48. Nanobody 27 was not bound in the crystals and crystals grown in the presence of nanobodies 30 and 48 diffracted poorly. Despite the failure of these nanobodies to improve protein crystallization, the six thermostabilising nanobodies may be useful for future crystallization and NMR studies. In addition complexes of PglB bound to one or a number of nanobodies may be useful for cryo-electron microscopy studies, where increasing the size of the complex and adding addition features can facilitate higher resolution.

References

1. Hamers-Casterman, C. *et al.* Naturally occurring antibodies devoid of light chains. *Nature* **363,** 446–448 (1993).
2. Rasmussen, S. G. F. *et al.* Structure of a nanobody-stabilized active state of the b 2 adrenoceptor. *Nature* **469,** (2011).
3. Löw, C. *et al.* Nanobody Mediated Crystallization of an Archeal Mechanosensitive Channel. *PLoS One* **8,** (2013).
4. Loris, R. *et al.* Crystal Structure of the Intrinsically Flexible Addiction Antidote MazE. (2003). doi:10.1074/jbc.M302336200
5. Steyaert, J. & Kobilka, B. K. Nanobody stabilization of G protein-coupled receptor conformational states. *Curr. Opin. Struct. Biol.* **21,** 567–572 (2011).
6. Pardon, E. *et al.* A general protocol for the generation of Nanobodies for structural

biology. *Nat. Protoc.* **9,** 674–693 (2014).
7. Padlan, E. A. Anatomy of the Antibody molecule. *Mol. Immunol.* **31,** 169–217 (1994).
8. Muyldermans, S. Nanobodies: Natural Single-Domain Antibodies. *Annu. Rev. Biochem.* **82,** 775–797 (2013).
9. Nguyen, V. K., Hamers, R., Wyns, L. & Muyldermans, S. Loss of splice consensus signal is responsible for the removal of the entire CH1 domain of the functional camel IGG2A heavy-chain antibodies. *Mol. Immunol.* **36,** 515–524 (1999).
10. Nguyen, V. K., Desmyter, A. & Muyldermans, S. Functional heavy-chain antibodies in Camelidae. *Adv. Immunol.* **79,** 261–296 (2001).
11. Muyldermans, S. *et al.* Camelid immunoglobulins and nanobody technology. *Vet. Immunol. Immunopathol.* **128,** 178–183 (2009).
12. Vu, K. B., M, G., Wyns, L., Muyldermans, S. & Ghahroudi, M. A. Comparison of llama VH sequences from conventional and heavy chain antibodies. *Mol. Immunol.* **34,** 1121–1131 (1997).
13. Conrath, K. E., Wernery, U., Muyldermans, S. & Nguyen, V. K. Emergence and evolution of functional heavy-chain antibodies in Camelidae. *Dev. Comp. Immunol.* **27,** 87–103 (2003).
14. Flajnik, M. F., Deschacht, N. & Muyldermans, S. A case of convergence: why did a simple alternative to canonical antibodies arise in sharks and camels? *PLoS Biol.* **9,** e1001120 (2011).
15. Desmyter, A. *et al.* Crystal structure of a camel single-domain VH antibody fragment in complex with lysozyme. *Nat. Struct. Biol.* **3,** 284–289 (1996).
16. Chothia, C., Novotný, J., Bruccoleri, R. & Karplus, M. Domain association in immunoglobulin molecules. The packing of variable domains. *J. Mol. Biol.* **186,** 651–663 (1985).
17. Muyldermans, S., Atarhouch, T., Saldanha, J., Barbosa, J. & Harmes, R. Sequence and Structure of v-H Domain From Naturally-Occurring Camel Heavy-Chain Immunoglobulins Lacking Light-Chains. *Protein Eng.* **7,** 1129–1135 (1994).
18. Ghahroudi, M. A., Desmyter, A., Wyns, L., Hamers, R. & Muyldermans, S. Selection and identification of single domain antibody fragments from camel heavy-chain antibodies. *FEBS Lett.* **414,** 521–526 (1997).
19. De Genst, E. *et al.* Molecular basis for the preferential cleft recognition by dromedary heavy-chain antibodies. *Proc. Natl. Acad. Sci. U. S. A.* **103,** 4586–4591 (2016).
20. Lauwereys, M. *et al.* Potent enzyme inhibitors derived from dromedary heavy-chain antibodies. *EMBO J.* **17,** 3512–3520 (1998).
21. Perez, C. *et al.* Structural basis of inhibition of lipid-linked oligosaccharide flippase PglK by a conformational nanobody. *Sci. Rep.* **7,** 46641 (2017).
22. Mireku, S. A., Sauer, M. M., Glockshuber, R. & Locher, K. P. Structural basis of nanobody-mediated blocking of BtuF, the cognate substrate-binding protein of the Escherichia coli vitamin B12 transporter BtuCD. *Sci. Rep.* **7,** 14296 (2017).
23. Lizak, C. *et al.* A catalytically essential motif in external loop 5 of the bacterial oligosaccharyltransferase PglB. *J. Biol. Chem.* **289,** 735–746 (2014).
24. Geertsma, E. R., Nik Mahmood, N. A. B., Schuurman-Wolters, G. K. & Poolman, B. Membrane reconstitution of ABC transporters and assays of translocator function. *Nat. Protoc.* **3,** 256–266 (2008).
25. Schaffner, W. & Weissmann, C. A rapid, sensitive, and specific method for the determination of protein in dilute solution. *Anal. Biochem.* **56,** 502–514 (1973).
26. Maass, D. R., Sepulveda, J., Pernthaner, A. & Shoemaker, C. B. Alpaca (Lama pacos) as a convenient source of recombinant camelid heavy chain antibodies (VHHs). *J Immunol Methods* **324,** 13–25 (2007).

27. Schenck, S. *et al.* Generation and Characterization of Anti-VGLUT Nanobodies Acting as Inhibitors of Transport. *Biochemistry* **56,** 3962–3971 (2017).
28. Geertsma, E. R. & Dutzler, R. A Versatile and Efficient High-Throughput Cloning Tool for Structural Biology. *Biochemistry* **50,** 3272–3278 (2011).
29. Lizak, C., Gerber, S., Numao, S., Aebi, M. & Locher, K. P. X-ray structure of a bacterial oligosaccharyltransferase. *Nature* **474,** 350–355 (2011).
30. Gerber, S. *et al.* Mechanism of bacterial oligosaccharyltransferase: In vitro quantification of sequon binding and catalysis. *J. Biol. Chem.* **288,** 8849–8861 (2013).
31. Napiórkowska, M. *et al.* Molecular basis of lipid-linked oligosaccharide recognition and processing by bacterial oligosaccharyltransferase. *Nat. Struct. Mol. Biol.* **24,** 1100–1106 (2017).
32. Robert, X. & Gouet, P. Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res.* **42,** 320–324 (2014).

Chapter 3: Engineering an acceptor sequon into PglB to facilitate structural studies

Abstract

The main enzyme of the N-glycosylation pathway is oligosaccharyltransferase (OST), an integral membrane protein that catalyzes the transfer of a glycan from a lipid linked oligosaccharide (LLO) onto an acceptor asparagine that is located within a conserved sequon (N-X-T/S). Although structural and biochemical studies have provided insight into sequon recognition and the interactions of OST with LLO, the transfer mechanism remains poorly understood. Here we describe a unique strategy to trap the glycosylation reaction in an intermediate state by crystallizing a ternary complex of PglB with an acceptor peptide and a non-reactive LLO. In order to increase the local concentration of the acceptor peptide and further stabilize PglB, we designed and screened different PglB constructs carrying a glycosylation sequon (DQNAT) inserted into either the periplasmic domain of PglB, or the region linking transmembrane helix 13 and the periplasmic domain. We identified a more thermostable, fully functional construct that was selected as the best candidate for crystallisation trials. Crystallization experiments resulted in crystals that diffracted to 4.5 Å, providing a basis to obtain future higher resolution structures of ternary complexes of PglB.

Introduction

N-linked protein glycosylation is an essential post-translational modification, that is present in all domains of life[1–3]. The transfer of a sugar moiety from a lipid-linked oligosaccharide (LLO) to an acceptor asparagine is catalyzed by oligosaccharyltransefase (OST), an integral membrane enzyme. Whereas eukaryotic OST is a multisubunit complex, in some kinetoplasts, eubacteria and archaea, it is a single subunit enzyme (ssOST) [2–7]. The catalytic subunit of the eukaryotic complex, Stt3, is homologous to ssOSTs and recent cryo-EM structures of mammalian and yeast OST complexes showed that it is also structurally conserved[3,4,8–10]. The ssOST from *Campylobacter spp* is a well-characterized model system that can therefore be used to study the chemistry of the glycosylation reaction.

The first X-ray structure of a complete bacterial OST enzyme, *Campylobacter lari* PglB[6], and three subsequent structures of archeal OST, *Archaeoglobus fulgidus* AglB[7,11], revealed the fold of the ssOST enzymes and gave more insight into peptide specificity and recognition. Peptide specificity has also been investigated functionally[12–14]. *In vitro* studies showed the DQNAT sequence to be the optimal acceptor peptide substrate for PglB[12], and later the structure of DQNAT-bound PglB[6] revealed how this substrate specificity is achieved. Subsequent quantitative fluorescence-based *in vitro* assays were used to determine binding affinities for DQNAT, with a K_d of 1.0 μM and a K_M of 2.6 μM, as well as glycosylation rates[13,14].

Due to the absence of any structural data for OST in an LLO-bound state, the interactions of OST with LLO, and the subsequent transfer mechanism, remain poorly understood. In order to gain insight into these interactions, we tried to trap the glycosylation reaction in an intermediate state where both substrates, LLO and acceptor peptide, are bound to PglB. The main challenge in trapping a ternary-complex intermediate of OST was to ensure that glycan transfer did not occur. This could be achieved by generating non-functional PglB mutants or by using substrate analogs that can bind PglB, but would not allow for glycan transfer. All non-functional PglB mutants, studied previously, had considerably lower binding affinities for the peptide substrates[13–15]. In this study, we therefore focused on using the optimal DQNAT sequon as an

acceptor substrate, and a non-hydrolyzable LLO analog nerylneryl-PPCH$_2$-GlcNAc ((ωZZZ)-PPCH$_2$-GlcNAc)[16]. This strategy is described in detail in Chapter 4. A high-resolution ternary complex structure of PglB, with a wild- type peptide bound, could also give more insight into the activation mechanism of the carboxamide group of the acceptor asparagine. In order to facilitate crystallization experiments of the PglB ternary complex, we designed different constructs with the DQNAT glycosylation sequon inserted into either the periplasmic domain of PglB, or the region linking transmembrane helix 13 and the periplasmic domain. The presence of the inserted sequon would likely stabilize the protein and increase the local concentration of the acceptor peptide during crystallization trials. In this chapter the design, screening and initial crystallization experiments of the new PglB construct with non-hydrolyzable LLO analogs are described.

Experimental procedures

Materials

All materials were obtained from AppliChem or Sigma unless stated otherwise. Restriction enzymes were purchased from New England Biolabs and Fermentas.

Construction of plasmids

Christian Lizak generated the PglB$_{opt}$ construct

In all experiments the modified PglB gene from *Campylobacter lari*, strain RM2100 was used. The resulting cysteineless construct with glycosylation sites removed and carrying the mutations: K2E, C17A, C30A, A108T, C360L, N535Q, Q536K, K549P, D550N, F553I, N556P, A600P, A602D, T606K, T607Q, V610I, M611T, I619S, F622Y, A624S, V627I, A630N, F663Y, and F670Y was referred to as PglB$_{opt}$[16]. All subsequent changes to generate Inseq- constructs were based on PglB$_{opt}$. The synthetic gene of PglB$_{opt}$ (in pUC57 vector, GenScript) with a C-terminal decahistidine affinity tag carried the following silent mutations: NcoI (1bp), XhoI (367 bp), AfeI (701 bp), NheI (713 bp), SpeI (1010 bp), BsiWI (1324 bp), MluI (1674 bp), BstBI (1964 bp), EcoRI (2211 bp). The Inseq constructs (Inseq1-Inseq7) with a glycosylation sequon inserted into the periplasmic domain of PglB (between Asp613 and Asn614) were generated using gene fragments (gBlocks, Intergarted DNA Technologies) and cloned with the MluI and BstBI restriction enzymes, while the Inseq-constructs (Inseq8-Inseq11) with a glycosylation sequon inserted into the region linking transmembrane helix 13 and the periplasmic domain of PglB (between Lys434 and Val438) were generated using gene fragments (gBlocks, Intergarted DNA Technologies) and SpeI and BsiWI restriction sites. New constructs were validated by DNA sequencing (Microsynth) and sub-cloned into the pBAD expression vector (Invitrogen).

Overexpression and purification of PglB

Overexpression and purification of PglB and the Inseq-constructs was performed as previously described[6,16]. Briefly, PglB was overexpressed in *Escherichia coli* BL21-Gold cells (DE3) (Strategene) at 37 °C in five-liter flasks using TB (Terrific Broth) medium supplemented with 1% glycerol (w/v). The cells were induced at 37 °C for 4 h at an optical density (OD), A_{600} of 3.0 by adding 0.1% arabinose (w/v). Cells were harvested by centrifugation and cell pellets were stored at −80 °C. All following steps were carried out at 4 °C. Cells were resuspended in 25 mM Tris-HCl, pH 8.0, 250 mM NaCl and disrupted in a M-110-l microfluidizer (Microfluidics) at 15,000 p.s.i. chamber pressure. Membranes were pelleted by centrifugation at 100,000 g for 30 min and solubilized in 25 mM Tris-HCl, pH 8.0, 250 mM NaCl, 10 %

glycerol (v/v) and 1% N-dodecyl-β-D-maltopyranoside (w/v) (DDM, Anatrace) for 1.5 h. All subsequent purification buffers contained 0.016 % DDM. PglB was purified on a nickel-nitrilotriacetic acid (Ni-NTA) Superflow affinity column (Qiagen) and desalted into 10 mM MES-NaOH, pH 6.5, 100 mM NaCl, 3 % glycerol (v/v), and 0.016 % DDM. For crystallization experiments, desalted protein was concentrated to 6 mg mL^{-1} in an Amicon Ultra-15 concentrator (Milipore) with a molecular mass cutoff of 100 kDa and purified by size exclusion chromatography (Superdex 200 10/300, GE Healthcare) in desalting buffer. Collected fractions were pooled and concentrated to 12 mg mL^{-1}.

Test expression and solubilisation of Inseq4

For test expression of Inseq4, the construct was transformed into *E. coli* BL21-Gold (DE3) (Strategene) cells. 25 mL of LB (Lysogeny broth) medium precultures were inoculated from a single colony and grown over night at 37 °C. Main cultures were inoculated to OD of 0.05 in 500 mL TB (Terrifific broth) medium supplemented either with 1 % glucose (w/v) or 1 % glycerol (v/v), or in 500mL LB medium in 2 L shaking flasks. The cells in TB media were grown to OD of 3.0 and in LB media to OD of 0.8 at 37 °C. Cultures were induced by addition of arabinose to a final concentration of 0.1 %, 0.02 %, 0.01 % or 0.002 % for 4 h at 37 °C or overnight at 18 °C. Cells were harvested by centrifugation at 9,000 g for 4 min at 4 °C. All following steps were carried out at 4 °C. The pellet was resuspended in 25 mM Tris-HCl, pH 8.0, 250 mM NaCl at the ratio 1:10 (cell weight to buffer volume) and the cells were disrupted on ice by sonication (3 times 2 min, 50% duty cycle, output control 6). The buffer was supplemented with 1 % DDM (Anatrace) and PglB was solubilized for 1 h. Samples were centrifuged at 100,000 g for 30 min. The supernatant was transferred to a fresh tube and the pellet was resuspended in the equivalent amount of buffer. The samples were analyzed by immunoblot following SDS-PAGE. Immunodetection of *C. lari PglB* was performed with anti-poly-histidine-HRP serum (Sigma).

Size-exclusion chromatography thermostability (SEC-Ts) assay

Sequence-optimized PglB or Inseq4 were overexpressed and purified as described above. Protein was desalted into 10 mM MES-NaOH, pH 6.5, 100 mM NaCl, 3% glycerol, 0.016 % DDM. 5 μM PglB was then preincubated (or not) with 1 mM $MnCl_2$ in the presence or absence of 100 μM acceptor peptide at 4 °C. For Inseq4 and thermostabilizing nanobody 64 (Nb64), protein was preincubated with a 2-fold molar excess of nanobody at 4 °C. Mixtures were incubated at different temperatures (4-70 °C) for 10 min. The samples were cooled on ice and centrifuged for 30 min at 120, 000 g. The supernatants were then analyzed by size-exclusion chromatography using a TSK-gel G3000SWXL column equilibrated with 10 mM MES-NaOH, pH 6.5, 100 mM NaCl, 3 % glycerol, 0.016 % DDM and supplemented with 1 mM $MnCl_2$ for the experiments with acceptor peptide. For each temperature, the height of the main peak was plotted against temperature. The data was then fitted using the 'sigmoidal dose-response (variable slope)' function in Graph Pad Prism 7.0 and the thermostability (Ts) of each sample was calculated.

***In vitro* glycosylation assay**

Jérémy Boilevin synthesized the reactive, synthetic LLO analog, Tamis Darbre and Jean-Louis Reymond supervised the chemical synthesis of the LLO analog

In vitro glycosylation assays were performed as described previously[13,16], using a labeled acceptor peptide 5-carboxyfluorescein-GS-DQNATF-NH2 (10 μM) as the acceptor substrate and wild-type LLO (50 μM) or nerylneryl-PP-GlcNAc (1 mM) as the donor substrate. Reaction mixtures with wild-type LLO and nerylneryl-PP-GlcNAc contained 1 nM or 100 nM PglB,

respectively. For turnover-rate determination, a total of at least six samples were taken at different time intervals so that the reaction was in the linear range. Data was fitted by linear regression in GraphPad PRISM 7.0.

***In vitro* autoglycosylation assay**

The purified protein was incubated at different concentrations (from 0.3-0.03 mg mL^{-1}) in buffer containing 10 mM MES-NaOH, pH 6.5, 100 mM NaCl, 3 % glycerol (v/v), and 0.016 % DDM supplemented with 1 % Triton X-100, 10 uM wild type LLO and 10 mM $MnCl_2$. The reaction mixtures were incubated at 30 °C for 2 h and stopped by the addition of 4 x SDS loading buffer.

Glycosylation of the Inseq variants was analysed by immunoblot following SDS–PAGE. Immunodetection was performed with anti-glycan serum hR6[6,17] (Aebi laboratory) to observe glycosylated PglB. Immunodetection of *C. lari* PglB was performed with anti-poly-histidine-HRP serum (Sigma).

Screening of the detergents

Inseq4 was desalted into 25 mM Tris-HCl, pH 8.0, 250 mM NaCl, 5 % glycerol, 0.016% DDM. Thermostabilizing Nb64 was added to Inseq4 at a 2-fold molar excess. The complex was rebound to a preequilibrated nickel-nitrilotriacetic acid (Ni-NTA) column. The NiNTA-bound protein was firstly washed with buffers containing 25 mM Tris-HCl, pH 8.0, 250 mM NaCl, 5 % glycerol, 0.016 % DDM supplemented with different detergents at 3 × CMC (critical micelle concentration) (Table 1). Samples were then washed with buffers containing only the desired detergent at 3 × CMC, and the complex was eluted with 25 mM Tris-HCl, pH 8.0, 250 mM NaCl, 5 % glycerol and 3 × CMC of detergent. Samples were subsequently analyzed by size-exclusion chromatography using a TSK-gel G3000SWXL column equilibrated with 10 mM MES-NaOH pH 6.5, 100 mM NaCl, 3 % glycerol, 0.016 % DDM. Promising detergents were used during the full purification of Inseq4. To stabilize Inseq4, Nb64 was added to the solubilization step at a 2-fold molar excess over PglB.

Detergent	**Abbreviation**	**CMC(mM)**
n-Dodecyl-β-D-Maltoside	DDM	0.17
n-Undecyl-β-D-Maltopyranoside	UDM	0.59
n-Nonyl-β-D-Maltopyranoside	NDM	6
n-Decyl-β-D-Maltopyranoside	DM	1.80
n-Octyl-β-D-Glucopyranoside	OG	18-20
n-Nonyl-β-D-Glucopyranoside	NG	6.5
Lauryl-DimethylAmine-*N*-Oxide	LDAO	1-2
Lauryl-Maltose-Neopentyl-Glycol	LMNG	0.01
Octyl-Glucose-Neopentyl-Glycol	OGNG	1.02
5-Cyclohexyl-1-Pentyl-β-D-Maltoside	Cymal5	2.4-5
6-Cyclohexyl-1-Hexyl-β-D-Maltoside	Cymal6	0.56

Table 1. Tested detergents and their CMC values.

Crystallization of Inseq4

Jérémy Boilevin synthesized the non-hydrolyzable LLO analog, Tamis Darbre and Jean-Louis Reymond supervised the chemical synthesis of the LLO analog

Inseq4 was overexpressed and purified as described above. Protein was desalted into 10 mM MES-NaOH, pH 6.5, 100 mM NaCl, 3 % glycerol, 0.016 % DDM. For crystallization experiments, desalted protein was concentrated to 6 mg mL^{-1} in an Amicon Ultra-15 concentrator (Milipore) with a molecular mass cutoff of 100 kDa and purified by size exclusion chromatography (Superdex 200 10/300, GE Healthcare) in desalting buffer. Collected fractions were pooled and concentrated to 12 mg mL^{-1}. For Inseq4-Nb64 crystallization, Nb64 was added to the desalt sample before the concentration steps. Inseq4/ Inseq4+Nb64 was incubated with 1 mM $MnCl_2$, 2 mM nerylneryl-PPC-GlcNAc for 15 min at 4 °C. Vapor-diffusion crystallization experiments were performed at 20 °C with home-made and commercial 96-well screens using 1:1 and 2:1 protein to well solution ratios. Crystals were reproduced in 24-well screens in hanging and sitting drops. Crystals were cryoprotected by gently increasing the concentration of the cryoprotectant (up to 30 % PEG) and flash frozen by immersion in liquid nitrogen before data collection. The crystals were screen at the Swiss Light Source (SLS) at PSI, Villigen.

Results

Thermostabilizing effect of an acceptor peptide on PglB

A construct of $PglB_{opt}$ (cloned by C. Lizak) was originally generated to improve protein crystallization, including lattice contacts. The optimized construct was less prone to aggregation and exhibited an increased thermostability (Ts) (5 °C more stable than wild- type (psf2) PglB (Ts_{psf2} = 30 °C, A. Ramírez)) (Fig. 1a, Table 2). Furthermore, it's *in vitro* glycosylation activity was indistinguishable from the turnover rate of 1.50 ± 0.04 peptides per second determined previously for wild-type PglB[13].

Subsequently, we tested the effect of an acceptor peptide Ac-DQNATF{NO2}-NH_2 on $PglB_{opt}$ thermostability. Knowing that divalent ions are essential for peptide binding[6,13,14], we firstly checked the effect of $MnCl_2$ on $PglB_{opt}$ stability (Fig. 1b). Although the size exclusion chromatography (SEC) profile of $PglB_{opt}$ in the presence of 1 mM $MnCl_2$ looked broader, the presence of the divalent ion did not influence the thermostability of PglB itself. However, upon the addition of an acceptor peptide, PglB stability increased by another 5 °C (Fig 1c and d, Table 2).

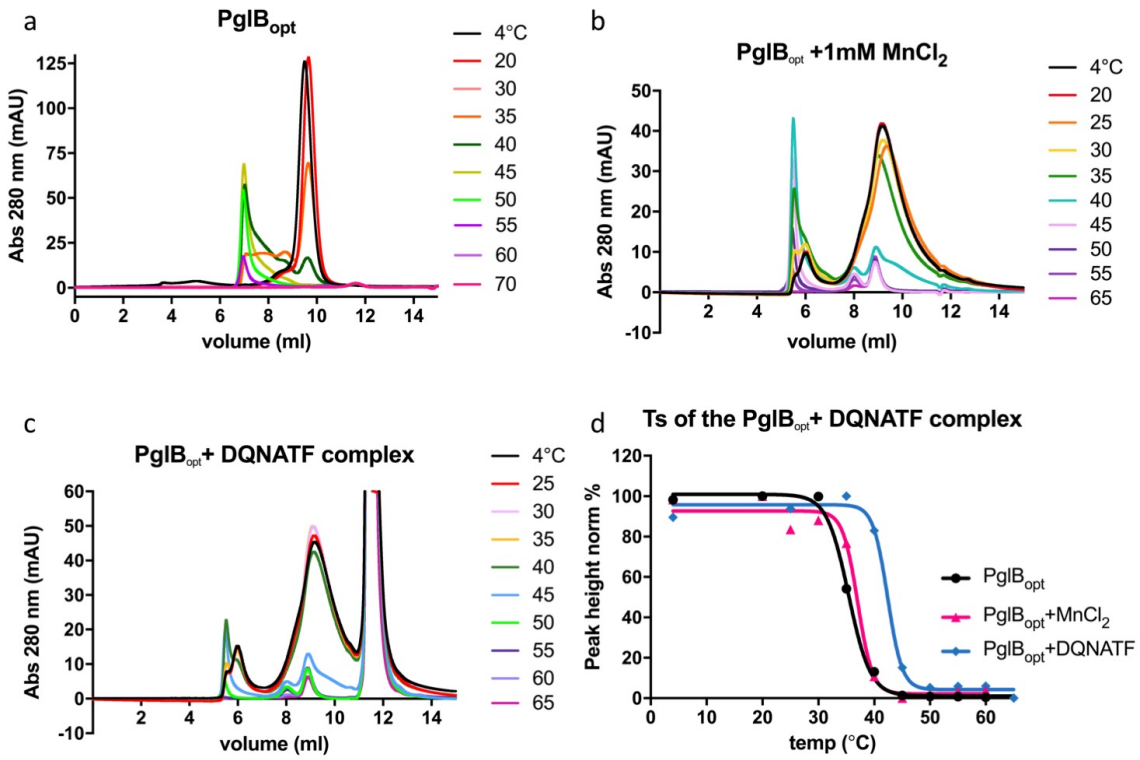


Figure 1. Thermostability of PglB. The samples were incubated at different temperatures for 10 min, cooled on ice and analyzed by SEC using a TSK-gel G3000SWXL column (a-c). (**a**) The SEC profiles of $PglB_{opt}$ in 10 mM MES-NaOH, pH 6.5, 100 mM NaCl, 3 % glycerol and 0.016 % DDM. (**b**) The SEC profiles of $PglB_{opt}$ preincubated with 1 mM $MnCl_2$. Samples were run in buffer containing: 10 mM MES-NaOH pH 6.5, 100 mM NaCl, 3 % glycerol, 0.016 % DDM and 1 mM $MnCl_2$. (**c**) The SEC profiles of $PglB_{opt}$ preincubated with 1 mM $MnCl_2$ and 100 µM acceptor peptide. Samples were run in buffer containing: 10 mM MES-NaOH, pH 6.5, 100 mM NaCl, 3 % glycerol, 0.016 % DDM and 1 mM $MnCl_2$. (**d**) SEC-Ts of all samples. PglB is stabilized by the presence of an acceptor peptide.

Sample	**Ts (°C)**
$PglB_{opt}$	35.4 ± 0.2
+1 mM $MnCl_2$	37.0 ± 0.6
+1 mM $MnCl_2$ +DQNATF	42.4 ± 0.4

Table 2. Thermostability (Ts) values obtained from the plots in Figure 1.

Engineering of constructs with an inserted sequon to improve protein crystallization

In order to increase the local concentration of acceptor peptide, and increase protein stability during the crystallization experiments, we designed different constructs carrying a glycosylation sequon inserted into either the periplasmic domain (Fig. 2) or into the region linking transmembrane helix 13 (TM13) and the periplasmic domain of PglB (Fig. 3).

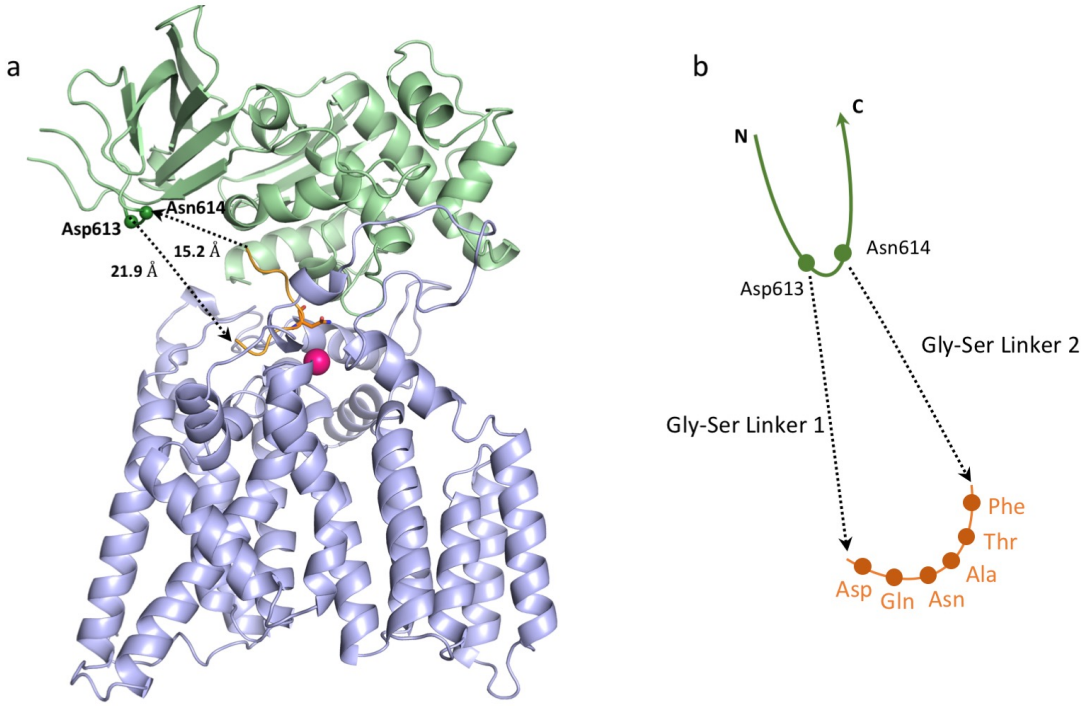


Figure 2. Strategy for insertion of the sequon into the periplasmic domain. (**a**) Cartoon representation of PglB with the insertion positions (Asp613 and Asn614) marked as green spheres. (**b**) Close-up schematic view of (a), with periplasmic residues shown in green and acceptor sequon in orange.

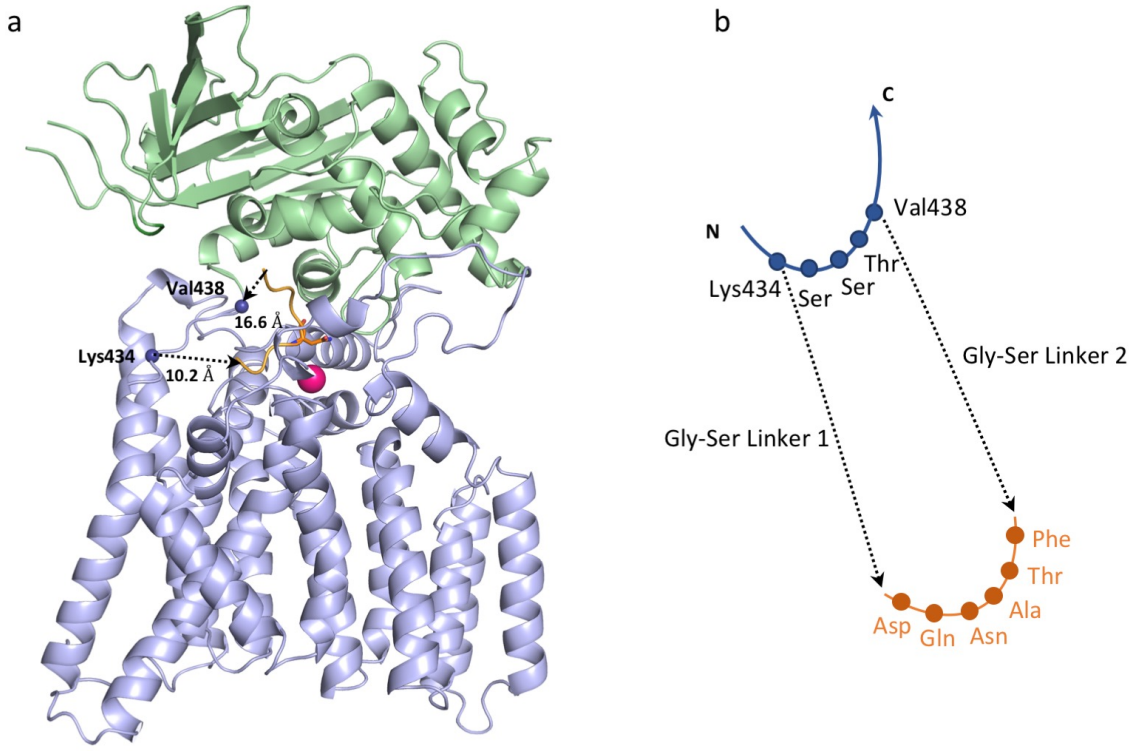


Figure 3. Strategy for the insertion of the sequon into the region linking the periplasmic domain and TM13. (**a**) Cartoon representation of PglB with the insertion positions (Lys434 and Val438) marked as blue spheres. (**b**) Close-up schematic view of (a), with transmembrane residues shown in blue and acceptor sequon in orange.

The constructs with the glycosylation sequon inserted into the periplasmic domain and into the linking region are shown in Fig. 4 and Fig 5, respectively. Almost all constructs carried the DQNATF sequon except Inseq11, which contained DQNATY. We screened 11 different constructs (Fig. 4 and Fig. 5) by expressing and fully purifying them and then verifying their autoglycosylation and glycosylation activities. In this study, we quantified the *in vitro* glycosylation activity of Inseq-variants using an external fluorescently labelled peptide, DQNATF, and wild type or synthetic (nerylneryl-PP-GlcNAc) LLO[13,16], and compared the obtained values with that of wild type or $PglB_{opt}$. Additionally, we qualitatively analyzed the *in vitro* autoglycosylation activity of the Inseq-constructs by incubating the protein with wild type LLO and $MnCl_2$ and detecting the glycosylation of an internal sequon present in the Inseq-variants with an anti-glycan immunoblot, using the hR6 antibody (provided by the Aebi laboratory)[18]. We ran the assay at very low concentrations of the Inseq-variants to avoid detecting glycosylation activity of sequons from other Inseq molecules.
All constructs showed lower expression levels compared to sequence optimized $PglB_{opt}$ (whose yield is 0.27 mg of protein after desalt per g of cells). The first construct screened, Inseq1, contained a 10 aa-linker on both sides of the sequon (Fig. 4a, Fig. 6). Although Inseq1 displayed a slightly impaired *in vitro* glycosylation activity of 0.56 ± 0.007 peptides per second (Fig. 6f), and was prone to aggregation upon concentration (Fig. 6a and b), it was a promising starting point for the design of further constructs with linkers of different lengths on both sides of the sequon (Fig. 4b and 6). Interestingly, we could detect autoglycosylation activity of Inseq1. (Fig. 6c-e).

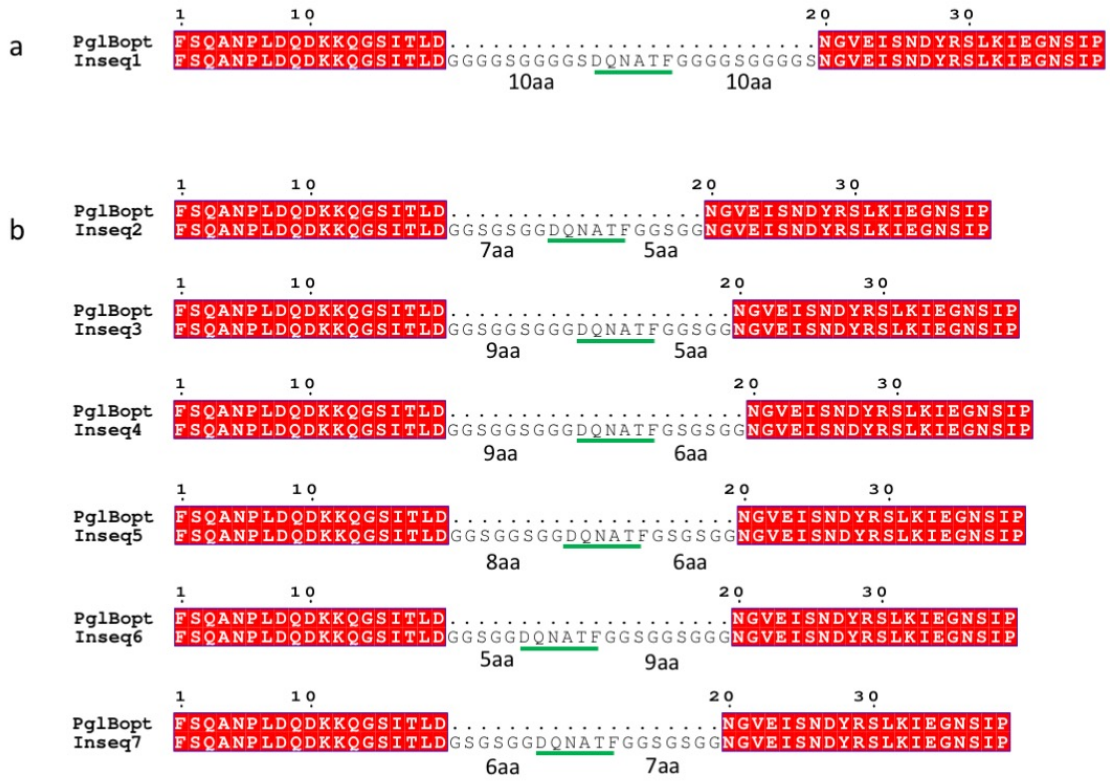


Figure 4. Amino acid sequence alignment of $PglB_{opt}$ and constructs carrying the sequon inserted into the periplasmic domain. ESPRIT server 3.0 was used to generate the alignments[19].

Figure 5. Amino acid sequence alignment of $PglB_{opt}$ and constructs carrying the sequon inserted into the region linking TM13 and the periplasmic domain. ESPRIT server 3.0 was used to generate the alignments[19].

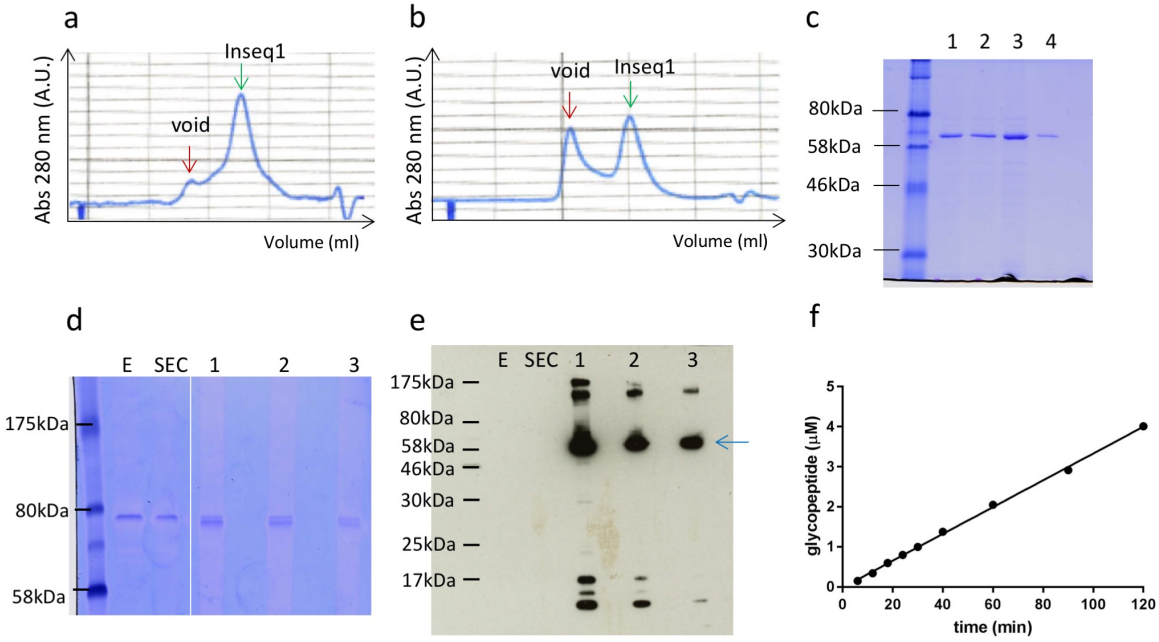


Figure 6. Purification and glycosylation activity of Inseq1. (**a**) Size exclusion chromatogram of Inseq1. (**b**) Size exclusion chromatogram of Inseq1 concentrated to 4 mg mL^{-1}. (**c**) SDS-PAGE analysis of Inseq1. Lane 1: sample after elution, lane 2: sample after desalt, lane 3: concentrated sample after desalt, lane 4: sample after preparative SEC. (**d**) SDS-PAGE analysis of the autoglycosylated Inseq1 samples. 'E' and 'SEC' depict the control samples of Inseq1 after elution and after SEC in the absence of wild type LLO. Lanes 1-3 are the reaction mixtures at different concentrations of Inseq1, supplemented with 10 mM $MnCl_2$ and 10 µM wild type LLO. Lane 1 corresponds to 0.11 mg mL^{-1} Inseq1, lane 2 to 0.055 mg mL^{-1} and lane 3 to 0.0275 mg mL^{-1}. (**e**) Immunoblot analysis of the samples described in (d). Glycosylated Inseq1 was detected in all 3 samples using the primary h6R antibody. The blue arrow indicates glycosylated Inseq1. (**f**) Determination of the *in vitro* glycosylation activity of Inseq1. Data was fitted with linear regression and the turnover rate was calculated from the slope of the fit (Graphpad PRISM 7.0).

Out of all the constructs that had the sequon inserted in the periplasmic region of PglB, Inseq4 gave the best yield (0.06 mg of protein after desalt per gram of cells), and was stable when concentrated to 6 mg mL^{-1} (Fig. 7a and b). While the glycosylation activity of Inseq4 was very similar to that of $PglB_{opt}$, (1.26 ± 0.045 peptide per second), we could also detect its autoglycosylation activity (Fig 7b and d), similar to Inseq1.
The best variant of PglB with the sequon inserted in the linking region was Inseq11, as it gave the highest yield from all of the constructs (0.1 mg of protein after desalt per g of cells), it eluted as a monodisperse peak during SEC (Fig. 8a and b), and we could detect its autoglycosylation activity (Fig. 8c and d). While the Inseq8 and Inseq9 variants, with longer linkers after the sequon and Lys434 replaced by arginine, eluted as monodisperse peaks during SEC (Fig 9a and b), we could not detect any *in vitro* glycosylation activity with either synthetic or wild type LLO. Their autoglycosylation activity could be detected on an immunoblot but only with high a concentration of protein (Fig. 9c and d). Either replacement of Lys434 to arginine or placement of the glycosylated sequon on the longer linker influenced stability and, directly or indirectly, the activity of the protein. In the case of the shorter linkers, Inseq11 and 10, the *in vitro* glycosylation activity with wild type LLO was significantly decreased (Fig. 10), but not abolished, with activities around 90-fold slower when compared to $PglB_{opt}$. We did not observe such a dramatic decrease in activity with synthetic LLO, when during the reaction only one sugar moiety is added. The *in vitro* glycosylation activity of Inseq11 with nerylneryl-PP-GlcNAc was 2.3 ± 0.07 peptides per minute, three-fold lower than the 6.63 ± 1.70 peptides per minute determined for $PglB_{opt}$. Introducing the bulky glycan moiety, close to the peptide binding site, possibly restricted the access of unglycosylated sequons to the active site, slowing the reaction rates.

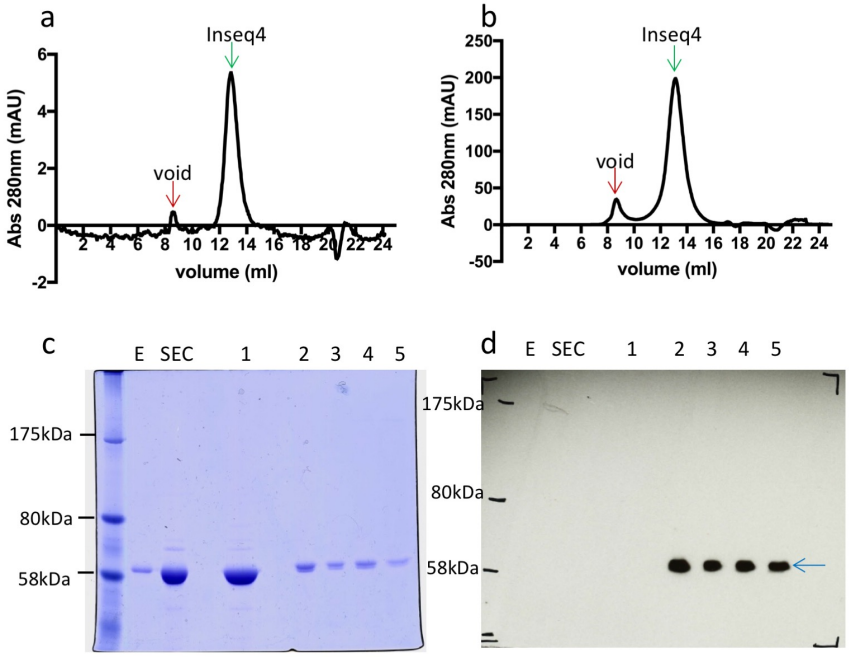


Figure 7. Purification and autoglycosylation activity of Inseq4. (**a**) Size exclusion chromatogram of Inseq4. (**b**) Size exclusion chromatogram of Inseq4 concentrated to 6 mg mL^{-1}. (**c**) SDS-PAGE analysis of the autoglycosylated Inseq4 samples. 'E' and 'SEC' depict the control samples of Inseq4 after elution and after SEC in the absence of wild type LLO. Lane 1 is a control sample of purified $PglB_{opt}$ in the absence of wild type LLO. Lanes 2-5 are the reaction mixtures at different concentrations of Inseq4 supplemented with 10 mM $MnCl_2$ and 10 µM wild type LLO. Lane 2 corresponds to 0.225 mg mL^{-1} Inseq4, 3 to 0.113 mg mL^{-1}, 4 to 0.056 mg $mL^{-1}$and 5 to 0.028 mg mL^{-1}. (**d**) Immunoblot analysis of

samples described in (c). Glycosylated Inseq4 was detected in all 3 samples using the primary h6R antibody. The blue arrow indicates glycosylated Inseq4.

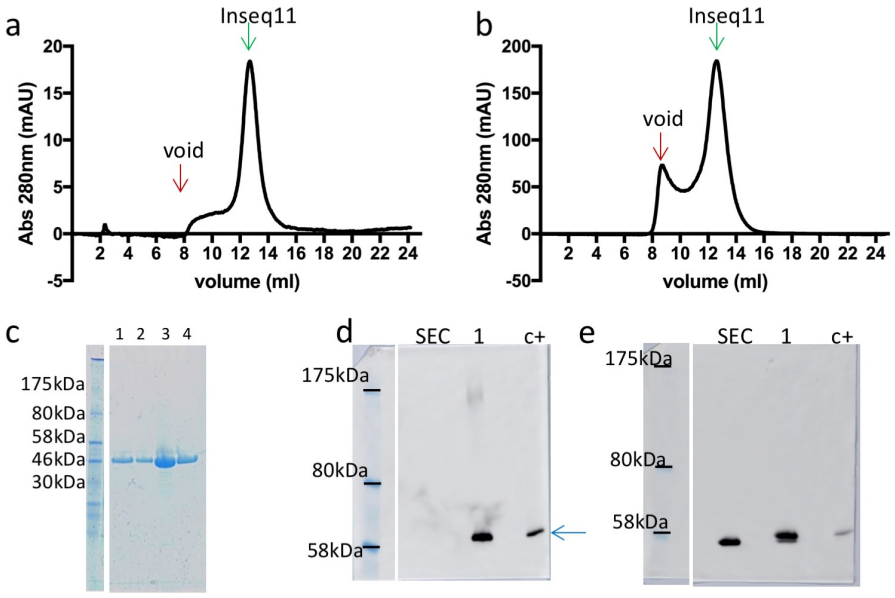


Figure 8. Purification and autoglycosylation activity of Inseq11. (**a**) Size exclusion chromatogram of Inseq11. (**b**) Size exclusion chromatogram of Inseq11 concentrated to 6 mg mL^{-1}. (**c**) SDS-PAGE analysis of Inseq11. Lane 1: sample after elution, lane 2: sample after desalt, lane 3: concentrated sample after desalt, lane 4: sample after preparative SEC. (**d**) Immunoblot analysis of the glycosylated Inseq11 sample. 'SEC' depicts the control sample of Inseq11 after SEC in the absence of 10 mM $MnCl_2$ and 10 µM LLO, and 'c+' depicts the control sample of glycosylated Inseg4 supplemented with 10 mM $MnCl_2$ and 10 µM LLO. Lane 1 is a reaction mixture of 0.03 mg mL^{-1} of Inseq11 supplemented with 10 mM $MnCl_2$ and 10µM wild type LLO. Glycosylation of Inseq11 was detected using the primary h6R antibody (**e**) Anti-poly-histidine immunoblot analysis of the samples described in (d). The immunoblot gave positive bands for all three samples.

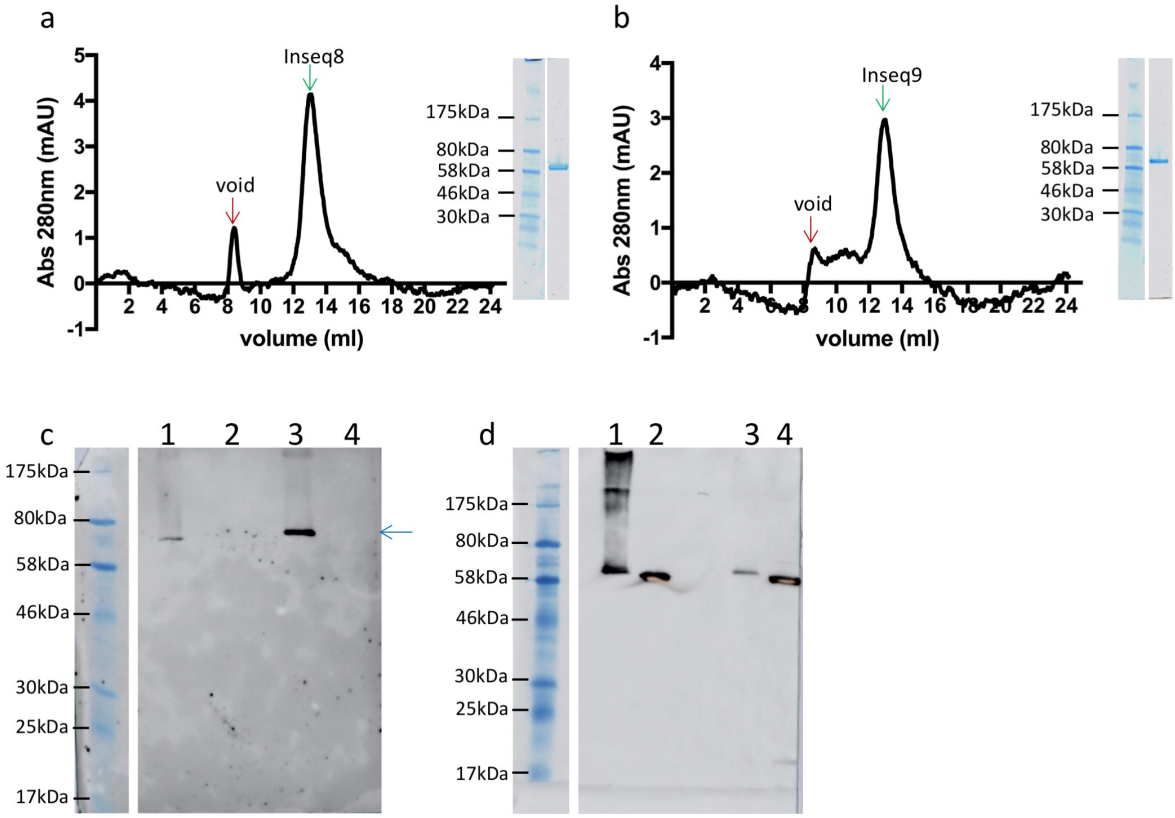


Figure 9. Purification and autoglycosylation activity of Inseq8 and Inseq 9. (**a**) Size exclusion chromatogram and SDS-PAGE analysis of the Inseq8 desalt sample. (**b**) Size exclusion chromatogram and SDS-PAGE analysis of the Inseq9 desalt sample. (**c**) Immunoblot analysis of the autoglycosylated Inseq8 and Inseq9 samples. Lane 1 and lane 3 depict reaction mixtures of 0.2 mg mL^{-1} of Inseq8 and Inseq9, respectively, supplemented with 10mM $MnCl_2$ and 10 µM wild type LLO. Lane 2 and lane 4 depict the control desalt samples, in the absence of 10 mM $MnCl_2$ and 10 µM LLO, of Inseq8 and Inseq9, respectively. The glycosylation of Inseq variants was detected using the primary h6R antibody. (**d**) Anti-poly-histidine immunoblot analysis of samples described in (c).

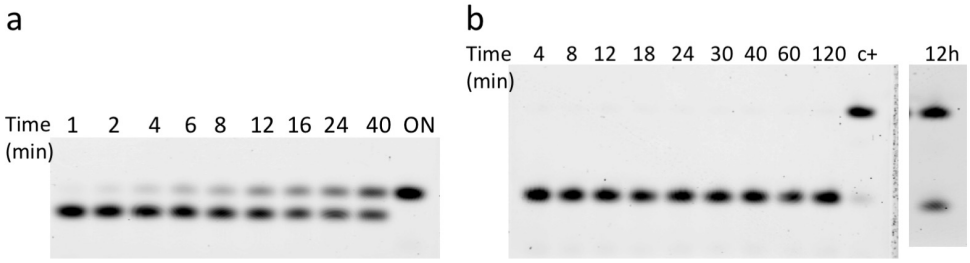


Figure 10. *In vitro* glycosylation activity of Inseq11. (**a**) Tricine-SDS-PAGE analysis of peptide glycosylation determined by quantification of fluorescently labeled substrate. The assays were performed at 30 °C and contained 0.1 µM PglB and 1mM synthetic LLO analog. 'ON' depicts a sample taken after overnight incubation. (**b**) Tricine SDS-PAGE analysis of peptide glycosylation determined by quantification of fluorescently labeled substrate. The assays were performed at 30 °C and contained 1 nM PglB and 50 µM wild type LLO. 'c+' depicts the control sample of the DQNATF peptide glycosylated by $PglB_{opt}$ with wild type heptasaccharide.

Optimizing expression, purification and stability of Inseq4 for crystallization experiments

Expression of Inseq 4

Inseq4 had a similar activity to wild type PglB and was stable when concentrated to 6 mg mL^{-1} (Fig. 7a and b). We therefore identified it as the most promising candidate for crystallization studies. However, the low yield of the Inseq4 protein posed a serious problem for high-throughput crystallization trials. We therefore rescreened the expression of Inseq4 in *E. coli* cell*s*, testing if production of Inseq4 was more efficient in different conditions than for wild type PglB. We tested various concentrations of inducing agent (arabinose), media and temperatures of induction and we confirmed that the highest amount and best quality protein was expressed when cells were grown in Terrific broth medium supplemented with 1 % glycerol and induced with 0.1 % arabinose for 4 h at 37 °C, the same as for wild type PglB and $PglB_{opt}$ (Fig. 11). The expression of Inseq4 overnight, even at low temperatures, resulted in protein degradation (red arrow, Fig. 11).

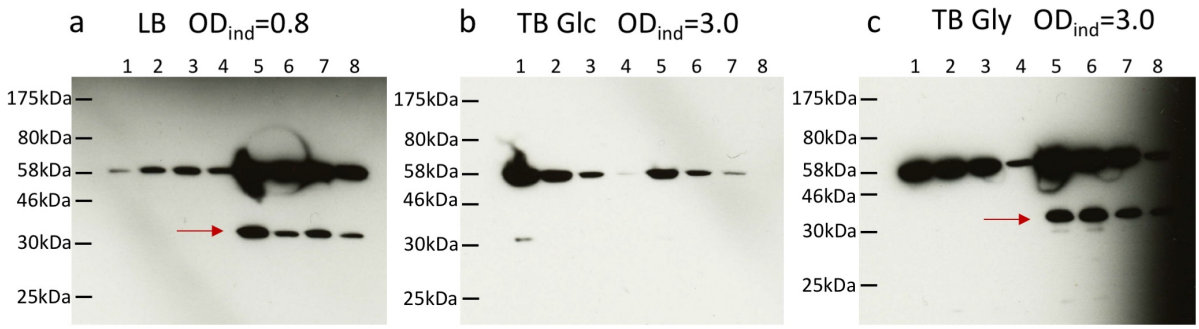


Figure 11. Screening expression conditions for Inseq 4. Expression of Inseq4 was tested in three different media (**a**) LB (Lysogeny broth), TB (Terrific broth) supplemented either with (**b**) 1 % glucose (Glc) or (**c**) glycerol (Gly). The cells were grown to OD= 0.8 for LB media or OD=3.0 for TB media and induced with different concentrations of arabinose and at different temperatures. Cells were disrupted and solubilized with 1% DDM. After centrifugation, the supernatant fractions were analyzed by SDS-PAGE. Lane 1: induction with 0.1 % arabinose for 4 h at 37 °C, lane 2: induction with 0.02 % arabinose for 4 h at 37 °C, lane 3: induction with 0.01 % arabinose for 4 h at 37 °C, lane 4: induction with 0.002 % arabinose for 4 h at 37 °C, lane 5: induction with 0.1 % arabinose overnight at 18 °C, lane 6: induction with 0.02 % arabinose overnight at 18 °C, lane 7: induction with 0.01 % arabinose overnight at 18 °C, lane 8: induction with 0.002% arabinose overnight at 18 °C. The samples loaded on the gel were normalized to grams of cells used. The expression was detected with an anti-polyhistidine antibody conjugated to HRP (horseradish peroxidase).

Subsequently, we investigated the solubilization efficiency of Inseq4 (Fig. 12). The cells were grown in TB medium supplemented with 1 % glycerol and induced with 0.1 % arabinose for 4 h at 37 °C. The cells were then disrupted by sonication and solubilized by the addition of 1 % DDM. These experiments revealed that the Inseq4 protein was not solubilized efficiently as approximately half of the protein stays in the pellet (Fig. 12a). The same experiment with $PglB_{opt}$ showed a much higher percentage of PglB in the solubilized fraction (Fig. 12b). Possibly, some of Inseq4 might be misfolded and therefore difficult to extract from the membranes.

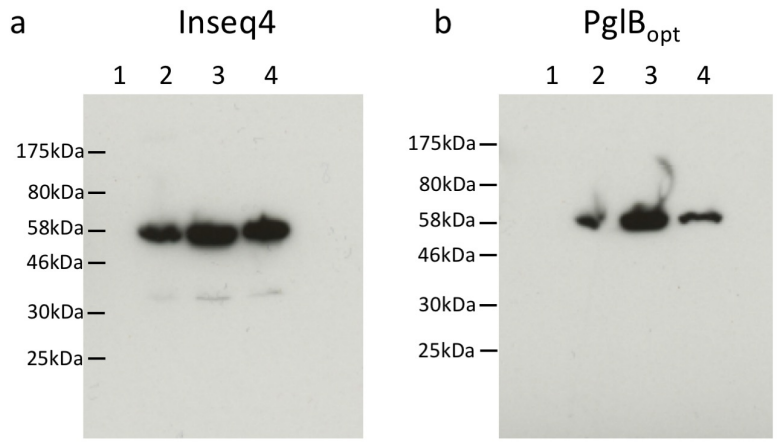


Figure 12. Comparison of the solubilisation efficiency of Inseq4 and PglB$_{opt}$. Both proteins were expressed in the same conditions, i.e. *E. coli* cells were grown in TB media supplemented with 1 % glycerol and induced with 0.1 % arabinose for 4 h at 37 °C. (**a**) Expression and solubilisation of Inseq4 (**b**) Expression and solubilization of PglB$_{opt}$. Lane 1: cells before induction (control), lane 2: whole cell control, lane 3: supernatant after solubilization, lane 4: resuspended pellet after solubilization.

Thermostability of Inseq4

The thermostability assay confirmed that the Inseq4 construct was 15 °C more stable than PglB$_{opt}$ (Table 3, Fig. 13). Although the Ts values calculated for Inseq4 and Inseq4-Nb64 are similar (Table 3), the plots shown in Figure 13c show that Nb64 did in fact have an additional thermostabilising effect on Inseq4. While there is a gradual decrease in the peak height of Inseq4 at temperatures above 35 °C, indicating that it is unfolding, the Inseq4-Nb64 complex is still stable. The Inseq4-Nb64 complex only begins to unfold at temperatures approaching 50 °C, at which point the signal decreases rapidly. Although $MnCl_2$ was not present in the Inseq-Nb64 samples, meaning direct comparisons cannot be made, Nb64 does seem to have a stabilizing effect on Inseq4.

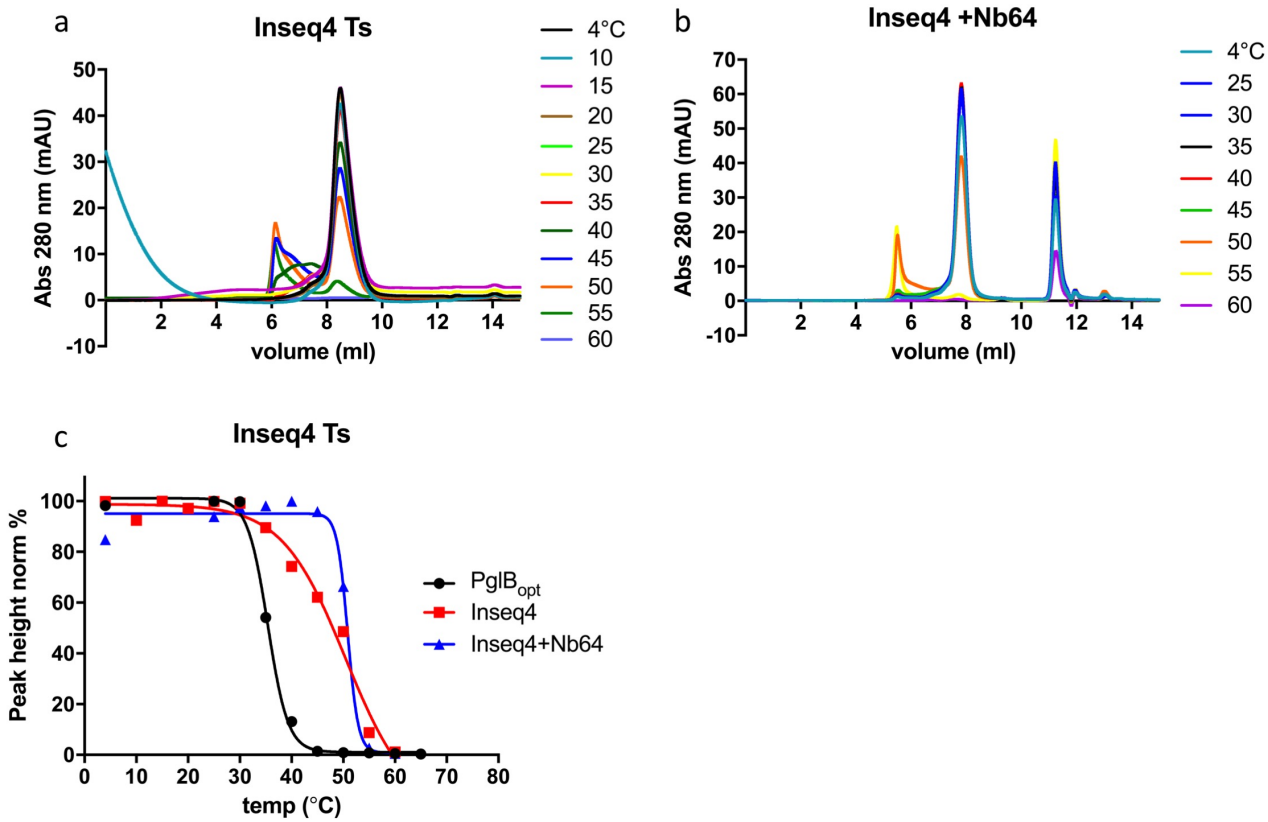


Figure 13. Thermostability of Inseq4. The samples were incubated at different temperatures for 10 min, cooled on ice and analyzed by size-exclusion chromatography using a TSK-gel G3000SWXL column. (**a**) Size-exclusion profiles of Inseq4 in 10 mM MES-NaOH pH 6.5, 100 mM NaCl, 3% glycerol, 1 mM $MnCl_2$ and 0.016 % DDM. (**b**) Size-exclusion profiles of Inseq4 and thermostabilizing nanobody 64. Samples were run in buffer containing: 10 mM MES-NaOH pH 6.5, 100 mM NaCl, 3 % glycerol, 0.016 % DDM. (**c**) SEC-Ts of all samples. Inseq4 is stabilized by the presence of an inserted acceptor sequence and by a PglB-specific nanobody.

Sample	**Ts (°C)**
$PglB_{opt}$	35.4 ± 0.2
Inseq4 +1mM $MnCl_2$	50.5± 2.6
Inseq4 +Nb64	50.9 ± 0.9

Table 3. Thermostability (Ts) values of the Inseq4 and Inseq4-Nb64 complexes obtained from the plots in Figure 13.

Screening of detergents

Nb64 was a valuable tool that we used for screening different detergents in order to facilitate Inseq4 crystallization. We added Nb64 to Inseq4, purified in DDM, and exchanged the buffers supplemented, with different detergents, on a Ni-NTA column. The stability and quality of Inseq4 was verified by SEC (Fig. 14). DM was the only detergent in which Inseq4 was as stable as in DDM. To further analyze DM-solublized Inseq 4 we fully purified it in DM-containing buffers. We added Nb64 during the solubilization step to stabilize Inseq4 at the start of the

purification. Although the fully purified sample looked monodisperse and we detected the Inseq4-Nb64 complex by SDS-PAGE, the sample did aggregate as it was concentrated (Fig. 15). The low stability of concentrated DM-solubilized Inseq4 suggested that further crystallization experiments should be carried out in DDM-purified Inseq4 in presence or absence of Nb64.

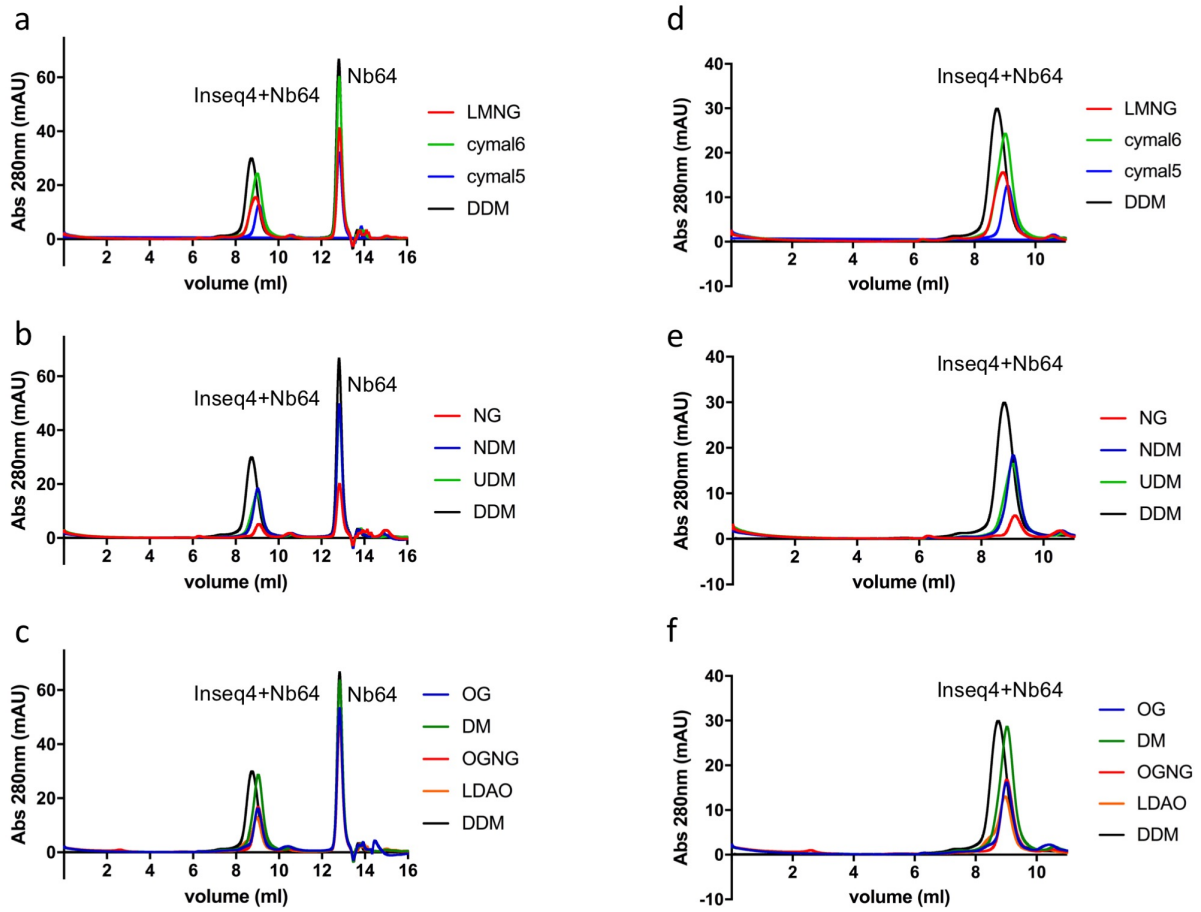


Figure 14. Detergent screening using SEC. Inseq4 stability was tested in different detergents (Table 1. (**a-c**) show full chromatograms. (**d-f**) show a close-up view of the Inseq4-Nb64 complex peak.

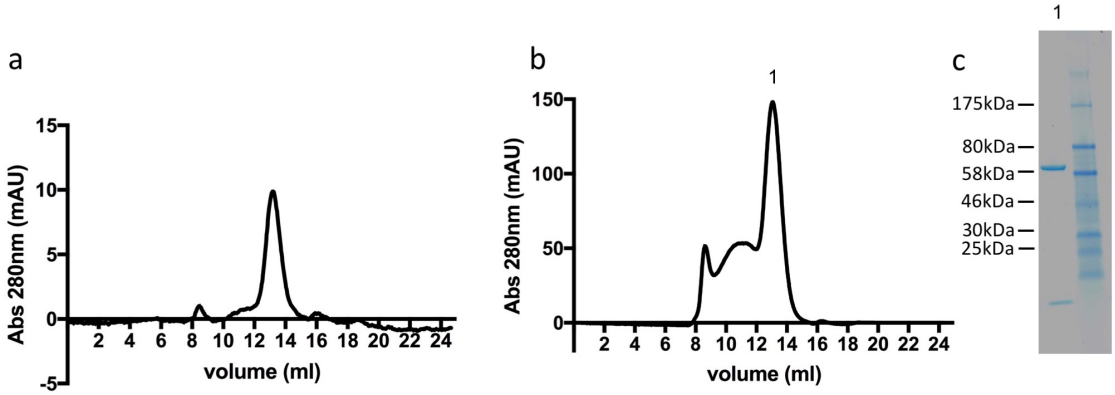


Figure 15. Purification of the Inseq4-Nb64 complex in DM. Inseq4 was purified in buffers containing DM. (**a**) Size-exclusion chromatogram of DM-solubilized Inseq4 purified in the presence of Nb64. (**b**) Size-exclusion chromatogram of Inseq4 purified in the presence of Nb64 and concentrated to 6 mg mL^{-1}. (**c**) SDS-PAGE analysis of the main peak.

Crystallization of Inseq4

Crystallization of Inseq4 with Nb64 and non-reactive LLO (Inseq4-Nb64-LLO) and with just non-reactive LLO (Inseq4-LLO) was set up with commercial and homemade screens. Many conditions yielded crystalline material. The crystals appeared after 14 days and were fully grown after 21 days. Interestingly, the Inseq4-Nb64-LLO crystals appeared in wells with high molecular weight poly(ethylene glycol)s (PEGs), such as PEG3350 and PEG8000 and a pH around 6-7. On the other hand, the Inseq4-LLO complex without nanobody present crystallized in conditions with low molecular weight PEGs, such as PEG400 or PEG300, and at low pH (4.5-5.5). All Inseq4 crystals were rather small (Fig. 16) and were further optimized in 96 and 24 well plates. However, despite these efforts, for Inseq-Nb64-LLO, only small crystals could be obtained (Fig 16b).

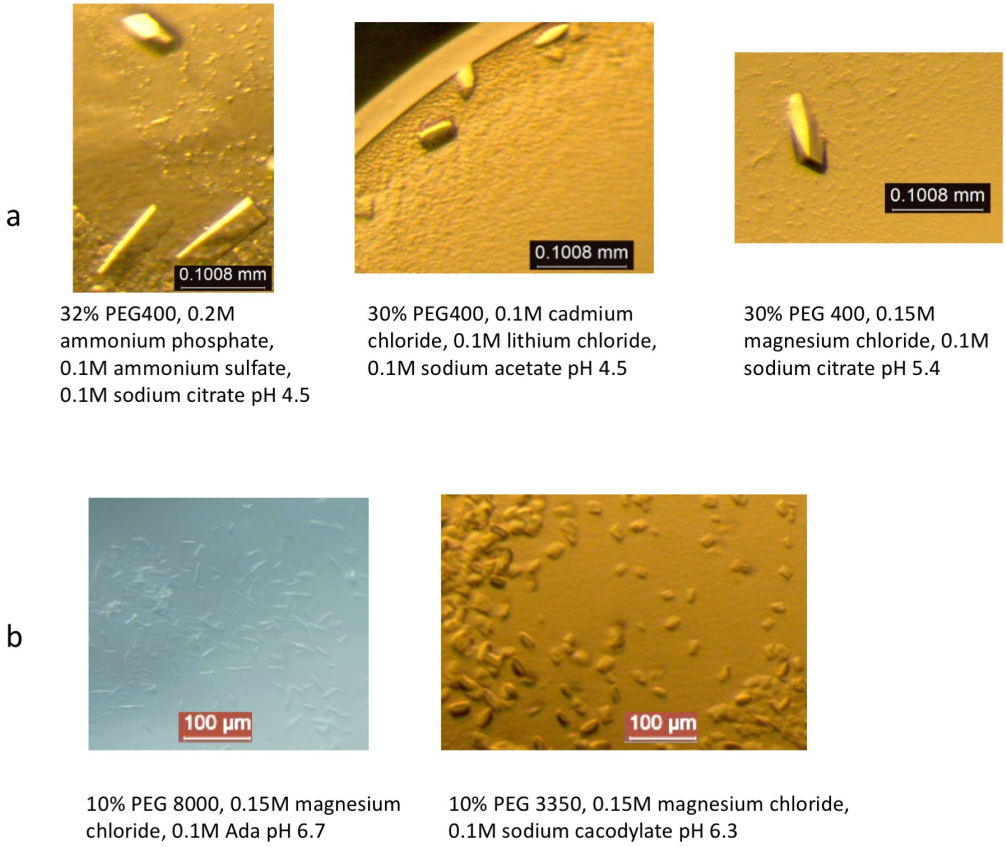


Figure 16. Crystals of Inseq4 in presence of a non-reactive LLO. (**a**) crystals of Inseq4 in the presence of LLO in low molecular weight PEG and low pH conditions. (**b**) crystals of the Inseq4 in the presence of LLO and Nb64 in high molecular weight PEG and neutral pH conditions.

Crystals were harvested and cryoprotected before freezing in liquid nitrogen. The crystals from the low pH conditions were very fragile and tended to break easily. Inseq4 crystals were further screened with a microfocus beam at the X06SA beamline, SLS, Villigen. While the crystals were not good enough to collect complete datasets, we were able to perform two 0° and 90° test shots. The Inseq4-Nb64-LLO crystals diffracted very poorly, to ~15 Å in both directions, while the best Inseq4-LLO crystal diffracted to 4.5 Å in both directions (Fig. 17). Further optimization of crystallization conditions, by screening additives or seeding experiments, may improve crystal growth, and may allow full datasets to be collected in order to determine the structure of Inseq4 bound to a non-reactive LLO.

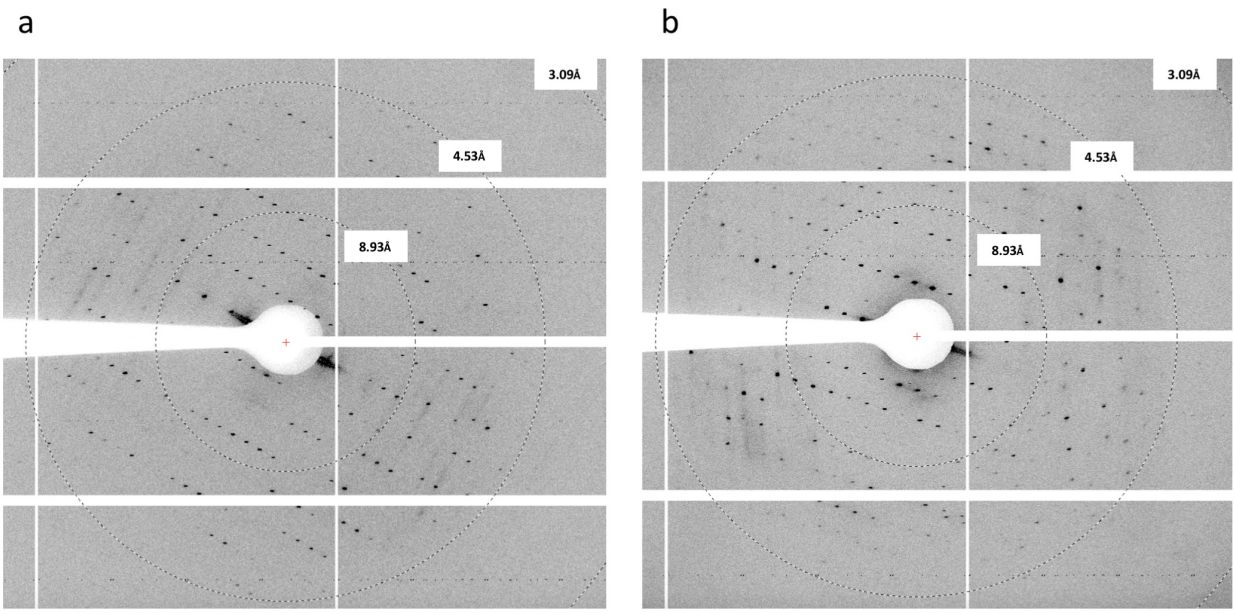


Figure 17. Diffraction patterns of initial Inseq4 crystals grown in the presence of LLO. The diffraction patterns were recorded at 1 Å wavelength with oscillation angle: 0.5°, exposure time: 0.5 s, beam transmission: 0.5 (**a**) Face image 0° (**b**) side image 90°.

Conclusions

Based on the structure of peptide-bound PglB, several amino acid mutations were introduced into wild type PglB from *C. lari* in order to improve protein crystallization, especially lattice contacts. While mutations did not influence protein crystallization, the construct displayed higher thermostability and was less prone to aggregation during concentration. Further experiments showed that the binding of the DQNAT acceptor sequon thermostabilized the protein by another 5 °C. The stabilizing effect of the peptide might have been a reason why only the peptide-bound state of PglB has been captured so far and crystallization of 'apo' PglB has not been successful [6].

In order to get more insight into the mechanism of glycan transfer and the critical interactions between the LLO and PglB, we aimed to structurally characterize PglB bound to both substrates. In order to facilitate structural studies of the ternary complex of PglB with an acceptor peptide and a non-hydrolyzable LLO analog, we designed different constructs carrying a glycosylation sequon inserted into either the periplasmic domain of PglB or the region linking TM13 and the periplasmic domain. Inserting the sequon into PglB would also increase the local concentration of the acceptor peptide, meaning the protein would not have to be diluted by adding an external acceptor peptide, therefore more crystallization conditions could be screened.

We successfully identified the best candidate, Inseq4, that was fully functional and 15 °C more thermostable than $PglB_{opt}$. We therefore used it in our crystallization trials. The complex of Inseq4 and non-hydrolyzable LLO analog (Inseq4-LLO) crystallized in conditions containing low molecular weight PEGs and at low pH, and the crystals diffracted to 4.5 Å. Further optimization of the crystallization of the Inseq4-LLO complex would be necessary to determine its structure. Although these results were very promising, we did not pursue this approach further. In parallel to screening the Inseq constructs, we also followed the strategy of directly co-crystallizing PglB with an external accetor peptide and a non-hydrolyzable LLO. This approach turned out to be more successful, as described in Chapter 4.

References

1. Larkin, A. & Imperiali, B. The expanding horizons of asparagine-linked glycosylation. *Biochemistry* **50,** 4411–4426 (2011).
2. Calo, D., Kaminski, L. & Eichler, J. Protein glycosylation in Archaea: Sweet and extreme. *Glycobiology* **20,** 1065–1076 (2010).
3. Kelleher, D. J. & Gilmore, R. An evolving view of the eukaryotic oligosaccharyltransferase. *Glycobiology* **16,** 47–62 (2006).
4. Wild, R. *et al.* Structure of the yeast oligosaccharyltransferase complex gives insight into eukaryotic N-glycosylation. *Science (80-.).* **359,** 545–550 (2018).
5. Ramírez, A. S. *et al.* Characterization of the single-subunit oligosaccharyltransferase STT3A from Trypanosoma brucei using synthetic peptides and lipid-linked oligosaccharide analogs. *Glycobiology* **27,** 525–535 (2017).
6. Lizak, C., Gerber, S., Numao, S., Aebi, M. & Locher, K. P. X-ray structure of a bacterial oligosaccharyltransferase. *Nature* **474,** 350–355 (2011).
7. Matsumoto, S. *et al.* Crystal structures of an archaeal oligosaccharyltransferase provide insights into the catalytic cycle of N-linked protein glycosylation. *Proc. Natl. Acad. Sci. U. S. A.* **110,** 17868–73 (2013).
8. Bai, L., Wang, T., Zhao, G., Kovach, A. & Li, H. The atomic structure of a eukaryotic oligosaccharyltransferase complex. *Nature* **555,** 328–333 (2018).
9. Braunger, K. *et al.* Structural basis for coupling protein transport and N-glycosylation at the mammalian endoplasmic reticulum. *Science (80-.).* **360,** 215–219 (2018).
10. Schwarz, F. & Aebi, M. Mechanisms and principles of N-linked protein glycosylation. *Curr. Opin. Struct. Biol.* **21,** 576–582 (2011).
11. Matsumoto, S., Taguchi, Y., Shimada, A., Igura, M. & Kohda, D. Tethering an N-Glycosylation Sequon-Containing Peptide Creates a Catalytically Competent Oligosaccharyltransferase Complex. *Biochemistry* **56,** 602–611 (2017).
12. Chen, M. M., Glover, K. J. & Imperiali, B. From peptide to protein: Comparative analysis of the substrate specificity of N-linked glycosylation in C. jejuni. *Biochemistry* **46,** 5579–5585 (2007).
13. Gerber, S. *et al.* Mechanism of bacterial oligosaccharyltransferase: In vitro quantification of sequon binding and catalysis. *J. Biol. Chem.* **288,** 8849–8861 (2013).
14. Lizak, C. *et al.* Unexpected reactivity and mechanism of carboxamide activation in bacterial N-linked protein glycosylation. *Nat. Commun.* **4,** 2627 (2013).
15. Lizak, C. *et al.* A catalytically essential motif in external loop 5 of the bacterial oligosaccharyltransferase PglB. *J. Biol. Chem.* **289,** 735–746 (2014).
16. Napiórkowska, M. *et al.* Molecular basis of lipid-linked oligosaccharide recognition and processing by bacterial oligosaccharyltransferase. *Nat. Struct. Mol. Biol.* **24,** 1100–1106 (2017).
17. Alaimo, C. *et al.* Two distinct but interchangeable mechanisms for flipping of lipid-linked oligosaccharides. *EMBO J.* **25,** 967–976 (2006).
18. Lizak, C., Fan, Y.-Y., Weber, T. C. & Aebi, M. *N* -Linked Glycosylation of Antibody Fragments in *Escherichia coli*. *Bioconjug. Chem.* **22,** 488–496 (2011).
19. Robert, X. & Gouet, P. Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res.* **42,** 320–324 (2014).

Chapter 4: Molecular basis of lipid-linked oligosaccharide recognition and processing by oligosaccharyltransferase PglB

This study was published in *Nature Structural and Molecular biology*.
Authors: Maja Napiórkowska[1], Jérémy Boilevin[2], Tina Sovdat[2,†], Tamis Darbre[2], Jean-Louis Reymond[2], Markus Aebi[3], Kaspar P. Locher[1]
[1] Institute of Molecular Biology and Biophysics, ETH Zurich, Switzerland
[2] Department of Chemistry and Biochemistry, University of Berne, Switzerland
[3] Institute of Microbiology, ETH Zurich, Switzerland
[†] Present address, Rolic Technologies Ltd., Allschwil, Switzerland

Contributions:
I performed the over-expression, purification, disulfide cross-linking, functional characterization, and crystallization of PglB. I performed X-ray data collection, structure determination, and model building and refinement with some assistance from K.P.L. J.B. and T.S. synthesized LLO analogs. T.D. and J.-L.R. supervised the chemical synthesis of the LLO analogs. M.N., M.A., and K.P.L. devised experiments and analyzed the data. K.P.L. and M.N. wrote the manuscript with input from all authors.

Abstract

Oligosaccharyltransferase (OST) is a membrane-integral enzyme that catalyzes the transfer of glycans from lipid-linked oligosaccharides (LLOs) onto asparagine side chains, the first step in protein *N*-glycosylation. Here, we present the X-ray structure of a single-subunit OST, the PglB protein from *Campylobacter lari*, trapped in an intermediate state bound to acceptor peptide and a synthetic LLO analog. We found that substrate recognition depends on the external loop EL5 present in all OSTs. Whereas the N-terminal half of EL5 binds LLO, the C-terminal half interacts with the peptide. The glycan moiety of the LLO must thread under EL5 to access the active site. Reducing EL5 mobility decreases the catalytic rate of OST when full-size heptasaccharide LLO is provided, but not for a monosaccharide-containing LLO analog. Our results define the chemistry of a ternary complex state, assign functional roles to conserved OST motifs, and provide opportunities for glyco-engineering by rational design of PglB.

Introduction

N-linked protein glycosylation is one of most abundant post-translational modifications of proteins that is present in all domains of life[1–4]. It is essential in eukaryotes, facilitating many biological processes such as protein folding, quality control or host-pathogen interactions[5–7], and influences virulence in bacteria[8,9]. The molecular mechanism of the glycan transfer from lipid-linked oligosaccharide (LLO) to the acceptor asparagine located in glycosylation sequons N-X-(S/T) is conserved[10,11] and catalyzed by the membrane-integral oligosaccharyltransferase (OST). OST is a member of the C family of glycosyltransferases (GT-C), which consists of membrane proteins that are located in the endoplasmic reticulum (ER) or the plasma membrane

and that contain 8-13 predicted transmembrane helices[12,13]. OST of higher eukaryotes is a multi-protein complex with STT3 as a catalytic subunit. In contrast, archea, kinetoplastids and some proteobacteria contain single-subunit OST enzymes (ssOST) that are homologous to STT3[7,14–16]. Eukaryotic and bacterial LLOs differ in the length and configuration of the polyprenyl tail and in the nature of the attached glycans. However, all contain a pyrophosphate or phosphate moieties that serve as the leaving groups of the nucleophilic substitution reaction[1,3,10].
The ssOST from the Gram-negative, pathogenic bacterium *Campylobacter jejuni* is the PglB protein. It is a well-studied, *bona fide* model system that allows the reaction mechanism of the OST-catalyzed process to be investigated without the complexity of additional subunits present in multi-subunit OSTs. The previously reported structures of PglB from *Campylobacter lari*[17] revealed the fold of STT3 proteins. The same fold was subsequently found in the related AglB protein from *Archaeoglobus fulgidus*[18,19]. Although structural and biochemical studies have provided insight into the mechanism of sequon recognition[17,20], the interactions of OST with LLO and the transfer mechanism remain poorly understood due to the absence of structural insight. Here we describe the crystal structure of PglB bound to synthetic substrate analogs, an acceptor peptide and a non-hydrolyzable LLO. This allowed a ternary complex to be trapped and its structure to be determined at 2.7 Å resolution. By combining the structural insight with chemo-enzymatic approaches and functional studies, we can define a key intermediate in the reaction mechanism of OST.

Experimental procedures

No statistical methods were used to predetermine sample size. The experiments were not randomized. The investigators were not blinded to allocation during experiments and outcome assessment.

Construction of plasmids
Christian Lizak generated PglB optimized construct.
A construct of PglB used in all experiments contained several amino acid changes generated to improve protein crystallization, i.e. lattice contacts. The optimized construct was fully functional, less prone to aggregation and exhibited an increased thermostability compared to wild type PglB (data not shown). Mutations in the *pglB* gene of *C. lari* were generated by the QuikChange method on the pSF2 plasmid[17]. The resulting cysteine-less construct with removed glycosylation sites and carrying mutations: K2E, C17A, C30A, A108T, C360L, N535Q, Q536K, K549P, D550N, F553I, N556P, A600P, A602D, T606K, T607Q, V610I, M611T, I619S, F622Y, A624S, V627I, A630N, F663Y and F670Y was referred to as PglB. All subsequent mutations in the LLO binding site were based on this construct.

Over-expression and purification of PglB
Overexpression and purification of PglB and PglB mutants was performed as previously described[17]. Shortly, PglB was overexpressed in *E. coli* BL21-Gold (DE3) (Strategene) cells at 37 °C in 5 l flasks using Terrific Broth medium supplemented with 1 % glycerol (w/v). The cells were induced at 37 °C for four hours at A_{600} of 3.0 by adding 0.1 % arabinose (w/v). Cells were harvested by centrifugation and cell pellets were stored at -80 °C. All following steps were carried out at 4 °C. Cells were resuspended in 25 mM Tris-HCl, pH 8.0, 250 mM NaCl and disrupted in a M-110-L microfluidizer (Microfluidics) at 15,000 p.s.i. chamber pressure. Membranes were pelleted by centrifugation at 100,000g for 30 min and solubilized in 25 mM Tris-HCl, pH 8.0, 250 mM NaCl, 10 % glycerol (v/v) and 1 % *N*-dodecyl-β-D-maltopyranoside

(w/v) (DDM, Anatrace) for 1.5 h. All subsequent purification buffers contained 0.016 % DDM. PglB was purified on a nickel-nitrilotriacetic acid (Ni-NTA) Superflow affinity column (Qiagen) and desalted into 10 mM MES-NaOH, pH 6.5, 100 mM NaCl, 3% glycerol (v/v), 0.016% DDM. For crystallization experiments desalted protein was concentrated to 6 mg mL^{-1} in an Amicon Ultra-15 concentrator (Milipore) with a molecular mass cutoff of 100 kDa and purified by size exclusion chromatography (Superdex 200 10/300, GE Healthcare) in desalting buffer. Collected fractions were pooled and concentrated to 12 mg mL^{-1}.

Purification of the PglB V294C/R371C mutant

The PglB V294C/R371C mutant was purified as described above with addition of 10 mM β-mercaptoethanol to resuspension, solubilization, washing and elution buffers. The mutant was firstly desalted into 25mM Tris-HCl, pH 8.0, 250mM NaCl, 10 % glycerol (v/v), 0.016 % DDM supplemented with 1 mM $CuCl_2$ and incubated for 30 min at 4 °C, then desalted into 10 mM MES-NaOH, pH 6.5, 100 mM NaCl, 3 % glycerol (v/v), 0.016 % DDM. A small fraction of noncross-linked V294C/R371C PglB was used as a control for quantification of the cross-linking efficiency and glycosylation activity. For glycosylation turnover measurements of the reduced V294C/R371C PglB mutant, cross-linked samples were treated with 10 mM dithiothreitol (DTT).

Quantification of cross-linking efficiency

The estimation of the cross-linking efficiency was performed by labeling free cysteines with fluoresceine 5-maleimide. The cross-link and control sample were incubated with a 10-fold molar excess of fluoresceine 5-maleimide over cysteines for 10 min at 4 °C. The reaction was quenched with 2 mM *N*-ethylmaleimide. The same amount (0.5 μg) of labeled samples was resolved by SDS-PAGE and fluorescent bands were visualized using the Typhoon Trio Plus imager (GE Healthcare). Band intensities were calculated in ImageJ.

Protein crystallization

Ac-DQNATF{p-NO_2}-NH2, nerylneryl-PPC-GlcNAc and $MnCl_2$ were added to the concentrated protein to the final concentration of 0.75 mM, 2 mM, 1 mM, respectively and incubated for 15 min at 4 °C. PglB was crystallized by vapor diffusion in hanging drops at 20 °C against a reservoir containing 100 mM glycine-NaOH, pH 9.4, 120 mM Li_2SO_4, 29-31 % PEG400 at the protein to reservoir ratio 2:1. Crystals usually appeared after 10-12 days and grew to full size within 3-4 weeks. Crystals were cryoprotected by stepwise addition of cryoprotectant PEG400 up to 30 % final concentration and directly flash frozen in liquid nitrogen before data collection.

Data collection and structural determination

The X-ray data was collected at the microfocus X06SA beamline at the Swiss Light Source (Villigen). The wavelength for data collection was 0.9998 Å. All data were processed using XDS[21]. There was no indication of any severe diffraction anisotropy. The ternary complex crystallized in $P2_12_12_1$ space group, containing one PglB molecule per asymmetric unit. The structure was solved by molecular replacement using peptide-only PglB (PDB: 3RCE) as a model. The iterative model building and refinement was performed in Coot[22] and Phenix[23]. Final refinement statistics and data collection are summarized in Table 3. The average B factors were calculated with Baverage from the CCP4 program suite. Ramachandran analysis: Ramachandran favored 93.23 %, Ramachandran allowed 5.78 %, Ramachandran outliers 0.99 %. Structural figures were drawn using PyMOL[24].

C. jejuni wild type LLO extraction

Extraction of the *C. jejuni* heptasaccharide wild type LLO (WT LLO) was done as described before[25]. Shortly, LLO was extracted from SCM6 *E.coli* cells carrying a *C. jejuni* $pglB_{mut}$ cluster, containing an inactivated *pglB* gene. LLO was extracted with a mixture of chloroform: methanol: water (10:10:3 v:v:v) followed by rotary evaporation. The LLO was reconstituted in 10 mM MES-NaOH, pH 6.5, 100 mM NaCl, 1 % Triton X-100 (w/v) buffer and its concentration was determined by titrating different amounts of LLO against constant amounts of fluorescently labeled peptide in an *in vitro* glycosylation assay (see below).

In vitro glycosylation assay

In vitro glycosylation assay were performed as described previously[20] using a labeled acceptor peptide 5-carboxyfluorescein-GS-DQNATF-NH_2 (10 μM) as an acceptor substrate and wild type LLO (50 μM) or nerylneryl-PP-GlcNAc (500 μM) as a donor substrate. Reaction mixtures with wild type LLO and nerylneryl-PP-GlcNAc contained 1 nM or 100 nM PglB, respectively. 1 % Triton X-100 (w/v) was added only to wild type LLO reaction mixtures. For turnover rate determination, a total of 6 samples were taken at different time intervals so that the reaction was in the linear range. Data was fitted by linear regression in GraphPad PRISM7.

Determination of the Michaelis- Menten constants

Michaelis-Menten kinetics of the monosaccharide and disaccharide LLO analogs were determined using the *in vitro* glycosylation assay described above. 100 nM PglB was used for assays containing monosaccharide LLO analogs, whereas 1 μM PglB was used for assays containing disaccharide LLO analogs. The concentrations of LLO analogs varied from 1-1000 μM. Samples were taken at different time points and initial turnover rates were calculated for each concentration of LLO analog using linear regression. The resulting rates were plotted against LLO concentration and the data was fitted by nonlinear regression using the Michaelis-Menten analysis tool in GraphPad PRISM7.

Determination of the apparent IC_{50} values

Apparent IC_{50} values of non-hydrolyzable LLO analogs were determined using the *in vitro* glycosylation assay described above. For the kinetics of inhibitory LLO analogs, reaction mixtures contained 17 μM nerylneryl-PP-GlcNAc and 10-5000 μM of competitive LLO inhibitors. PglB was added at a final concentration of 100 nM to start the reaction. Samples were taken at different time points and initial turnover rates were calculated using linear regression. The resulting rates were plotted against the concentration of the inhibitory LLO and the data was fitted by nonlinear regression in GraphPad PRISM7.

Synthesis of LLO analogs

Six LLO substrate analogs were synthesized according to previously reported procedures with slight modifications[26–28]. Five of them were composed of a *N*-acetyl-D-glucosamine (GlcNAc) residue linked to polyprenyl pyrophosphates of different chain length and stereochemistry (*E*-geraniol, *Z*-nerol, *E,E*-farnesol, *E,E,E*-genanylgeraniol and *Z,Z,Z*-nerylnerol) while the last one had a chitobiose moiety coupled to *E,E*-farnesol. The synthetic strategy included three steps: the synthesis of GlcNAc or chitobiose α-phosphate, the preparation of the lipid phosphate precursors and the coupling of both monophosphates to the desired final pyrophosphates. Nerylneryl-PP-GlcNAc ((ωZZZ)-PP-GlcNAc), geranylgeranyl-PP-GlcNAc ((ωEEE)-PP-GlcNAc), farnesyl-PP-GlcNAc ((ωEE)-PP-GlcNAc), neryl-PP-GlcNAc ((ωZ)-PP-GlcNAc), geranyl-PP-GlcNAc ((ωE)-PP-GlcNAc) and farnesyl-PP-chitobiose ((ωEE)-PP-GlcNAc-1,4-β-GlcNAc) were isolated as final compounds. Phosphonate LLO analogs nerylneryl-PPC-GlcNAc ((ωZZZ)-PPC-GlcNAc) and nerylneryl-PPC(OH)-GlcNAc ((ωZZZ)-PPC(OH)-

GlcNAc) were obtained using previously described procedures with modifications[27,29]. The synthesis required 10 and 8 steps respectively from GlcNAc, where the hydroxyl-phosphonate analog was an intermediate in the phosphonate synthesis. Nerylneryl-phosphate (Fig. 1c, 4) (1.00 eq) and CDI (5.00 eq) were dissolved in *N,N*-dimethylformamide (DMF) and stirred for three hours at room temperature under an argon atmosphere. MeOH (10.00 eq) was added and the reaction mixture was stirred for one hour to quench the unreacted CDI. MeOH was removed *in vacuo*. A solution of the corresponding saccharide precursor (Fig. 1c, 1-3) (3.00-4.00 eq) in DMF was added to the reaction mixture which was then stirred for six days. The reaction mixture was concentrated *in vacuo* and the residue was dissolved in MeOH and a large excess of NH_4OH was added. The reaction mixture was stirred at room temperature for twelve hours, then concentrated at the rotavapor and freeze-dried. The residue was purified by flash chromatography to yield the corresponding LLOs (Fig. 1c, 5-7) as colorless lyophilisats. All synthetic compounds but nerylneryl-PPC(OH)-GlcNAc were shown to be pure by ^{1}H, ^{13}C, ^{31}P NMR and ESI-HRMS. nerylneryl-PPC(OH)-GlcNAc could not be obtained at a purity of more than 90 %.

Chemoenzymatic synthesis of farnesyl-PP-GlcNAc-1,3α-GalNAc

Farnesyl-PP-GlcNAc-1,3α-GalNAc was obtained by the chemoenzymatic approach previously described, where GalNAc was enzymatically added to synthetic undecaprenyl-PP-GlcNAc by PglA[30,31]. The obtained product could be verified by Tricine SDS-PAGE analysis.

Results

***In vitro* functional assays and inhibitory LLO analogs**

The main challenge for trapping a ternary complex intermediate of OST was to ensure that the glycan transfer did not proceed. This could in principle be achieved by generating non-functional PglB mutants or by using substrate analogs that could bind PglB but did not allow glycan transfer. We did not use PglB mutants because for all active site mutations that were investigated in the past, binding of the peptide substrate was also severely reduced. We next considered the use of non-reactive peptide substrates. Arguably, the least-reactive substitution of the acceptor asparagine residue in the sequon is 2,4-diaminobutanoic acid (Dab). Dab-containing peptides were previously found to act as competitive inhibitors both of prokaryotic[32] and eukaryotic OST[33–35]. While we had earlier shown that a Dab-containing sequon also inhibited PglB[32], the absence of the carboxamide group reduced the affinity of the corresponding peptide to PglB approximately 10-fold relative to an Asn-containing sequon. Also, the activation mechanism of the carboxamide group cannot be studied using a Dab-containing peptide. We therefore synthesized various LLO analogs (Fig. 1a) with the goal of identifying a useful compound for structural studies. It was previously shown that water-soluble, monosaccharide-containing LLOs can serve as substrates for PglB[26]. We followed a similar synthetic approach and generated LLO analogs containing one *N*-acetyl-D-glucosamine (GlcNAc) moiety, a pyrophosphate group, and polyprenyl tails of different lengths in either *cis* or *trans* configuration.

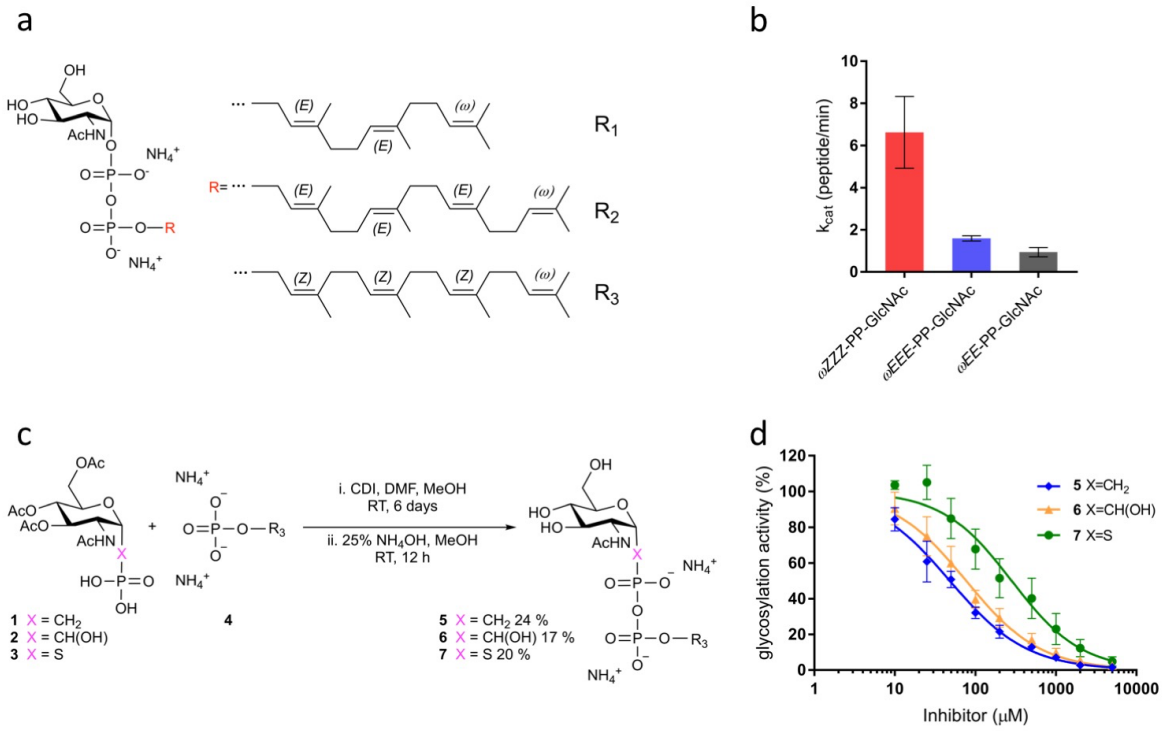


Figure 1. *In vitro* inhibition of PglB with synthetic LLO analogs. (**a**) Structure of synthetic lipid-linked monosaccharides. Red "R" denotes the polyprenyl tails used: R_1: farnesyl or (EEω), R_2: geranylgeranyl or (EEEω), R_3: nerylneryl or (ZZZω). (**b**) Maximal turnover rates of PglB-catalyzed glycosylation reaction using fluorescently labeled peptide (DQNATF) and synthetic donor LLOs shown in a. The data represent three independent cell cultures (error bars indicate s.d. n≥6) (**c**) Synthesis of inhibitory LLO analogs **5-7** by a coupling reaction between saccharide precursors **1-3** and lipid phosphate **4**. (**d**) Kinetic analysis of inhibitory LLO analogs shown in (c). PglB-catalyzed reaction using 17 μM ωZZZ-PP-GlcNAc with various concentrations of inhibitory LLO analogs. Deduced IC_{50} values are given in Table 2. Each point represents the average initial turnover rate from three independent protein preparations. Error bars indicate s.d (n≥4).

To determine the rates and Michaelis-Menten constants of the synthetic LLOs, we used a fluorescently labeled peptide carrying the glycosylation sequon to monitor the glycosylation reaction by in-gel fluorescence[20]. Although it was previously suggested that at least 4 isoprenyl units are required for PglB activity[26], we found that compounds with as few as two isoprenyl units allowed the glycosylation reaction to proceed, albeit with lower rates (Fig. 2). We tested two compounds that contained 3 or 4 isoprenyl units in trans configuration, (ωEE)-PP-GlcNAc (or farnesyl-PP-GlcNAc) and (ωEEE)-PP-GlcNAc (or geranylgeranyl-PP-GlcNAc), and one compound containing 4 isoprenyl units in cis configuration, (ωZZZ)-PP-GlcNAc (or nerylneryl-PP-GlcNAc). The latter recapitulates the stereochemistry of the wild type LLO lipid moiety. Although all synthetic LLO analogs were recognized as substrates by PglB, their turnover rates (Fig. 1b, Table 1) were considerably lower than the 1.55 +/- 0.32 peptide/s (error denotes s.d. of six independent cell cultures, n=6) we observed with full-length, *C. jejuni* heptasaccharide LLO (wild type LLO). Intriguingly, whereas the K_M of the (ωEE)-PP-GlcNAc compound was 6-fold higher than that of the (ωEEE)-PP-GlcNAc compound, the turnover rate was only 2-fold lower (Table 1, Fig. 3). This suggests that while longer polyprenyl tails indeed increase the affinity of LLO analogs for PglB, the steps following LLO binding also determine the rate of the reaction. While the stereochemistry of the double bonds of polyprenyl tail had no impact on LLO affinity (unaltered K_M values), it affected the turnover rates (Fig. 1b, Fig. 3, Table 1), indicating an impact on the transfer reaction. Similar specificity was observed for the

O-OST PglL from *Neisseria meningitides*, where the stereochemistry of the lipid carrier strongly influenced the glycosylation reaction[36]. We identified (*ω*ZZZ)-PP-GlcNAc as a PglB substrate with a high turnover rate (Fig. 1b) while having sufficiently high solubility, which is required for co-crystallization experiments.

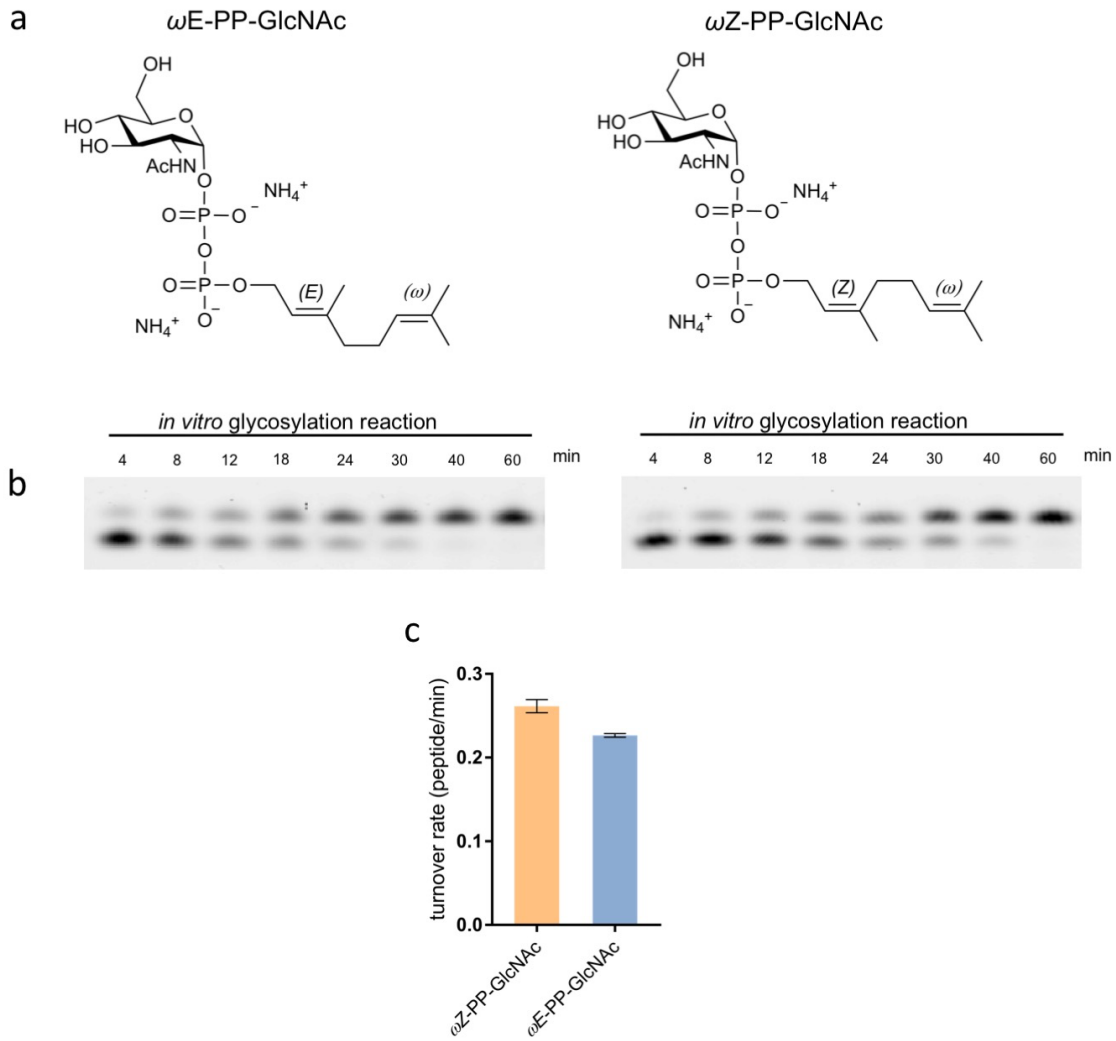


Figure 2. *In vitro* activity of PglB with synthetic LLO analogs. PglB activity was measured *in vitro* using synthetic LLO analogs with distinct polyprenyl tails. (**a**) Structures of synthetic LLO analogs featuring *cis* (*ω*Z-PP-GlcNAc) or *trans* (*ω*E-PP-GlcNAc) configuration of double bonds. (**b**) Tricine SDS-PAGE analysis of peptide glycosylation determined by quantification of fluorescently labeled substrate. The assays were performed at 30 °C and contained 1 μM PglB and 100 μM synthetic LLO analog. (**c**) Turnover numbers of glycosylation assays shown in (b). Data represent three independent protein preparations (error bars indicate s.d., n=3).

This compound then served as a template for the design of non-reactive LLOs. It has been shown that non-hydrolyzable substrate analogs act as inhibitors rather than donors for glycosyltransferases that use nucleotide-activated[37] or lipid-linked glycans[27]. In part following earlier procedures[27,29,38] three non-hydrolyzable LLOs were synthesized, (*ω*ZZZ)-PPS-GlcNAc, (*ω*ZZZ)-PPC-GlcNAc and (*ω*ZZZ)-PPC(OH)-GlcNAc, which contain unreactive thiophosphate, phosphonate, or hydroxy-phosphonate groups, respectively, replacing the first phosphate group connected to the anomeric C1 atom of the reducing-end GlcNAc. The GlcNAc phosphonate[29] (Fig. 1c, **1**), GlcNAc hydroxy phosphonate[29] (Fig. 1c, **2**) and GlcNAc

thiophosphate[39,40] (Fig. 1c, **3**) were individually coupled to nerylneryl phosphate[26] (Fig. 1c, **4**) using 1,1'-carbonyldiimidazole (CDI), followed by deacetylation with NH_4OH and purification on basified silica gel[27]. Apparent IC_{50} of these compounds were determined in PglB-catalyzed glycosylation containing (*ω*ZZZ)-PP-GlcNAc as the functional LLO (Fig. 1d). The results showed similar inhibitory potencies for the phosphonate- and hydroxy-phosphonate-containing LLO analogs, but a lower potency (higher IC_{50} value) for the thiophosphate-containing analog (Table 2). All IC_{50} values were considerably higher than the K_M value of the corresponding, functionally active (*ω*ZZZ)-PP-GlcNAc) compound, suggesting that the oxygen atom of the natural pyrophosphate moiety directly or indirectly (e.g. due to a different bond angle or bond length[41]) contributes to LLO binding.

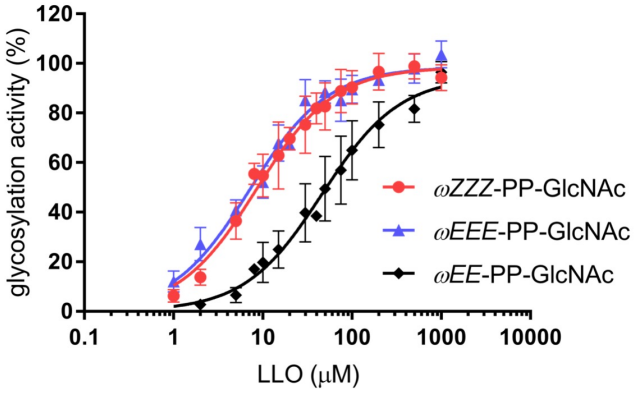


Figure 3. Kinetic analysis of synthetic LLO substrates. Normalized PglB activities as a function of varying concentrations of three LLO analogs (structures shown in Fig. 1a). Activities were normalized to the maximal turnover rate for each LLO analog independently to allow the comparison of the Michaelis constants K_M. Each data point represents three independent cell cultures, calculated from the slope of the linear regression fit (error bars indicate s.d., n≥4).

LLO	**K_M (μM)**	**k_{cat} (peptide/min)**
(*ω*EE)-PP-GlcNAc	46.3 +/- 5.8	0.94 +/- 0.22
(*ω*EEE)-PP-GlcNAc	7.2 +/- 0.7	1.60 +/- 0.12
(*ω*ZZZ)-PP-GlcNAc	8.5 +/- 0.8	6.63 +/- 1.70
(*ω*EE)-PP-GlcNAc-1,3α-GalNAc	20.8 +/- 2.0	1.64 +/- 0.04
(*ω*EE)-PP-GlcNAc-1,4β-GlcNAc	46.0 +/- 3.0	0.0053 +/- 0.0001

Table 1. K_M and k_{cat} values of synthetic LLO analogs used for PglB-catalyzed glycosylation, calculated from data shown in Fig. 3 and Figs. 9c and 9d. For Michaelis-Menten kinetics of the monosaccharide LLO analogs and disaccharide LLO analogs, 0.1 μ M and 1 μM PglB was used respectively. For monosaccharide-containing LLO analogs error denotes s.d of three independent cell cultures n≥6, whereas for disaccharide-containing LLO analogs error denotes s.d. of three technical replicates n≥3.

Inhibitory LLO	IC_{50} (μM)
(ωZZZ)-PPS-GlcNAc	283 +/- 35
(ωZZZ)-PPC(OH)-GlcNAc	77.2 +/- 5.8
(ωZZZ)-PPC-GlcNAc	48.5 +/- 2.8

Table 2. Apparent IC_{50} values of different inhibitory LLO analogs calculated from data shown in Fig. 1d. Error denotes s.d of four independent protein preparations n≥4.

Structure of the ternary complex and peptide binding pocket

We co-crystallized a sequence-optimized variant of *C. lari* PglB with the hexapeptide DQNATF and the synthetic, non-hydrolyzable LLO analog (ωZZZ)-PPC-GlcNAc and determined its structure at 2.7Å resolution (Table 3, Fig. 4a). The PglB construct contained several amino acid changes introduced with the aim of improving lattice contacts. The construct exhibited an increased thermostability (data not shown), was less prone to aggregation, but its *in vitro* glycosylation activity was indistinguishable from the 1.50 +/- 0.04 peptide/s determined previously for wild type PglB[20].

The architecture of PglB is similar to the previously determined, peptide-bound PglB structure[17], where the C- terminal fragment of EL5 (C-EL5, located between transmembrane helices TM9 and TM10a) pins the bound acceptor peptide against the periplasmic domain of PglB. However, we found that the previously disordered, N-terminal segment of the external loop EL5 (N-EL5) now adopts a defined three-dimensional conformation, folding over and significantly contributing to the binding of LLO (Fig. 4a, 5 and 6). Notably, N-EL5 forms an α-helix located at the membrane boundary, with the helix axis forming an angle of ~30° relative to the membrane plane. A similar helix was observed in one of the apo-structures of AglB[18], suggesting that N-EL5 has a propensity for forming this secondary structure element. However, this is functionally relevant only upon LLO binding.

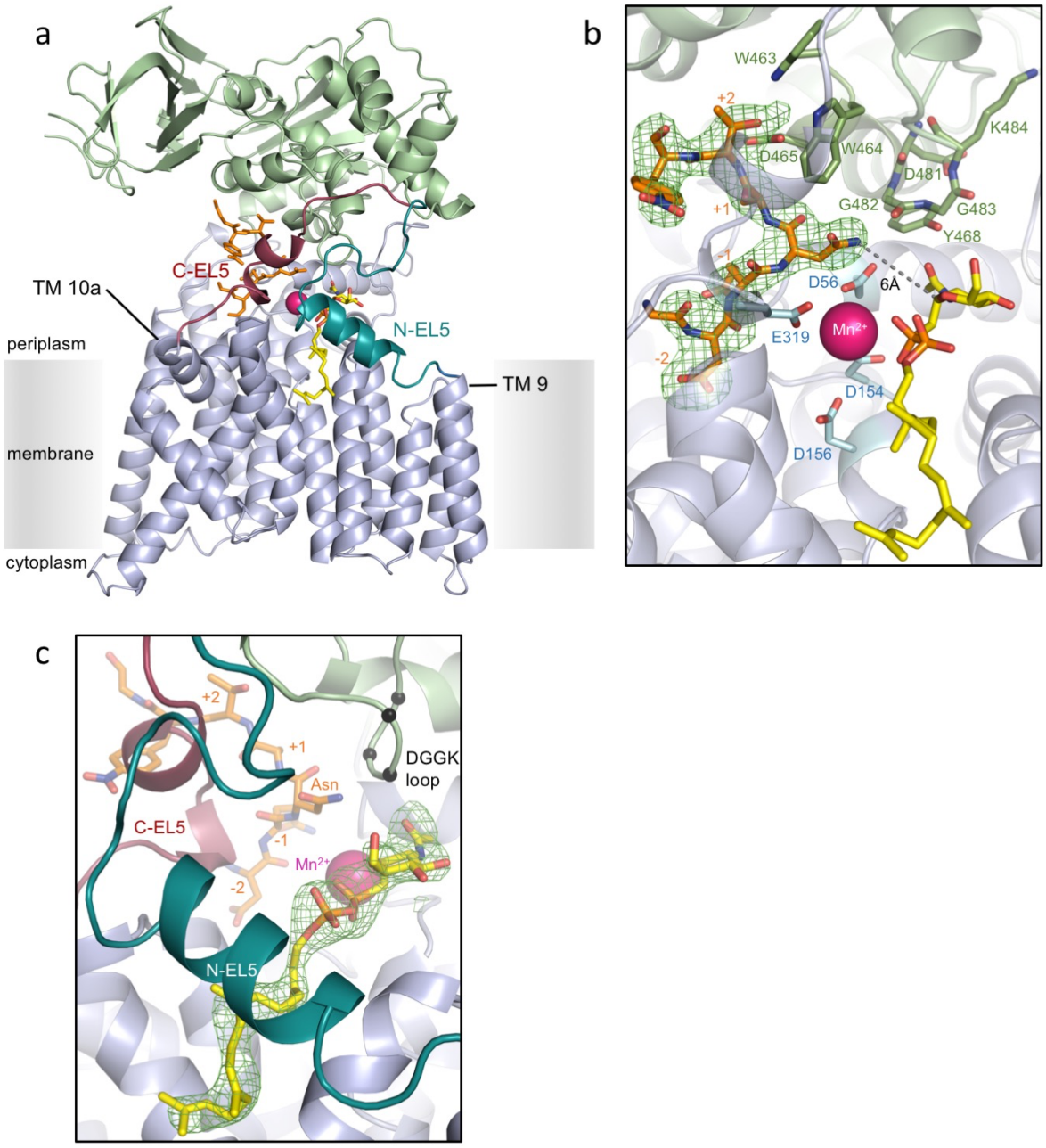


Figure 4. Structure of the ternary complex. (**a**) Ribbon diagram of PglB with transmembrane domain in light blue and periplasmic domain in light green. TM helices flanking EL5 are indicated. N-terminal part of EL5 (N-EL5) is colored in turquoise, C-terminal part of EL5 (C-EL5) in dark purple. Mn^{2+} is shown as a pink sphere. Bound acceptor peptide and inhibitory LLO are shown as orange and yellow sticks, respectively. (**b**) Close-up view of the active site, with PglB shown as a ribbon diagram and colored as in (a). The green mesh depicts an omit Fo-Fc map contoured at σ=3.0 and only shown around bound acceptor peptide, which is shown as orange sticks, with numbers indicating sequon positions relative to the acceptor Asn (zero). Relevant residues of the WWD and DGGK motifs or contributing to the active site are shown as sticks and labeled. Bound LLO is shown as yellow sticks. EL5 was removed for clarity. The dashed line depicts the distance between the carboxamide nitrogen of the acceptor Asn and the C1 atom of the reducing-end GlcNAc. (**c**) Close-up view of the active site similar to (a), but depicting N-EL5 (colored turquoise) and C-EL5 (colored purple). The green mesh depicts the same omit map as in b but shown around bound LLO. Black spheres represent the Cα atoms of the DGGK motif.

	Ternary complex[1]
Data collection	
Space group	$P2_12_12_1$
Cell dimensions	
a, *b*, *c* (Å)	83.94, 116.18, 172.02
(°)	90.0, 90.0, 90.0
Resolution (Å)	47.77-2.70 (2.80-2.70)*
R_{merge}	0.087 (1.36)*
I / σ	21.42 (2.18)*
Completeness (%)	100 (100)*
Redundancy	12.8 (13.5)*
Refinement	
Resolution (Å)	47.7-2.7
No. reflections	601398 (62190)*
R_{work} / R_{free}	0.217/0.244
No. non hydrogen atoms	5941
Protein	5816
Peptide	59
LLO	43
Mn^{2+} ion (I)	1
Mn^{2+} ion (II)	1
Na^{+} ion	1
Water	20
B-factors (Å^2)	
Protein	79.2
Peptide	58.3
LLO	77.9
Mn^{2+} ion (I)	48.8
Mn^{2+} ion (II)	70.0
Na^{+} ion	55.2
Water	57.6
R.m.s. deviations	
Bond lengths (Å)	0.009
Bond angles (°)	1.12

Table 3. Data collection and refinement statistics (molecular replacement).
[1] data collected from one crystal
*Values in parentheses are for highest-resolution shell.

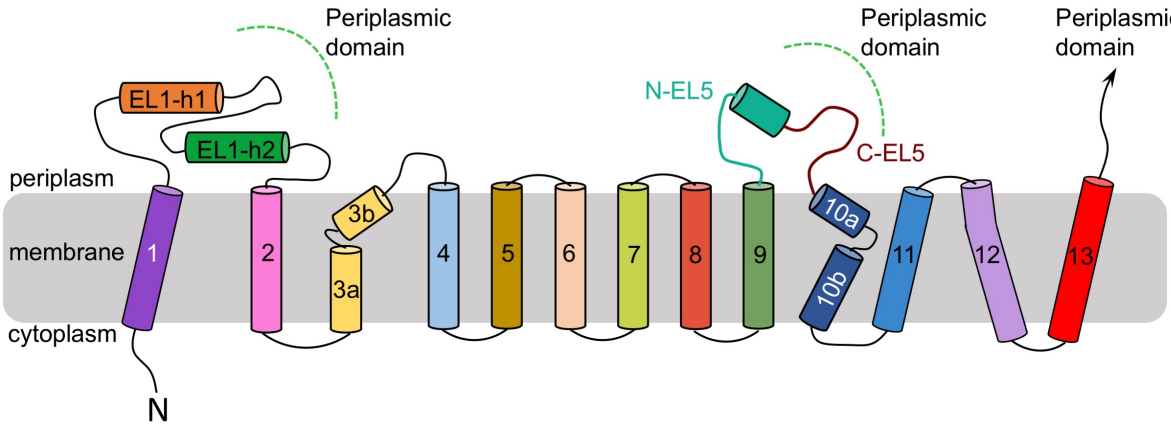


Figure 5. PglB membrane topology. Topological scheme of PglB transmembrane domain, with TM helices depicted as cylinders and numbered. Dashed green lines indicate non-covalent contacts to the periplasmic domain. The N-terminal and C-terminal segments of the external loop EL5 are indicated as N-EL5 and C-EL5, respectively.

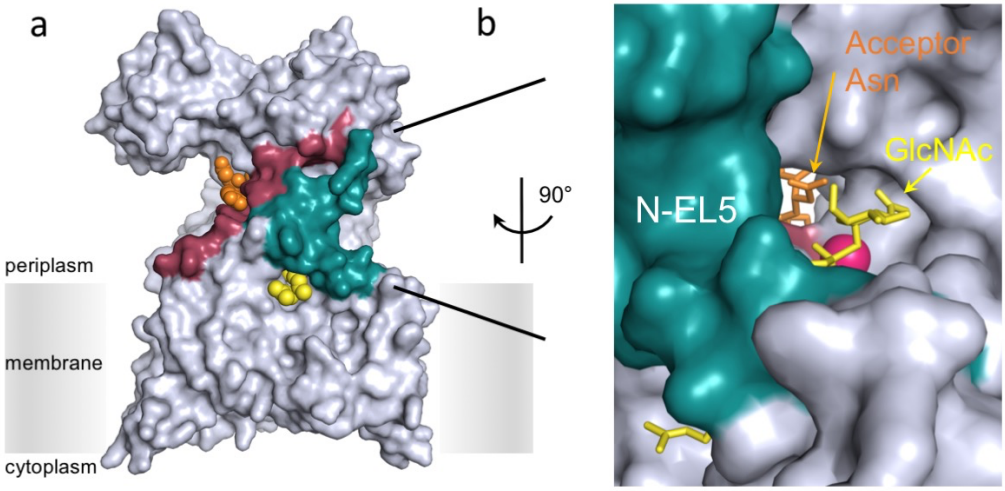


Figure 6. Surface representation of the ternary complex structure. PglB- acceptor peptide-LLO complex. (**a**) Surface representation of the ternary complex structure of PglB bound to acceptor peptide and synthetic, inhibitory LLO analog. PglB is colored grey, N-EL5 in turquoise, C-EL5 in purple, bound peptide in orange, bound LLO in yellow. (**b**) Close-up view of (a). The acceptor peptide and inhibitory LLO are shown in orange and yellow sticks, respectively. The substrates are bound to cavities at opposite entrances / exits of the PglB tunnel.

High quality electron density was found for both substrates (Fig. 4b and c), which are bound on opposite sides of the tunnel connecting the active site of PglB, indicated by the catalytic Mn^{2+} ion, to the peptide-binding pocket. The location of bound peptide in the ternary complex structure is unaltered relative to the peptide-only structure[17], with the side chain of the acceptor asparagine located in a tunnel formed between the C-terminal segment of EL5 (C-EL5) and the core of PglB (Fig. 4b). The carboxamide group of the acceptor Asn is in close proximity to the catalytically essential residues D56 and E319 as well as the catalytic Mn^{2+} ion (Fig. 7a). While the exact rotamer of the carboxamide group of the acceptor Asn side chain cannot be defined, we find continuous electron density to the nearby D56 residue (but not to E319), suggesting a strong contact in the form of a hydrogen bond. No bound water could be identified near the carboxamide group of the acceptor Asn, but three water molecules could be placed near the Mn^{2+} ion. Combined with the contacts to the conserved residues D56, E319 and D154, this leads to a distorted octahedral coordination of the metal (Fig. 7b). Importantly, the nitrogen atom of the acceptor Asn side chain is approximately 6 Å apart from the C1 atom of the reducing-end GlcNAc of the LLO analog (Fig. 4b), suggesting that in order to reach the transition state of the reaction, further closing of the distance is required.

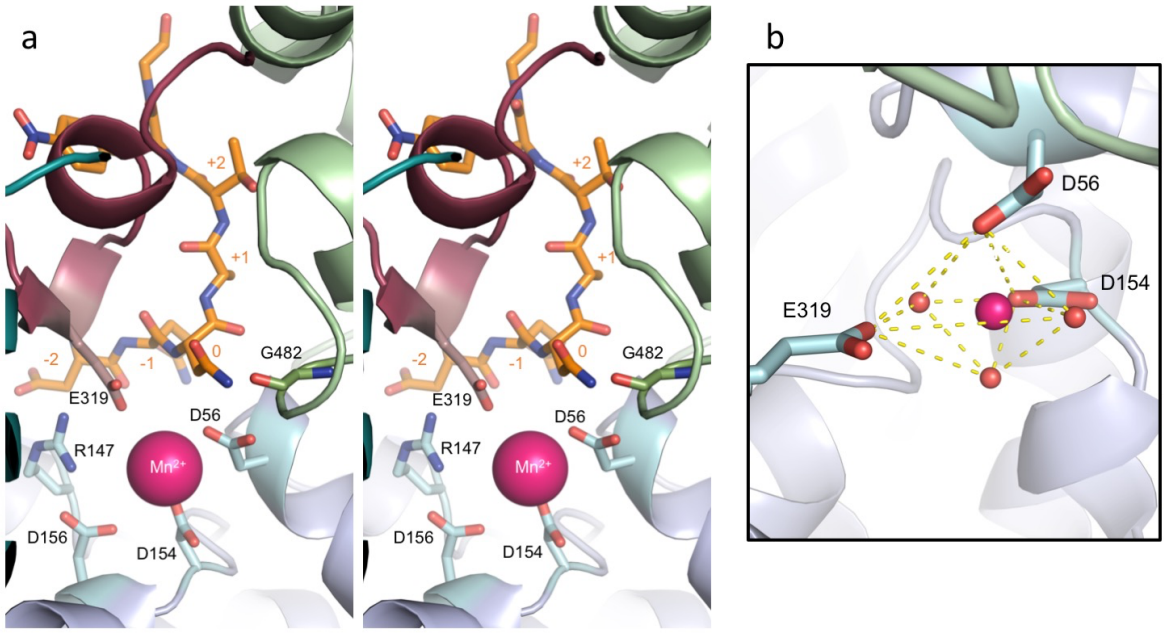


Figure 7. Catalytic site and acceptor asparagine binding. (**a**) Stereo view of the catalytic site and acceptor peptide binding. PglB is colored as in Fig. 4a, selected residues are shown as sticks and labeled. The acceptor peptide is shown as orange sticks, with residues numbered relative to their position in the acceptor sequon (Asn at position zero). The bound Mn^{2+} ion is shown as a pink sphere and labeled. (**b**) Schematic view of coordination geometry of Mn^{2+} ion, yellow dashed lines depict octahedral shape. The six ligands are the side-chain atoms of D56, D154, E319 and three water molecules, shown as red spheres.

Our structure reveals that a short loop containing the DGGK [^{481}DGGK] motif that is conserved in many bacteria, is present near the active site and interacts both with bound peptide and with the LLO molecule (Fig 4b, 4c). The backbone carbonyl of G482 forms a hydrogen bond with the amide nitrogen of the acceptor asparagine (2.9Å distance), but not with the GlcNAc moiety of the LLO. The closest contact between the DGGK motif and the GlcNAc moiety of the LLO is formed by two backbone atoms (N and Cα) of G483, with a distance of 3.8 Å, representing weak van der Waals interactions. The side chains of D481 and K484 point away from the

substrates and are engaged in salt bridges with residues from the periplasmic domain and therefore probably have structural roles. The DGGK loop had previously been investigated and found to be essential for the function of *C. jejuni* PglB[42,43]. Given the proximity of this loop to the reducing-end GlcNAc of the bound LLO analog, it is likely to impose steric constraints during the transfer reaction.

LLO binding site and interactions with PglB

The density for the synthetic (ωZZZ)-PPC-GlcNAc molecule was of excellent quality, allowing unambiguous building of its structure (Fig. 4c). The polyprenyl tail reaches approximately halfway across the membrane (Fig. 8a) and has largely defined conformation. It is tightly embedded in a hydrophobic groove of the transmembrane domain of PglB but also forms van der Waals contacts with hydrophobic residues on N-EL5 located ~4 Å away. Our finding that electron density of the entire nerylneryl tail was resolved suggests that at least 4 isoprenyl units contribute to the binding of wild type LLO to PglB.

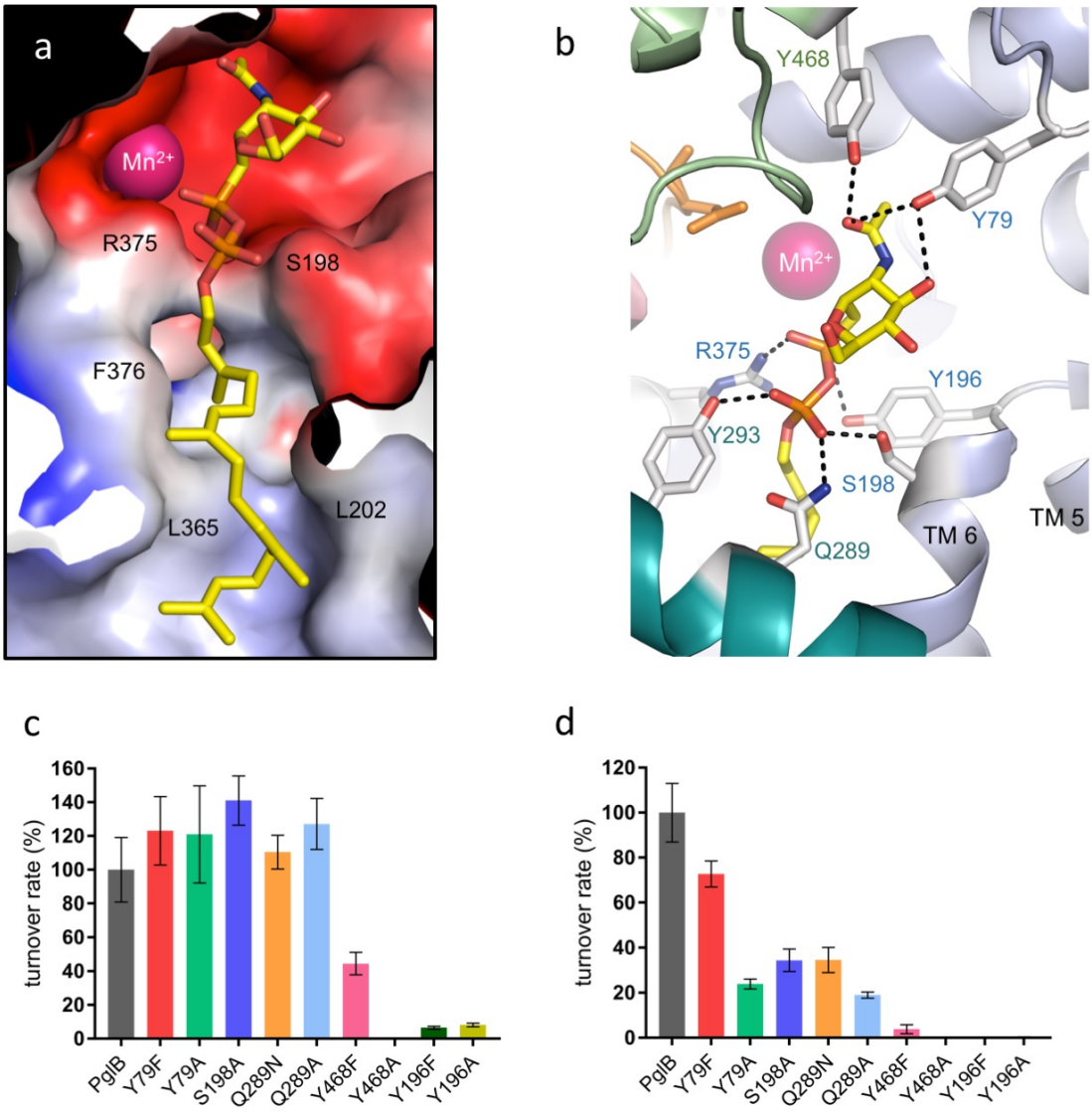


Figure 8. LLO binding site. (**a**) Electrostatic surface potential of LLO binding site of PglB, color-coded ranging from blue (most positive) to red (most negative). Residues interacting with LLO are indicated. (**b**) Close–up view of the LLO binding site, with PglB colored as in Fig. 2a. Acceptor peptide is shown as orange sticks, LLO as yellow sticks, and bound Mn^{2+} as a pink sphere. Residues interacting with bound LLO are shown as sticks and labeled. Black dashed lines indicate hydrogen bond or salt

bridge interactions. (**c**) Normalized initial turnover rates of selected PglB mutants using 50 μM wild type LLO as a donor substrate. (**d**) Normalized initial turnover rates of selected PglB mutants using 500 μM synthetic (ωZZZ)-PP-GlcNAc as a donor substrate. Data in (c) and (d) represents three independent cell cultures (error bars indicate s.d., n≥3).

The pyrophosphate and GlcNAc moieties are bound in the active site, near the Mn^{2+} ion. However, there is no direct contact between the pyrophosphate oxygens and the metal ion, as the shortest distance is ~5Å. Instead, the phosphate connected to GlcNAc forms a salt bridge with R375, a conserved residue whose essential role was previously reported[20]. The second phosphate group (connected to the polyprenyl tail) forms multiple hydrogen bonds, one of which with Y293, a residue from N-EL5 that was previously identified as essential. Intriguingly, the contacts between PglB and the pyrophosphate and GlcNAc moieties of the LLO are all of polar or electrostatic nature. The above interactions can explain why the previously identified R375 and Y293 residues are essential for glycosylation activity of PglB, but not for sequon binding. A mutation of R375 to alanine had earlier been reported to abolish turnover of PglB, without affecting acceptor peptide binding[20]. In contrast, a R375K mutant had revealed only a moderate reduction in activity, in line with the functional importance of the salt bridge provided by this residue. Based on the functional analysis of Y293 mutants, we had earlier proposed that it might form a stacking interaction with the reducing-end sugar[44]. Our structure disproves this interpretation, clearly showing a hydrogen bond with a pyrophosphate oxygen. While a Y293A mutation had been found to abolish PglB function, a Y293F mutation, despite being unable to form a hydrogen bond with the pyrophosphate moiety, retained some activity. This suggests that in addition to a direct contact with the LLO, the aromatic ring of Y293 has a structural role in stabilizing the conformation of EL5 when LLO is bound.
There are several additional residues interacting with the LLO. These include Y196, S198 and Q289, all of which form hydrogen bonds with the pyrophosphate moiety of LLO (Fig. 8b). In addition, the hydroxyl group of Y468 forms a hydrogen bond to the *N*-acetyl substituent of the reducing-end GlcNAc moiety. Y468 is part of the extended WWD motif and is strictly conserved in OSTs that accept HexNAc reducing-end sugars and had therefore been proposed to provide interactions with the *N*-acetyl group[17]. To analyze the relevance of the interactions observed in the structure, we mutated the involved residues and determined turnover rates of PglB both using full-length, wild type LLO and the synthetic (ωZZZ)-PP-GlcNAc analog (Fig. 8c and d). Mutations of Y468 and Y196 cause a strong reduction or complete loss of PglB activity, whereas mutations of Y79, S198 and Q289 did not affect the activity equally strongly. The activity of a Y468A mutant was reduced by 1000-fold using wild type LLO and 600-fold using synthetic LLO. In contrast, the mutation Y468F showed only 50-fold and 2.3-fold reduction in turnover rate when using synthetic or wild type LLO, respectively. This suggests that in addition to the hydrogen bond provided by the hydroxyl group, the aromatic ring of Y468 has a structural role in the active site. There was an intriguing difference in the impact of the mutations on PglB function with full-length, wild type LLO versus the shortened, synthetic analog. The rates observed with the synthetic LLO analog were generally more strongly reduced, indicating that additional interactions exist between PglB and the glycan or undecaprenol moieties in the wild type LLO.

Functional studies with disaccharide-containing LLO analogs
Our structure reveals that both the C3 and C4 atoms of the GlcNAc moiety are exposed to the solvent, and that non-reducing extensions with both α-1,3 and β-1,4 glycosidic linkages would be compatible without steric clashes. To further explore the impact of additional saccharides on LLO binding and specificity, we used synthetic farnesyl-PP-chitobiose (containing a GlcNAc-β-1,4-GlcNAc glycan)[27] and chemo-enzymatically generated farnesyl-PP-GlcNAc-α-1,3-

GalNAc, which we produced from synthetic farnesyl-PP-GlcNAc using purified PglA protein[30,31] (Fig. 9a). We found that both disaccharide LLOs could serve as glycan donors for PglB, and the resulting glycopeptides containing two saccharides could indeed be observed by tricine SDS-PAGE (Fig. 9b). Compared to farnesyl-PP-GlcNAc, the chemo-enzymatically extended farnesyl-PP-GlcNAc-α-1,3-GalNAc revealed a 2-fold increase in turnover and a reduction of K_M value by a factor of two (Table 1). Intriguingly, whereas the K_M of the chitobiose-containing analog was only twice that of farnesyl-PP-GlcNAc-α-1,3-GalNAc, the observed turnover rate was reduced ~300-fold (Fig. 9c and d, Table 1). This suggests that whereas the α-1,3-linked second saccharide unit contributes little to the initial binding of LLO, it is compatible with the subsequent transfer reaction, whereas the chitobiose-containing LLO does not allow PglB to reach the transition state effectively, which could be due to steric clashes or missing interactions with the enzyme.

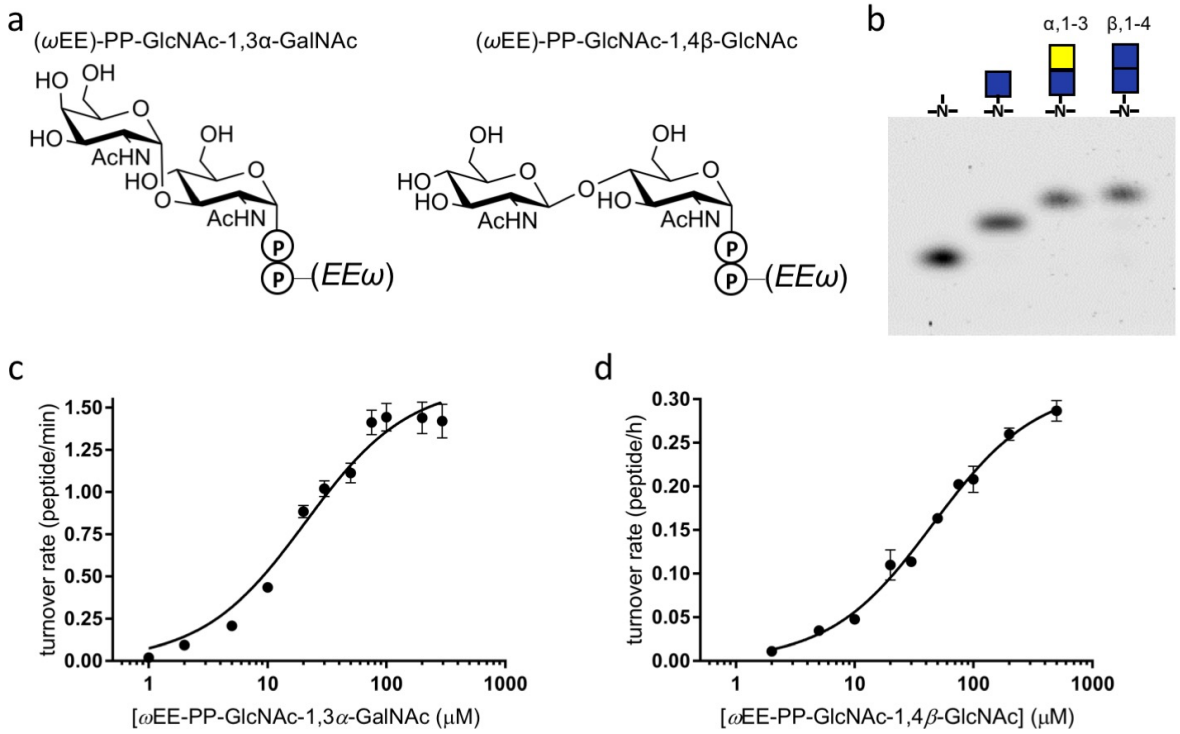


Figure 9. Effect of disaccharide linkage on PglB activity. (**a**) Structure of the synthetic or chemo-enzymatically generated disaccharide LLO analogs. (**b**) Tricine SDS-PAGE analysis of the *in vitro* glycosylation activity using fluorescently labeled peptide and different LLOs. The transferred glycans are shown schematically for each lane: Yellow squares, N-acetyl-galactosamine (GalNAc); blue squares, N-acetyl-glucosamine. (**c**) and (**d**) Kinetic analysis of PglB activity using LLO donors (ωEE)-PP-GlcNAc-1,3α-GalNAc and (ωEE)-PP-GlcNAc-1,4β-GlcNAc, respectively. 1 μM PglB and varying concentrations of disaccharide LLOs were used. The deduced K_M and k_{cat} values are indicated in Table1. Data in (c) and (d) represent technical triplicate (error bars indicate s.d., n≥3).

Immobilization of the External loop 5 by disulfide cross-linking

The structure reveals that in order to access the active site, the glycan of LLO diffusing to PglB in the lipid bilayer will have to "dive" under the external loop 5, which subsequently becomes ordered. We hypothesized that for efficient LLO binding, full disordering at least of the N-EL5 segment would have to occur. We therefore generated a mutant of PglB that contained two cysteins, one in N-EL5 (V294C) and another in the transmembrane domain of PglB (R371C). Upon addition of an oxidizing agent, a disulfide cross-link could be formed, tethering N-EL5 to the transmembrane domain and restricting the opening for a glycan to thread under EL5 (Fig.

10a). The cross-linking yield was determined to be 80 +/- 2.9% (Fig. 11). The *in vitro* activity of this PglB mutant was unaltered compared to the wild type protein both for full-length, wild type LLO and for the synthetic LLO analog nerylneryl-PP-GlcNAc (1.42+/-0.12 peptide/s and 2.8 +/-0.7 peptide/min, respectively, error denotes s.d. of three independent cell cultures, n=3). However, upon disulfide cross-linking, a difference was evident. Whereas the cross-link did not affect the activity with the monosaccharide-containing LLO analog, the turnover observed with full-length, wild type LLO was reduced 3.8-fold (Fig. 10b). This suggests that LLO binding, including glycan passage through the loop EL5, was at least partially rate-limiting.

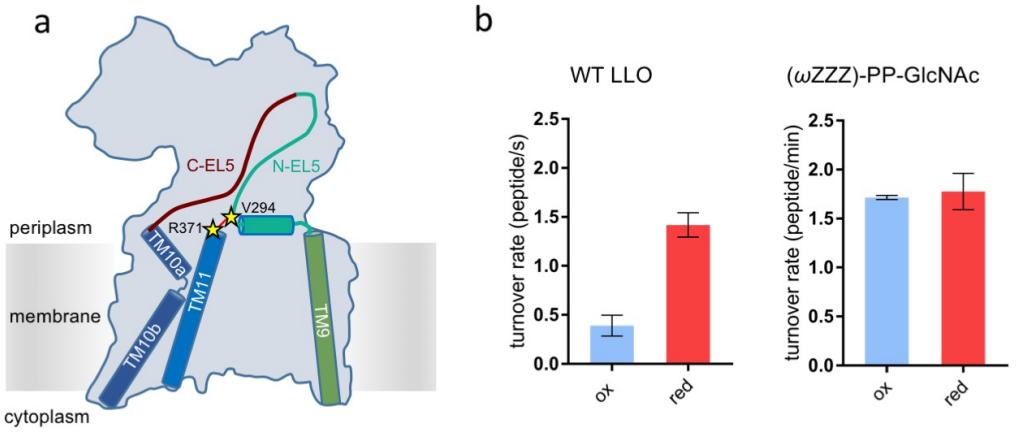


Figure 10. Disulfide cross-linking of EL5 to the TM domain of PglB. (**a**) Schematic representation of PglB (grey surface) showing the localization of the engineered disulfide cross-link. EL5 is depicted as a dark purple (C-EL5) and turquoise (N-EL5) line and labeled. Helices are shown as cylinders, TM helices are labeled. Residues V294 (EL5) and R371 (TM11) were mutated to cysteines and are depicted as yellow asterisks. (**b**) Glycosylation activity of the PglB V194C/R371C double mutant using 50 μM wild type LLO (left panel) and 500 μM synthetic (ωZZZ)-PP-GlcNAc (right panel). Bars labeled "ox" (oxidized) depict turnover rates of disulfide cross-linked samples, whereas bars labeled "red" (reduced) depict turnover rates of samples containing 10 mM DTT, breaking the disulfide bond. Turnover rates of the disulfide cross-linked samples were corrected taking cross-linking efficiency into account. Data in (b) represent 3 independent cell cultures (error bars indicate s.d., n=3).

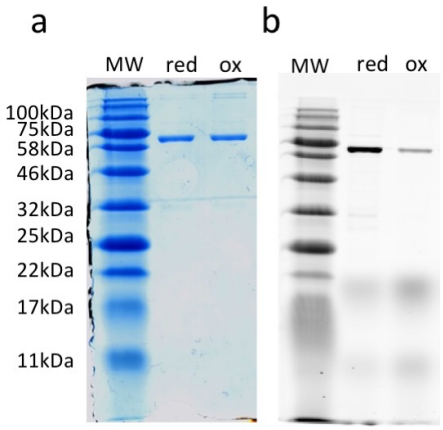


Figure 11. Disulfide cross-linking efficiency of the cysteine double mutant. The ratio of disulfide cross-linking was determined by labeling unreacted, free cysteines with fluorescein maleimide and quantitation of in-gel fluorescence. (**a**) Coomassie-stained gel; (**b**) fluorescence-scanned gel. The oxidative disulfide bond formation was determined to be 80 +/- 2.9% from three independent cell

cultures (error denotes s.d., n=3). Lanes labeled "ox" (oxidized) indicates a sample that was cross-linked with $CuCl_2$ during the purification process, whereas "red" (reduced) indicates a control sample that contained 10 mM β-mercaptoethanol during purification, which was removed by desalting. MW denotes marker proteins, with masses indicated on the side.

Discussion

Our results provide the first structural insight into a STT3 protein (and to the best of our knowledge any GT-C family member) with both the acceptor peptide and the lipid-linked donor substrates bound. We have captured PglB in an intermediate state, prior to glycan transfer from the LLO to the peptide. To reach its observed binding site, LLO has to be pulled out of the membrane. PglB facilitates this step by providing a hydrophobic groove that binds the polyprenyl chain and by providing several specific contacts to the hydrophilic moiety of the LLO, including a salt bridge from R375 to the pyrophosphate group.

The distance between the atoms that would ultimately form a *N*-glycosidic bond (carboxamide nitrogen of the acceptor Asn and C1 atom of the reducing-end sugar) is ~6Å in our structure, a distance that is too long for the reaction to proceed. It is possible that the use of a phosphonate LLO analog for our structural studies might have contributed to the observed distance to the acceptor peptide, because phosphonates have a slightly different geometry than pyrophosphates (longer P–C than P–O bonds and slightly different bond angle), and the methylene group cannot form a strong contact with the Mn^{2+} ion. However, these differences are small and cannot explain the gap that needs to be bridged to reach the transition state of the reaction. This requires a conformational change, during which the LLO most likely will have to shift towards the peptide, because the peptide is more firmly embedded in its binding site. LLO shifting would involve a repositioning of the pyrophosphate moiety, whereby the phosphate group currently in contact with the conserved R375 would replace water molecules and interact with the Mn^{2+} ion, in agreement with the essential role of the metal ion for reactivity[12,13,20,42,45]. The second phosphate would then take the place of the first in forming a salt bridge with R375. Importantly, in our structure, steric clashes would prevent such a repositioning. In particular, the loop containing the DGGK motif blocks a further approaching of the LLO to the acceptor asparagine. The importance of this motif for PglB function was previously recognized: In *C. jejuni* PglB, mutations of the aspartate side chain of the DGGK motif to alanine or glutamate significantly decreased glycosylation activity, although the effect was partially reversed when higher LLO concentrations were used[42]. Molecular dynamics simulations implicated the motif in correctly orienting the carboxamide group of the acceptor Asn residue[46]. However, beyond the requirement for a conformational change to allow the acceptor Asn and the LLO to approach, the exact role of the motif cannot be defined at present. Our results also cannot clarify the exact mechanism of asparagine carboxamide group activation (the twisted amide hypothesis) further. Nevertheless, our results provide a promising starting point for computational approaches aimed at exploring the molecular events of the transfer step.

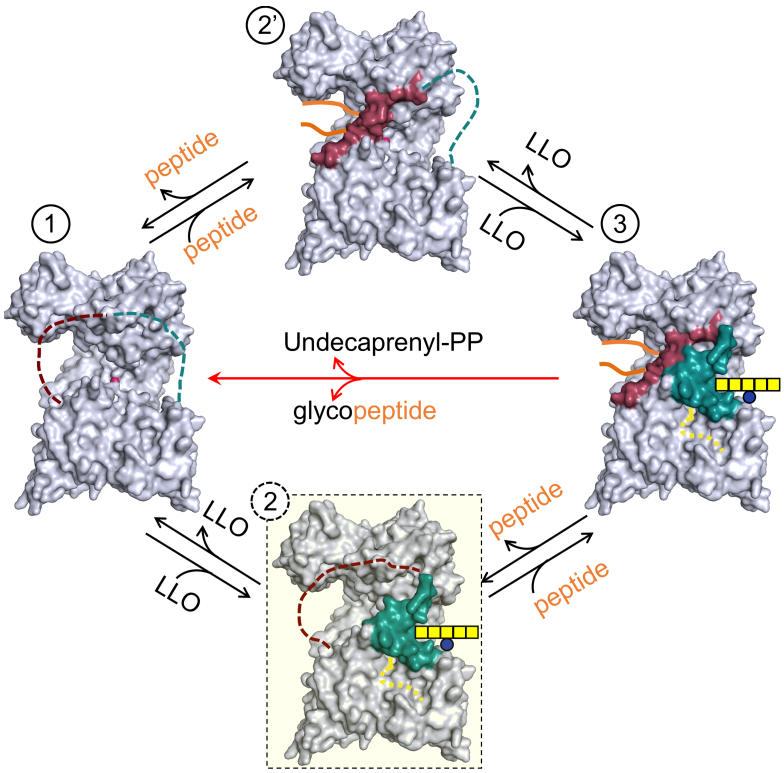


Figure 12. Glycosylation mechanism. PglB is shown as a grey-colored molecular surface, with N-EL5 and C-EL5 colored turquoise and purple, respectively. States are indicated with circled numbers. In state 1 no substrates are bound and EL5 is disordered, as indicated by a dashed line. Depending on whether LLO or peptide bind first, states 2 or 2' are reached. Bound LLO is represented by a yellow line for isoprenoid moieties, yellow squares for *N*-acetyl-galactosamine (GalNAc), and blue circles for glucose (Glc). The pyrophosphate moiety is hidden behind N-EL5. State 2 is shown in a yellow-shaded box surrounded by dashed lines to indicate that no structure exists of this state. The ordering of N-EL5 in state 2 is putative. The observed crystal structure of the ternary complex reflects state 3, with acceptor peptide and LLO bound and the entire EL5 ordered.

Our results also define the role of EL5, a loop that is present in all STT3 proteins and that has a dual function in binding peptide (C-EL5) and LLO (N-EL5). Because enzymatic cleavage of EL5 has been observed to result in only a slight reduction in activity[44], the two functions of EL5 are independent of each other and thus modular. In contrast to recent molecular dynamics simulations that proposed that EL5 becomes disordered upon LLO binding[47], our structural data demonstrate that the opposite is correct. Our disulfide cross-linking experiment also demonstrates that the larger the glycan moiety, the slower the diffusion of the donor LLO to the active site given that the glycan must dive or thread under / into EL5. This has consequences for the mechanism of PglB, as outlined in Fig. 12. In the absence of substrates (state 1), EL5 is flexible, disordered, and disengaged from the OST core. This matches one of the apo-AglB structures[18]. The binding of the substrates can in principle occur in either order. If peptide binds first (state 2'), C-EL5 becomes ordered, as demonstrated by the peptide-only structure of PglB[17]. If LLO binds first (state 2), N-EL5 might be engaged, but no structure is currently available of this state. Given that the eukaryotic LLO has a larger, more branched glycan than the bacterial counterpart, but eukaryotic OST does not have a longer EL5, it is tempting to speculate that LLO binding to OST before peptide could be a more efficient route, because both N-EL5 and C-EL5 would be disengaged and disordered, allowing more efficient access of the large glycan moiety to the active site.

Not only does the ternary complex structure of PglB provide a mechanistic understanding of

bacterial OSTs, but also gives insight into structural changes that may be required for the function of all OSTs, as the glycosylation mechanism is thought to be broadly conserved in all domains of life. It also provides opportunities for engineering the substrate specificity with respect to the LLO. This may be applied in the development and production of glycoconjugate therapeutics by inducing relaxed substrate specificity of PglB in order to synthesize various new bacterial and eukaryotic peptidoglycans to use as novel vaccines.

References

1. Schwarz, F. & Aebi, M. Mechanisms and principles of N-linked protein glycosylation. *Curr. Opin. Struct. Biol.* **21,** 576–582 (2011).
2. Nothaft, H. & Szymanski, C. M. Protein glycosylation in bacteria: sweeter than ever. *Nat. Rev. Microbiol.* **8,** 765–778 (2010).
3. Calo, D., Kaminski, L. & Eichler, J. Protein glycosylation in Archaea: Sweet and extreme. *Glycobiology* **20,** 1065–1076 (2010).
4. 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.* **32,** 1022–1030 (1999).
5. Helenius, A. & Aebi, M. Roles of N-linked glycans in the endoplasmic reticulum. *Annu. Rev. Biochem.* **73,** 1019–49 (2004).
6. Kornfeld, R. & Kornfeld, S. Assembly of Asparagine-Linked Oligosaccharides. *Annu. Rev. Biochem.* **54,** 631–664 (1985).
7. Cherepanova, N., Shrimal, S. & Gilmore, R. N-linked glycosylation and homeostasis of the endoplasmic reticulum. *Curr. Opin. Cell Biol.* **41,** 57–65 (2016).
8. Nothaft, H. & Szymanski, C. M. Bacterial protein n-glycosylation: New perspectives and applications. *J. Biol. Chem.* **288,** 6912–6920 (2013).
9. Valguarnera, E., Kinsella, R. L. & Feldman, M. F. Sugar and Spice Make Bacteria Not Nice: Protein Glycosylation and Its Influence in Pathogenesis. *J. Mol. Biol.* **428,** 3206–3220 (2016).
10. Wacker, M. *et al.* Substrate specificity of bacterial oligosaccharyltransferase suggests a common transfer mechanism for the bacterial and eukaryotic systems. *Proc. Natl. Acad. Sci. U. S. A.* **103,** 7088–93 (2006).
11. Marshall, R. D. Glycoproteins. *Annu. Rev. Biochem.* **41,** 673–702 (1972).
12. Lairson, L. L., Henrissat, B., Davies, G. J. & Withers, S. G. Glycosyltransferases: Structures, Functions, and Mechanisms. *Annu. Rev. Biochem.* **77,** 521–555 (2008).
13. Liu, J. & Mushegian, A. Three monophyletic superfamilies account for the majority of the known glycosyltransferases. *Protein Sci.* **12,** 1418–31 (2003).
14. Kelleher, D. J. & Gilmore, R. An evolving view of the eukaryotic oligosaccharyltransferase. *Glycobiology* **16,** 47R–62R (2005).
15. Yan, Q. & Lennarz, W. J. Studies on the function of oligosaccharyl transferase subunits: a glycosylatable photoprobe binds to the luminal domain of Ost1p. *Proc. Natl. Acad. Sci. U. S. A.* **99,** 15994–9 (2002).
16. Nasab, F. P., Schulz, B. L., Gamarro, F., Parodi, A. J. & Aebi, M. All in one: Leishmania major STT3 proteins substitute for the whole oligosaccharyltransferase complex in Saccharomyces cerevisiae. *Mol. Biol. Cell* **19,** 3758–68 (2008).
17. Lizak, C., Gerber, S., Numao, S., Aebi, M. & Locher, K. P. X-ray structure of a bacterial oligosaccharyltransferase. *Nature* **474,** 350–355 (2011).
18. Matsumoto, S. *et al.* Crystal structures of an archaeal oligosaccharyltransferase provide insights into the catalytic cycle of N-linked protein glycosylation. *Proc. Natl. Acad.*

Sci. U. S. A. **110,** 17868–73 (2013).
19. Matsumoto, S., Taguchi, Y., Shimada, A., Igura, M. & Kohda, D. Tethering an N-Glycosylation Sequon-Containing Peptide Creates a Catalytically Competent Oligosaccharyltransferase Complex. *Biochemistry* **56,** 602–611 (2017).
20. Gerber, S. *et al.* Mechanism of bacterial oligosaccharyltransferase: In vitro quantification of sequon binding and catalysis. *J. Biol. Chem.* **288,** 8849–8861 (2013).
21. Kabsch, W. Xds. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66,** 125–132 (2010).
22. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Cryst* **60,** 2126–2132 (2004).
23. Adams, P. D. *et al.* PHENIX: building new software for automated crystallographic structure determination. *Res. Pap. PHENIX Acta Cryst* (1948).
24. W. L.Delano. The PyMOL Molecular Graphics System. (2002).
25. Kowarik, M. *et al.* N-linked glycosylation of folded proteins by the bacterial oligosaccharyltransferase. *Science* **314,** 1148–50 (2006).
26. Liu, F. *et al.* Rationally Designed Short Polyisoprenol-Linked PglB Substrates for Engineered Polypeptide and Protein N-Glycosylation. *J. Am. Chem. Soc.* **136,** 566–569 (2014).
27. Ramírez, A. S. *et al.* Characterization of the single-subunit oligosaccharyltransferase STT3A from Trypanosoma brucei using synthetic peptides and lipid-linked oligosaccharide analogs. *Glycobiology* **16,** 1–11 (2017).
28. Perez, C. *et al.* Structure and mechanism of an active lipid-linked oligosaccharide flippase. *Nature* **524,** 433–438 (2015).
29. Hajduch, J. *et al.* A convenient synthesis of the C-1-phosphonate analogue of UDP-GlcNAc and its evaluation as an inhibitor of O-linked GlcNAc transferase (OGT). *Carbohydr. Res.* **343,** 189–95 (2008).
30. Weerapana, E., Glover, K. J., Chen, M. M. & Imperiali, B. Investigating bacterial N-linked glycosylation: synthesis and glycosyl acceptor activity of the undecaprenyl pyrophosphate-linked bacillosamine. *J. Am. Chem. Soc.* **127,** 13766–7 (2005).
31. 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.* **102,** 14255–9 (2005).
32. Lizak, C. *et al.* Unexpected reactivity and mechanism of carboxamide activation in bacterial N-linked protein glycosylation. *Nat. Commun.* **4,** 2627 (2013).
33. Bause, E., Breuer, W. & Peters, S. Investigation of the active site of oligosaccharyltransferase from pig liver using synthetic tripeptides as tools. *Biochem. J* **312,** 979–985 (1995).
34. Imperiali, B., Shannon, K. L. & Rickert, K. W. Role of peptide conformation in asparagine-linked glycosylation. *J. Am. Chem. Soc.* **114,** 7942–7944 (1992).
35. Imperiali, B. & Tai, V. W. Chemistry and Biochemistry of Asparagine-Linked Protein Glycosylation. 281–303 (2003).
36. Musumeci, M. A. *et al.* In vitro activity of Neisseria meningitidis PglL O-oligosaccharyltransferase with diverse synthetic lipid donors and a UDP-activated sugar. *J. Biol. Chem.* **288,** 10578–87 (2013).
37. Compain, P. & Martin, O. R. Carbohydrate mimetics-based glycosyltransferase inhibitors. *Bioorg. Med. Chem.* **9,** 3077–3092 (2001).
38. Knapp, S. & Myers, D. S. α-GlcNAc Thioconjugates. *Org. Chem.* **66**, 3636–3638 (2001).
39. Knapp, S. & Ajayi, K. The anomeric Pudovik rearrangement. *Tetrahedron Lett.* **48,** 1945–1949 (2007).
40. Knapp, S., Gonzalez, S., Myers, D.S., Eckman, L.L. & Bewley, C.A. Shortcut to

mycothiol analogues. *Org. Lett.* **4**, 4337–4339 (2002).
41. Engel, R. Phosphonates as Analogues of Natural Phosphates. *Chem. Rev.* 77, 349–367 (1977).
42. Jaffee, M. B. & Imperiali, B. Exploiting topological constraints to reveal buried sequence motifs in the membrane-bound N-linked oligosaccharyl transferases. *Biochemistry* **50,** 7557–67 (2011).
43. Ihssen, J. *et al.* Structural insights from random mutagenesis of Campylobacter jejuni oligosaccharyltransferase PglB. *BMC Biotechnol.* **12,** (2012).
44. Lizak, C. *et al.* A catalytically essential motif in external loop 5 of the bacterial oligosaccharyltransferase PglB. *J. Biol. Chem.* **289,** 735–746 (2014).
45. Igura, M. *et al.* Structure-guided identification of a new catalytic motif of oligosaccharyltransferase. *EMBO J.* **27,** 234–43 (2008).
46. Pedebos, C., Arantes, P. R., Giesel, G. M. & Verli, H. *In silico* Investigation of the PglB Active Site Reveals Transient Catalytic States and Octahedral Metal Ion Coordination. *Glycobiology* **25,** 1183–1195 (2015).
47. Sun Lee, H. & Im, W. Transmembrane Motions of PglB Induced by LLO are Coupled with EL5 Loop Conformational Changes Necessary for OST Activity. *Glycobiology* (2017). doi:10.1093/glycob/cwx052

Chapter 5: Structure of bacterial oligosaccharyltransferase PglB bound to a reactive LLO and an inhibitory peptide

This study was submitted to *Scientific Reports*.
Authors: Maja Napiórkowska M[1], Jérémy Boilevin[2], Tamis Darbre[2], Jean-Louis Reymond[2] & Kaspar P. Locher[1]

[1]Institute of Molecular Biology and Biophysics, ETH Zurich, Zurich, Switzerland.
[2]Department of Chemistry and Biochemistry, University of Bern, Bern, Switzerland

Contributions:
I performed all experiments, collected X-ray data, determined the structure and built and refined the model under the supervision of K.P.L. J.B. performed chemical synthesis of the LLO. T.D. and J.-L.R. supervised the chemical synthesis of the LLO analog. I and K.P.L. devised experiments and wrote the manuscript with contribution from all authors.

Abstract

Oligosaccharyltransferase (OST) is a key enzyme of the *N*-glycosylation pathway, where it catalyzes the transfer of a glycan from a lipid-linked oligosaccharide (LLO) to an acceptor asparagine within the conserved sequon N-X-T/S. A previous structure of a ternary complex of bacterial single subunit OST, PglB, bound to a non-hydrolyzable LLO analog and a wild type acceptor peptide showed how both substrates bind and how an external loop (EL5) of the enzyme provided specific substrate-binding contacts. However, there was a relatively large separation of the substrates at the active site. Here we present the X-ray structure of PglB bound to a reactive LLO analog and an inhibitory peptide, revealing previously unobserved interactions in the active site. We found that the atoms forming the *N*-glycosidic bond (C-1 of the GlcNAc moiety of LLO and the $-NH_2$ group of the peptide) are closer than in the previous structure, suggesting that we have captured a conformation closer to the transition state of the reaction. We find that the distance between the divalent metal ion and the glycosidic oxygen of LLO is now 4Å, suggesting that the metal stabilizes the leaving group of the nucleophilic substitution reaction. Further, the carboxylate group of a conserved aspartate of PglB mediates an interaction network between the reducing-end sugar of the LLO, the asparagine side chain of the acceptor peptide, and a bound divalent metal ion. The interactions identified in this novel state are likely to be relevant in the catalytic mechanisms of all OSTs.

Introduction

N-protein glycosylation is a post-translational modification that is conserved in all domains of life[1–7]. The addition of a glycan to a protein can profoundly affect its structure, function, and targeting. In eukaryotes *N*-glycosylation can be important in protein trafficking, cellular signalling or self-non-self interactions[1,8,9]. In bacteria, including the food pathogen *Campylobacter jejuni*, *N*-glycans improve host cell- adhesion and colonization, indicating their role in virulence[5,10].
Although bacteria and eukaryotes display different glycan structures attached to asparagine residues, their *N*-glycosylation pathways are thought to be homologous and follow a conserved

mechanism[11]. The oligosaccharide is first assembled on a pyrophosphate-lipid carrier and then transferred *en bloc* to conserved sequons (N-X-S/T) of secretory proteins. The reaction is catalyzed by an integral-membrane oligosaccharyltransferase (OST). Whereas the OST of higher eukaryotes is a multiprotein complex with STT3 being the catalytic subunit, archaea, kinetoplastids and bacteria have single-subunit OST (ssOST) enzymes that are homologous to STT3[4,12–14]. Although OSTs have been found to show some relaxed specificity with respect to the oligosaccharide donor, bacterial and archaeal enzymes can transfer a much vaster array of glycans compared to their eukaryotic counterparts[3,12,15–18].

X-ray structures of ssOSTs of *Campylobacter lari* (PglB) and *Archaeoglobus fulgidus* (AglB) have revealed the fold of the enzyme[13,14,19,20]. Extensive *in vitro* studies and a high resolution structure of PglB bound to a wild type peptide and a non-hydrolyzable LLO provided many mechanistic details of peptide and LLO recognition[15,16,21–23]. In particular, it was found that the external loop 5 (EL5), present in all OST enzymes, provided crucial interactions to bound substrates: Whereas the N-terminal half primarily contacted bound LLO, the C-terminal half provided key contacts to bound acceptor peptide. Engagement and disengagement of EL5 appeared essential for substrate binding and product release[19,23].

Despite the available structural data, the exact mechanism of glycan transfer is not fully understood. For example, details of how the substrates interact and how the amido group of the acceptor asparagine is activated, are lacking. Specifically, the previously solved ternary complex structure revealed an arrangement of bound substrates that was evidently not very close to the transition state given that the distance between the atoms that would form the glycosidic bond was ~6Å. Here we present the X-ray structure of PglB in complex with an inhibitory peptide and a synthetic reactive LLO analog. The structure captures a new intermediate of the glycosylation reaction where the substrates are significantly closer, identifying new contacts with conserved residues in the catalytic site. The new structural evidence reveals previously unknown interactions between PglB, LLO, acceptor peptide and the bound divalent metal ion.

Experimental procedures

Overexpression and purification of PglB

Overexpression and purification of PglB was performed as previously described[19]. Briefly, PglB was overexpressed in *Escherichia coli* BL21-Gold cells (DE3) (Strategene) at 37 °C in five-liter flasks using Terrific Broth medium supplemented with 1 % glycerol (w/v). The cells were induced at 37 °C for 4 h at A600 of 3.0 by adding 0.1 % arabinose (w/v). All following steps were carried out at 4 °C. Cells were harvested by centrifugation. Cells were resuspended in 25 mM Tris-HCl, pH 8.0, 250 mM NaCl and disrupted in a M-110-l microfluidizer (Microfluidics) at 15,000 p.s.i. chamber pressure. Membranes were pelleted by centrifugation at 100,000 g for 30 min and solubilized in 25 mM Tris-HCl, pH 8.0, 250 mM NaCl, 10 % glycerol (v/v) and 1 % N-dodecyl-β-D-maltopyranoside (w/v) (DDM, Anatrace) for 1.5 h. All subsequent purification buffers contained 0.016 % DDM. PglB was purified on a nickel-nitrilotriacetic acid (Ni-NTA) Superflow affinity column (Qiagen) and desalted into 10 mM MES, pH 6.5, 100 mM NaCl, 3 % glycerol (v/v), and 0.016 % DDM. For crystallization experiments, desalted protein was concentrated to 6 mg mL^{-1} in an Amicon Ultra-15 concentrator (Milipore) with a molecular mass cutoff of 100 kDa and purified by size exclusion chromatography (Superdex 200 10/300, GE Healthcare) in desalting buffer. Collected fractions were pooled and concentrated to 12 mg mL^{-1}.

Protein crystallization

Ac-DQ(Dab)ATF(p-NO_2)-NH2, nerylneryl-PP-GlcNAc and $MnCl_2$ were added to the concentrated protein to the final concentration of 0.75 mM, 1.5 mM, 2 mM, respectively and incubated for 15 min at 4 °C. PglB was crystallized by vapor diffusion in hanging drops at 20 °C against a reservoir containing 100 mM glycine, pH 9.4, 50-200 mM magnesium acetate, and 27–31 % PEG400 at the protein-to-reservoir ratio 2:1. Crystals usually appeared after 10–14 days and grew to full size within 3–4 weeks. Crystals were cryoprotected by stepwise addition of cryoprotectant PEG400 up to 30 % final concentration and directly flash frozen in liquid nitrogen before data collection.

Data collection and structural determination

The X-ray data was collected at the microfocus X06SA beamline at the Swiss Light Source (Villigen). The wavelength for data collection was 1.000 Å. All data were processed using XDS package[24]. Ellipsoidal truncation and anisotropic correction were applied to the data using the diffraction anisotropy server[25]. The ternary complex crystallized in the $P2_12_12_1$ space group, containing one PglB molecule per asymmetric unit. The structure was solved by molecular replacement using PDB 5OGL as a model. The iterative model building and refinement was performed in Coot[26] and Phenix[27]. Final refinement statistics and data collection are summarized in Table 1. Structural figures were drawn using PyMOL[28].

Quantification assay

The LLO and PglB were incubated in the presence or absence of Dab peptide for 48 h at 20 °C. The mixture conditions replicated the crystallization conditions, where the 10 mg mL^{-1} PglB, 1.5 mM LLO (nerylneryl-PP-GlcNAc), 0.75 mM inhibitory peptide (Ac-DQ(Dab)ATF{p-NO2}-NH2) in 10 mM MES pH 6.5, 3 % glycerol, 100 mM NaCl, 0.016 % DDM was added to 100 mM glycine pH 9.4, 28% PEG 400, 200 mM magnesium acetate in 2:1 ratio. The samples were diluted and the LLO was quantified against fluorescently labeled wild type peptide (DQNATF). The LLO dilutions were incubated for 12 h at 25 °C with 10 μM labeled peptide and purified protein in reaction buffer containing 10 mM MES pH 6.5, 100 mM NaCl, 3 % glycerol 0.016 % DDM, 10 mM $MnCl_2$. The samples were analyzed by in-gel fluorescence and data was fitted by linear regression in GraphPad PRISM 7.0.

Synthesis of nerylneryl-PP-GlcNAc ((ωZZZ)-PP-GlcNAc

The nerylneryl-PP-GlcNAc ((ωZZZ)-PP-GlcNAc) analog was synthesized according to previously reported procedures[19,29]. The synthetic strategy included three steps: the synthesis of GlcNAc α-phosphate, preparation of the nerylneryl-phosphate precursor, and finally the coupling of both monophosphates to the desired final pyrophosphate. Nerylneryl-PP-GlcNAc was shown to be pure by ^{1}H, ^{13}C, ^{31}P NMR and ESI-HRMS.

Results

Strategy to trap a distinct ternary complex

Two strategies can in principle be pursued to trap ternary complex states of OST (Fig. 1). The first, which was the basis of the previously published structure[19], was to co-crystallize PglB with wild type peptide (DQNAT(pNO_2-F)) and a non-hydrolyzable, phosphonate-containing LLO analog (ωZZZ-PPCH_2-GlcNAc). DQNATF is an optimal acceptor sequon for *C. jejuni* PglB and the phosphonate-containing LLO analog was identified as a competitive inhibitor (Fig. 1)[14,19,30]. In the resulting ternary complex structure we found that

the phosphonopyrophosphate group was not coordinated by the catalytic Mn^{2+} ion and that the substrates were distantly located (~6 Å apart). We hypothesized that this gap was due to the use of a phosphonopyrophosphate rather than a pyrophosphate group and that using a reactive LLO might reveal a bound state closer to the transition state of the reaction. We therefore pursued a second strategy to form a ternary complex by incubating PglB with a functionally competent, synthetic LLO analog and an inhibitory peptide (Fig.1). We chose 2,4-diaminobutanoic acid (Dab) to replace the acceptor asparagine, because Dab-containing peptides were previously shown to act as competitive inhibitors of bacterial and eukaryotic OSTs at physiological pH (Fig. 1a)[21,31,32]. We investigated whether the Dab-containing peptide and the functional LLOs showed activity with PglB at pH values used to crystallize it (pH ~8.9) and did not detect any glycosylation of the Dab-containing peptide, demonstrating that no glycan transfer occurs even after very long incubations. We therefore used this Dab-containing peptide for crystallization experiments. We used the reactive synthetic LLO analog ωZZZ-PP-GlcNAc, which is chemically very similar to the wild type LLO except that it contains a GlcNAc rather than di-*N*-acetyl-bacillosamine as a reducing-end sugar (Fig. 1)[19]. This LLO compound exhibited the highest turnover rate of all of the water-soluble synthetic LLO analogs tested[19,29]. The resulting ternary complex was crystallized and the structure determined at 3.4 Å resolution (Fig. 2 and Table 1).

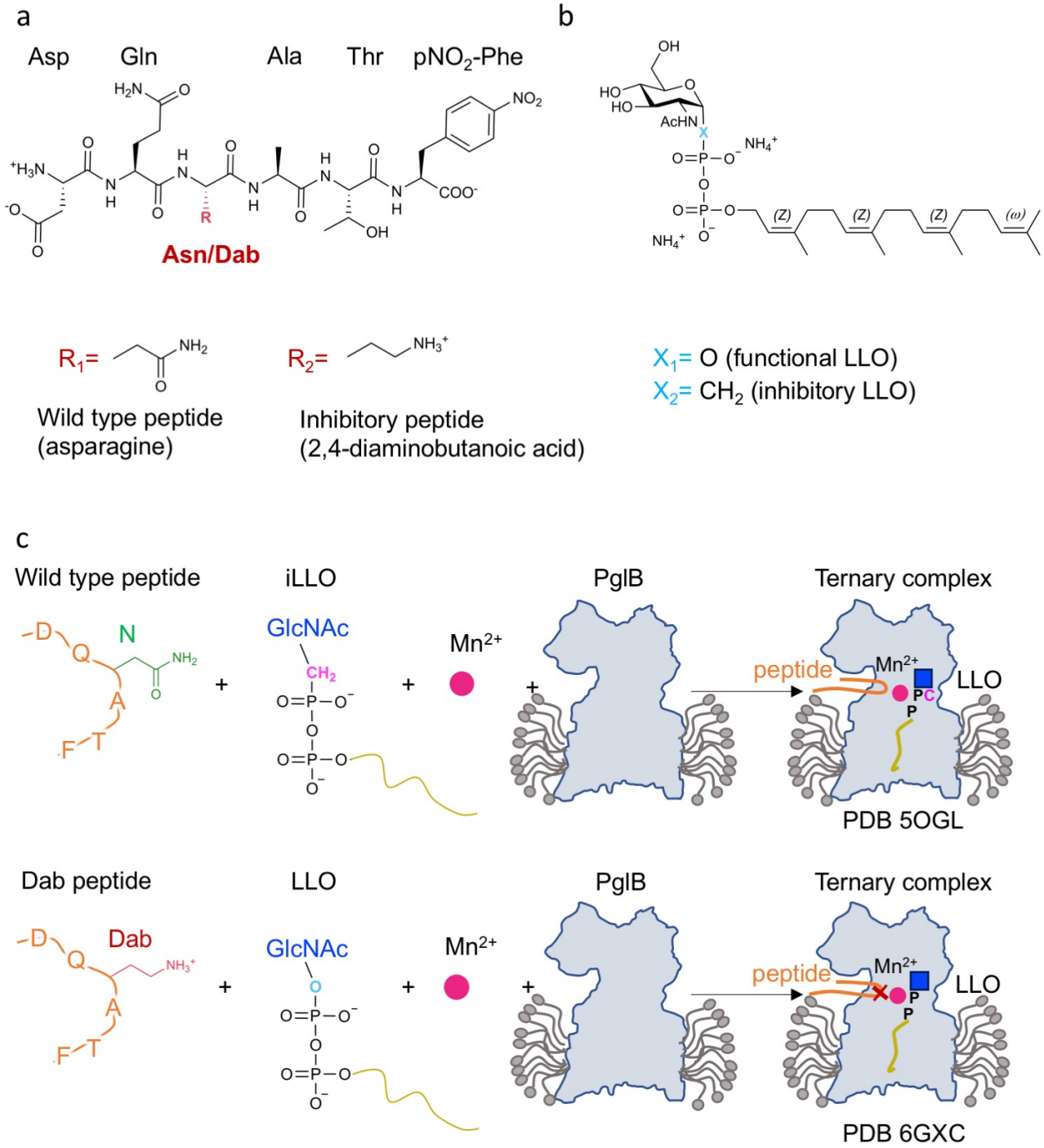


Figure 1. Strategies to trap distinct intermediate states of the PglB-catalyzed glycosylation reaction. (**a**) Structure of the substrate and inhibitory peptides used in crystallization experiments. The red 'R' denotes a side chain of either asparagine or 2,4-diaminobutanoic acid (Dab). (**b**) Structure of the synthetic LLO analogs used in crystallization experiments. The blue 'X' denotes the linkage between C-1 of the GlcNAc moiety and the phosphorous atom of the first phosphate group. (**c**) Schematic of two strategies pursued to trap ternary complexes of PglB using different peptide and LLO analogs. Top: method used previously to co-crystallize PglB with a wild type peptide and a non-hydrolyzable LLO analog (iLLO)[19]. Bottom: method used in this study to co-crystallize PglB with an inhibitory, Dab-containing peptide and a reactive LLO analog. The outline of *C. lari* PglB in a detergent micelle is shown in grey, peptide in orange and the LLO analogs are represented by a blue square for the *N*-acetyl-glucosamine (GlcNAc) moiety, a black 'P' for the phosphate group and a gold line for the lipid tail. The non-hydrolyzable LLO analog and the inhibitory peptide are marked with a pink 'C' and a red 'X', respectively.

	Ternary complex[1]
Data collection	
Space group	$P2_12_12_1$
Cell dimensions	
a, *b*, *c* (Å)	83.81, 116.54, 173.89
α, β, γ (°)	90.0, 90.0, 90.0
Resolution (Å)	40.74 - 3.4 (3.49 - 3.4)*
R_{meas}	0.069 (1.44)*
I / σ	23.25 (1.95)*
$CC_{1/2}$	0.999 (0.676)*
Completeness (%)	92.1 (36.7)*
Redundancy	12.4 (4.8)*
Refinement	
Resolution (Å)	40.74 - 3.4
No. reflections	24002 (2377)*
R_{work} / R_{free}	0.2492/0.2831
No. non hydrogen atoms	5931
Protein	5866
Ligands	65
B-factors (Å^2)	
Protein	110.17
Ligands	103.73
R.m.s. deviations	
Bond lengths (Å)	0.008
Bond angles (°)	1.13

Table 1. Data collection and refinement statistics (molecular replacement).
Values shown correspond to statistics after anisotropy correction.
[1] data collected from one crystal
*Values in parentheses are for highest-resolution shell.

LLO binding and the role of Mn^{2+}

Whereas the overall structure of PglB bound to the reactive LLO and inhibitory peptide is very similar to the previously determined ternary complex, there are important differences (Fig. 2 and Fig. 3)[19]. The key difference is that the two substrates are closer in the new ternary complex, with a distance of ~3.4 Å between carbon C-1 of the GlcNAc moiety of reactive LLO and the $–NH_2$ group of the Dab residue, representative of the atoms that would form an *N*-glycosidic bond if Dab was replaced by Asn. In contrast, the equivalent distance between the GlcNAc of non-reactive LLO and the acceptor asparagine was ~6 Å in the previously determined structure[19]. The closing of the distance is mainly due to a shift of the reducing-end GlcNAc and is probably due to new contacts formed. Whereas the lipid tail of the LLO analog has an indistinguishable conformation to the previous ternary complex, the pyrophosphate and GlcNAc moieties have shifted and interact differently with the surface of PglB in the new structure (Fig. 4). We now observe continuous electron density between the pyrophosphate group of LLO and the catalytic Mn^{2+}, with the shortest distance (~4 Å) observed between Mn^{2+} and the oxygen linking C1 of GlcNAc and the pyrophosphate moiety (Fig. 4a and Fig. 5). This indicates that the divalent metal ion, instead of being coordinated by water molecules, interacts with the pyrophosphate group. The structure also revealed

different interactions between the pyrophosphate of the LLO and two essential catalytic residues, R375 and Y196. They appear to form hydrogen bonds with both phosphate groups, not just with the phosphate directly attached to the GlcNAc, as was observed for the ternary complex with non-reactive LLO (Fig. 4). This may explain why mutation of either of these residues to alanine abolished the glycosylation activity of PglB, but not binding of acceptor peptide[19,22]. Our structure also revealed contacts between the reducing-end sugar and PglB that were not observed before. In the previous ternary complex, the *N*-acetyl group of the reducing-end GlcNAc was at a distance of ~3.1Å from Y468 but was not close enough to interact with D56 (distance of 5.2Å). In the present structure, the *N*-acetyl group can form hydrogen bonds with the side chains of both D56 and Y468. The carbonyl oxygen is placed at a distance of 2.2Å from the hydroxyl oxygen of Y468, whereas the nitrogen is at a distance of 3.3Å from the carboxyl group of D56. Both D56 and Y468 have previously been demonstrated to be essential for PglB activity, but D56 was only shown to be involved in Mn^{2+} coordination and peptide binding[14,19]. Because both D56 and Y468 are strictly conserved catalytic residues, the newly identified interactions are likely relevant for all OST enzymes.

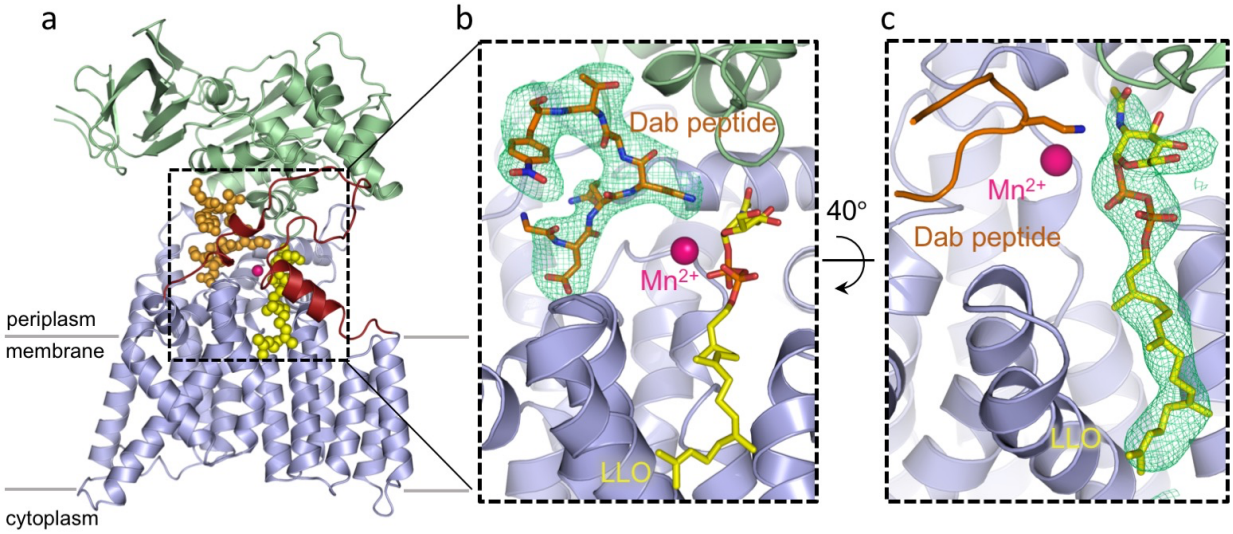


Figure 2. Structure of the ternary complex with inhibitory Dab peptide and reactive LLO. (**a**) cartoon representation of PglB with the transmembrane domain colored in light blue, periplasmic domain in light green, and N-terminal and C-terminal part of EL5 in turquoise and dark red, respectively. The LLO and Dab peptide are shown as yellow and orange spheres, respectively. (**b**) close-up view of the PglB active site. EL5 was removed for clarity. The LLO and Dab peptide are shown as yellow and orange sticks and the divalent metal ion in pink. A polder omit map contoured at 4.0 σ around bound Dab peptide is shown as green mesh. (**c**) close-up view of the PglB active site rotated 40° relative to that in (b). A polder omit map contoured at 4.0 σ around bound LLO is shown as green mesh. The Dab peptide is shown as ribbon.

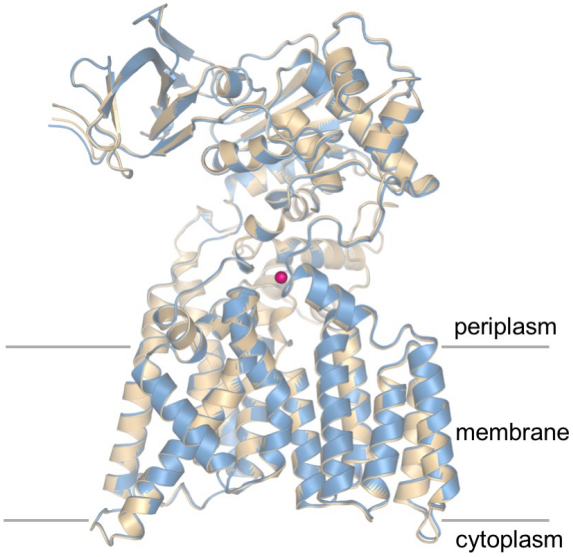
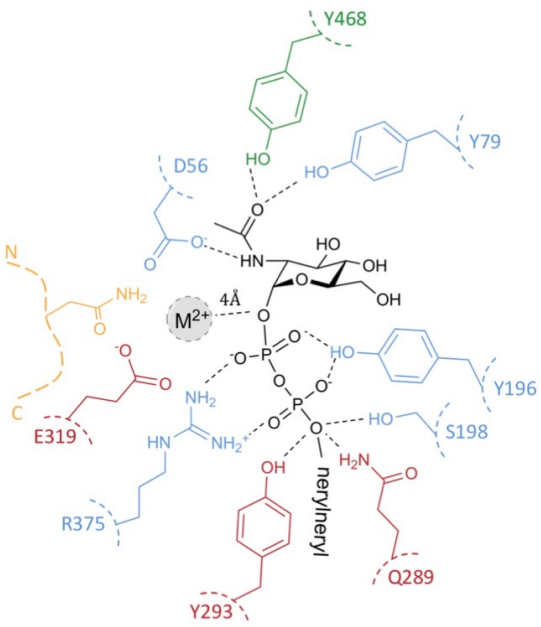


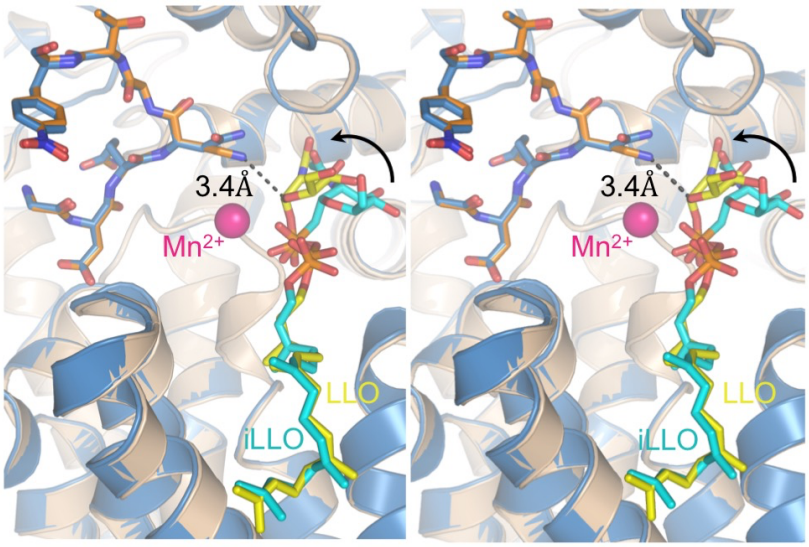
Figure 3. Superposition of the structures of both PglB ternary complexes. The previously reported ternary complex (PDB 5OGL), shown in blue, aligned with the new ternary complex PglB–LLO–Dab-peptide, shown in light brown. The divalent metal ion is shown in pink. No significant conformational changes were observed.

Figure 4. Interactions between PglB and bound LLO. (**a**) Schematic of the interactions. Periplasmic, transmembrane and EL5 residues are shown in green, blue and red, respectively, and labeled. The acceptor peptide is shown in orange. The conserved R375 and Y196 residues interact with both phosphate groups. A previously unobserved interaction between the side chain of D56 and the *N*-acetyl group of the GlcNAc moiety is suggested by the structure. (**b**) Stereo view of the superposition of both ternary complexes. The new ternary complex PglB–LLO–Dab-peptide is shown in light brown with the reactive LLO and inhibitory Dab peptide shown in orange and yellow sticks, respectively. The previously reported ternary complex (PDB 5OGL) is shown in blue with the non-hydrolyzable LLO analog (iLLO) and WT acceptor peptide shown in cyan and dark blue sticks, respectively. The divalent metal ion is shown as a pink sphere. The black dashed line represents the distance between the reactive LLO and the amino group of the Dab residue. (**c**) Stereo view of the LLO binding site rotated 90° relative to that of (a). The conserved residues in the catalytic site are shown as sticks. The GlcNAc moiety is positioned closer to the Mn^{2+} ion and residue D56.

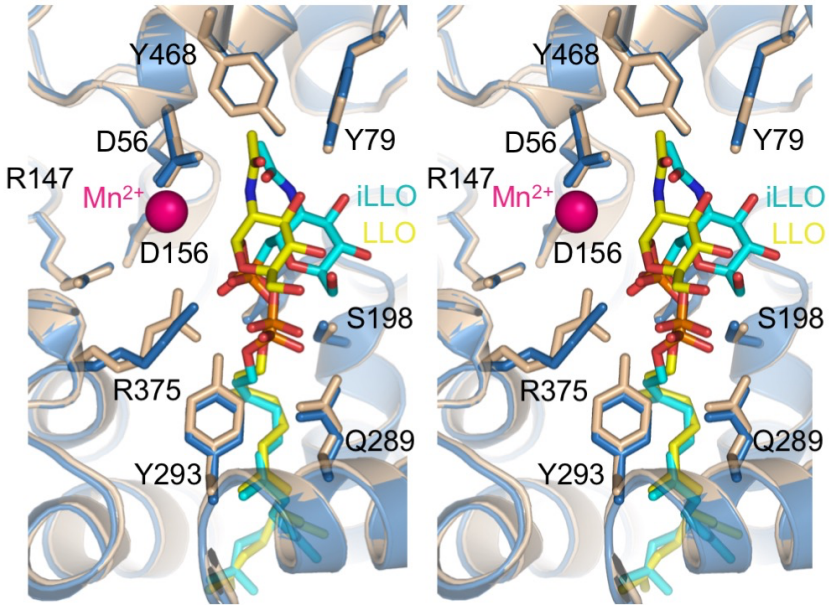
a


Y468
Y79
D56
HO
HO
HN
HO
OH
O
OH
N
NH2
4Å
M2+
O
O
O
O
P
O
O
HO
Y196
C
E319
NH2
O
P
O
HO
S198
HN
NH2
OH
H2N
O
nerylneryl
R375
Q289
Y293

b


3.4Å
Mn2+
LLO
iLLO
3.4Å
Mn2+
LLO
iLLO

c


Y468
D56
Y79
R147
Mn2+
iLLO
LLO
D156
S198
R375
Q289
Y293
Y468
D56
Y79
R147
Mn2+
iLLO
LLO
D156
S198
R375
Q289
Y293

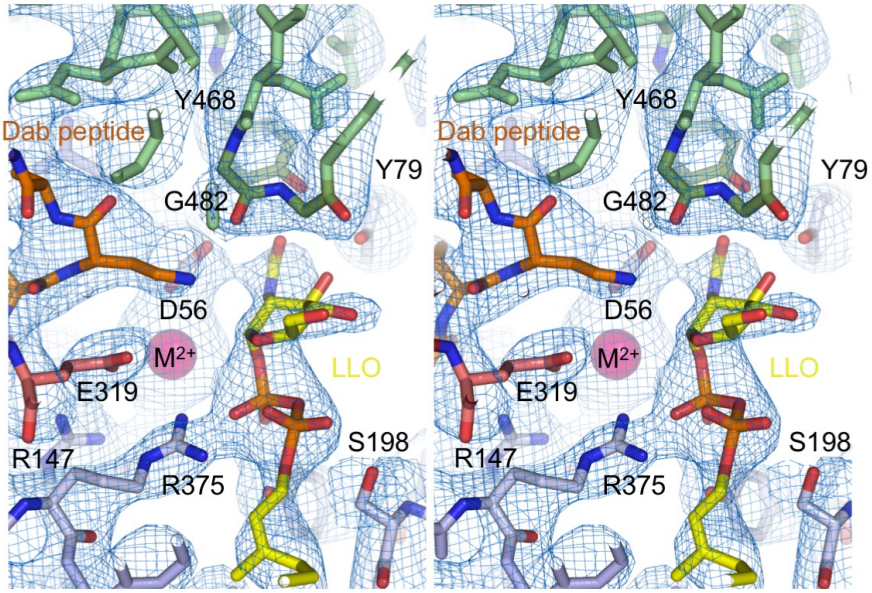


Figure 5. Stereo view of the electron density map of the catalytic site. 2Fo-Fc electron density map is shown at 1.3 σ level. The LLO and Dab peptide are shown as yellow and orange sticks, the divalent metal ion is shown in pink. The periplasmic, transmembrane and residues in EL5 are colored in green, light blue and red, respectively. Unlike the previously described ternary complex structure (PDB 5OGL), continuous electron density was observed between the divalent metal ion and the pyrophosphate group.

Mobility of the reducing-end GlcNAc in the catalytic site

The electron density for the lipid tail and the pyrophosphate group of reactive LLO was stronger than that for the GlcNAc moiety, unlike in the previous ternary complex with non-reactive LLO where the quality of the density for the LLO was similar throughout[19]. This suggests that either the sugar moiety of the reactive LLO molecule is more mobile, or that the LLO may have been hydrolyzed in the crystallization conditions, which would result in a mixture of LLO and lipid-pyrophosphate. Both PglB and yeast octameric OST were previously reported to exhibit hydrolytic activity and produce free oligosaccharides (fOS) in the absence of the acceptor sequon[33–35]. We therefore tested the stability of LLO by incubating it in crystallization conditions in the presence of PglB and the presence or absence of Dab peptide, before quantifying the amount of LLO left using an *in vitro* glycosylation assay (Fig. 6a)[19,21,22]. We did not detect any loss of LLO in the crystallization conditions over a period of 48 hours, which suggests that the intact LLO analog, and not its hydrolyzed derivative, are present in the co-crystal (Fig. 6b). The discrepancy between our results from *in vitro* experiments and previous studies remains unresolved[33,34]. Because there is no LLO hydrolysis in our experiment, the slightly weaker density for GlcNAc could be explained by an increased mobility of the GlcNAc moiety of the reactive LLO. We speculate that for wild type heptasacccharide-containing LLO the presence of extra sugar moieties would form additional interactions with PglB, likely decreasing the mobility of the reducing-end sugar.

Binding of the inhibitory peptide

The position of the backbone of the inhibitory peptide is unaltered relative to that of the wild type peptide in the previous ternary complex, validating the use of the Dab-containing peptide as a substrate mimic[19]. The Dab side chain reaches the catalytic site where it can form hydrogen bonds with three conserved and catalytically essential residues, the side

chains of D56 and E319 and the main chain carbonyl oxygen of G482 (Fig. 4 and Fig. 5)[14,21,22,36]. It was previously shown that replacing the acceptor asparagine with Dab resulted in a 10- fold decrease in peptide affinity[21]. This suggests that the presence of the carbonyl oxygen in the acceptor asparagine side chain, and the absence of a charge on the $-NH_2$ group, produce favourable interactions with the enzyme that contribute to the binding affinity but also allow glycan transfer. However, when comparing the Dab-peptide- to the wild type-peptide-bound structure of PglB, there are no specific interactions (hydrogen bonds) between PglB and the carbonyl group of the acceptor asparagine that would contribute significantly to the increased peptide affinity, as only van der Waals interactions are observed (Fig. 4b)[19]. The exact chemical mechanism of increased binding and transfer activity, and the precise reason for why Dab is inhibiting, are therefore unknown.

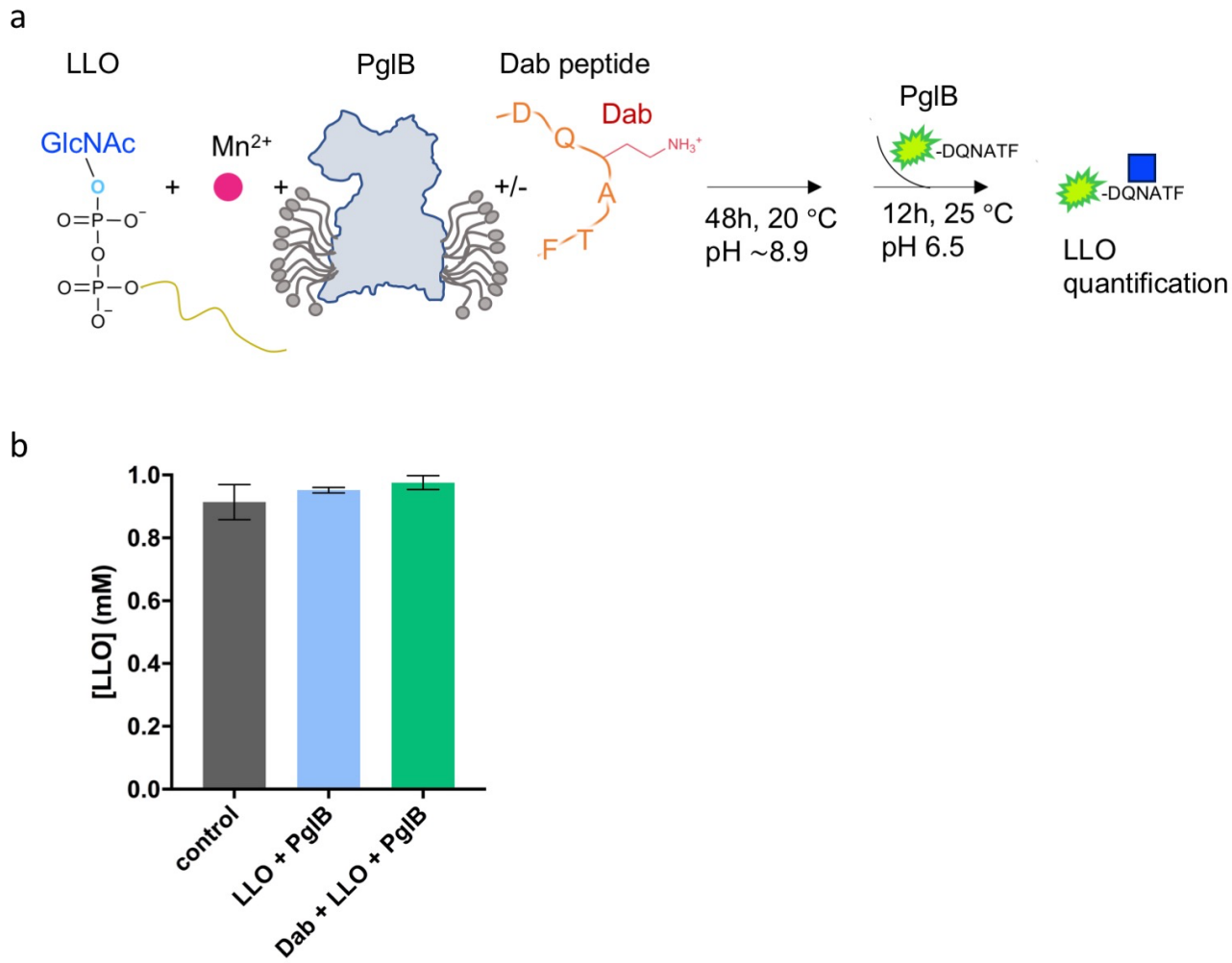


Figure 6. Stability of LLO at buffer conditions used for crystallization. (**a**) Schematic representation of the *in vitro* assay used to verify the stability of the reactive LLO in the crystallization conditions. LLO was incubated with PglB for 48 h at 20 °C in the presence or absence of Dab peptide and the amount of LLO was quantified. As a control, non-incubated LLO was used. (**b**) Quantification of the reactive LLO using an *in vitro* glycosylation assay. Data represents technical replicates and error bars indicate s.d. (n =3).

Discussion

The ternary complex presented here provides a first structural view into how OST interacts with a reactive LLO molecule. The captured intermediate is closer to the transition state of the glycosylation reaction than the previous ternary complex, because we observe an interaction between the pyrophosphate group of LLO and the Mn^{2+} ion, and the C-1 of GlcNAc is closer to the Dab side chain of the peptide than to the amide nitrogen in the previous structure[19]. The new structure also provides direct evidence that one of the roles of the divalent metal ion is the stabilization of the pyrophosphate, the leaving group of the substitution reaction. Similar roles were attributed to divalent metal ions in other glycosyltransferases (GTs) and were expected / predicted for OST[22,37–40]. Unexpectedly, however, the Mn^{2+} ion seems to be coordinated by the glycosidic oxygen atom rather than the negatively charged phosphate oxygens as observed in structures of GT-A superfamily members[37]. This suggests that the divalent metal ion might directly activate the glycosidic oxygen during the reaction by generating a reactive electrophile. This could compensate for the poor nucleophilicity of the carboxamide group of the acceptor asparagine and might even allow glycan transfer without amide activation. This hypothesis is in line with the observed inhibitory effect of Dab-containing peptides that cannot be glycosylated *in vitro*, which is probably due to the absence of a free electron pair in the protonated form. In contrast, it has been shown that a homoserine-containing peptide, which has the same number of methylene groups and therefore the same side chain length as Dab, is O-glycosylated[21]. The hydroxyl group of the homoserine analog, unlike the protonated Dab side chain, readily provides free electron pairs that can act as a nucleophile during the substitution reaction.

An *N*-acetyl group at the C-2 position of the reducing-end sugar is highly favoured by PglB and is strictly required for glycan transfer in eukaryotic OSTs[1,11,12,15,41,42]. Biochemical studies of yeast OST showed that minor modifications to this substituent, such as replacement of hydrogens with fluorines, reduced LLO binding and abolished glycosylation activity probably because the electron-withdrawing trifluoromethyl group made the neighbouring carbonyl oxygen a less efficient hydrogen bond acceptor[16]. The new ternary complex of PglB reveals hydrogen bonding interactions between the *N*-acetyl group of LLO and residues D56 and Y468 in the catalytic site of PglB. Because D56 also interacts with the acceptor peptide and the bound divalent metal ion, modification of the *N*-acetyl substituent is likely to affect the entire hydrogen bonding network, impairing glycan transfer[16].

Although the new ternary complex reveals novel interactions, future functional and molecular dynamics studies are required to give insight into the exact role of the divalent metal ion in activation of the glycosidic oxygen of LLO. The recent cryo-EM structures of eukaryotic octameric OSTs revealed that the fold and key side chains of the catalytic STT3 subunit are conserved, demonstrating that PglB is an excellent model system to study the glycosylation reaction[19,42–44]. The structure presented here is therefore likely to prove valuable for future mechanistic interpretations not only of bacterial ssOSTs, but of OSTs from all domains of life.

References

1. Larkin, A. & Imperiali, B. The expanding horizons of asparagine-linked glycosylation. *Biochemistry* **50,** 4411–4426 (2011).
2. Schwarz, F. & Aebi, M. Mechanisms and principles of N-linked protein glycosylation. *Curr. Opin. Struct. Biol.* **21,** 576–582 (2011).

3. Calo, D., Kaminski, L. & Eichler, J. Protein glycosylation in Archaea: Sweet and extreme. *Glycobiology* **20,** 1065–1076 (2010).
4. Kelleher, D. J. & Gilmore, R. An evolving view of the eukaryotic oligosaccharyltransferase. *Glycobiology* **16,** 47–62 (2006).
5. Nothaft, H. & Szymanski, C. M. Bacterial protein n-glycosylation: New perspectives and applications. *J. Biol. Chem.* **288,** 6912–6920 (2013).
6. Eichler, J. Extreme sweetness: protein glycosylation in archaea. *Nat. Rev. Microbiol.* **11,** 151–156 (2013).
7. Nothaft, H. & Szymanski, C. M. Protein glycosylation in bacteria: sweeter than ever. *Nat. Rev. Microbiol.* **8,** 765–778 (2010).
8. Helenius, A. & Aebi, M. Roles of N-linked glycans in the endoplasmic reticulum. *Annu. Rev. Biochem.* **73,** 1019–49 (2004).
9. Cherepanova, N., Shrimal, S. & Gilmore, R. N-linked glycosylation and homeostasis of the endoplasmic reticulum. *Curr. Opin. Cell Biol.* **41,** 57–65 (2016).
10. Szymanski, C. M., Burr, D. H. & Guerry, P. Campylobacter Protein Glycosylation Affects Host Cell Interactions Campylobacter Protein Glycosylation Affects Host Cell Interactions. *Infect. Immun.* **70,** 2242–2244 (2002).
11. Wacker, M. *et al.* Substrate specificity of bacterial oligosaccharyltransferase suggests a common transfer mechanism for the bacterial and eukaryotic systems. *Proc. Natl. Acad. Sci. U. S. A.* **103,** 7088–93 (2006).
12. Ramírez, A. S. *et al.* Characterization of the single-subunit oligosaccharyltransferase STT3A from Trypanosoma brucei using synthetic peptides and lipid-linked oligosaccharide analogs. *Glycobiology* **27,** 525–535 (2017).
13. Matsumoto, S. *et al.* Crystal structures of an archaeal oligosaccharyltransferase provide insights into the catalytic cycle of N-linked protein glycosylation. *Proc. Natl. Acad. Sci. U. S. A.* **110,** 17868–73 (2013).
14. Lizak, C., Gerber, S., Numao, S., Aebi, M. & Locher, K. P. X-ray structure of a bacterial oligosaccharyltransferase. *Nature* **474,** 350–355 (2011).
15. Ihssen, J. *et al.* Increased efficiency of Campylobacter jejuni N-oligosaccharyltransferase PglB by structure-guided engineering. *Open Biol.* **5,** 140227 (2015).
16. Tai, V. W. F. & Imperiali, B. Substrate specificity of the glycosyl donor for oligosaccharyl transferase. *J. Org. Chem.* **66,** 6217–6228 (2001).
17. Feldman, M. F. *et al.* Engineering N-linked protein glycosylation with diverse O antigen lipopolysaccharide structures in Escherichia coli. *Proc. Natl. Acad. Sci. U. S. A.* **102,** 3016–21 (2005).
18. Ramírez, A. S. *et al.* Chemo-enzymatic synthesis of lipid-linked GlcNAc2Man5oligosaccharides using recombinant Alg1, Alg2 and Alg11 proteins. *Glycobiology* **27,** 726–733 (2017).
19. Napiórkowska, M. *et al.* Molecular basis of lipid-linked oligosaccharide recognition and processing by bacterial oligosaccharyltransferase. *Nat. Struct. Mol. Biol.* **24,** 1100–1106 (2017).
20. Matsumoto, S., Taguchi, Y., Shimada, A., Igura, M. & Kohda, D. Tethering an N-Glycosylation Sequon-Containing Peptide Creates a Catalytically Competent Oligosaccharyltransferase Complex. *Biochemistry* **56,** 602–611 (2017).
21. Lizak, C. *et al.* Unexpected reactivity and mechanism of carboxamide activation in bacterial N-linked protein glycosylation. *Nat. Commun.* **4,** 2627 (2013).
22. Gerber, S. *et al.* Mechanism of bacterial oligosaccharyltransferase: In vitro quantification of sequon binding and catalysis. *J. Biol. Chem.* **288,** 8849–8861 (2013).
23. Lizak, C. *et al.* A catalytically essential motif in external loop 5 of the bacterial

oligosaccharyltransferase PglB. *J. Biol. Chem.* **289,** 735–746 (2014).
24. Kabsch, W. Xds. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66,** 125–132 (2010).
25. Strong, M. *et al.* Toward the structural genomics of complexes: crystal structure of a PE/PPE protein complex from Mycobacterium tuberculosis. *Proc. Natl. Acad. Sci. U. S. A.* **103,** 8060–8065 (2006).
26. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Cryst* **60,** 2126–2132 (2004).
27. Adams, P. D. *et al.* PHENIX: building new software for automated crystallographic structure determination. *Res. Pap. PHENIX Acta Cryst* (1948).
28. W. L.Delano. The PyMOL Molecular Graphics System. (2002).
29. Liu, F. *et al.* Rationally Designed Short Polyisoprenol-Linked PglB Substrates for Engineered Polypeptide and Protein N-Glycosylation. *J. Am. Chem. Soc.* **136,** 566–569 (2014).
30. Chen, M. M., Glover, K. J. & Imperiali, B. From peptide to protein: Comparative analysis of the substrate specificity of N-linked glycosylation in C. jejuni. *Biochemistry* **46,** 5579–5585 (2007).
31. Hendrickson, T. L., Spencer, J. R., Kato, M. & Imperiali, B. Design and evaluation of potent inhibitors of asparagine-linked protein glycosylation. *J. Am. Chem. Soc.* **118,** 7636–7637 (1996).
32. Bause, E., Breuer, W. & Peters, S. Investigation of the active site of oligosaccharyltransferase from pig liver using synthetic tripeptides as tools. *Biochem. J* **312,** 979–985 (1995).
33. Nothaft, H., Liu, X., McNally, D. J., Li, J. & Szymanski, C. M. Study of free oligosaccharides derived from the bacterial N-glycosylation pathway. *Proc. Natl. Acad. Sci. U. S. A.* **106,** 15019–15024 (2009).
34. Dwivedi, R., Nothaft, H., Reiz, B., Whittal, R. M. & Szymanski, C. M. Generation of free oligosaccharides from bacterial protein N-linked glycosylation systems. *Biopolymers* **99,** 772–783 (2013).
35. Harada, Y. *et al.* Eukaryotic oligosaccharyltransferase generates free oligosaccharides during N-glycosylation. *J. Biol. Chem.* **288,** 32673–32684 (2013).
36. Barre, Y. *et al.* A conserved DGGK motif is essential for the function of the PglB oligosaccharyltransferase from Campylobacter jejuni. *Glycobiology* **27,** 978–989 (2017).
37. Lairson, L. L., Henrissat, B., Davies, G. J. & Withers, S. G. Glycosyltransferases: Structures, Functions, and Mechanisms. *Annu. Rev. Biochem.* **77,** 521–555 (2008).
38. Liu, J. & Mushegian, A. Three monophyletic superfamilies account for the majority of the known glycosyltransferases. *Protein Sci.* **12,** 1418–31 (2003).
39. Gloster, T. M. Advances in understanding glycosyltransferases from a structural perspective. *Curr. Opin. Struct. Biol.* **28,** 131–141 (2014).
40. Albesa-Jové, D. & Guerin, M. E. The conformational plasticity of glycosyltransferases. *Curr. Opin. Struct. Biol.* **40,** 23–32 (2016).
41. Garcia-Quintanilla, F., Iwashkiw, J. A., Price, N. L., Stratilo, C. & Feldman, M. F. Production of a recombinant vaccine candidate against Burkholderia pseudomallei exploiting the bacterial N-glycosylation machinery. *Front. Microbiol.* **5,** 1–10 (2014).
42. Wild, R. *et al.* Structure of the yeast oligosaccharyltransferase complex gives insight into eukaryotic N-glycosylation. *Science* **359,** 545–550 (2018).
43. Bai, L., Wang, T., Zhao, G., Kovach, A. & Li, H. The atomic structure of a eukaryotic oligosaccharyltransferase complex. *Nature* **555,** 328–333 (2018).
44. Braunger, K. *et al.* Structural basis for coupling protein transport and N-glycosylation at the mammalian endoplasmic reticulum. *Science* **360,** 215–219 (2018).

Chapter 6: Conclusion and Outlook

N-linked protein glycosylation is a post-translation modification that is essential in eukaryotes, facilitating many biological processes such as protein folding, quality control, and host–pathogen interactions, as well as influencing virulence in bacteria[1–7]. Oligosaccharyltransferase (OST) is a key enzyme of the *N*-glycosylation pathway, where it catalyzes the transfer of a glycan from a lipid-linked oligosaccharide (LLO) to an acceptor asparagine within the conserved sequon N-X-T/S[8,9]. The objective of this thesis was to investigate the molecular basis of glycan transfer by a bacterial single subunit OST called PglB.

The single-subunit OST from the Gram-negative pathogenic bacterium *Campylobacter jejuni* is a well-studied, *bona fide* model system that allows the reaction mechanism of the OST-catalyzed process to be investigated without the complexity of additional subunits present in multisubunit OSTs. The previously reported structure of PglB from *Campylobacter lari* revealed the fold of STT3 proteins[8]. The same fold was subsequently found in the related AglB protein from *Archaeoglobus fulgidus*[10,11] and in yeast and mammalian OST complexes[12–14]. These structures showed, despite low sequence similarity, that the fold of the catalytic subunit is conserved in all OSTs, highlighting how PglB is an excellent model system to understand the mechanism of the glycosylation reaction. Structural and functional studies of PglB are therefore likely to prove valuable for interpretation of current and future complexes of OST.

At the beginning of this study, the structure of the PglB was known, and further functional studies gave insight into acceptor sequon binding and recognition[8,15–17]. However, owing to the absence of structural insight, the interactions of OST with LLO and the glycan transfer mechanism remained poorly understood. Moreover, little is known about how the amido group of the acceptor asparagine is activated for nucleophilic attack on the C1 carbon of the LLO substrate. Amides are poor nucleophiles because the free electron pair of the nitrogen is conjugated to the carbonyl group. It was proposed that the acceptor asparagine might form hydrogen bonds with two catalytically essential residues, D56 and E319[8,17]. This would require a rotation of the N–C bond of the amido group, which breaks the conjugation of the nitrogen electrons with the carbonyl group, activating the amide nitrogen, and results in nucleophilic attack[8,17]. However, a higher resolution structure would be required to identify the residues and interactions involved in the activation of the amide nitrogen of the acceptor asparagine, i.e. reliably measure the lengths of the potential hydrogen bonds formed with catalytic residues and to detect the presence of water molecules in the catalytic site.

In order to obtain high-quality crystals, and therefore a higher resolution X-ray structure of PglB, we pursued two approaches. The first was to generate conformation specific nanobodies and use these during crystallization trials. We hoped this would increase the hydrophilic surface area for protein-protein contacts, and restrict the flexibility of external loop 5 (EL5). Despite an exceptionally good immune response from the alpaca in the heavy-chain class, and the identification of 29 PglB-specific nanobodies, the binders did not improve protein crystallization. However, six nanobodies were thermostabilizing and may therefore be useful for future crystallization and NMR studies. In addition, complexes of PglB bound to one or a number of nanobodies may be useful for cryo-electron microscopy studies, where increasing the size of the complex and adding addition features can facilitate higher resolution. The second approach was to engineer PglB by introducing point mutations and generating a sequence optimized PglB construct, to both improve protein stability and crystallization. The sequence

optimized construct contained several amino acid changes aimed at improving lattice contacts. It was also fully functional and less prone to aggregation. Because of the thermostabilizing effect of peptide binding on PglB, we also designed and screened different PglB constructs carrying a glycosylation sequon inserted into either the periplasmic domain of PglB or the region linking transmembrane helix 13 and the periplasmic domain (Inseq constructs). Binding of the inserted sequon would likely stabilize the protein and increase the local concentration of the acceptor peptide during crystallization trials. The best Inseq construct, Inseq 4, showed much higher stability and crystallization efforts resulted in crystals that diffracted to 4.5 Å. This provides a basis to obtain future higher resolution structures of binary and ternary complexes of PglB in the future.

In this study, we pursued functional and structural studies of PglB ternary complexes in order to get insight into the interactions between OST, LLO, peptide and the divalent metal ion, and ultimately understand the molecular mechanism of glycan transfer. The main challenge in trapping a ternary-complex intermediate of OST was to prevent catalysis without compromising the binding affinity of the peptide and LLO substrates. Co-crystallization of such complexes was also very challenging due to the hydrophobic nature and low yields of wild-type LLO extracts. Crude extracts of wild-type LLO need large amounts of detergents for solubilization, which is was not optimal for crystallization experiments. Synthetic chemistry was a key approach to overcome both of these problems. Prior work by the Davis laboratory showed that synthetically accessible monosaccharide LLO analogs with shorter lipidic tails are indeed substrates for PglB[18]. The same approach was successfully used in structural and functional studies of *C. jejuni* PglH indicating the importance of synthetic chemistry in studies on glycosyltransferases[19]. In collaboration with the Reymond laboratory, we therefore tested PglB specificity for different water-soluble LLO analogs. We identified the best candidates for structural studies and subsequently determined X-ray structures of two ternary complexes of PglB. The first structure of PglB, bound to a wild type peptide and a non-hydrolyzable LLO analog, was determined at 2.7 Å resolution. The second structure bound to an inhibitory peptide and a reactive LLO analog was resolved at 3.4 Å. By combining structural insight from these ternary complexes with the previously determined structures of peptide-bound PglB and apo-AglB, as well as chemo-enzymatic approaches and functional studies, we defined key intermediates in the reaction mechanism of OST[8,10]. One of the main differences between the structures was the conformation of EL5, a flexible loop present in all OSTs that contains residues involved in peptide binding and catalysis[8,15]. We found that the previously disordered N-terminal segment of EL5 adopts an α-helix in the ternary complexes of PglB, folding over and contributing to the binding of LLO. In order to position the reducing end sugar close to the catalytic site LLO has to be pulled out of the membrane. PglB facilitates this step by providing a hydrophobic groove that binds the polyprenyl chain and makes several specific contacts to the GlcNAc and pyrophosphate moieties.

Disulfide cross-linking experiments demonstrated that the larger the glycan moiety, the slower the diffusion of the donor LLO into the active site, suggesting that the glycan must dive under or thread into EL5. The C terminal half of EL5 binds peptide while the N terminal half binds LLO, and its engagement and disengagement is necessary for substrate binding and product release. This has consequences for the mechanism of PglB, as outlined in chapter 4. In the absence of substrates, EL5 is flexible, disordered, and disengaged from the core of OST. This matches one of the apo-AglB and apo- yeast and mammalian STT3 structures[10,12–14]. We propose that binding of the substrates can in principle occur in either order. If peptide binds first, the C terminal half of EL5 becomes ordered, as demonstrated by the peptide-only

structure of PglB[8]. If LLO binds first, the N terminal half of EL5 might be engaged, but no structure of this state is currently available. Given that eukaryotic OSTs do not have a longer EL5 than their bacterial counterparts, but bind to larger more branched LLOs, we can speculate that the binding of LLO before peptide would be a more efficient mechanistic route, because when all of EL5 is disengaged the large glycan moiety may access the active site more easily.

The most recent structure of PglB bound to an inhibitory peptide and reactive LLO has probably captured the glycosylation reaction closer to the transition state than the ternary complex with non-hydrolyzable LLO. This is because we observe an interaction between the pyrophosphate group of LLO and the Mn^{2+} ion, and the C-1 of GlcNAc is closer to the Dab side chain of the peptide. In addition, the amine group of the Dab residue and the *N*-Acetyl group of the reducing end sugar are in close contact with D56. The use of a non-hydrolyzable phosphonate LLO analog in the previous ternary complex may have altered the interactions between the substrates. The ternary complex bound to the reactive LLO also provides structural evidence that the metal ion is necessary for OST activity by stabilizing the pyrophosphate-lipid, the leaving group during the glycosylation reaction, as it was shown in other glycosyltransferases and concluded for OST[16,20,21].

During this PhD, we have succeeded in generating the first ternary complex structures, not only of PglB, but any member of the Glycosyltransferase-C (GT-C) superfamily. Recently, crystal structures of another member of the GT-C superfamily, arabinose transferase, ArnT, from *Cupriavidus metallidurans*, alone and in complex with undecaprenyl-phosphate, was reported[22]. In these structures, periplasmic loop 4 only formed an α-helix in the presence of substrate, which resembles the rearrangement of EL5 upon LLO binding in PglB[22]. Future comparisons between the ternary complexes of PglB and other GT-C superfamily members will reveal key mechanistic similarities and differences.

Although our results give unprecedented mechanistic insight, they cannot clarify the exact mechanism of activation of the asparagine carboxamide group (cannot confirm the twisted amide hypothesis). They do however provide a promising starting point for computational approaches aimed at exploring the glycan transfer step in molecular detail. This work also gives insight into structural changes that may be required for the function of all OSTs, as the glycosylation mechanism is likely to be broadly conserved in all domains of life. In addition, these structures provide opportunities for engineering substrate specificity with respect to the LLO. This may be applied in the development and production of glycoconjugate therapeutics by inducing relaxed substrate specificity of PglB in order to synthesize various new bacterial and eukaryotic peptidoglycans to use as novel vaccines. During a time when structural biology is experiencing a drastic shift towards cryo-electron microscopy, the generation of conformational, thermostabilizing binders for PglB may be valuable for future studies of a protein that is currently at the size limit for high resolution structure determination of membrane proteins by this technique.

References

1. Helenius, A. & Aebi, M. Roles of N-linked glycans in the endoplasmic reticulum. *Annu. Rev. Biochem.* **73,** 1019–49 (2004).
2. Cherepanova, N., Shrimal, S. & Gilmore, R. N-linked glycosylation and homeostasis of the endoplasmic reticulum. *Curr. Opin. Cell Biol.* **41,** 57–65 (2016).

3. Shrimal, S., Cherepanova, N. A. & Gilmore, R. Cotranslational and posttranslocational N-glycosylation of proteins in the endoplasmic reticulum. *Semin. Cell Dev. Biol.* **41,** 71–8 (2015).
4. Szymanski, C. M., Burr, D. H. & Guerry, P. Campylobacter Protein Glycosylation Affects Host Cell Interactions Campylobacter Protein Glycosylation Affects Host Cell Interactions. *Infect. Immun.* **70,** 2242–2244 (2002).
5. Nothaft, H. & Szymanski, C. M. Protein glycosylation in bacteria: sweeter than ever. *Nat. Rev. Microbiol.* **8,** 765–778 (2010).
6. Larkin, A. & Imperiali, B. The expanding horizons of asparagine-linked glycosylation. *Biochemistry* **50,** 4411–4426 (2011).
7. Valguarnera, E., Kinsella, R. L. & Feldman, M. F. Sugar and Spice Make Bacteria Not Nice: Protein Glycosylation and Its Influence in Pathogenesis. *J. Mol. Biol.* **428,** 3206–3220 (2016).
8. Lizak, C., Gerber, S., Numao, S., Aebi, M. & Locher, K. P. X-ray structure of a bacterial oligosaccharyltransferase. *Nature* **474,** 350–355 (2011).
9. Kelleher, D. J. & Gilmore, R. An evolving view of the eukaryotic oligosaccharyltransferase. *Glycobiology* **16,** 47R–62R (2005).
10. Matsumoto, S. *et al.* Crystal structures of an archaeal oligosaccharyltransferase provide insights into the catalytic cycle of N-linked protein glycosylation. *Proc. Natl. Acad. Sci. U. S. A.* **110,** 17868–73 (2013).
11. Matsumoto, S., Taguchi, Y., Shimada, A., Igura, M. & Kohda, D. Tethering an N-Glycosylation Sequon-Containing Peptide Creates a Catalytically Competent Oligosaccharyltransferase Complex. *Biochemistry* **56,** 602–611 (2017).
12. Wild, R. *et al.* Structure of the yeast oligosaccharyltransferase complex gives insight into eukaryotic N-glycosylation. *Science* **359,** 545–550 (2018).
13. Bai, L., Wang, T., Zhao, G., Kovach, A. & Li, H. The atomic structure of a eukaryotic oligosaccharyltransferase complex. *Nature* **555,** 328–333 (2018).
14. Braunger, K. *et al.* Structural basis for coupling protein transport and N-glycosylation at the mammalian endoplasmic reticulum. *Science* **360,** 215–219 (2018).
15. Lizak, C. *et al.* A catalytically essential motif in external loop 5 of the bacterial oligosaccharyltransferase PglB. *J. Biol. Chem.* **289,** 735–746 (2014).
16. Gerber, S. *et al.* Mechanism of bacterial oligosaccharyltransferase: In vitro quantification of sequon binding and catalysis. *J. Biol. Chem.* **288,** 8849–8861 (2013).
17. Lizak, C. *et al.* Unexpected reactivity and mechanism of carboxamide activation in bacterial N-linked protein glycosylation. *Nat. Commun.* **4,** 2627 (2013).
18. Liu, F. *et al.* Rationally Designed Short Polyisoprenol-Linked PglB Substrates for Engineered Polypeptide and Protein N-Glycosylation. *J. Am. Chem. Soc.* **136,** 566–569 (2014).
19. Ramírez, A. S. *et al.* Structural basis of the molecular ruler mechanism of a bacterial glycosyltransferase. *Nat. Commun.* **9,** (2018).
20. Lairson, L. L., Henrissat, B., Davies, G. J. & Withers, S. G. Glycosyltransferases: Structures, Functions, and Mechanisms. *Annu. Rev. Biochem.* **77,** 521–555 (2008).
21. Liu, J. & Mushegian, A. Three monophyletic superfamilies account for the majority of the known glycosyltransferases. *Protein Sci.* **12,** 1418–31 (2003).
22. Petrou, V. I. *et al.* Structures of aminoarabinose transferase ArnT suggest a molecular basis for lipid A glycosylation. *Science* **351,** (2016).

Curriculum Vitae

Acknowledgments

I would like to thank my doctoral thesis supervisor, Prof. Kaspar Locher, for giving me an opportunity to work on this exciting project and for providing an excellent research environment. I am very grateful for his support and guidance throughout my PhD thesis.

I would like to thank my co-supervisors, Prof. Jean-Louis Reymond and Prof. Markus Aebi, for reviewing this PhD thesis and for their support and expertise.

I wish to thank all of my collaborators for their hard work and assistance throughout the project: Prof. Markus Aebi, Prof. Jean-Louis Reymond, Dr. Tamis Darbre, Dr. Jérémy Boilevin, Dr. Tina Sovdat, Giorgio Pesciullesi and Dr. Saša Štefanić.

I am thankful to the beamline staff at SLS, PSI Villingen, for their support during data collection.

I wish to thank all members of the Locher lab for their help. I would especially like to thank Dr. Scott Jackson, Dr. Ana Ramírez, Dr. Ioannis Manolaridis and Joël Bloch.

I am eternally grateful to my family for their love and support.

Appendix

Protein sequences:

>PglBopt
MELQQNFTDNNSIKYTAILILIAFAFSVLARLYWVAWASEFYEFFFNDQLMITTNDGY
AFAEGARDMIAGFHQPNDLSYFGSSLSTLTYWLYSILPFSFESIILYMSTFFASLIVVPII
LIAREYKLTTYGFIAALLGSIANSYYNRTMSGYYDTDMLVLVLPMLILLTFIRLTINKD
IFTLLLSPVFIMIYLWWYPSSYSLNFAMIGLFGLYTLVFHRKEKIFYLTIALMIIALSML
AWQYKLALIVLLFAIFAFKEEKINFYMIWALIFISILILHLSGGLDPVLYQLKFYVFKAS
DVQNLKDAAFMYFNVNETIMEVNTIDPEVFMQRISSSVLVFILSFIGFILLLKDHKSML
LALPMLALGFMALRAGLRFTIYAVPVMALGFGYFLYAFFNFLEKKQIKLSLRNKNILL
ILIAFFSISPALMHIYYYKSSTVFTSYEASILNDLKNKAQREDYVVAWWDYGYPIRYY
SDVKTLIDGGKHLGKDNFFSSFVLSKEQIPAANMARLSVEYTEKSFKENYPDVLKAM
VKDYQKTSAKDFLESLNDPNFKIDTPKTRDVYIYMPYRMLRIMPVVAQFANTNPDN
GEQEKSLFFSQANPLDQDKKQGSITLDNGVEISNDYRSLKIEGNSIPLKAFVDIESITNG
KFYYNEIDSKAQIYLLYLREYKSYVILDESLYNSSYIQMFLLNQYDQDLFEQITNDTR
AKIYRLKREFHHHHHHHHHH

>Nb8
MKYLLPTAAAGLLLLAAQPAMAQLQLVESGGGLVQAGGSLRLSCAASGRTFS
NYAMAWFRQAPGNEREFVAVHSWIGTSTYYADSVKGRFTISRDNAKNTLYLQMNSL
KPEDTAVYYCAATPPAMRTTGWGTRVPEYDYWGKGTPVTVSSHHHHHH
>Nb9
MKYLLPTAAAGLLLLAAQPAMAQMQLVESGGGLVQAGGSLRLSCAASGLT
LSTYTMTWFRQAPGKEREFVAAINRTGAGTYYADSMKGRFTISRDNTKNSLYLQMN
SLKPEDTAVYYCAAAKYGYGSDVVRARAYEHWGQGTPVTVSSHHHHHH
>Nb13
MKYLLPTAAAGLLLLAAQPAMAQLQLVESGGGLVQAGGSLRLSCVASGSIF
NLYATGWYRQTPGKERELVATIVRGGTPDYADSVKGRFTISSANPKNTYLQMNSLKP
EDTAVYYCNARQLGGGYYWGKGTRVTVSSHHHHHH
>Nb14
MKYLLPTAAAGLLLLAAQPAMAQWQLVESGGGLVQPGGSLRLSCAASGFTL
GYYRIGWFRQAPGKEREAVSCMRSPGDSTYYVDSVRGRFTISKDNAKNMVYLQMN
SLKPEDTAVYYCAADRRPYDCSNYGLNRNNQYDHWGQGTPVTVSSHHHHHH
>Nb16
MKYLLPTAAAGLLLLAAQPAMAQRQLVESGGGLAQVGDSLTLSCTASGLTFSRFNN
AWFRQAPGKEREFVAGLTWSGKNVHYSSSVEGRFTISQDGAKNTLYLQMNNLNPDD
TAVYYCAAGNNYPPTTYWGQGTRVTVSSHHHHHH
>Nb17
MKYLLPTAAAGLLLLAAQPAMAQLQLVESGGGVVQAGRSLRLSCAASGRTF
SSRYMGWFRQAPGKEREFVAAISWISGSTYYRNSVKGRFTISRDNAENTLYLEMNSL
EPEDTAIYYCAAPVSLRYGIGRSPSEYDHWSKGTRVTVSSHHHHHH
>Nb20
MKYLLPTAAAGLLLLAAQPAMAQWQLVESGGGLVQAGGSLRLSCTASGGIF
GLVDMGWYRQAPGKQRELVATISNRENTNYADSVKGRVTISRDNAKNTVYLQMNS
LKPEDTGVYFCHCNYNSRFVTGRRDYWGKGTPVTVSSHHHHHH

>Nb22
MKYLLPTAAAGLLLLAAQPAMAQGQFVESGGGLVQAGGSLRLSCAASDSMLTWNS
VGWYRQAPGKQRDLLASITGDGTANYADFLKGRFTISRDNAKKTVYLQMNNLKPED
TAVYYCKYIAFGSWSQGTPVTVSSHHHHHH
>Nb25
MKYLLPTAAAGLLLLAAQPAMAQWQFVESGGGYVQAGGSLRLSCAASGLTFTDYN
MAWFRQAPGKEREFVASINDIGITATHYADSVKGRFTISRDNAKNALYLQMNSLKPE
DTAVYYCAAGKRISLPFTGAHGMDYWGQGTPVTVSSHHHHHH
>Nb27
MKYLLPTAAAGLLLLAAQPAMAQLQLVESGGGLVQAGGSLRLSCAASGIIFG
INHMGWYRQAPGKQRELVATITSGGSTNYADSVKGRFTISRDNAKNTVYLQMNNLK
YEDTALYYCVVMRTHNGMDYWGKGTPVTVSSHHHHHH
>Nb30
MKYLLPTAAAGLLLLAAQPAMAQLQLVESGGGLVQAGGSLRLSCAASGRTGG
SYAMGWFRQAPGKERVFVAAINPSGGGTYYGNFVKGRFTISRDNAKNTVYLQMNSL
KPDDTAVYYCAAREFFSTRYTSTDDYNYWSKGTPVTVSSHHHHHH
>Nb34
MKYLLPTAAAGLLLLAAQPAMAQLQLVESGGGLVQAGASLRLSCVAPRNIFDI
GINTMGWYRRAPGKQRELVASITSDGRTNYIDSVKGRFTIAGDNAKNAVYLQMNSL
KTEDTGVYYCNAAPRPLLNYWGKGTPVTVSSHHHHHH
>Nb35
MKYLLPTAAAGLLLLAAQPAMAQWQLVESGGGVVQAGRSLRLSCAASGSIFSLGD
MGWYRQAPGKQREMVATITSRDNTDYADSVKGRVTISRDNAKNTVYLQMNSLNPE
DTGVYYCNLKYNSPLISGRRDYWGKGTPVTVSSHHHHHH
>Nb43
MKYLLPTAAAGLLLLAAQPAMAQGQLVESGGGLVQAGGSLRLSCTASGGI
FSINPMGWFRQAPGKQREVVAQIFSVGTVKYADSVKGRFTISRENAKKTLYLQMNSL
KPEDTAVYYCRGAFLSQEYWGQGTPVTVSSHHHHHH
>Nb44
MKYLLPTAAAGLLLLAAQPAMAQGQLVESGGGLVQAGGSLRLSCAASGSLFN
LYAMAWYRQAPGKERELVATSTLGGTPDYADSVKGRFTISRYNGAKNTYLQMNSL
KPEDTAVYYCNVRQLGGSSYWSQGTRVTVSSHHHHHH
>Nb46
MKYLLPTAAAGLLLLAAQPAMAQLQLVESGGGLVQAGGSLRLSCAASGRTFS
SYAMAWFRQAPGKEREFVAVNSWIGSSTYYADSVKGRFTISRDNAKNMLYLQMNS
LKPEDTAVYYCGSAFGSGYWYTSLYWSQGTPVTVSSHHHHHH
>Nb48
MKYLLPTAAAGLLLLAAQPAMAQGQLVESGGGVVQPGGSLRLSCTASGSIFSLAV
MGWYRQAPGKQRELVATITNQDHTSYTDSVKGRVTISRDNAKNTVYLQMNSLKPE
DTGVYYCNLNYNSRFITGRTDYWGKGTPVTVSSHHHHHH
>Nb49
MKYLLPTAAAGLLLLAAQPAMAQVQLVESGGTLVQAGGSLRLSCIASGRRISSY
NMAWFRRPPGKEREFVAAISYSGGGTGYGNSVQGRFTISRDNAKNTLYLQMNSLKS
EDTAVYYCAAMSATAFTTSADAYDYWGKGTPVTVSSHHHHHH
>Nb51
MKYLLPTAAAGLLLLAAQPAMAQLQFVESGGGLVRPGGSLRLSCVTSGFDL
DYYAVAWFRQAPGNKREGVAFITSSGYGTGYADSLKGRFTISRDNAKNAVYLQMNS
LKPEDTAVYYCAASGSIATLTPKYYRNWGKGTPVTVSSHHHHHH
>Nb52
MKYLLPTAAAGLLLLAAQPAMAQRQLVESGGGLVQAGGSLRLSCAASGRTF

SNYAMGWFRQAPGKEREFVGVDSWIGSSTYYADSVKGRFTISRDNTKNTLYLQMNS
LKPEDTAMYYCAATPPGMRTTGWGTRVSEYDYWGKGTPVTVSSHHHHHH
>Nb58
MKYLLPTAAAGLLLLAAQPAMAQLQLVESGGGLVQAGGSLRLSCEVSLLTTRTYNM
GWFRQAPGKEREFVAGRAWIGETTSYSDSVKGRFTVSKDNDKSILYLQMNNLRPED
TAVYFCAASNKVGDWSDPTRYNYWGKGTRVTVSSHHHHHH
>Nb59
MKYLLPTAAAGLLLLAAQPAMAQGQLVESGGGLAQAGGSLRLSCAASGRTFSSR
YMGWFRQAPGKEREFVAAISWISGSTYYRNSVKGRFTISRDNAENTLYLEMNSLEPE
DTAIYYCAAPVSLRYGIGRSPSEYDHWGQGTPVTVSSHHHHHH
>Nb62
MKYLLPTAAAGLLLLAAQPAMAQRQLVESGGGLVQVGGSLRLSCAVSGFTLDHY
QIGWFRKAPGKEREGVSCISSSDGSTHYAASVRGRFTISRDNAKNTMYLEMNNLEPE
DTADYYCAAARPVNSFGAIYYCEYYHAMDYWGKGTRVTVSSHHHHHH
>Nb63
MKYLLPTAAAGLLLLAAQPAMAQVQLVESGGGLVQAGGSLRLSCAASGRTFGNYA
AGWFRQAPGKEREFVSGLNYAGAATRYADSVKGRFTISRDNAKKMLYLQMNSLKP
EDTAVYYCAASNYAGVTVTMGEYDYWSQGTPVTVSSHHHHHH
>Nb64
MKYLLPTAAAGLLLLAAQPAMAQVQFVESGGGLVQAGGSLRLSCAASGRTFSNSV
MAWFRQAPGKEREFVGAFTWANIFYVDSVKGRFTISGDNAKNTLYLQMNSLKPEDT
AVYYCAAREYYRGSYDRASDYAYWGQGTRVTVSSHHHHHH
>Nb65
MKYLLPTAAAGLLLLAAQPAMAQRQLVESGGGLVQTGGSLRLSCAASGIIFSLADM
GWYRQAPGKQRELVATITSRESTNYADSVKGRVTISRDNAKNTMYLQMSSLKPEDT
GVYYCRLTYNSQFITGRRDYWSQGTPVTVSSHHHHHH
>Nb68
MKYLLPTAAAGLLLLAAQPAMAQVQLVESGGGLVQAGGSLRLSCVASANIFSVA
PVHWYRQAPGKQRELVATLHAGDVTNYADSVRGRFTGSRDNAKKTVYLQMNSLKP
EDTAVYYCNYFNYWGKGTPVTVSSHHHHHH
>Nb69
MKYLLPTAAAGLLLLAAQPAMAQWQLVESGGGLVQAGGSLRLSCAASGFTFNNYD
MGWFRQAPGKEREFVAHINWLGGNPNYADSVKGRFTISRDNAKNTLYLQMTSLKPE
DTAVYNCARRRSLGSVVRRDAAAYDYWSQGTRVTVSSHHHHHH
>Nb75
MKYLLPTAAAGLLLLAAQPAMAQGQLVESGGGLVQAGGSLRLSCLASTTILTHSTV
GWYRQAPGKERELVAHIRTTGTTNYADSVKGRFTISRDSAKNTVYLQMNSLKPEDT
AVYYCNLKTPFWVGSPGDSWGKGTPVTVSSHHHHHH
>Nb79
MKYLLPTAAAGLLLLAAQPAMAQRQLVESGGGLVQAGGSLRLSCTASGRTFSRD
LMGWFRQAPGKEREFVGAISWTGGGIGYADSVKGRFTISRDNAKNTLYLQMNSLKP
EDTAVYYCAASSLGAFRADVQAYNYWSKGTPVTVSSHHHHHH
>Nb81
MKYLLPTAAAGLLLLAAQPAMAQMQLVESGGGLVQAGDSLRLSCAASGRTFTNYN
MGWFRQAPGKQREFVAAITKVSGSTGYADSVKGRFTISRDNAKNSVYLDMNSLKSA
DTAVYYCAAAAAAFGSSARQYDYWSQGTPVTVSSHHHHHH
>Nb82
MKYLLPTAAAGLLLLAAQPAMAQMQLVESGGGMVQAGGSLRLSCTASGSIFGLAV
MSWYRQAPGNQRELVATITHGDNTNYADSVKGRVTISRDNAKDTVYLQMNSLKPE
DTGVYYCNLNYNSPMFAGRRDYWSKGTPVTVSSHHHHHH

>Nb87
MKYLLPTAAAGLLLLAAQPAMAQMQLVESGGGLVQSGGSLRLSCAASGFIFSNIPM
NWVRQAPGKGLEWVSEISGYGGTTYYQESVKGRFTISRDNAKNMLYLQMNSLKPED
TAVYYCALNRVGSRRGQGTPVTVSSHHHHHH
>Nb93
MKYLLPTAAAGLLLLAAQPAMAQWQFVESGGGVVQAGRSLRLSCVASGSIFSLADM
GWYRQAPGKQREMVATITSRDNTDYADSVKGRVTISRDNAKNTVYLQMNSLNPED
TGVYYCNLKYNSPLISGRRDYWSQGTPVTVSSHHHHHH
>Nb100
MKYLLPTAAAGLLLLAAQPAMAQMQLVESGGGLVQAGGSLRLSCAASGRTFSNYA
MGWFRQAPGKEREFITAINWIGGSTYYADSVKGRFTISRDNAKNTLYLQMNSLKPED
TAVYYCAARPVGVWSDLMKFDYWGKGTPVTVSSHHHHHH