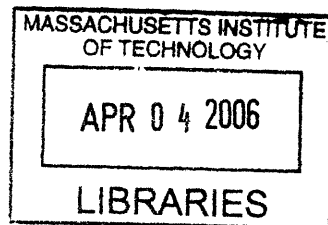# Investigating Asparagine-Linked Protein Glycosylation in Eukaryotic and Prokaryotic Systems

by

Eranthie Weerapana

B.S. Chemistry
Yale University, 2000

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

at the

Massachusetts Institute of Technology

February 2006

Signature of Author: _____
Department of Chemistry
December 6, 2005

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

Accepted by:_____
Robert W. Field
Haslam and Dewey Professor of Chemistry
Chairman, Departmental Committee on Graduate Students

This doctoral thesis has been examined by a committee of the Department of Chemistry as follows:

Professor Sarah O'Connor_____
Chair

Professor Barbara Imperiali_____
Thesis Supervisor

Professor Stephen L. Buchwald_____

Investigating Asparagine-Linked Protein Glycosylation in Eukaryotic and Prokaryotic Systems

by

Eranthie Weerapana

Submitted to the Department of Chemistry
on December 6, 2005 in Partial Fulfillment of the
Requirements for the Degree of Doctor of Philosophy

ABSTRACT

N-linked protein glycosylation is characterized by the formation of a β-glycosylamine linkage to an asparagine residue within the Asn-Xaa-Ser/Thr consensus sequence. This modification is found in organisms from eukaryotic, archaeal and bacterial domains and is implicated in numerous cellular processes.

Recently, a system of N-linked glycosylation was characterized in a gram-negative bacterium, *Campylobacter jejuni*. Glycosylation in this organism involves the transfer of a heptasaccharide from an undecaprenyl-pyrophosphate (Und-PP) carrier onto the asparagine side-chain of a protein. The genes in the '*pgl* gene cluster' encode all of the proteins necessary for the biosynthesis of the glycan donor and its ultimate transfer to protein. The heptasaccharide donor has been characterized as GalNAc-α1,4- GalNAc-α1,4-(Glcβ1,3)-GalNAc-α1,4-GalNAc-α1,4-GalNAc-α1,3-Bac-α1,PP-Und, where Bac is bacillosamine (2,4-diacetamido-2,4,6-trideoxyglucose). A synthetic route was developed to access bacillosamine-phosphate, which was incorporated into UDP-bacillosamine (UDP-Bac) and undecaprenyl-pyrophosphate-bacillosamine (Und-PP-Bac), which are substrates for the Pgl enzymes. Using the synthetic UDP-Bac, the role of the PglC glycophosphoryltransferase was elucidated *in vitro*. The activities of the Pgl glycosyltransferases, PglA, PglJ, PglH and PglI were validated using the synthetic Und-PP-Bac substrate and it was discovered that PglH is a polymerase that catalyzes the transfer of the three terminal GalNAc residues. PglB is the oligosaccharyl transferase of the bacterial system. Using the synthetic glycan donor, PglB was shown to act *in vitro* on a short peptide substrate. This *in vitro* system enables detailed mechanistic investigations into the action of this intriguing enzyme.

N-linked glycosylation in eukaryotes is catalyzed by oligosaccharyl transferase (OT), a multimeric protein complex localized in the lumen of the endoplasmic reticulum. The Stt3p protein of the eukaryotic OT cluster is homologous to the bacterial PglB enzyme. With the goal of inhibiting OT in a cellular environment, a family of peptidomimetic inhibitors with nanomolar affinity was synthesized. These inhibitors were evaluated for cellular inhibition of OT using a novel, high-throughput assay that monitors the production of a reporter glycoprotein, secreted alkaline phosphatase. The results from the screening yielded a hydrophobic peptidomimetic compound as a potential candidate for further studies into the *in vivo* inhibition of OT.

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

Acknowledgements

Table of Contents

List of Figures

9

List of Tables

List of Abbreviations

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

2-AB	2-aminobenzamide
Ac	acetyl
Alloc	allyl oxycarbonyl
Bac	bacillosamine (2,4,6-trideoxy-2,4-diacetamido-D-glucopyranose)
Bac-6-OH	6-hydroxy-bacillosamine (2,4-dideoxy-2-4-diacetamido-D-glucopyranose)
Bn	benzyl
Boc	*tert*-butyloxycarbonyl
Bz	benzoyl
calcd	calculated
CDI	1,1'-carbonyldiimidazole
CHO	Chinese hamster ovary
C. jejuni	*Campylobacter jejuni*
Dab	diaminobutyric acid
DCM	dichloromethane
DIPEA	N,N-diisopropylethylamine
DMEM	Dulbecco's modified eagle media
DMF	N,N-dimethylformamide
DMSO	dimethylsulfoxide
Dol-P	dolichyl monophosphate
Dol-PP	dolichyl pyrophosphate
DPM	disintegrations per minute
E. coli	*Escherichia coli*
ε	extinction coefficient or molar absorptivity
ER	endoplasmic reticulum
ESI	electrospray ionization mass spectrometry
FBS	fetal bovine serum
Fmoc	fluoren-9-ylmethyloxycarbonyl
GDP-Man	guanosine 5'-diphospho-α-D-mannose
GalNAc	N-acetyl-D-galactosamine
GlcNAc	N-acetyl-D-glucosamine
Gal	D-galactose
Glu/Glc	D-glucose
HATU	O-(7-azabenzotriazol-1-yl)-1,1,3,3-tetramethyluronium
HEPES	N-(2-hydroxyethyl)-piperazine-N'-(2-ethane-sulfonic acid)
HPLC	high-performance liquid chromatography
HRMS	high-resolution mass spectrometry
Hz	hertz
K_i	inhibition constant
λ_{em}	emission wavelength
λ_{ex}	excitation wavelength
LiHMDS	lithium hexamethyl-disilazide
LRMS	low resolution mass spectrometry

MALDI-MS	Matrix assisted laser desorption ionization mass spectrometry
Man	D-mannose
MeCN	acetonitrile
min	minutes
MUP	4-methylumbelliferyl phosphate
NBD	7-nitrobenz-2-oxa-1,3-diazole
NMR	nuclear magnetic resonance
Nph	*p*-nitrophenylalanine
Nva	norvaline
NVoc	nitroveratryloxycarbonyl
OT	oligosaccharyl transferase
PAL	5-(4'-aminomethyl-3',5'-dimethoxyphenoxy)valeric acid
PEG-PS	polyethyleneglycol-grafted polystyrene
PBS	phosphate-buffered saline
ppm	parts per million
PyBOP	benzotriazole-1-yl-oxy-tris-pyrrolidino-phosphoniumhexafluorophosphate
RP-HPLC	reverse-phase HPLC
SDS	sodium dodecyl sulfate
S. cerevisiae	*Saccharomyces cerevisiae*
SeAP	secreted alkaline phosphatase
SPPS	solid-phase peptide synthesis
tBu	*tert*-butyl
TFA	trifluoroacetic acid
THF	tetrahydrofuran
TIS	triisopropylsilane
TMS	tetramethylsilane
Tn	tunicamycin
t_R	retention time
TLC	thin layer chromatography
Tris	tris(hydroxymethyl)aminomethane
TUP	theoretical upper phase
UDP-Bac	uridine 5'-diphosphobacillosamine
UDP-Bac-6-OH	uridine 5'-diphopho-6-hydroxybacillosamine
UDP-GalNAc	uridine 5'diphospho-*N*-acetyl-α-D-galactosamine
UDP-GlcNAc	uridine 5'diphospho-*N*-acetyl-α-D-glucosamine
UDP-Glc	uridine 5'diphospho-α-D-glucose
UMP	uridine 5'-monophosphate
Und-P	undecaprenyl monophosphate
Und-PP	undecaprenyl diphosphate
Und-PP-Bac	undecaprenyl-diphospho-bacillosamine
UV-Vis	ultraviolet-visible
Xaa	used to denote any amino acid

Chapter 1

Introduction

Glycosylation is a complex, co- or post-translational protein modification that serves to expand the diversity of the proteome. A vast array of carbohydrate units, together with a variety of glycan-protein linkages, have been identified in glycoproteins originating from eukaryotic, archaeal and bacterial organisms.[1] Eukaryotic glycoproteins have been implicated in a multitude of cellular processes including the immune response, intracellular targeting, intercellular recognition and protein folding and stability.[2] The biological role of prokaryotic glycoproteins requires further exploration, but it is evident that glycosylation plays a vital role in pathogenicity and evasion of the host immune system.[3]

This thesis focuses on N-linked protein glycosylation, which is characterized by a β-glycosylamine linkage to asparagine.[4, 5] This system has been extensively investigated in eukaryotes, particularly in the yeast, *Saccharomyces cerevisiae*. Recently, a system of N-linked glycosylation was discovered in a gram-negative bacterium, *Campylobacter jejuni*, which was the first observation of this protein modification in the bacterial domain.[6, 7] Although the β-glycosylamine-linkage to protein is conserved, the structure of the glycan that is transferred is strikingly different between eukaryotic and bacterial systems.[8, 9] The biosynthetic machinery responsible for this elaborate protein modification follows the same overall progression, whereby an oligosaccharide is assembled step-wise on a polyisoprenyl-pyrophosphate carrier and ultimately transferred to protein. Here, we provide a detailed comparison of the N-linked glycosylation machinery of eukaryotic organisms, exemplified by *Saccharomyces cerevisiae*, with the parallel process in the gram-negative bacterium, *Campylobacter jejuni*.

Chapters 2-4 of this thesis describes the synthesis of intermediates in the N-linked glycosylation pathway in *C. jejuni* and the *in vitro* validation of the biosynthetic enzymes

involved in this process. In chapters 5 and 6, the design and evaluation of inhibitors for the eukaryotic glycosylation process is described.

Eukaryotic glycosylation

Overview

N-linked glycosylation in eukaryotes is a co-translational process that is catalyzed by oligosaccharyl transferase (OT), a protein complex localized in the lumen of the endoplasmic reticulum (ER).[10] Proteins in the secretory pathway encode a signal sequence that is recognized by the signal recognition particle (SRP) (Figure 1-1).[11, 12] The SRP directs the growing polypeptide chain to the translocon machinery, whereby transport across the ER membrane occurs.[13, 14] The signal peptidase complex, containing the essential Sec11p protein, then cleaves the signal sequence,[15] which moves the polypeptide to the compartment where OT-mediated glycosylation takes place. OT transfers a tetradecasaccharide 'core' unit ($GlcNAc_2Man_9Glc_3$) to the polypeptide chain. Approximately 14-residues of the newly-translated polypeptide have to clear the luminal surface of the ER-membrane for glycosylation to occur.[16] Since the protein is still being translated by the ribosome during this process, global folding and tertiary structure of the protein are not important determinants in the recognition events leading to glycosylation. However, as discussed later, the local secondary structure around the site of glycosylation is a vital determinant in this enzymatic process.
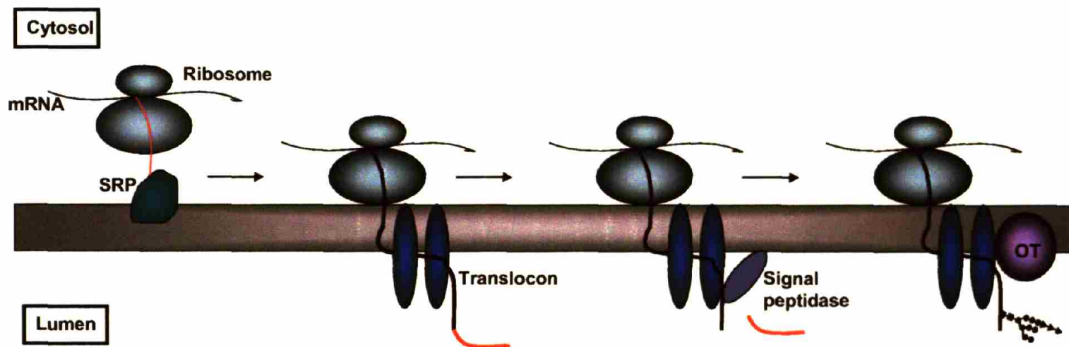
Figure 1-1. Co-translational glycosylation of proteins in the secretory pathway.

The process of protein translocation and glycosylation is a well-characterized eukaryotic phenomenon. Significant work has been done on the role of the translocon, the signal peptidase and oligosaccharyl transferase in this protein modification process. The exact machinery involved in bacterial *N*-linked glycosylation is poorly defined. The glycosylation process is thought to occur in the periplasm of bacteria, which is the functional equivalent of the ER in eukaryotes. Currently, there is no conclusive evidence to demonstrate either the co-translational or post-translational nature of *N*-linked glycosylation in the bacterial periplasm. If the *C. jejuni* machinery functions post-translationally, on fully-folded proteins, it could potentially be one of the most significant differences between the eukaryotic and prokaryotic processes of *N*-linked glycosylation.

After OT-catalyzed glycosylation, the newly-synthesized glycoproteins undergo a series of processing reactions including the enzymatic cleavage of the glucose residues and several of the mannose residues on the tetradecasaccharide.[17] Upon trimming of the two terminal glucose residues, by glucosidase I and II, the protein enters a folding cycle mediated by two ER-resident lectins with chaperone function, calnexin (membrane-bound) and calreticulin (soluble) (Figure 1-

17

2).[18-20] These lectins function together with the co-chaperone Erp57 (a thiol-oxidoreductase), to facilitate the folding process.[21, 22] Dissociation of the glycoprotein-calnexin/calreticulin complex is mediated by glucosidase II, by cleavage of the remaining glucose residue. In the case of misfolded proteins, a UDP-glucose:glycoprotein glucosyltransferase (UGGT) acts to re-glucosylate the protein, which can re-enter the calnexin/calreticulin cycle. The fully-folded proteins that are released by the glycoprotein chaperones are acted on by ER mannosidases, before transport to the Golgi with the aid of the ERGIC-53 lectin. Quality control at this stage, ensures that non-native conformations are recognized by mannosidase-like lectins (Mnl1p, EDEM) to initiate the ER-associated degradation (ERAD) pathway.[23] The fully-folded glycoproteins that are transported to the Golgi, undergo further transformation involving glycan trimming and elaboration, catalyzed by a series of glycosidases/glycosyltransferases, to afford the plethora of diverse carbohydrate units that are present on eukaryotic glycoproteins.

The calnexin/calreticulin cycle exemplifies the intricate mechanisms by which eukaryotic cells maintain protein quality control to prevent the release of misfolded proteins into the extra-cellular milieu. Such a complex system of glycoprotein folding has not been demonstrated in the bacterial system, and the processes by which these organisms maintain glycoprotein quality is a prevailing question. The glycan structures displayed on all N-linked glycoproteins of C. jejuni are identical, lacking the immense diversity of eukaryotic N-linked glycoproteins. This is due to the lack of the glycan trimming/elaboration steps that occur post-glycosylation in the Golgi of eukaryotic cells. Since bacterial cells lack the extensive compartmentalization present in eukaryotic cells, there is no functional equivalent of the Golgi, where such elaboration steps can take place.
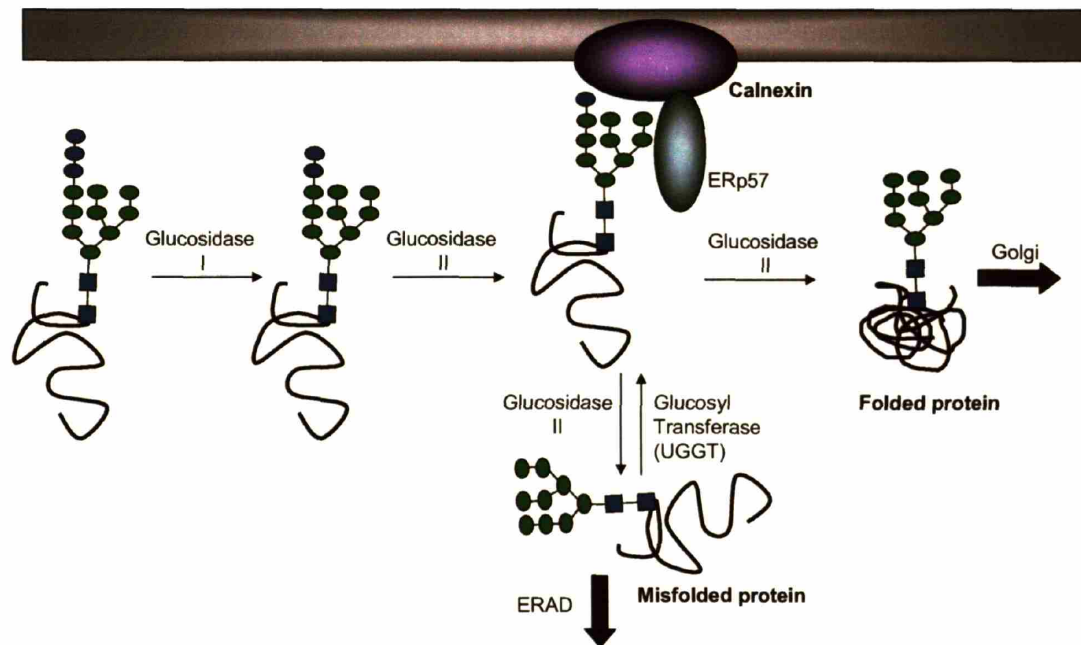
Figure 1-2. The calnexin-calreticulin cycle.

The Dolichol Pathway

The glycan that is transferred to the asparagine side chain of a nascent protein is a tetradecasaccharide ($Glc_3Man_9GlcNAc_2$) that is highly conserved in higher order eukaryotic systems (Figure 1-3).[5] In trypanosomes, truncated analogs ($Man_{6,7,9}GlcNAc_2$) have been shown to be transferred to protein.[24] The tetradecasaccharide is assembled in the ER membrane on a dolichyl-pyrophosphate carrier by a series of glycosyltransferases, in a process known as the dolichol pathway.[8] The pathway begins with dolichyl-phosphate (Dol-P); the dolichols constitute a family of α-saturated, (S)-polyisoprenyl-phosphates containing 14-17 isoprene units that are biosynthesized from farnesyl-pyrophosphate on the cytoplasmic face of the ER.[25] Dol-P is then elaborated on the cytoplasmic side to form a dolichyl-pyrophosphate-linked heptasaccharide ($Dol-PP-GlcNAc_2Man_5$) by a series of glycosyltransferases that utilize the nucleotide-activated

sugar donors, UDP-N-acetylglucosamine (UDP-GlcNAc) and GDP-mannose (GDP-Man). This heptasaccharide is then flipped to the luminal face of the ER membrane by the flippase, Rft1p,[26] where further elaboration occurs to yield the tetradecasaccharide (Dol-PP-GlcNAc$_2$Man$_9$Glc$_3$).
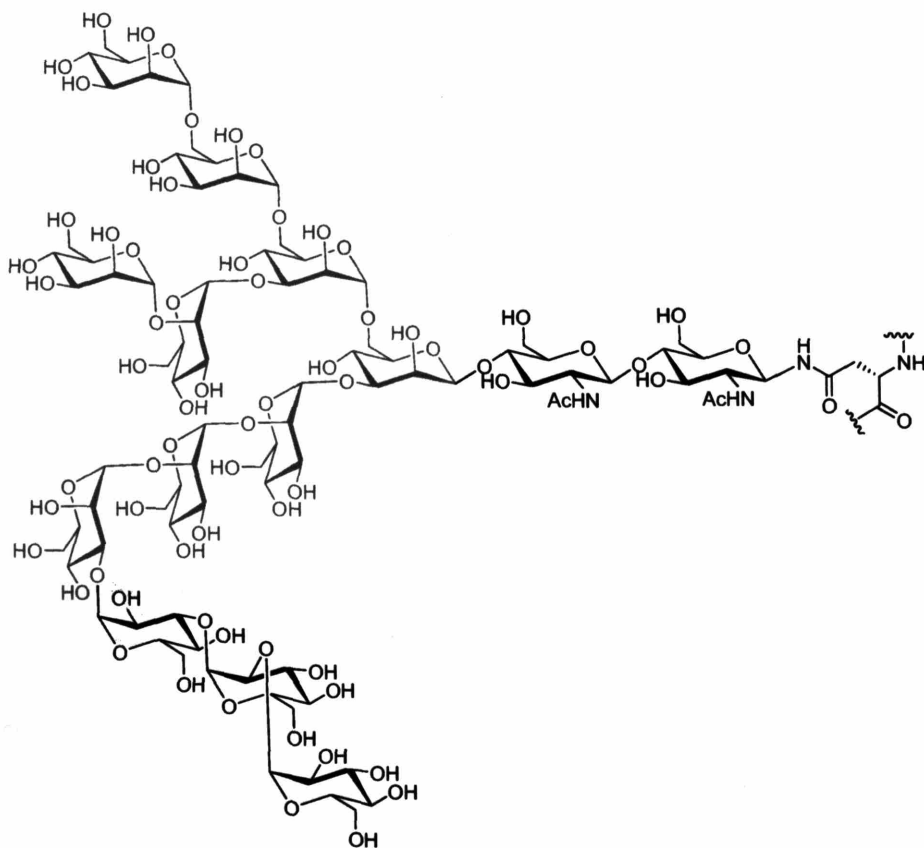
Figure 1-3. Structure of the tetradecasaccharide (Glc$_3$Man$_9$GlcNAc$_2$) transferred to protein in eukaryotic N-linked glycosylation.

The reactions on the cytoplasmic face of the ER begin with Alg7, an N-acetylglucosamine-phosphate transferase that elaborates Dol-P to Dol-PP-GlcNAc (Figure 1-4).[27] Alg7 is inhibited by tunicamycin, which is a microbial natural product that acts as a transition-state inhibitor for this step and is currently the only cellular inhibitor that affects N-

linked protein glycosylation.[28] Recent bioinformatics studies have determined that the addition of

the second GlcNAc residue is catalyzed by a hetero-oligomeric protein classified as Alg13/14.[29]

Further *in vivo* experiments have demonstrated that the transmembrane-bound Alg14 recruits the

soluble Alg13, containing a predicted catalytic domain, to the ER membrane for catalysis.[30, 31]

The role of Alg1, which catalyzes the first mannosylation step, has been extensively

characterized *in vitro* due to the advantageous formation of the chemically challenging $\beta(1-4)$-

mannosidic linkage, which may be useful in chemoenzymatic syntheses.[32, 33] Two mannosyl-

transferases, Alg2 and Alg11 are implicated in the remaining cytosolic glycosylations in the

dolichol pathway,[34-36] yet their precise roles in this process remains to be rigorously validated.

Biochemical and genetic assays have shown that Alg1 interacts with both Alg2 and Alg11,

suggesting a role for protein-complex formation in maintaining the fidelity of the early steps in

the dolichol pathway.[37]

On the luminal face of the ER membrane, the glycan donors are the dolichyl-phosphate

linked Dol-P-Man and Dol-P-Glc. These donors are synthesized from the corresponding

nucleotide-linked sugar donors, GDP-Man or UDP-Glc, by a Dol-P-Man synthase (DPM1)[38] and

a Dol-P-Glc synthase (Alg5),[39] which act on the cytoplasmic face of the ER membrane and

flipped into the lumen. Four mannosyltransferases and three glucosyltransferases are involved in

the completion of the dolichyl-pyrophosphate-linked tetradecasaccharide. All of the 'luminal'

glycosyltransferases in the dolichol pathway have now been identified and are all highly

hydrophobic, basic proteins (MW 65-75 kDa) that include multiple transmembrane domains with

small hydrophilic loops.[8] Recently, the bi-functional nature of Alg9 was demonstrated, whereby

Alg9 catalyzes the addition of both the seventh and the ninth mannose residue (both α-1,2-

mannosyl linkages).[40] Once the entire tetradecasaccharide has been assembled in the lumen of the ER, it is utilized by OT, which catalyzed transfer of the oligosaccharide to protein.[41]
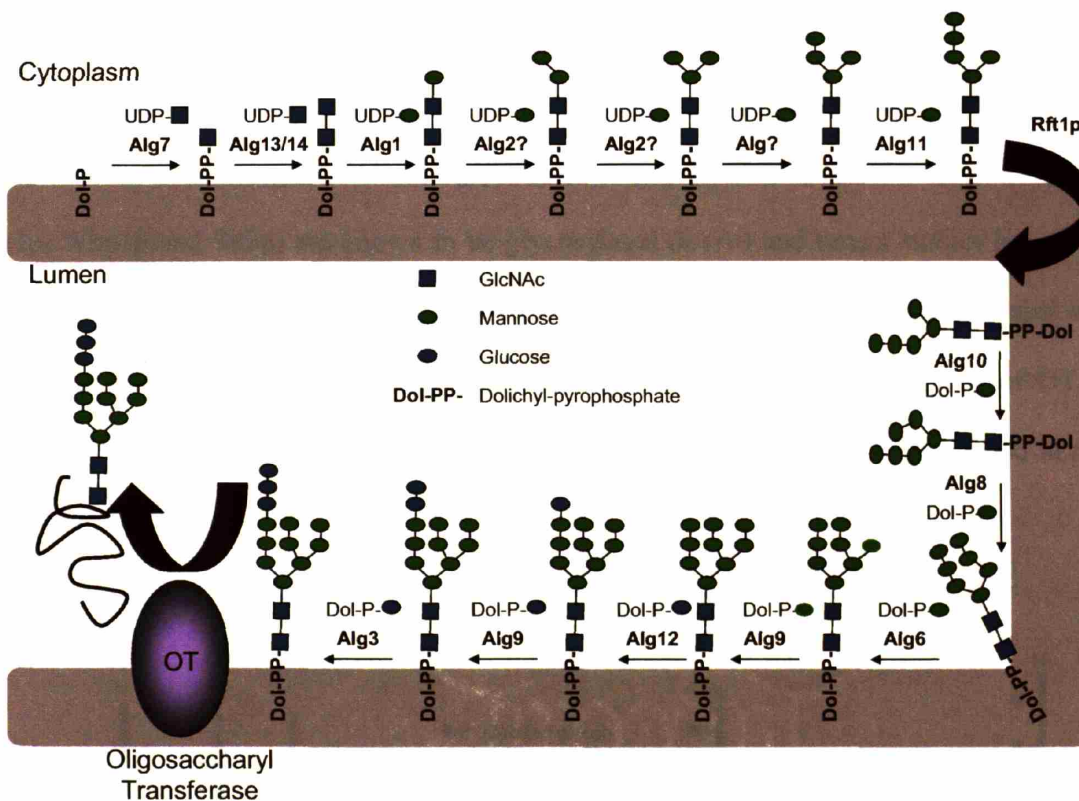
Figure 1-4. The dolichol pathway of *N*-linked glycosylation in *Saccharomyces cerevisiae*.

Oligosaccharyl Transferase (OT)

The OT complex is a multimeric, membrane-associated enzyme that is localized in the membrane of the lumen of the endoplasmic reticulum, with the active site disposed to the luminal compartment.[41, 42] This enzyme complex has been most extensively investigated in the yeast *Saccharomyces cerevisiae* and comprises at least eight membrane-bound protein subunits (Figure 1-5) that exist in three sub-complexes, Ost1p-Ost5p, Ost2p-Swp1p-Wbp1p, and Stt3p-

Ost4p-Ost3p/Ost6p. Genetic-knockout experiments have revealed that five of these subunits (Ost2p, Ost1p, Stt3p, Swp1p and Wbp1p) are absolutely essential for yeast viability.[43-46] Recent data indicate that there are two, functionally-distinct OT complexes *in vivo*.[47, 48] These complexes differ in one of the subunits, containing either Ost3p or Ost6p, which are highly homologous proteins. The two complexes vary in specificity and function and are thought to be responsible for fine-tuning the N-linked glycosylation process.[48] Three of the essential subunits (Ost1p, Wbp1p and Stt3p) are known to be glycosylated *in vivo* and recent studies have shown that glycosylation of Ost1p and Wbp1p is not essential for activity.[49] Stt3p is the largest of the essential proteins in the complex and is hypothesized to be the catalytic subunit of OT.[50, 51] Disrupting the two adjacent glycosylation sites on Stt3p ($N^{535}NT$ and $N^{539}NT$) resulted in major growth defects.[49]
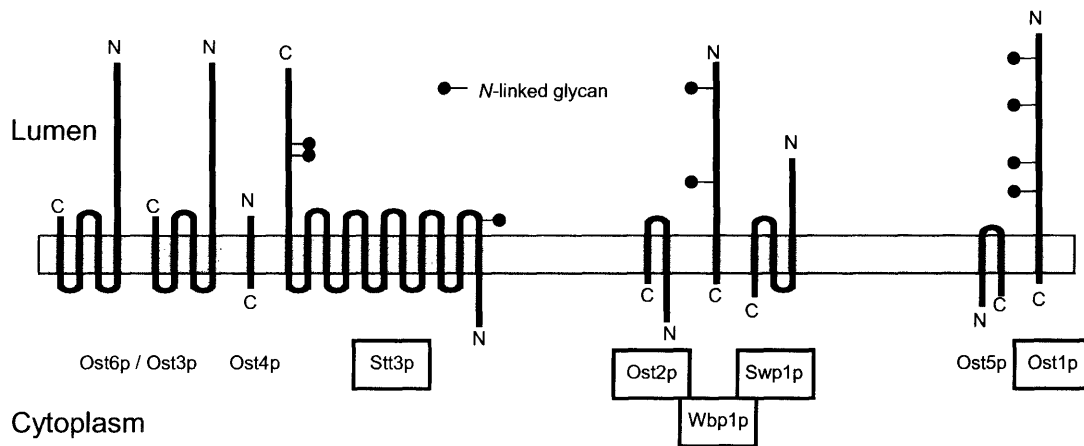
Figure 1-5. Subunit composition of the yeast (*S. cerevisiae*) oligosaccharyl transferase complex. The boxed subunits are essential for yeast viability.

Stt3p is a highly-conserved transmembrane protein that is found in all eukaryotic organisms. Site–directed mutagenesis combined with photo-crosslinking experiments show that Stt3p is directly involved in the catalytic process.[51, 52] The *S. cerevisiae* Stt3p comprises 11-13 transmembrane segments and a hydrophilic C-terminal domain. A detailed topology mapping study using engineered glycosylation sites, established that the *N*-terminus of Stt3p is in the cytosol and the C-terminus in the lumen, and places most of the highly-conserved residues on the luminal face of the ER membrane, which is in accord with the lumen being the site of catalysis.[53] A highly-conserved WWDYGY amino acid sequence is present in all homologs of Stt3p in organisms that contain a system for *N*-linked glycosylation. A bacterial protein from *Campylobacter jejuni*, PglB, is homologous to Stt3p, and *N*-linked glycosylation in this organism is abolished with the loss of PglB.[7, 54] Mutations within this conserved WWDYGY motif result in the loss of glycosylation activity, suggesting that this sequence includes residues that are essential for catalysis in PglB and Stt3p.[54]

Mechanism of OT

The glycosylation step catalyzed by OT involves the transfer of a pre-assembled tetradecasaccharide from a dolichyl-pyrophosphate carrier onto the side chain of asparagine, within the Asn-Xaa-Ser/Thr sequon (Scheme 1-1), where Xaa can be any amino acid except proline.[55]
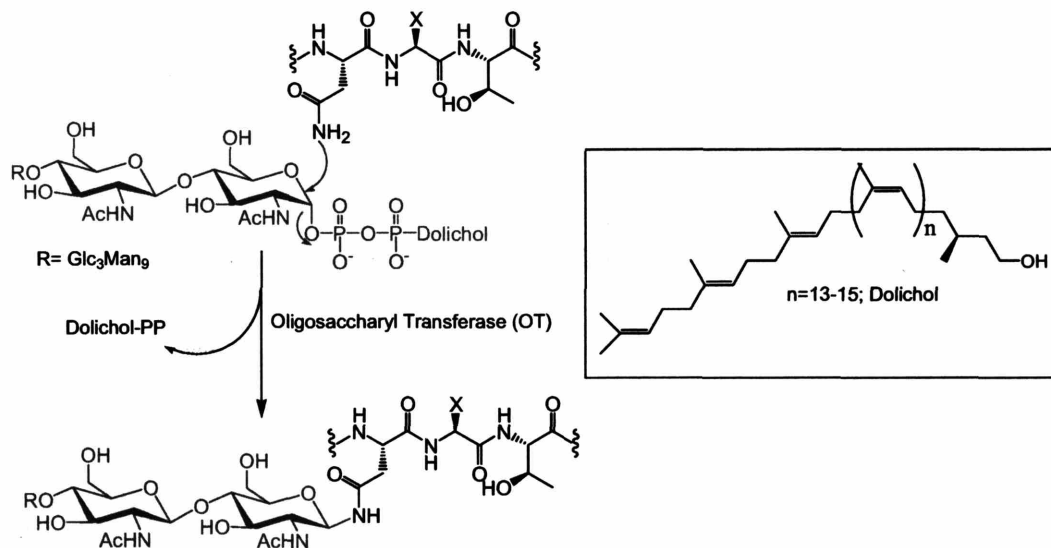
Scheme 1-1. Reaction catalyzed by oligosaccharyl transferase (OT).

The enhanced nucleophilicity of the amide nitrogen in this glycosylation reaction is a prevailing mechanistic question and several mechanisms have been proposed to explain this amide reactivity. The first model by Marshall (Figure 1-6 A), proposes that a hydrogen bond between the hydroxyl group of the Ser/Thr residue and the carbonyl of the Asn side chain promotes deprotonation of the amide nitrogen, resulting in nucleophilic attack.[56] In a second model, Bause (Figure 1-6 B) suggests that the amide group of asparagine functions as a hydrogen-bond donor and the hydroxyl group as a hydrogen-bond acceptor, resulting in amide activation.[57] In the final model by Imperiali (Figure 1-6 C), an Asx-turn motif is believed to be an important element in the glycosyl transfer mechanism. The hydrogen-bonding network in the Asx-turn may facilitate deprotonation of the nitrogen to afford a neutral imidol species. This intermediate could then react with the dolichyl-pyrophosphate-linked oligosaccharide donor to form the β-linked glycopeptide.[58, 59]
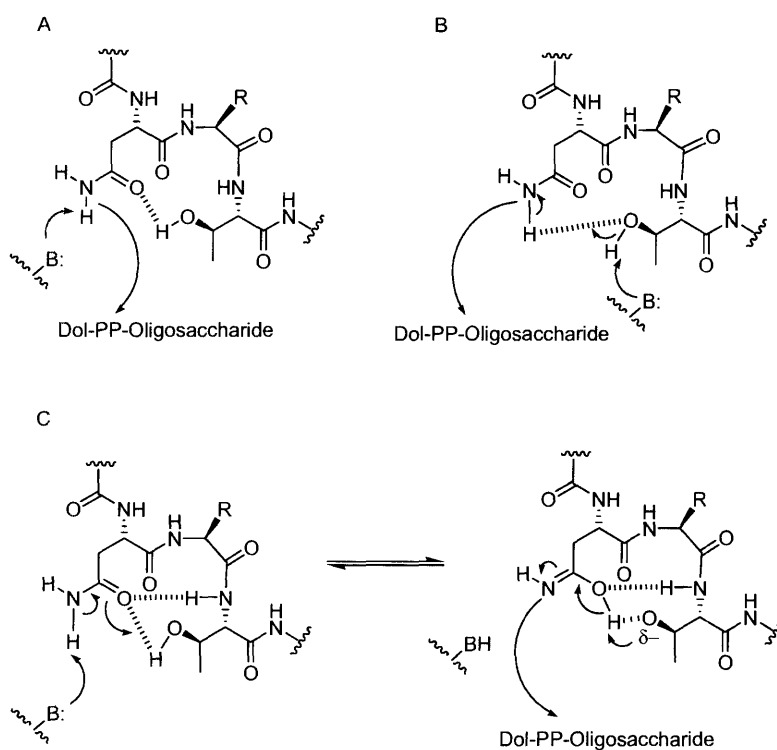
Figure 1-6. The proposed mechanisms of OT. (A) Marshall. (B) Bause. (C) Imperiali.

The fact that proline is not accepted at the Xaa site within the Asn-Xaa-Ser/Thr sequon, coupled with the observation that 10-30% of potential glycosylation sites are not glycosylated, suggests that local conformational effects play an important role in the process of *N*-linked glycosylation.[60] The peptide can bind to the OT active site in two distinct conformations – an Asx-turn or a β-turn conformation. A β-turn is characterized by hydrogen bonding between backbone amide groups.[61] On the other hand, in an Asx-turn, the side-chain amide of the asparagine partakes in hydrogen-bonding interactions with the backbone resulting in a more open peptide conformation (Figure 1-7).[61] Using NMR studies it was demonstrated that an unglycosylated peptide based on a short sequence of hemagglutinin, adopts an Asx-turn

conformation, but upon glycosylation, undergoes a chain reversal to induce a compact type-I-β-turn conformation.[62, 63]

β–Turn   Asx-Turn
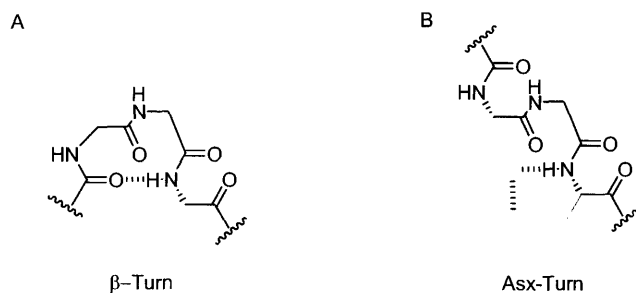
Figure 1-7. Structure of (A) β-turn motif. (B) Asx-turn motif.

In vitro, a short peptide sequences containing the Asn-Xaa-Ser/Thr motif can be glycosylated by OT. In support of the Asx-turn motif being an important determinant in *N*-linked glycosylation site-specificity, a tripeptide substrate was synthesized with a carbon macrocyle pre-organizing the sequence into an Asx-turn. This resulted in a substrate with a significantly greater affinity for OT (Figure 1-8).[64]
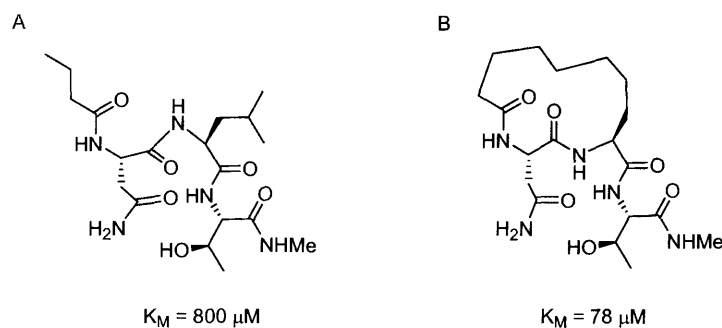
$K_M = 800 \ \mu M$ $K_M = 78 \ \mu M$

Figure 1-8. Effect of cyclization on enzyme affinity. (A) Linear substrate K_M 800 μM. (B) Cyclized substrate K_M 78 μM.

In order to derive a mechanistic picture, several peptides in which the asparagine was replaced by various isosteres, were synthesized and assayed for activity against OT (Table 1-1).[59, 65, 66] These asparagine analogs demonstrate different ionization properties, as well as varying hydrogen-bonding capacities and therefore add insight into the mechanism of the glycosylation process. Compounds 2-5 (Table 1-1) were synthesized in the Imperiali group.[59] Replacing the Asn with a negatively charged Asp moiety (2) resulted in complete loss of glycosylation with no competitive binding, suggesting that the active-site residues do not tolerate a negative charge. The incorporation of a methyl ester (3) also diminishes binding, suggesting the importance of a proton-donor at the Asn-position. The thioasparagine-containing peptide (4) is a substrate for OT, but with a lower rate of turnover, possibly due to the reduced basicity and hydrogen-bonding capacity of sulfur. Finally, replacing the Asn with a diaminobutyric-acid residue (Dab) (5) competitively inhibits OT with a K_i that is comparable to the K_M of the substrate. Compounds 6-8, were synthesized by Coward *et al*. The diazoketone (6) was proposed to target a catalytic nucleophile such as a cysteine thiolate, but showed no catalytic turnover or inhibition of OT.[66] The methyl ketone (7) and sulfoxide (8) could be substrates for *N*-glycosylation if the mechanism involved amide deprotonation, however neither of these compounds acted as substrates or inhibitors. Finally, compounds 9-12 were synthesized by Bause *et al*,[65] and demonstrated that the hydroxamide (10) group resulted in a weak inhibitor and the hydroxyasparagine (12) was a poor substrate for OT. The biological characterization of all of these asparagine analogs, provides clues to the mechanism of OT catalyzed glycosylation, but does not conclusively support either of the proposed mechanisms.
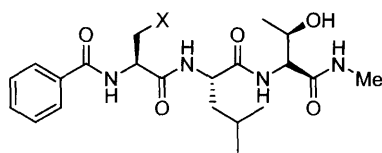
X = 1.	NH_2 structure	good substrate	7.	ketone structure	no binding
2.	O^- structure	no binding	8.	sulfoxide structure	no binding
3.	OMe structure	no binding	9.	SO_2NH_2 structure	no binding
4.	thioamide NH_2 structure	poor substrate	10.	N-OH structure	weak inhibitor
5.	NH_3^+ structure	good inhibitor	11.	OH structure	no binding
6.	N_2 structure	no binding	12.	HO, NH_2 structure	poor substrate

Table 1-1. Kinetic analysis of substrates/inhibitors containing asparagine analogs.

Inhibitors of OT

The discovery that a tripeptide containing the Dab amino acid is a competitive inhibitor of OT, sparked the field of OT inhibitor development. This observation, coupled with the pre-organization of potential inhibitors into the bio-active Asx-turn conformation, resulted in inhibitors with low-nanomolar affinity for OT. The most potent of these inhibitors are illustrated in Figure 1-9.[64, 67]

Although potent inhibitors of OT exist, none of these compounds function within a cellular environment. An inhibitor that can selectively target OT in cells can enable investigations into the downstream effects of glycosylation on various cellular functions. The

focus of chapters 5 and 6 of this thesis is the modification of the current inhibitors to afford

peptidomimetic structures as potential *in vivo* inhibitors of OT.[68]
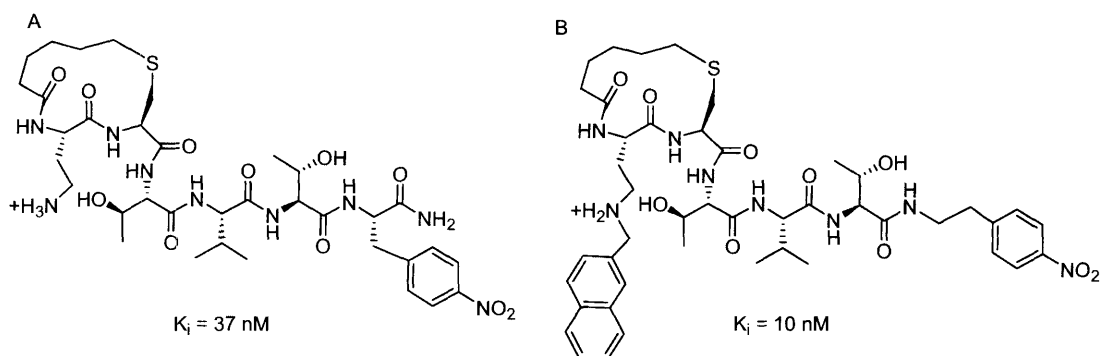
Figure 1-9. The most potent OT inhibitors to date.

N-linked glycosylation in eukaryotic systems is an extremely complex process involving

a multitude of enzymes acting in concert to biosynthesize the dolichyl-pyrophosphate-linked

tetradecasaccharide and facilitate the ultimate transfer of the glycosyl moiety to protein. All of

the proteins involved in this process are highly hydrophobic, membrane-associated proteins,

which complicate detailed *in vitro* biochemical and biophysical characterization. In order to

understand this complex process of *N*-linked protein glycosylation, it would be ideal to have

access to a parallel system that is more amenable to biochemical characterization. The recent

discovery of a system of *N*-linked protein glycosylation in the gram-negative bacterium,

Campylobacter jejuni, has provided us with a parallel, yet simpler system that is more suitable

for in-depth biochemical characterization, and can potentially shed light on the more complex

eukaryotic process.

Prokaryotic glycosylation

Overview

Investigations into glycosylation systems in eukaryotic organisms have prevailed since the late 1930's, yet for many decades it was assumed that bacteria and archaea were devoid of this important protein modification.[69] The discovery of surface layer (S-layer) glycoproteins in the gram-negative halophile, *Halobacterium salinarium*, was the first such system to be found outside of the eukaryotic domain.[70] These S-layer glycoproteins in archaea have the unique feature of assembling into two-dimensional crystalline arrays on the cell wall of halobacteria and are characterized by a variety of glycans and a diverse array of linkages to protein.[71] Since this initial report of S-layer glycoproteins in halobacteria, several characterizations of similar glycosylated proteins in the bacterial domain have also surfaced.[72] These glycoproteins are integrated into cell-surface appendages, such as pili and flagella.[69] The pili of pathogenic bacteria, such as *Neisseria meningitidis* and *Campylobacter jejuni/coli* contain *O*-linked glycans that involve unusual sugars such as pseudaminic acid (a nine-carbon sugar that resembles sialic acid) and 2,4-diacetamido-2,4,6-trideoxyhexose (DATDH) (Figure 1-10).[73, 74] The flagella of gram-negative bacteria have also been shown to include *O*-linked pseudaminic acid analogs.[75, 76]

Pseudaminic acid   2,4-diacetamido-2,4,6-trideoxyhexose
                                                    (DATDH)
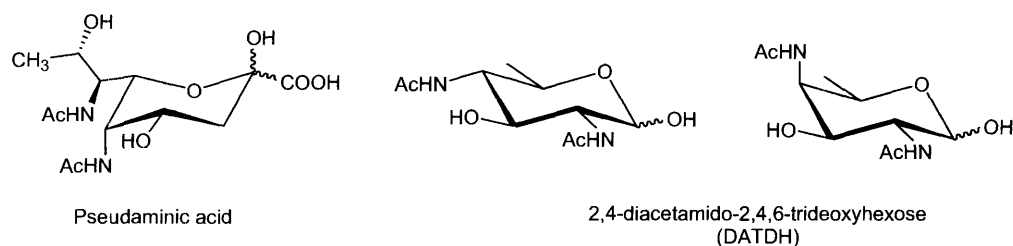
Figure 1-10. Structures of pseudaminic acid and DATDH sugars.

The first system of *N*-linked protein glycosylation to be discovered in gram-negative bacteria comes from *Campylobacter jejuni*, a human-gut mucosal pathogen that is implicated in gastroenteritis (Figure 1-11).[7] *Campylobacter* enteritis is characterized by acute abdominal pain and inflammatory diarrhea,[77] hence, understanding the pathogenicity of *C. jejuni* could potentially lead to better prevention and infection-control strategies. The sequencing of the *C. jejuni* genome, together with detailed genetic maps have facilitated genetic characterization of various strains of this organism and the majority of the work described herein specifically focuses on the strain *C. jejuni* NCTC 11168.[78, 79 80]

Figure 1-11. Electron micrograph of *Campylobacter jejuni* bacterial cells.

(Figure taken from **www.niwa.co.nz/.../ bacteria2_large.jpg/view**).

In 1999, it was discovered that *C. jejuni* contains a gene locus that is involved in the biosynthesis of a number of highly immunogenic glycoproteins.[6] This cluster was termed the '*pgl* gene cluster' and contained the genes *pglA* to *pglG*, which demonstrate significant homology to enzymes involved in bacterial lipopolysaccharide (LPS) and capsular polysaccharide (CPS) biosynthesis. Mutagenesis of key residues in this cluster resulted in no

discernible effect on CPS or LPS levels but caused a dramatic reduction in the immunoreactivity of numerous *C. jejuni* proteins.

Soybean agglutinin (SBA) is a plant lectin known to bind terminal GalNAc residues. The highly immunogenic *C. jejuni* proteins affected by mutations in the *pgl* gene cluster bind strongly to the SBA lectin. This allowed the identification of PEB3 and CgpA, two highly immunoreactive glycoproteins in *C. jejuni*.[81] The glycan attached to these proteins was not affected by β-elimination, which generally removes *O*-linked glycans, thus suggesting a linkage *via* a glycosyl amide to an asparagine residue.[9] Through the action of specific exoglycosidases, the oligosaccharide was shown to include one or more α-linked GalNAc residues.[81] The PEB3 glycoprotein was partially purified and analyzed by mass spectrometry and shown to be modified *via* an Asn-linked glycan with a mass of 1406 Da. Using nano-NMR techniques on very small quantities of isolated glycan, the structure of the glycan was determined to be the heptasaccharide, GalNAc-α-1,4-GalNAc-α1,4-(Glcβ1,3)-GalNAc-α1,4-GalNAc-α1,4-GalNAc-α1,3-Bac-β1,N-Asn where Bac is bacillosamine (2,4-diacetamido-2,4,6-trideoxyglucose) (Figure 1-12).[9] Furthermore, this heptasaccharide structure was shown to be conserved throughout all *C. jejuni* strains.[82]
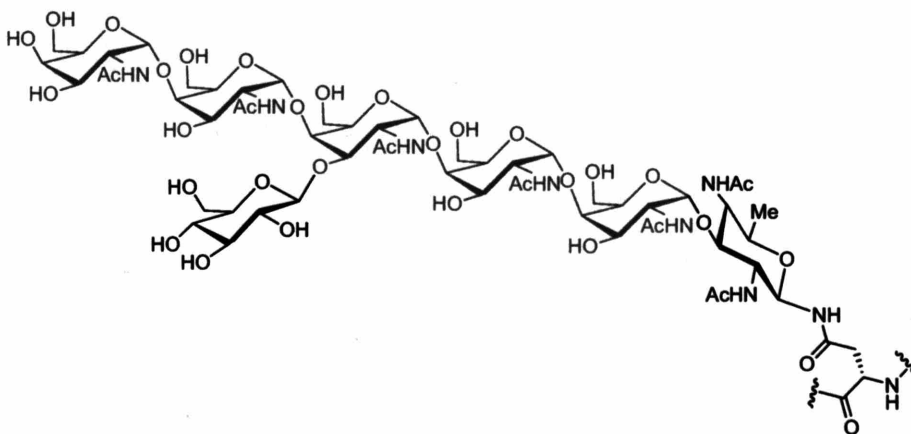
Figure 1-12. The glycan transferred to protein in *C. jejuni N*-linked glycosylation.

The *C. jejuni* heptasaccharide is structurally very different from the tetradecasaccharide

transferred in eukaryotic *N*-linked glycosylation. Bacteria utilize a wide variety of amino- and

deoxy-sugars that are not found in eukaryotic systems.[83] This is exemplified by the *N*-linked

glycan in *C. jejuni* that incorporates bacillosamine, a diacetamido-trideoxy-sugar found in

several bacterial strains such as *Neisseria* and *Pseudomonas*. Chapter 2 of this thesis is dedicated

to the chemical synthesis of bacillosamine and various analogs that can be used to investigate the

Pgl pathway *in vitro*.[84]

The Pgl pathway

Computational analysis of the *pgl* gene cluster (Figure 1-13), suggested that the locus

encodes five putative glycosyltransferases (PglA, PglC, PglH, PglI and PglJ), three putative

integral membrane proteins (PglB, WlaB and PglG) and three enzymes involved in sugar

biosynthesis (PglD, PglE and PglF). The *wlaB* gene encodes a putative ABC transport protein

and the *galE* gene encodes a UDP-glucose 4-epimerase that converts UDP-glucose to UDP-

galactose.[81]
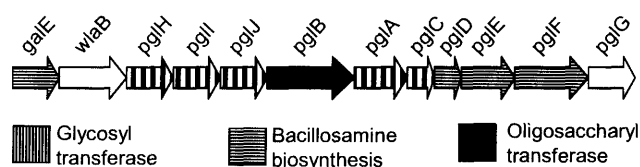
Figure 1-13. The *pgl* gene cluster from *Campylobacter jejuni*.

This *pgl* gene cluster in *C. jejuni* is very similar to a cluster found in the genome of

Nesseria meningitides that is known to be responsible for the *O*-linked glycosylation of pilin

(Figure 1-14). Pilin glycosylation involves an *O*-modified serine with a Gal-β1,3-Gal-α1,3-

DATDH modification (DATDH = 2,4-diacetamido-2,4,6-trideoxyhexose).[73] The stereochemistry of the DATDH sugar in pilin glycosylation has not been unambiguously determined, but is most likely to be bacillosamine. Bioinformatic analysis of the *pgl* gene cluster was greatly facilitated by the fact that several homologous genes in the *N. meningitides* cluster were already functionally annotated. Analogs of the sugar modifying enzymes, PglD, PglE and PglF are present in the *N. meningitides* cluster, and are attributed to the biosynthesis of bacillosamine. There is no homolog of the *pglB* gene in *N. meningitides* but a homologous protein to *pglC* (Nm *pglB*) is a glycosyltransferase responsible for transferring the first sugar phosphate onto a polyisoprene-phosphate carrier (Nm *pglB* codes for a bi-functional protein demonstrating both glycosyltransferase and acetyltransferase activity).[73] The *pglA* gene in *C. jejuni* is homologous to Nm *pglA*, which is responsible for the Gal-α1,3-Bac linkage. The other putative glycosyltransferase genes in *C. jejuni* are *pglH*, *pglI* and *pglJ*, but the bioinformatics data is insufficient to assign GalNAcα1,4- or Glcβ1,3-transferase function to these genes.[9]
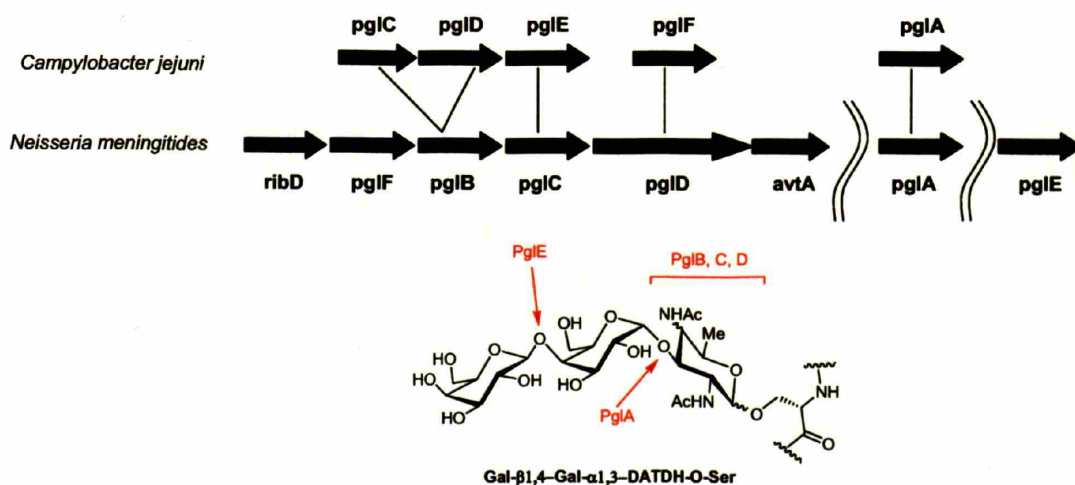
Figure 1-14. The pilin glycosylation locus of *Neisseria meningitides* and its comparison to the *Campylobacter jejuni pgl* gene cluster.

In critical work by Aebi and coworkers, the *pgl* gene cluster was functionally transferred to *E. coli* and a *C. jejuni* periplasmic protein, AcrA, was shown to be glycosylated in this modified *E. coli* system.[54] This suggests that the *pgl* cluster contains all of the genes necessary for the biosynthesis of the polyisoprenyl-pyrophosphate-linked heptasaccharide and its eventual transfer to protein. It is postulated that the prokaryotic oligosaccharide is constructed on a polyisoprenyl-pyrophosphate in a manner similar to the assembly of dolichyl-pyrophosphate linked oligosaccharide in eukaryotes. The polyisoprene used is undecaprenol (also known as bactoprenol), and contains 11 isoprene units, where the α-isoprene is unsaturated, in contrast to the α-saturated nature of dolichol.[54] Analysis of *Campylobacter* isolates using the SBA lectin, which binds GalNAc residues, resulted in the isolation of up to 38 proteins that were identified as possibly containing this *N*-linked glycan.[9] These glycoproteins are predominantly annotated as periplasmic proteins, which suggest that the glycosylation machinery is specific for periplasmic substrates. An AcrA mutant that lacks the periplasmic signal sequence is not glycosylated, further supporting the identification of the periplasm as the site of modification.[85]

Through mutational studies of the *pgl* gene cluster in *E. coli*, the exact roles of various *pgl* genes were explored using structural analysis of the glycan transferred to protein.[86] As predicted by bioinformatics, the *pglA*, *pglJ*, *pglH* and *pglI* genes were shown to encode specific glycosyltransferases responsible for sequential addition of monosaccharides to form the ultimate heptasaccharide donor. The *pglA* mutant, showed transfer of monosaccharide to protein, verifying the earlier observation that PglA transfers the α1,3-GalNAc to bacillosamine. The *pglJ* mutant showed transfer of disaccharide, suggesting that PglJ is responsible for the first α-1,4-GalNAc linkage to afford the trisaccharide. The *pglH* mutant, showed transfer of a trisaccharide to protein, suggesting a role for PglH in the transfer of the second α1,4-GalNAc sugar. Finally,

the *pglI* mutant showed transfer of a linear hexasaccharide, suggesting its role as a glucosyltransferase, adding the final branching glucose residue (Figure 1-15). While this study provided crucial information on the role of several Pgl glycosyltransferases, it did not provide information on the identity of the transferases responsible for the addition of the two terminal α-1,4-GalNAc residues. Therefore, the suggested scenarios were that PglH added all three terminal GalNAc residues or that PglH and PglJ acted alternately, adding two GalNAc residues each to form the hexasaccharide.

The role of each of these enzymes was unambiguously validated through *in vitro* biochemical analysis, using chemically-synthesized Und-PP-Bac and purified Pgl glycosyltransferases (Chapter 3).[84, 87] These data provided further evidence to support the bioinformatics and mutational analyses above and also demonstrated that PglH is a sugar polymerase, adding three GalNAc residues to the undecaprenyl-pyrophosphate-linked glycan. Reconstitution of the sequence of enzymatic steps *in vitro* also provided valuable insight into how these enzymes function together at a membrane interface.[87]
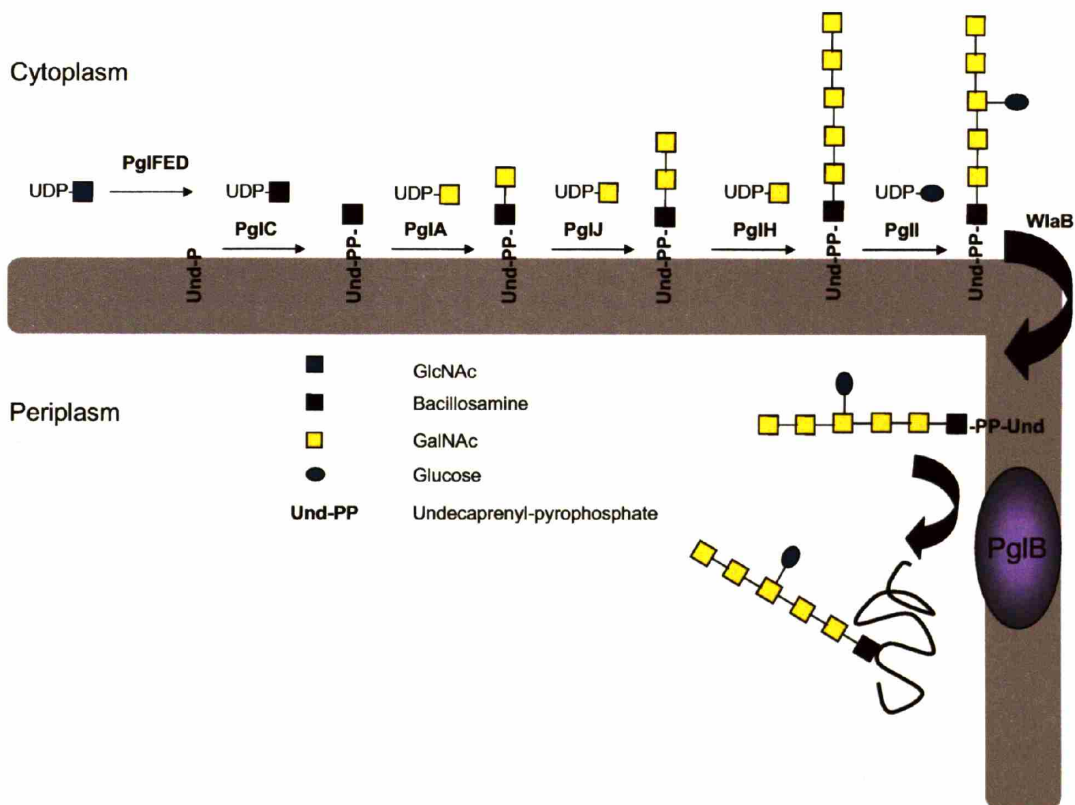
Figure 1-15. The Pgl pathway of *N*-linked glycosylation in *Campylobacter jejuni*.

The Pgl pathway shares striking similarities with the eukaryotic dolichol pathway. The oligosaccharide substrate for the oligosaccharyl transferase is built up sequentially on a polyisoprenyl-pyrophosphate carrier (undecaprenol in *C. jejuni*, dolichol in *S. cerevisiae*) by a series of glycosyltransferase that utilize nucleotide-activated sugar donors or dolichol-phosphate activated donors. This sequence of biosynthetic transformations occurs in the periplasmic membrane of *C. jejuni*, which is the functional equivalent of the ER membrane in yeast. Interestingly, in both systems, the glycan is built up to a heptasaccharide structure on the cytoplasmic face and then flipped to the other side of the membrane, either the ER lumen or the periplasm, by a flippase. The flippase in the *C. jejuni* system, WlaB, has been annotated to be an ABC (ATP-binding cassette) transporter, whereas the Rft1p protein in the dolichol pathway is

38

non-ATP driven. In the dolichol pathway, this heptasaccharide is further elaborated to the tetradecasaccharide, whereas in *C. jejuni*, no further elaboration occurs. Both pathways contain at least one enzyme that catalyzes the transfer of multiple glycans (PglH in *C. jejuni*, Alg9 in yeast). One striking difference is that the *ALG* genes in yeast encode highly hydrophobic proteins, which all include at least one transmembrane domain. Although the Pgl glycosyltransferases function on similar isoprene-bound intermediates, they contain no predicted transmembrane domains. This renders the Pgl enzymes more amenable to detailed biochemical analysis. Both pathways are examples of multistep enzymatic transformations that occur at a membrane interface. Studies devoted to understanding the interactions that occur amongst the Pgl glycosyltransferases at the membrane interface, can provide clues on how the corresponding Alg enzymes maintain the fidelity of the dolichol pathway.

PglB: The oligosaccharyl transferase of *C. jejuni*

The *PglB* gene shares significant homology with the *STT3* gene that codes for the largest subunit of the yeast, *S. cerevisiae*, oligosaccharyl transferase (OT) cluster.[51, 54] PglB contains a highly conserved amino acid motif WWDYGY that is present in all putative OT homologs. This conserved sequence is located on the hydrophilic C-terminal portion of PglB. In a *pglB* mutant strain, PEB3 and AcrA, both known glycoproteins from *C. jejuni*, were found to be unglycosylated.[9, 54] When functionally reconstituted in *E. coli*, the *pgl* cluster containing a mutation in the [457]WWDYGY[462] motif of PglB (W458A, D459A), resulted in unglycosylated protein.[54] This suggests the direct involvement of PglB in the glycosylation process whereby PglB facilitates the transfer of the heptasaccharide onto the side-chain of asparagine.

There are significant similarities between the Pgl pathway and the biosynthesis of the *O*-antigen lipopolysaccharide (LPS), where sequential addition of glycans results in an isoprenyl pyrophosphate-bound oligosaccharide that is transferred to the Lipid A core.[88] When the *O*-antigen ligase in *E. coli* was replaced with PglB, various *O*-antigen glycans were transferred to acceptor proteins (Figure 1-16).[89] This illustrates the substrate flexibility of PglB, which can accept a diverse array of undecaprenyl-linked oligosaccharide substrates.[89]
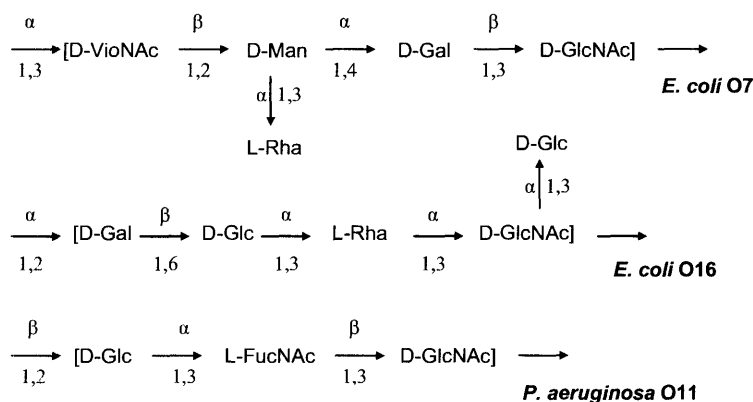
Figure 1-16. Diverse O-antigen glycans transferred to protein by PglB.

The OT cluster in *S. cerevisiae* and PglB in *C. jejuni*, both catalyze a similar reaction, the transfer of an oligosaccharide from a polyisoprenyl-pyrophosphate-linked glycan to an asparagine side chain. Yet, the *S. cerevisiae* system requires at least 8 proteins to efficiently catalyze this process, whereas the bacterial system appears to use only a single protein. In the yeast system, OT is required to interact with multiple other protein complexes such as the translocon and the signal peptidase, for efficient co-translational glycosylation. It is hypothesized that some of the OT subunits play a role in these interactions. A predominant area of research

currently is to purify PglB to homogeneity in order to discern if it is solely responsible for catalysis. Regardless, the simplicity of the PglB mediated process provides us a great opportunity to investigate the mechanism of this intriguing enzymatic reaction in more depth.

The OT complex in *S. cerevisiae* shows a high degree of specificity with regards to the dolichyl-pyrophosphate-linked glycan substrate and accepts very few truncated and non-native structures.[90] PglB on the other hand, appears to display much greater glycan flexibility by accepting various *O*-antigen structures as well as glycans of varying length and structure (Chapter 4). The glycans are however limited to those containing a C-2 *N*-acetamido-group on the proximal sugar, which suggests a role of the *N*-acetamido group in the catalytic mechanism of both OT and PglB. This substrate promiscuity of PglB suggests great promise for the potential of using the bacterial glycosylation system in engineering humanized glycoproteins.

Mechanism of PglB

All the glycosylated proteins identified in *C. jejuni*, were shown to contain the Asn-Xaa-Ser/Thr sequon.[81] Other proteins in the genome also contain this sequon but do not appear to be glycosylated. Hence, similar to the eukaryotic system, it appears that the N-X-S/T sequon is a necessary but not absolute determinant of glycosylation. The Thr-sequon in *C. jejuni* occurs at 43% of the sites.[9] Detailed investigation of the glycosylation sequon illustrated that similar to the eukaryotic process, proline is not accepted as the X amino acid, hence indicating the importance of peptide conformation in the glycosylation process.[85] *In vitro* studies using an *E. coli* cell membrane fraction expressing PglB, showed that similar to the yeast OT system, PglB can accept a truncated peptide substrate in place of a full length protein. Initial studies indicate that

the recognition sequence for PglB may require determinants in addition to the canonical tripeptide substrate for yeast OT (Chapter 4).

Inhibitors of PglB

The glycosyl modifications synthesized by the *pgl* genes are highly immunogenic.[6] Mutations in *pglB* and *pglE* resulted in a significant reduction in adherence to, and invasion of, INT407 cells *in vitro*, and a reduced ability to colonize the intestinal tract of mice, suggesting a role for these *N*-linked glycans in *Campylobacter* virulence.[91] Recently it was demonstrated that the *N*-linked glycans in *C. jejuni* plays a direct role in complex protein assembly. VirB10 is an *N*-linked glycoprotein that is present in the type IV secretion system (T4SS) of *C. jejuni*. Lack of VirB10 glycosylation results in *C. jejuni* cells containing a competence defect due to lack of protein complexation.[92] Interestingly, the closest homolog of the VirB10 glycoprotein is found in *Wolinella succinogenes*, which is the only other bacterium known to contain a putative *N*-linked glycosylation system similar to the *pgl* system.[92]

Due to the essential role played by the *C. jejuni* *N*-glycans in bacterial adherence and pathogenicity, PglB and the Pgl pathway as a whole appear to be interesting potential targets for antibacterial therapeutics. The extensive work done on the synthesis of inhibitors for the eukaryotic OT system[93] can now be applied toward the design of potent inhibitors of PglB. The periplasmic location of PglB also makes it a much more accessible target than the OT complex that is located in the ER lumen.

Conclusion

Due to the similarities between the dolichol (Alg) pathway and the Pgl pathway, as well as the parallels between the Stt3p-catalyzed glycosylation with the PglB reaction, the eukaryotic and prokaryotic systems are greatly intertwined. Our knowledge accumulated over decades of research devoted to understanding eukaryotic *N*-linked glycosylation can now be applied to the recently discovered prokaryotic system. Hopefully, the reduced complexity of the *C. jejuni* glycosylation process will allow for detailed biochemical and biophysical characterization that is currently virtually impossible with the eukaryotic system. Hence, the knowledge that can be gained from understanding the prokaryotic process will be invaluable in shedding light on the mechanism and function of the eukaryotic glycosylation system.

References

1. Spiro, R.G. Protein Glycosylation: Nature, Distribution, Enzymatic Formation, and Disease Implications of Glycopeptide Bonds. *Glycobiology*, **2002**, *12*, 43r-56r.

2. Varki, A. Biological Roles of Oligosaccharides - All of the Theories Are Correct. *Glycobiology*, **1993**, *3*, 97-130.

3. Upreti, R.K., Kumar, M., and Shankar, V. Bacterial Glycoproteins: Functions, Biosynthesis and Applications. *Proteomics*, **2003**, *3*, 363-379.

4. Imperiali, B. Protein Glycosylation: The Clash of the Titans. *Acc. Chem. Res.*, **1997**, *30*, 452-459.

5. Imperiali, B., O'Connor, S.E., Hendrickson, T., and Kellenberger, C. Chemistry and Biology of Asparagine-Linked Glycosylation. *Pure Appl. Chem.*, **1999**, *71*, 777-787.

6. Szymanski, C.M., Yao, R.J., Ewing, C.P., Trust, T.J., and Guerry, P. Evidence for a System of General Protein Glycosylation in Campylobacter Jejuni. *Mol. Microbiol.*, **1999**, *32*, 1022-1030.

7. Szymanski, C.M., and Wren, B.W. Protein Glycosylation in Bacterial Mucosal Pathogens. *Nat. Rev. Microbiol.*, **2005**, *3*, 225-237.

8. Burda, P., and Aebi, M. The Dolichol Pathway of N-Linked Glycosylation. *Biochim. Biophys. Acta*, **1999**, *1426*, 239-257.

9. Young, N.M., Brisson, J.R., Kelly, J., Watson, D.C., Tessier, L., Lanthier, P.H., Jarrell, H.C., Cadotte, N., Michael, F.S., Aberg, E., and Szymanski, C.M. Structure of the N-Linked Glycan Present on Multiple Glycoproteins in the Gram-Negative Bacterium, Campylobacter Jejuni. *J. Biol. Chem.*, **2002**, *277*, 42530-42539.

10. Imperiali, B., and Hendrickson, T.L. Asparagine-Linked Glycosylation: Specificity and Function of Oligosaccharyl Transferase. *Bioorg. Med. Chem.*, **1995**, *3*, 1565-1578.

11. Shan, S.O., and Walter, P. Co-Translational Protein Targeting by the Signal Recognition Particle. *FEBS Lett.*, **2005**, *579*, 921-926.

12. Egea, P.F., Tsuruta, H., Shan, S.O., Napetschnig, J., Savage, D.F., Walter, P., and Stroud, R.M. The Signal Recognition Particle and Structural Basis of Protein Targeting to Membranes. *Biophys. J.*, **2005**, *88*, 399a-399a.

13. Chavan, M., Yan, A.X., and Lennarz, W.J. Subunits of the Translocon Interact with Components of the Oligosaccharyl Transferase Complex. *J. Biol. Chem.*, **2005**, *280*, 22917-22924.

14.   Karamyshev, A.L., and Johnson, A.E. Protein Targeting to the Bacterial Translocon: Role of SecA and SRP. *Mol. Biol. Cell*, **2004**, *15*, 203a-203a.

15.   Karla, A., Lively, M.O., Paetzel, M., and Dalbey, R. The Identification of Residues That Control Signal Peptidase Cleavage Fidelity and Substrate Specificity. *J. Biol. Chem.*, **2005**, *280*, 6731-6741.

16.   Nilsson, I., and Vonheijne, G. Determination of the Distance between the Oligosaccharyltransferase Active-Site and the Endoplasmic-Reticulum Membrane. *J. Biol. Chem.*, **1993**, *268*, 5798-5801.

17.   Herscovics, A. Importance of Glycosidases in Mammalian Glycoprotein Biosynthesis. *Biochim. Biophys. Acta*, **1999**, *1473*, 96-107.

18.   Roth, J. Protein N-Glycosylation Along the Secretory Pathway: Relationship to Organelle Topography and Function, Protein Quality Control, and Cell Interactions. *Chem. Rev.*, **2002**, *102*, 285-303.

19.   Helenius, A., and Aebi, M. Roles of N-Linked Glycans in the Endoplasmic Reticulum. *Annu. Rev. Biochem.*, **2004**, *73*, 1019-1049.

20.   Helenius, A., and Aebi, M. Intracellular Functions of N-Linked Glycans. *Science*, **2001**, *291*, 2364-2369.

21.   Ellgaard, L., and Frickel, E.M. Calnexin, Calreticulin, and Erp57 - Teammates in Glycoprotein Folding. *Cell. Biochem. Biophys.*, **2003**, *39*, 223-247.

22.   Pollock, S., Kozlov, G., Pelletier, M.F., Trempe, J.F., Jansen, G., Sitnikov, D., Bergeron, J.J.M., Gehring, K., Ekiel, I., and Thomas, D.Y. Specific Interaction of Erp57 and Calnexin Determined by NMR Spectroscopy and an ER Two-Hybrid System. *EMBO J.*, **2004**, *23*, 1020-1029.

23.   Meusser, B., Hirsch, C., Jarosch, E., and Sommer, T. Erad: The Long Road to Destruction. *Nat. Cell Biol.*, **2005**, *7*, 766-772.

24.   Parodi, A.J. N-Glycosylation in Trypanosomatid Protozoa. *Glycobiology*, **1993**, *3*, 193-199.

25.   Schenk, B., Fernandez, F., and Waechter, C.J. The Ins(Ide) and Outs(Ide) of Dolichyl Phosphate Biosynthesis and Recycling in the Endoplasmic Reticulum. *Glycobiology*, **2001**, *11*, 61r-70r.

26.   Helenius, J., Ng, D.T.W., Marolda, C.L., Walter, P., Valvano, M.A., and Aebi, M. Translocation of Lipid-Linked Oligosaccharides across the ER Membrane Requires Rft1 Protein. *Nature*, **2002**, *415*, 447-450.

27. Kukuruzinska, M.A., and Robbins, P.W. Protein Glycosylation in Yeast - Transcript Heterogeneity of the Alg7 Gene. *Proc. Natl. Acad. Sci. U. S. A.*, **1987**, *84*, 2145-2149.

28. Heifetz, A., Keenan, R.W., and Elbein, A.D. Mechanism of Action of Tunicamycin on the UDP-GlcNAc-Dolichyl-Phosphate GlcNAc-1-Phosphate Transferase. *Biochemistry*, **1979**, *18*, 2186-2192.

29. Chantret, I., Dancourt, J., Barbat, A., and Moore, S.E.H. Two Proteins Homologous to the N- and C-Terminal Domains of the Bacterial Glycosyltransferase MurG Are Required for the Second Step of Dolichyl-Linked Oligosaccharide Synthesis in Saccharomyces Cerevisiae (Vol 280, Pg 9236, 2005). *J. Biol. Chem.*, **2005**, *280*, 18551-18552.

30. Gao, X.D., Tachikawa, H., Sato, T., Jigami, Y., and Dean, N. Alg14 Recruits Alg13 to the Cytoplasmic Face of the Endoplasmic Reticulum to Form a Novel Bipartite UDP-N-Acetylglucosamine Transferase Required for the Second Step of N-Linked Glycosylation. *J. Biol. Chem.*, **2005**, *280*, 36254-36262.

31. Bickel, T., Lehle, L., Schwarz, M., Aebi, M., and Jakob, C.A. Biosynthesis of Lipid-Linked Oligosaccharides in Saccharomyces Cerevisiae: Alg13p and Alg14p Form a Complex Required for the Formation of GlcNAc(2)-PP-Dolichol. *J. Biol. Chem.*, **2005**, *280*, 34500-34506.

32. Watt, G.M., Revers, L., Webberley, M.C., Wilson, I.B.H., and Flitsch, S.L. The Chemoenzymatic Synthesis of the Core Trisaccharide of N-Linked Oligosaccharides Using a Recombinant Beta-Mannosyltransferase. *Carbohydr. Res.*, **1997**, *305*, 533-541.

33. Wilson, I.B.H., Webberley, M.C., Revers, L., and Flitsch, S.L. Dolichol Is Not a Necessary Moiety for Lipid-Linked Oligosaccharide Substrates of the Mannosyltransferases Involved in in-Vitro N-Linked-Oligosaccharide Assembly. *Biochem. J.*, **1995**, *310*, 909-916.

34. Thiel, C., Schwarz, M., Peng, J.H., Grzmil, M., Hasilik, M., Braulke, T., Kohlschutter, A., von Figura, K., Lehle, L., and Korner, C. A New Type of Congenital Disorders of Glycosylation (CDG-II) Provides New Insights into the Early Steps of Dolichol-Linked Oligosaccharide Biosynthesis. *J. Biol. Chem.*, **2003**, *278*, 22498-22505.

35. Cipollo, J.F., Trimble, R.B., Chi, J.H., Yan, Q., and Dean, N. The Yeast Alg11 Gene Specifies Addition of the Terminal Alpha 1,2-Man to the Man(5)GlcNAc(2)-PP-Dolichol N-Glycosylation Intermediate Formed on the Cytosolic Side of the Endoplasmic Reticulum. *J. Biol. Chem.*, **2001**, *276*, 21828-21840.

36. Yamazaki, H., Shiraishi, N., Takeuchi, K., Ohnishi, Y., and Horinouchi, S. Characterization of Alg2 Encoding a Mannosyltransferase in the Zygomycete Fungus Rhizomucor Pusillus. *Gene*, **1998**, *221*, 179-184.

37.   Gao, X.D., Nishikawa, A., and Dean, N. Physical Interactions between the Alg1, Alg2, and Alg11 Mannosyltransferases of the Endoplasmic Reticulum. *Glycobiology*, **2004**, *14*, 559-570.

38.   Beck, P.J., Orlean, P., Albright, C., Robbins, P.W., Gething, M.J., and Sambrook, J.F. The Saccharomyces-Cerevisiae DPM1 Gene Encoding Dolichol-Phosphate-Mannose Synthase Is Able to Complement a Glycosylation-Defective Mammalian-Cell Line. *Mol. Cell. Biol.*, **1990**, *10*, 4612-4622.

39.   Teheesen, S., Lehle, L., Weissmann, A., and Aebi, M. Isolation of the Alg5 Locus Encoding the UDP-Glucose-Dolichyl-Phosphate Glucosyltransferase from Saccharomyces-Cerevisiae. *Eur. J. Biochem.*, **1994**, *224*, 71-79.

40.   Frank, C.G., and Aebi, M. Alg9 Mannosyltransferase Is Involved in Two Different Steps of Lipid-Linked Oligosaccharide Biosynthesis. *Glycobiology*, **2005**, *15*, 1156-1163.

41.   Dempski, R.E., Jr., and Imperiali, B. Oligosaccharyl Transferase: Gatekeeper to the Secretory Pathway. *Curr. Opin. Chem. Biol.*, **2002**, *6*, 844-850.

42.   Yan, Q., and Lennarz, W.J. Oligosaccharyltransferase: A Complex Multisubunit Enzyme of the Endoplasmic Reticulum. *Biochem. Bioph. Res. Commun.*, **1999**, *266*, 684-689.

43.   Heesen, S.T., Knauer, R., Lehle, L., and Aebi, M. Yeast Wbp1p and Swp1p Form a Protein Complex Essential for Oligosaccharyl Transferase-Activity. *EMBO J.*, **1993**, *12*, 279-284.

44.   Zufferey, R., Knauer, R., Burda, P., Stagljar, I., Heesen, S.T., Lehle, L., and Aebi, M. Stt3, a Highly Conserved Protein Required for Yeast Oligosaccharyl Transferase-Activity in-Vivo. *EMBO J.*, **1995**, *14*, 4949-4960.

45.   Yan, Q., and Lennarz, W.J. Structure-Function Studies on Ost1p, One of the Essential Subunits of Oligosaccharyl Transferase. *Glycobiology*, **2000**, *10*, 1105-1105.

46.   Dempski, R.E., and Imperiali, B. Heterologous Expression and Biophysical Characterization of Soluble Oligosaccharyl Transferase Subunits. *Arch. Biochem. Biophys.*, **2004**, *431*, 63-70.

47.   Yan, A., and Lennarz, W.J. Two Oligosaccharyl Transferase Complexes Exist in Yeast and Associate with Two Different Translocons. *Glycobiology*, **2005**.

48.   Spirig, U., Bodmer, D., Wacker, M., Burda, P., and Aebi, M. The 3.4 KDa Ost4 Protein Is Required for the Assembly of Two Distinct Oligosaccharyltransferase Complexes in Yeast. *Glycobiology*, **2005**.

49.   Li, G., Yan, Q., and Lennarz, W.J. Studies on the N-Glycosylation of Oligosaccharyl Transferase Subunits in Saccharomyces Cerevisiae. *Mol. Biol. Cell*, **2004**, *15*, 444a-444a.

50.   Yan, G., and Lennarz, W.J. Studies on the Function of Oligosaccharyl Transferase Subunits - Stt3p Is Directly Involved in the Glycosylation Process. *J. Biol. Chem.*, **2002**, *277*, 47692-47700.

51.   Yan, Q., and Lennarz, W.J. Studies on the Function of Stt3p, an Essential Subunit of Oligosaccharyl Transferase. *Glycobiology*, **2002**, *12*, 659-659.

52.   Nilsson, I., Kelleher, D.J., Miao, Y.W., Shao, Y.L., Kreibich, G., Gilmore, R., von Heijne, G., and Johnson, A.E. Photocross-Linking of Nascent Chains to the Stt3 Subunit of the Oligosaccharyltransferase Complex. *J. Cell Biol.*, **2003**, *161*, 715-725.

53.   Kim, H., von Heijne, G., and Nilsson, I. Membrane Topology of the Stt3 Subunit of the Oligosaccharyl Transferase Complex. *J. Biol. Chem.*, **2005**, *280*, 20261-20267.

54.   Wacker, M., Linton, D., Hitchen, P.G., Nita-Lazar, M., Haslam, S.M., North, S.J., Panico, M., Morris, H.R., Dell, A., Wren, B.W., and Aebi, M. N-Linked Glycosylation in Campylobacter Jejuni and Its Functional Transfer into E. Coli. *Science*, **2002**, *298*, 1790-1793.

55.   Bause, E. Structural Requirements of N-Glycosylation of Proteins. Studies with Proline Peptides as Conformational Probes. *Biochem. J.*, **1983**, *209*, 331-336.

56.   Marshall, R.D. Glycoproteins. *Annu. Rev. Biochem.*, **1972**, *41*, 673-702.

57.   Bause, E., and 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.*, **1981**, *195*, 639-644.

58.   Imperiali, B., Shannon, K.L., and Rickert, K.W. Role of Peptide Conformation in Asparagine-Linked Glycosylation. *J. Am. Chem. Soc.*, **1992**, *114*, 7942-7944.

59.   Imperiali, B., Shannon, K.L., Unno, M., and Rickert, K.W. A Mechanistic Proposal for Asparagine-Linked Glycosylation. *J. Am. Chem. Soc.*, **1992**, *114*, 7944-7945.

60.   Gavel, Y., and Vonheijne, G. Sequence Differences between Glycosylated and Nonglycosylated Asn-X-Thr Ser Acceptor Sites - Implications for Protein Engineering. *Protein Eng.*, **1990**, *3*, 433-442.

61.   Abbadi, A., Mcharfi, M., Aubry, A., Premilat, S., Boussard, G., and Marraud, M. Involvement of Side Functions in Peptide Structures - the Asx Turn - Occurrence and Conformational Aspects. *J. Am. Chem. Soc.*, **1991**, *113*, 2729-2735.

62.   OConnor, S.E., and Imperiali, B. Conformational Switching by Asparagine-Linked Glycosylation. *J. Am. Chem. Soc.*, **1997**, *119*, 2295-2296.

63. Imperiali, B., and O'Connor, S.E. Effect of N-Linked Glycosylation on Glycopeptide and Glycoprotein Structure. *Curr. Opin. Chem. Biol.* **1999**, *3*, 643-649.

64. Imperiali, B., Spencer, J.R., and Struthers, M.D. Structural and Functional-Characterization of a Constrained Asx-Turn Motif. *J. Am. Chem. Soc.*, **1994**, *116*, 8424-8425.

65. Bause, E., Breuer, W., and Peters, S. Investigation of the Active Site of Oligosaccharyltransferase from Pig Liver Using Synthetic Tripeptides as Tools. *Biochem. J.*, **1995**, *312*, 979-985.

66. Xu, T., Werner, R.M., Lee, K.C., Fettinger, J.C., Davis, J.T., and Coward, J.K. Synthesis and Evaluation of Tripeptides Containing Asparagine Analogues as Potential Substrates or Inhibitors of Oligosaccharyltransferase. *J. Org. Chem.*, **1998**, *63*, 4767-4778.

67. Ufret, M.D., and Imperiali, B. Probing the Extended Binding Determinants of Oligosaccharyl Transferase with Synthetic Inhibitors of Asparagine-Linked Glycosylation. *Bioorg. Med. Chem. Lett.*, **2000**, *10*, 281-284.

68. Weerapana, E., and Imperiali, B. Peptides to Peptidomimetics: Towards the Design and Synthesis of Bioavailable Inhibitors of Oligosaccharyl Transferase. *Org. Biomol. Chem.*, **2003**, *1*, 93-99.

69. Messner, P. Prokaryotic Glycoproteins: Unexplored but Important. *J. Bacteriol.*, **2004**, *186*, 2517-2519.

70. Mescher, M.F., and Strominger, J.L. Purification and Characterization of a Prokaryotic Glycoprotein from Cell-Envelope of Halobacterium-Salinarium. *J. Biol. Chem.*, **1976**, *251*, 2005-2014.

71. Schaffer, C., and Messner, P. Surface-Layer Glycoproteins: An Example for the Diversity of Bacterial Glycosylation with Promising Impacts on Nanobiotechnology. *Glycobiology*, **2004**, *14*, 31R-42R.

72. Messner, P. Bacterial Glycoproteins. *Glycoconjugate J.*, **1997**, *14*, 3-11.

73. Power, P.M., Roddam, L.F., Dieckelmann, M., Srikhanta, Y.N., Tan, Y.C., Berrington, A.W., and Jennings, M.P. Genetic Characterization of Pilin Glycosylation in Neisseria Meningitidis. *Microbiology*, **2000**, *146*, 967-979.

74. Szymanski, C.M., Logan, S.M., Linton, D., and Wren, B.W. Campylobacter--a Tale of Two Protein Glycosylation Systems. *Trends Microbiol.*, **2003**, *11*, 233-238.

75. Schirm, M., Soo, E.C., Aubry, A.J., Austin, J., Thibault, P., and Logan, S.M. Structural, Genetic and Functional Characterization of the Flagellin Glycosylation Process in Helicobacter Pylori. *Mol. Microbiol.*, **2003**, *48*, 1579-1592.

76. Thibault, P., Logan, S.M., Kelly, J.F., Brisson, J.R., Ewing, C.P., Trust, T.J., and Guerry, P. Identification of the Carbohydrate Moieties and Glycosylation Motifs in Campylobacter Jejuni Flagellin. *J. Biol. Chem.*, **2001**, *276*, 34862-34870.

77. Ketley, J.M. Pathogenesis of Enteric Infection by Campylobacter. *Microbiology*, **1997**, *143*, 5-21.

78. Taylor, D.E., Eaton, M., Yan, W., and Chang, N. Genome Maps of Campylobacter-Jejuni and Campylobacter-Coli. *J. Bacteriol.*, **1992**, *174*, 2332-2337.

79. Karlyshev, A.V., Henderson, J., Ketley, J.M., and Wren, B.W. An Improved Physical and Genetic Map of Campylobacter Jejuni NCTC 11168 (UA580). *Microbiology*, **1998**, *144*, 503-508.

80. Parkhill, J., Wren, B.W., Mungall, K., Ketley, J.M., Churcher, C., Basham, D., Chillingworth, T., Davies, R.M., Feltwell, T., Holroyd, S., Jagels, K., Karlyshev, A.V., Moule, S., Pallen, M.J., Penn, C.W., Quail, M.A., Rajandream, M.A., Rutherford, K.M., van Vliet, A.H.M., Whitehead, S., and Barrell, B.G. The Genome Sequence of the Food-Borne Pathogen Campylobacter Jejuni Reveals Hypervariable Sequences. *Nature*, **2000**, *403*, 665-668.

81. Linton, D., Allan, E., Karlyshev, A.V., Cronshaw, A.D., and Wren, B.W. Identification of N-Acetylgalactosamine-Containing Glycoproteins PEB3 and CgpA in Campylobacter Jejuni. *Mol. Microbiol.*, **2002**, *43*, 497-508.

82. Szymanski, C.M., St Michael, F., Jarrell, H.C., Li, J.J., Gilbert, M., Larocque, S., Vinogradov, E., and Brisson, J.R. Detection of Conserved N-Linked Glycans and Phase-Variable Lipooligosaccharides and Capsules from Campylobacter Cells by Mass Spectrometry and High Resolution Magic Angle Spinning NMR Spectroscopy. *J. Biol. Chem.*, **2003**, *278*, 24509-24520.

83. Maki, M., and Renkonen, R. Biosynthesis of 6-Deoxyhexose Glycans in Bacteria. *Glycobiology*, **2004**, *14*, 1r-15r.

84. Weerapana, E., Glover, K.J., Chen, M.M., and Imperiali, B. Investigating Bacterial N-Linked Glycosylation: Synthesis and Glycosyl Acceptor Activity of the Undecaprenyl Pyrophosphate-Linked Bacillosamine. *J. Am. Chem. Soc.*, **2005**, *127*, 13766-13767.

85. Nita-Lazar, M., Wacker, M., Schegg, B., Amber, S., and Aebi, M. The N-X-S/T Consensus Sequence Is Required but Not Sufficient for Bacterial N-Linked Protein Glycosylation. *Glycobiology*, **2005**, *15*, 361-367.

86. Linton, D., Dorrell, N., Hitchen, P.G., Amber, S., Karlyshev, A.V., Morris, H.R., Dell, A., Valvano, M.A., Aebi, M., and Wren, B.W. Functional Analysis of the Campylobacter Jejuni N-Linked Protein Glycosylation Pathway. *Mol. Microbiol.*, **2005**, *55*, 1695-1703.

87. Glover, K.J., Weerapana, E., and Imperiali, B. In Vitro Assembly of the Undecaprenylpyrophosphate-Linked Heptasaccharide for Prokaryotic N-Linked Glycosylation. *Proc. Natl. Acad. Sci. U S A*, **2005**, *102*, 14255-14259.

88. Raetz, C.R.H., and Whitfield, C. Lipopolysaccharide Endotoxins. *Annu. Rev. Biochem.*, **2002**, *71*, 635-700.

89. Feldman, M.F., Wacker, M., Hernandez, M., Hitchen, P.G., Marolda, C.L., Kowarik, M., Morris, H.R., Dell, A., Valvano, M.A., and Aebi, M. Engineering N-Linked Protein Glycosylation with Diverse O Antigen Lipopolysaccharide Structures in Escherichia Coli. *Proc. Natl. Acad. Sci. U S A*, **2005**, *102*, 3016-3021.

90. Tai, V.W., and Imperiali, B. Substrate Specificity of the Glycosyl Donor for Oligosaccharyl Transferase. *J. Org. Chem.*, **2001**, *66*, 6217-6228.

91. Szymanski, C.M., Burr, D.H., and Guerry, P. Campylobacter Protein Glycosylation Affects Host Cell Interactions. *Infect. Immun.*, **2002**, *70*, 2242-2244.

92. Larsen, J.C., Szymanski, C., and Guerry, P. N-Linked Protein Glycosylation Is Required for Full Competence in Campylobacter Jejuni 81-176. *J. Bacteriol.*, **2004**, *186*, 6508-6514.

93. Kellenberger, C., Hendrickson, T.L., and Imperiali, B. Structural and Functional Analysis of Peptidyl Oligosaccharyl Transferase Inhibitors. *Biochemistry*, **1997**, *36*, 12554-12559.

Chapter 2

Synthesis of bacillosamine and its derivatives for investigating the enzymes involved in *N*-linked

glycosylation in *Campylobacter jejuni*

Mark Chen scaled-up the synthesis of bacillosamine and provided NMR characterization of several intermediates.

Introduction

The process of *N*-linked glycosylation in *Campylobacter jejuni* displays significant similarities to the analogous process in eukaryotes.[1-3] A series of glycosyltransferases act sequentially to assemble a glycan donor on a polyisoprenyl carrier (undecaprenyl pyrophosphate (Und-PP)), which is transferred to the asparagine side chain of proteins at the Asn-Xaa-Ser/Thr (N-X-S/T) consensus sequence.[4] In *C. jejuni*, the glycan donor has been identified as the Und-PP linked heptasaccharide GalNAc-α1,4- GalNAc-α1,4-(Glcβ1,3)-GalNAc-α1,4-GalNAc-α1,4-GalNAc-α1,3-Bac-α1,PP-Und (**1**) (Figure 2-1), where Bac is the unusual sugar bacillosamine (2,4-diacetamido-2,4,6-trideoxyglucose) that is only found in specific bacterial systems.[5]
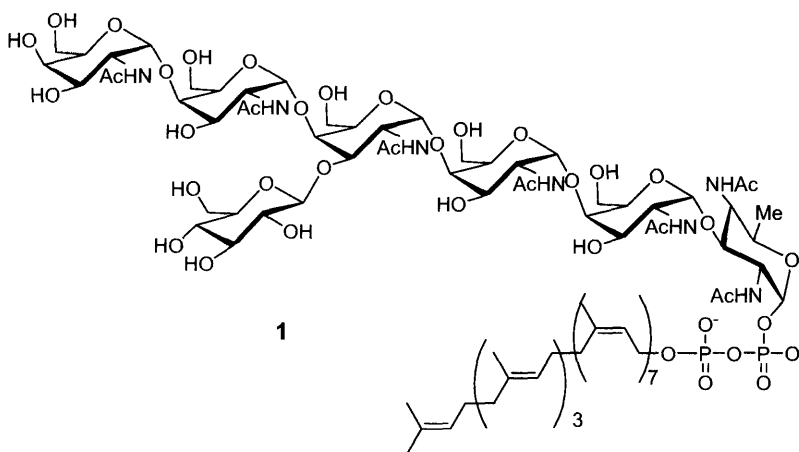
Figure 2-1. The undecaprenyl-pyrophosphate-linked heptasaccharide donor.

Bioinformatic and mutagenesis data suggest that the *pgl* gene cluster contains all of the enzymes required for the biosynthesis of the polyisoprenyl-linked heptasaccharide and its ultimate transfer to protein.[4, 6, 7] The proposed biosynthetic route to the heptasaccharide substrate begins in the cytoplasmic side of the periplasmic membrane with a UDP-HexNAc (either UDP-

GlcNAc or UDP-GalNAc) that is converted to UDP-Bacillosamine (UDP-Bac) by the action of three enzymes Pgl F, E and D, which code for a dehydratase, an aminotransferase and an acetyltransferase, respectively. PglC, a glycophosphoryltransferase then transfers the bacillosamine from UDP-Bac to Und-P to form Und-PP-Bac, which is the first membrane associated substrate in the pathway. This substrate is then elaborated to the heptasaccharide by the action of four glycosyltransferases, Pgl A, H, J and I and then flipped to the periplasm where it is utilized by PglB, which is the oligosaccharyl transferase of the bacterial system (Figure 2-2).[8, 9]
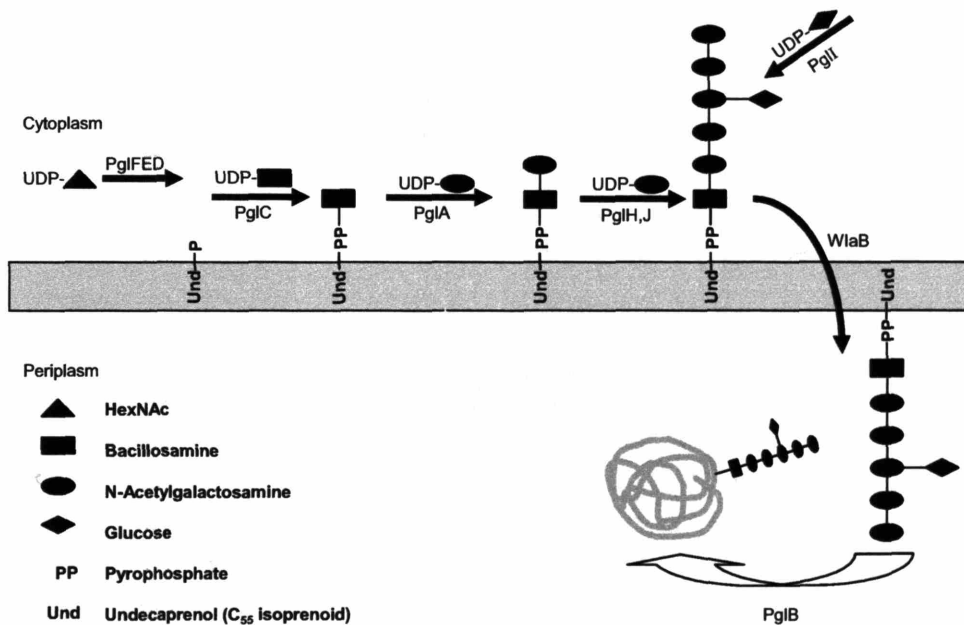
Figure 2-2. The Pgl pathway.

Our goal is to investigate the activity of the Pgl enzymes *in vitro* to validate their proposed activity and study their substrate specificity. Validation of enzyme activity *in vitro* involves the overexpression and purification of the Pgl proteins and the use of chemically-

54

defined substrates. The substrates for the Pgl enzymes all include bacillosamine, an unusual carbohydrate that is only found in specific bacterial systems. These compounds are only present in very small quantities in the native *C. jejuni* system, making purification from the natural source an unrealistic endeavor. Furthermore, extraction from *C. jejuni* is complicated by the fact that it is an extremely pathogenic strain of bacteria that requires specialized handling. Chemical synthesis provides an alternative to isolating the substrates from the natural source and allows access to milligram quantities of the desired substrates and other related derivatives in extremely pure form.

The substrates for the Pgl enzymes all incorporate bacillosamine either as the nucleotide diphosphate or polyisoprenyl-pyrophosphate forms. In order to study the activity of the glycophosphoryltransferase PglC, the nucleotide linked UDP-Bac substrate is required. To extend the study to the PglA, H, J and I glycosyltransferases, highly pure undecaprenyl pyrophosphate linked bacillosamine (Und-PP-Bac) is essential. Employing chemical synthesis, we can access the desired milligram quantities of UDP-Bac and Und-PP-Bac for biochemical studies of the glycosyltransferases. Here we report the synthesis of a protected bacillosamine-phosphate derivative (Bac-P) (**2**) and its subsequent coupling to UMP-morpholidate to form UDP-Bac (**3**) and to undecaprenyl phosphate to form Und-PP-Bac (**4**), both of which are key intermediates in the Pgl pathway of *N*-linked glycosylation (Figure 2-3).
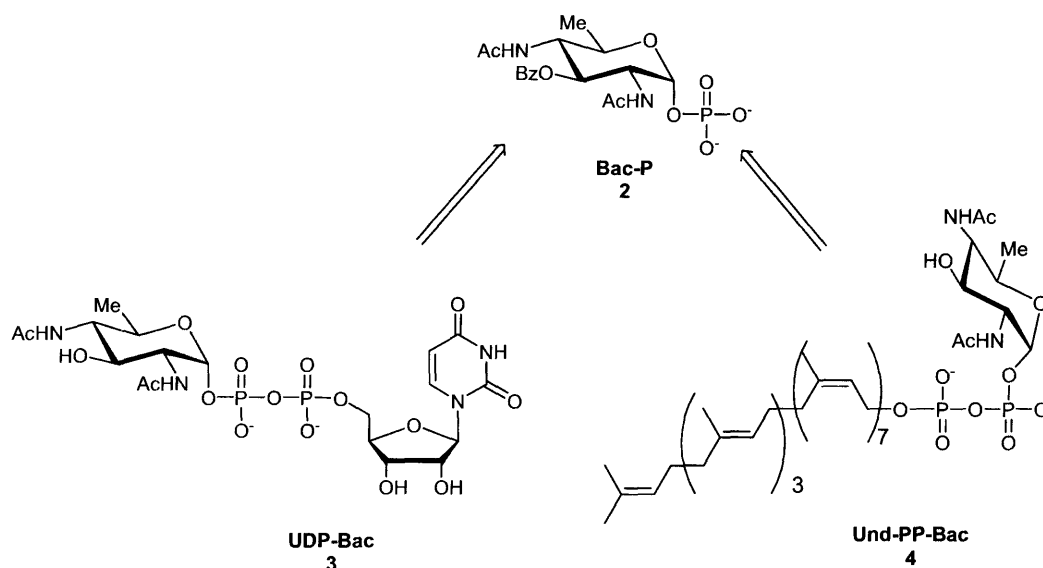
Figure 2-3. Retrosynthetic analysis of substrates for the Pgl enzymes.

Synthesis of Uridine-diphosphobacillosamine (UDP-Bac) (**3**) and undecaprenyl-pyrophosphate bacillosamine (Und-PP-Bac) (**4**) from 3-*O*-benzoyl bacillosamine-α-1-phosphate (Bac-P) (**2**).

Results and Discussion

Synthesis of 3-*O*-benzoyl-bacillosamine-α-1-phosphate (Bac-P) (2)

Several nucleotide linked derivatives of 6-deoxyhexose glycans similar to bacillosamine have been synthesized using enzyme-based approaches,[10-12] however the biosynthetic enzymes responsible for the synthesis of bacillosamine are yet to be validated and utilized in a preparative manner. The first chemical synthesis of bacillosamine was reported by Liav *et al.* from 2-acetamido-2-deoxy-glucose in a >15 step procedure.[13] Bundle *et al.* also reported a synthesis of a derivative of bacillosamine, methyl 2-acetamido-4-amino-2,4,6-trideoxy-β-D-glucopyranoside using a novel chlorosulfation procedure.[14] Other similar syntheses of 2,4-diacetamido-2,4,6-

trideoxy sugars have also been published[15-19], yet none of these synthetic routes are easily adapted towards the synthesis of bacillosamine phosphate. Here we describe the first chemical synthesis of the α-phosphate of bacillosamine at the anomeric center and its application to the synthesis of UDP-Bac and Und-PP-Bac, which are intermediates in the Pgl pathway (Scheme 2-1).

Our synthesis of bacillosamine began with benzyl 2-acetamido-2-deoxy-β-D-galactopyranoside (**5**), which was prepared from commercially available D-galactosamine hydrochloride using the method of Matta *et al.*[20] Briefly, this procedure involved selectively accessing the β-anomeric benzyl configuration *via* the oxazoline. Intermediate **5** was then modified via a selective benzoylation to produce benzyl 2-acetamido-3,6-di-*O*-benzoyl-2-deoxy-β-D-galactopyranoside (**6**) using 2.2 mol equivalents of benzoyl chloride in pyridine at -60 °C. Inversion of configuration at C-4 was achieved by triflation followed by subsequent displacement with sodium azide resulting in **7**. Debenzoylation of **7** with sodium methoxide yielded benzyl 2-acetamido-4-azido-2,4-deoxy-β-D-glucopyranoside (**8**). Selective tosylation of the resulting primary hydroxyl afforded **9**, which was followed by substitution of the tosyloxy group with iodide to afford the 6-iodo derivative **10**. Simultaneous reduction of the azido group and de-iodination occurred by catalytic hydrogenation on Pd(OH)$_2$/C. The presence of *N, N*-diisopropylethylamine in the hydrogenation reaction ensured that this reduction occurred without affecting the anomeric benzyl protecting group. Subsequent acetylation of both the amino and hydroxyl functionalities resulted in benzyl 2-4-acetamido-3-*O*-acetyl-2,4,6-trideoxyglucose (**11**).
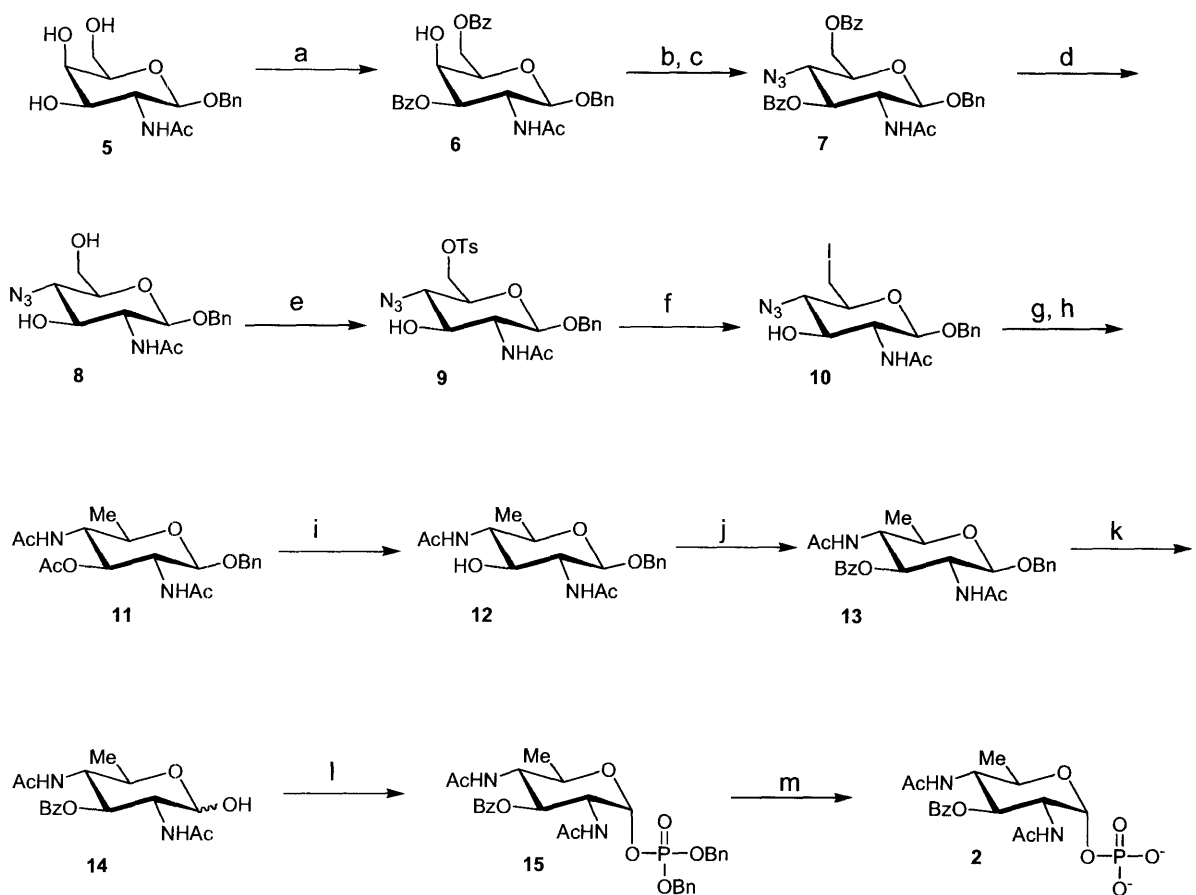
Scheme 2-1. Synthetic route to 3-*O*-benzoyl bacillosamine-α-1-phosphate (**2**).[a]

[a]Reagents and conditions: (a) benzoyl chloride, pyridine, -40 °C→0 °C, 70%; (b) Tf₂O, CH₂Cl₂/pyridine, 0 °C→rt; (c) NaN₃, DMF, rt, 62% over 2 steps; (d) NaOMe, MeOH, rt, 80%; (e) TsCl, pyridine, 0 °C→rt, 75%; (f) NaI, MeCN, 80 °C, 70%; (g) H₂, Pd(OH)₂/C, DIPEA, MeOH, 32 °C; (h) Ac₂O, pyridine, rt, 65% over 2 steps; (i) NaOMe, MeOH, 90%; (j) BzCl, pyridine, 60%; (k) H₂, Pd(OH)₂/C, MeOH, 32 °C, 86%; (l) LiHMDS, -68 °C; then [(BnO)₂P(O)]₂O, -68 °C→0 °C, 50%; (m) H₂, Pd/C, MeOH, 99%.

Deprotection of the anomeric benzyl group in **11** by hydrogenation resulted in a 1:1 α:β

mixture of 2-4-diacetamido-3-*O*-acetyl-2,4,6-trideoxyglucopyranoside. Attempts to

phosphorylate this compound using tetrabenzylpyrophosphate, resulted in very low yields due to

poor solubility. To circumvent this issue, the C-3 protection was changed to a benzoyl group to

afford **13**. This was achieved by sodium methoxide treatment of **11** to form intermediate **12**,

followed by treatment with benzoyl chloride at 0 °C to yield 2,4-diacetamido-3-O-benzoyl-2,4,6-

trideoxyglycopyranoside (**13**). Catalytic hydrogenation in the presence of $Pd(OH)_2/C$ resulted in

14 as an anomeric mixture. Phosphorylation of the more soluble benzoyl protected analog was

successfully achieved in 50% yield. This phosphorylation reaction proceeds with >16:1

selectivity in favor of the α-phosphate **15**. The benzyl protection on the phosphate was removed

by hydrogenation to yield the phosphorylated bacillosamine derivative **2**. A minor (<15%)

impurity in the product is observed, and this impurity cannot be separated by chromatography.

This minor product may result from epimerization at the anomeric center.

Synthesis of uridine diphospho-bacillosamine (UDP-Bac) (3)

PglC is the membrane-bound enzyme that facilitates the transfer of bacillosamine

phosphate from UDP-Bac to undecaprenyl phosphate to form Und-PP-Bac in *C. jejuni*.[8, 9] In

order to validate the activity of PglC *in vitro*, it is necessary to access the UDP-Bac substrate.

UDP-Bac was chemically synthesized via the coupling of the benzoyl-protected bacillosamine

phosphate derivative (Bac-P, **2**) to UMP-morpholidate. This coupling was performed in dry

pyridine in the presence of *1H*-tetrazole to yield the protected UDP-Bac derivative (**16**).[21, 22] The

benzoyl protection on this intermediate was then removed with sodium methoxide to afford the

fully deprotected UDP-Bac (**3**) substrate for PglC (Scheme 2-2).

59
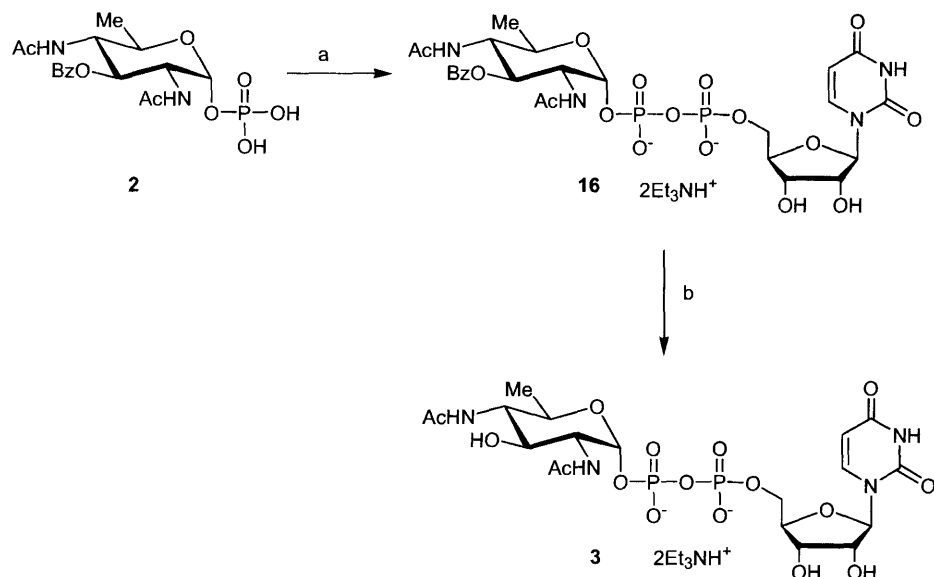
Scheme 2-2. Synthesis of UDP-Bac (**3**).[a]

[a]Reagents and conditions: (a) 4-morpholine-*N-N'*-dicyclohexylcarboxamidinium uridine 5'-monophosphormorpholidate; *1H*-tetrazole, pyridine, rt, 3 days, 50%, (b) NaOMe, MeOH, rt, 68%.

Synthesis of undecaprenyl-pyrophosphate-linked bacillosamine (Und-PP-Bac) (4)

The glycosyltransferases in the Pgl pathway function on polyisoprenyl-pyrophosphate-linked glycan substrates that are anchored to the membrane. Similar to the dolichol pathway of eukaryotic glycosylation, the glycans are bound to the isoprene via a pyrophosphate linkage and are elaborated on the isoprene by a series of glycosyltransferases. Undecaprenol is an abundant polyisoprene found in bacterial systems and consists of 11 isoprene units. In *C. jejuni*, the glycosyltranferases, Pgl A, J, H and I, all function on undecaprenyl pyrophosphate linked glycan substrates. In order to further investigate the activities of these glycosyltransferases, it is

necessary to access milligram quantities of the undecaprenyl-linked substrates. We chose to access Und-PP-Bac (**4**), which is the first polyisoprenyl-linked substrate in the pathway, and the substrate for the glycosyltransferase, PglA by chemical synthesis. Using this substrate, investigations into PglA activity as well as the later glycosyltransferases can be undertaken.

Und-PP-Bac is synthesized via the coupling of Bac-P to undecaprenyl phosphate (Und-P), which was synthesized from undecaprenol using phosphoramidite chemistry as previously reported.[23, 24] The coupling to Bac-P was afforded using 1,1'-carbonyldiimidazole (CDI) as a coupling reagent to yield the protected Und-PP-Bac derivative (**17**) (Scheme 2-3).[24, 25] In the final step, the C-3 benzoyl protection was removed with sodium methoxide to yield the final target molecule, Und-PP-Bac (**4**).
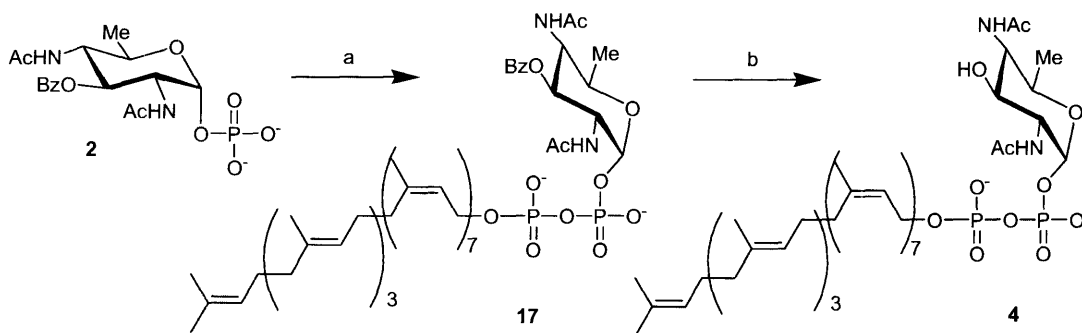
Scheme 2-3. Synthesis of Und-PP-Bac (**4**).[a]

[a]Reagents and conditions: (a) carbonyl diimidazole, DMF; then Und-P, 50%; (b) NaOMe, MeOH, 99%.

Exploring the glycan specificity of the Pgl enzymes

Since the synthesis of bacillosamine involves a multi-step reaction sequence, we were interested to determine if the Pgl enzymes would accept different glycans in place of

bacillosamine. If the enzymes recognized a synthetically simpler glycan substrate, it would greatly facilitate future studies into this system. Furthermore, studying the glycan specificity would also provide valuable insight into the interactions occurring at the enzyme active sites. In order to investigate the substrate specificity of the Pgl enzymes, two other undecaprenyl pyrophosphate linked glycans were synthesized. These were Und-PP-6-hydroxybacillosamine (Bac-6-OH), which differs from bacillosamine by the presence of a C-6 hydroxyl moiety (18) and Und-PP-GlcNAc, which additionally lacks the C-4 N-acetyl group (19) (Figure 2-3). GlcNAc is an abundant glycan present in most bacterial systems hence we were interested in studying the specificity of the Pgl enzymes for bacillosamine over the more abundant GlcNAc. The Bac-6-OH derivative is an intermediate that is accessed in fewer synthetic steps than bacillosamine, and could act as a useful Bac substitute in future studies. Moreover, this derivative would provide valuable information regarding the importance of the 4-acetamido group over the 6-deoxy moiety.
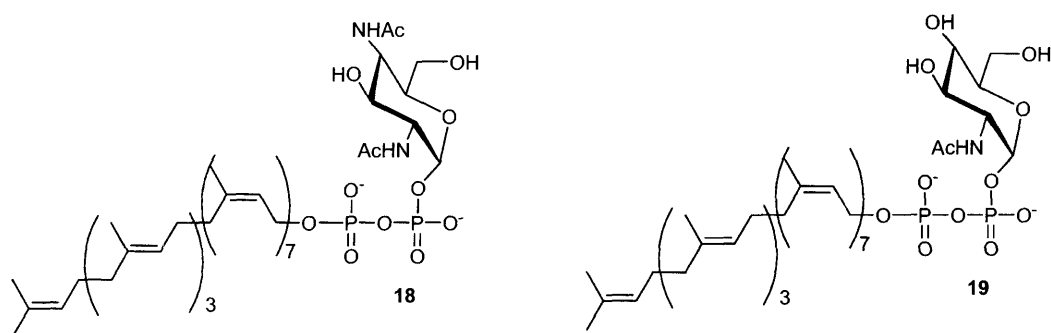
Figure 2-4. Substrates to explore the glycan specificity.

Und-PP-6-hydroxybacillosamine (Und-PP-Bac-6-OH) (18) and Und-PP-GlcNAc (19).

Compounds **18** and **19** were synthesized using a procedure similar to that for the synthesis of Und-PP-Bac (**4**). The synthesis involved the activation of the corresponding sugar phosphate (Bac-6-OH-P (**23**) or GlcNAc-P) with carbonyl diimidazole (CDI) followed by coupling to Und-P to form the pyrophosphate linkage. GlcNAc-P was synthesized according to a previous procedure[25] while Bac-6-OH-P (**23**) was synthesized from an intermediate in the synthesis of Bac-P. Intermediate **7** in the synthesis of Bac-P (**2**) (Scheme 2-1) was elaborated according to Scheme 2-4. A catalytic hydrogenation of **7** with Pd(OH)$_2$/C in the presence of diisopropylethylamine resulted in the reduction of the C-4 azido group without affecting the anomeric benzyl group. This intermediate was subsequently acetylated with acetic anhydride/pyridine to yield **20**. The anomeric benzyl protection was removed with catalytic hydrogenation to afford **21** and phosphorylated with lithium hexamethyldisilazide treatment followed by tetrabenzylpyrophosphate to yield the α-anomeric phosphate **22**. The benzyl protection on the phosphate was removed with catalytic hydrogenation using Pd/C to afford Bac-6-OH-P (**23**), which was coupled to Und-P without further purification.
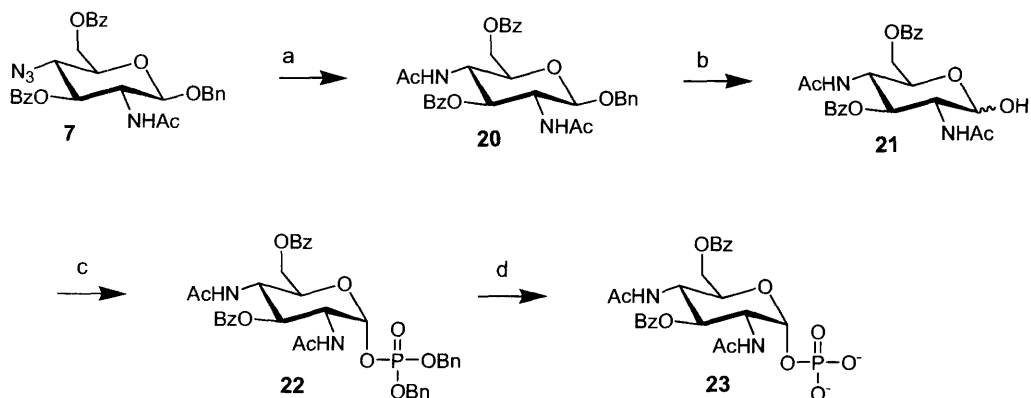
Scheme 2-4. Synthesis of 6-hydroxy-bacillosamine phosphate (**23**).[a]

[a] Reagents and conditions: (a) (i) H_2, Pd(OH)$_2$/C, DIPEA, MeOH, 32 °C; (ii) Ac$_2$O, pyridine, rt, 70% over 2 steps; (b) H_2, Pd(OH)$_2$/C, MeOH, 32 °C, 79%; (c) (i). LiHMDS, -68 °C; (ii). [(BnO)$_2$P(O)]$_2$O, -68 °C→0 °C, 65%; (d) H_2, Pd/C, MeOH, 99%.

Exploring the polyisoprene specificity of the Pgl enzymes

Some of the difficulties associated with studying the glycosyltransferases in the Pgl pathway are related to problems with obtaining and handling the hydrophobic undecaprenol. The C55 (11 isoprene units) carbon chain tends to aggregate and assays using undecaprenyl-pyrophosphate linked substrates require organic co-solvents, detergents and other additives. Undecaprenol is also a very expensive starting material ($250 for 5mg). For these reasons, it would be advantageous to identify a more convenient substrate for the Pgl enzymes. For example, if the enzymes accept polyisoprene-pyrophosphate-linked intermediates that are shorter and/or easier to access, studying systems such as these would be greatly facilitated.

In order to investigate the polyisoprene specificity of the Pgl enzymes, several different polyisoprene-pyrophosphate-linked analogs were synthesized (Figure 2-4). Commercially

available geranylgeraniol was used to afford geranylgeranyl-pyrophosphate-Bac-6-OH (24). This substrate includes a shorter chain isoprene than the native undecaprenol and would provide insight into the chain length specificity of the Pgl enzymes. Geranylgeraniol comprises 4 isoprene units (C20) and is far more readily available compared to its undecaprenol counterpart. The other substrate that was synthesized was the dolichyl-pyrophosphate-Bac-6-OH (25). Dolichol is a longer polyisoprene that comprises 16-18 isoprene units. Unlike undecaprenol and geranylgeraniol, which are fully unsaturated polyisoprenols, dolichol contains a saturated isoprene unit at the hydroxyl-terminus. This Dol-PP-Bac-6-OH derivative would provide insight into the ability of the Pgl enzymes to accept longer chain lengths as well as the effect of a terminal saturated isoprene unit.
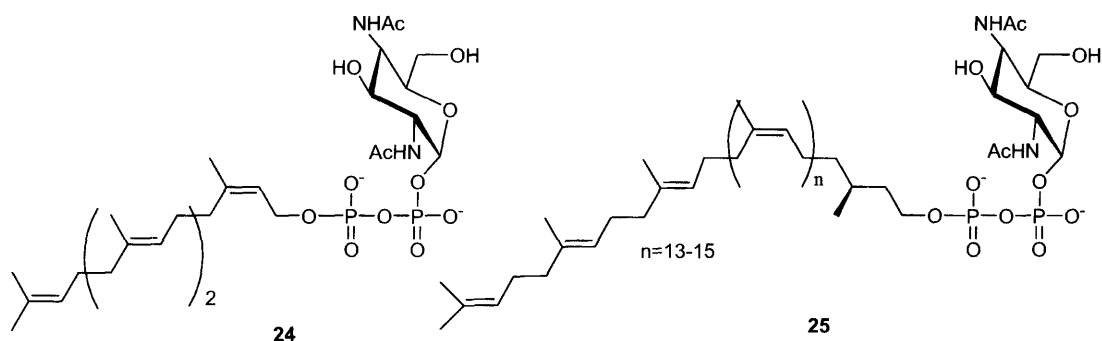
Figure 2-5. Substrates to explore the polyisoprene specificity.

Geranylgeranylpyrophosphate-Bac-6-OH (24) and Dolichylpyrophosphate-Bac-6-OH (25).

Conclusion

Bacillosamine is the first sugar found in the heptasaccharide moiety that is transferred to the asparagine side chain of proteins in *C. jejuni*. Studies into the activities of the enzymes involved in this *N*-linked glycosylation pathway have been hindered by the lack of substrates, since bacillosamine and its derivatives are difficult to isolate from natural sources. For this reason, it is important to devise a synthetic route that allows access to bacillosamine and its derivatives. In this chapter, the first chemical synthesis of bacillosamine phosphate is outlined. This crucial intermediate is then incorporated into two of the critical substrates for the Pgl enzymes, UDP-Bac and Und-PP-Bac. A bacillosamine derivative, Bac-6-OH (containing a C-6 hydroxyl group), is also synthesized and this analog, together with the more readily available GlcNAc is used to investigate the specificity of the Pgl enzymes. Strategies to link these different sugar phosphates to polyisoprene phosphates are discussed in the synthesis of undecaprenyl, geranylgeranyl and dolichol-linked glycan substrates, which can be used to explore the specificity of the Pgl enzymes towards the polyisoprene. Altogether, a plethora of native and non-native substrates for the enzymes in the Pgl pathway have been synthesized and can now be used to shed light on the elusive process of *N*-linked glycosylation in *C. jejuni*.

Acknowledgements

Experimental

General Procedures

All chemicals were purchased from Sigma/Aldrich unless otherwise noted. Undecaprenol and ^{3}H-UDP-GalNAc were purchased from American Radiolabeled Chemicals Inc. Analytical thin layer chromatography (TLC) was performed on Merck silica gel 60 F_{254} plates (0.25 mm). All compounds were visualized on TLC by UV irradiation or cerium sulfate-ammonium molybdate staining. Flash column chromatography was carried out using forced flow of the indicated solvent on AdTech Flash Silica Gel, 32-63 μm particle size, 60 Å pore size (Adedge technologies). Proton, carbon and phosphorous NMR spectra were recorded on a Bruker Avance 400 NMR Spectrometer (400 MHz). Chemical shifts (δ) are reported in parts per million (ppm) with chemical shifts referenced to internal standards: CDCl$_3$ (7.26 ppm for ^{1}H, 77.0 ppm for ^{13}C), CD$_3$OD (4.78 ppm for ^{1}H, 49.15 ppm for ^{13}C). ^{31}P spectra are reported in δ values relative to H$_3$PO$_4$ (0.0 ppm). Coupling constants (J) are reported in Hertz (Hz) and multiplicities are abbreviated as singlet (s), doublet (d), triplet (t), multiplet (m), broadened singlet (brs) and doublet of doublets (dd). The term apparent triplet (t$_{app}$) is used to denote a doublet of doublets (dd) with two similar coupling constants and apparent quartet (q$_{app}$) is used for a doublet of doublet of doublets (ddd) with three similar coupling constants. High resolution Mass Spectra (HRMS) were obtained at the Mass Spectrometry Facility at MIT (Cambridge, MA). The negative ion ESI mass spectrum of undecaprenyl-pyrophosphate bacillosamine was obtained at the Mass Spectrometry Facility, University of Illinois at Urbana-Champaign.

Synthesis of benzyl 2-acetamido-3,6-di-*O*-benzoyl-2-deoxy-β-D-galactopyranoside

6
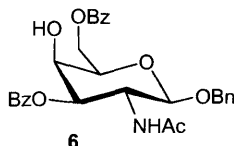
Benzyl 2-acetamido-2-deoxy-β-D-galactopyranoside (**5**) (1.8 g, 5.7 mmol, 1.0 eq) was dissolved in pyridine (10 mL) and the solution cooled to -40 °C in a dry ice/acetonitrile bath. Benzoyl chloride (1.4 mL, 12.5 mmol, 2.2 eq) was added dropwise and the solution was allowed to warm to room temperature overnight. The pyridine was removed under vacuum and the resulting crude mixture was purified by column chromatography (5% MeOH in CH_2Cl_2) to yield 2.1 g of **6** as a white solid (70%). R_f = 0.35 (CH_2Cl_2:MeOH 95:5). ^{1}H NMR (400 MHz, $CDCl_3$): δ 1.87 (s, 3H), 4.00 (m, 1H), 4.24 (brs, 1H), 4.46 (q_{app}, J = 8.8 Hz, 1H), 4.66 (m, 4H), 4.94 (d, J = 12.2 Hz, 1H), 5.34 (dd, J = 11.5, <1.0 Hz, 1H), 5.52 (d, J = 8.8 Hz, 1H), 7.50 (m, 6H), 7.61 (m, 3H), 8.07 (m, 4H), 8.12 (m, 2H). ^{13}C NMR (100 MHz, $CDCl_3$): δ 27.0, 52.7, 52.9, 53.1, 53.3, 53.5, 53.8, 54.5, 67.7, 70.7, 74.5, 76.9, 78.2, 104.0, 132.2, 132.4, 132.6, 132.8, 133.5, 133.9, 134.1, 137.6, 137.8, 141.3, 171.0. HRMS calcd for $[C_{29}H_{29}NO_8 + Na]^+$ requires *m/z* 542.1785. Found 542.1777 (ESI+).

Synthesis of benzyl 2-acetamido-4-azido-3,6-di-*O*-benzoyl-2,4-deoxy-β-D-glucopyranoside

7
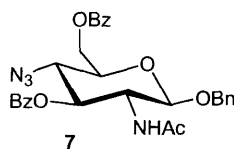
The dibenzoylated compound **6** (5.2 g, 10 mmol, 1 eq) was dissolved in CH_2Cl_2 (10 mL) with pyridine (1 mL). The solution was cooled to 0 °C and trifluoromethane sulfonic anhydride

(3.4 mL, 20 mmol, 2 eq) was added dropwise. The solution was stirred at 0 °C for 3 hours. The solvent was removed under vacuum and the resulting residue was dissolved in EtOAc. The solution was washed with H_2O (1 x 5 mL), 1M HCl (2 x 5 mL), H_2O (1 x 5 mL), saturated NaCl (1 x 5 mL) and dried over anhydrous $MgSO_4$. After filtering, the solvent was removed under vacuum to afford a yellow oil, which was used without further purification.

The triflate derivative from above was dissolved in DMF (10 mL) and sodium azide (3.3 g, 50 mmol, 5 eq) was added and stirred at room temperature overnight. The solvent was removed under vacuum and the residue dissolved in EtOAc, washed with H_2O (2 x 5 mL) and saturated NaCl (1 x 5 mL). After drying over anhydrous $MgSO_4$, the solvent was removed and the compound purified by recrystallization in EtOAc/Hexanes to afford 3.4 g of **5** as a white solid (62% over 2 steps). $R_f = 0.35$ (CH_2Cl_2:MeOH 95:5). ^1H NMR (400 MHz, $CDCl_3$): δ 1.79 (s, 3H), 3.77 (m, 1H), 3.99 (t_{app}, $J = 9.9$ Hz, 1H), 4.44 (q_{app}, $J = 9.3$ Hz, 1H), 4.64 (m, 2H), 4.73 (m, 2H), 4.88 (d, $J = 12.2$ Hz, 1H), 5.52 (t_{app}, $J = 10.1$ Hz, 1H), 6.60 (d, $J = 9.4$ Hz, 1H), 7.26 (m, 5H), 7.45 (m, 4H), 7.60 (m, 2H), 8.06 (d, $J = 7.7$ Hz, 2H), 8.13 (d, $J = 7.7$ Hz, 2H). ^{13}C NMR (100 MHz, $CDCl_3$): δ 23.5, 54.4, 61.6, 64.1, 70.9, 72.8, 74.9, 100.3, 128.4, 128.4, 128.8, 129.0, 129.1, 129.2, 130.0, 130.2, 130.4, 133.8, 134.2, 137.4, 166.6, 167.1, 170.9. HRMS calcd for $[C_{29}H_{28}N_4O_7 + Na]^+$ requires m/z 567.1850. Found 567.1848 (ESI+)

Synthesis of benzyl 2-acetamido-4-azido-2,4-deoxy-β-D-glucopyranoside
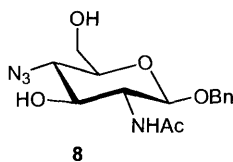
The azido derivative **7** (1.0 g, 1.84 mmol, 1 eq) was dissolved in MeOH (50 mL) and 100 μL of NaOMe (25 wt% in MeOH) (0.4 mmol, 0.08 eq) was added. The mixture was stirred at room temperature for 1 hour and the solvent neutralized with Amberlyte IR-120H ion exchange resin. The solvent was removed under vacuum and the product was purified by column chromatography (5% MeOH in CH_2Cl_2) to yield 500 mg of **8** as a white solid (80%). $R_f = 0.17$ (CH_2Cl_2:MeOH 95:5). ^{1}H NMR (400 MHz, CD_3OD): δ 1.99 (s, 3H), 3.21 (m, 1H), 3.48 (t_{app}, $J = 9.6$ Hz, 1H), 3.66 (t_{app}, $J = 9.9$ Hz, 1H), 3.77 (m, 2H), 3.88 (d, $J = 12.2$ Hz, 1H), 4.50 (d, $J = 8.3$ Hz, 1H), 4.61 (d, $J = 12.2$ Hz, 1H), 4.90 (d, $J = 12.3$ Hz, 1H), 7.30 (m, 5H). ^{13}C NMR (100 MHz, CD_3OD): δ 22.0, 56.7, 61.5, 62.9, 70.6, 74.2, 75.3, 100.7, 127.7, 127.8, 128.4, 128.5, 138.1, 172.8. HRMS calcd for $[C_{15}H_{20}N_4O_5 + Na]^+$ requires m/z 359.1326. Found 359.1324 (ESI+).

Synthesis of benzyl 2-acetamido-4-azido-6-tosyloxy-2,4-deoxy-β-D-glucopyranoside
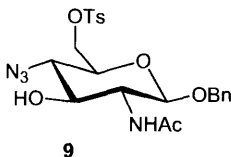
The deprotected compound **8** from the previous step (133 mg, 0.4 mmol, 1 eq) was dissolved in pyridine (10 mL) and cooled to 0 °C. p-Toluenesulfonyl chloride (83 mg, 0.44 mmol, 1.1 eq) was added to the solution and allowed to warm up to room temperature overnight. The pyridine was removed under vacuum and the resulting crude mixture was purified by column chromatography (5% MeOH in CH_2Cl_2) to afford **9** as a yellow oil (147 mg, 75% yield).

$R_f = 0.27$ (CH$_2$Cl$_2$:MeOH 95:5). ^{1}H NMR (400 MHz, CD$_3$OD): δ 1.97 (s, 3H), 2.44 (s, 3H), 3.37 (m,2H), 3.66 (m, 2H), 4.29 (m, 2H), 4.44 (d, = 7.8 Hz, 1H), 4.49 (d, J = 12.1 Hz, 1H), 4.75 (d, J = 12.2 Hz, 1H), 7.34 (m, 5H), 7.47 (d, J = 7.8 Hz, 2H), 7.87 (d, J = 8.0 Hz, 2H). ^{13}C NMR (100 MHz, CD$_3$OD): δ 22.0, 23.6, 30.1, 53.9, 57.6, 61.6, 62.4, 69.2, 71.0, 71.4, 72.2, 74.9, 99.7, 124.3, 127.9, 128.1, 128.4, 128.4, 128.9, 130.4, 132.8, 136.8, 137.3, 145.6, 149.8, 172.8. HRMS calcd for [C$_{22}$H$_{26}$N$_4$O$_7$S + Na]$^+$ requires m/z 513.1414. Found 513.1425 (ESI+).

Synthesis of benzyl 2-acetamido-4-azido-6-iodo-2,4,6-deoxy-β-D-glucopyranoside

10
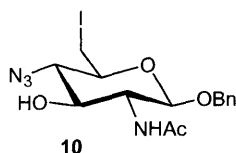
The tosylate derivative **9** (265 mg, 0.54 mmol, 1 eq) was dissolved in acetonitrile (10 mL) and sodium iodide (405 mg, 2.7 mmol, 5 eq) was added. The mixture was stirred with heating (80 °C) for 7 hours. The acetonitrile was removed under vacuum and the resulting residue was purified by column chromatography (5% MeOH in CH$_2$Cl$_2$) to yield 170 mg of **10** as a white solid (70%). $R_f = 0.27$ (CH$_2$Cl$_2$:MeOH 95:5). ^{1}H NMR (400 MHz, CD$_3$OD): δ 1.99 (s, 3H), 3.04 (m, 1H), 3.34 (m, 1H), 3.45 (m, 1H), 3.62 (d, J = 10.9 Hz, 1H), 3.72 (m, 2H), 4.53 (d, J = 8.2 Hz, 1H), 4.64 (d, J = 12.2 Hz, 1H), 4.89 (d, J = 12.1 Hz, 1H), 7.33 (m, 5H). ^{13}C NMR (100 MHz, CD$_3$OD): δ 5.4, 21.8, 37.0, 56.5, 67.1, 70.4, 72.9, 73.6, 99.9, 127.6, 127.8, 128.2, 137.6, 172.5. HRMS calcd for [C$_{15}$H$_{19}$IN$_4$O$_7$ + Na]$^+$ requires m/z 469.0343. Found 469.0352 (ESI+).

Synthesis of benzyl 2,4-acetamido-3-*O*-acetyl-2,4,6-deoxy-β-D-glucopyranoside

11
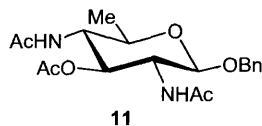
The iodo derivative **10** (220 mg, 0.5 mmol, 1 eq) was dissolved in MeOH (10 mL). *N,N*-di-isopropylethylamine (175 μL, 1.0 mmol, 2 eq) was added together with Pd(OH)$_2$/C (20 wt. %, wet) (100 mg, 0.14 mmol, 0.3 eq). The solution was stirred under a hydrogen atmosphere at 32 °C for 5 hours. The palladium was filtered off and the solvent removed under vacuum to afford a colorless oil. This crude mixture was acetylated without further purification, by dissolving in a 2:1 mixture of pyridine:acetic anhydride (10 mL) and stirred at room temperature overnight. The solvent was removed under vacuum and the resulting residue was purified by column chromatography (10% MeOH in CH$_2$Cl$_2$) to yield 120 mg of **11** as a white solid (65% over 2 steps). R$_f$ = 0.52 (CH$_2$Cl$_2$:MeOH 9:1). ^{1}H NMR (400 MHz, CDCl$_3$): δ 1.33 (d, *J* = 5.9 Hz, 3H), 1.91 (s, 3H), 1.96 (s, 3H), 2.03 (s, 3H), 3.54 (m, 1H), 3.70 (t$_{app}$, *J* = 10.0 Hz, 1H), 3.87 (t$_{app}$, *J* = 9.4 Hz, 1H), 4.60 (d, *J* = 11.9 Hz, 1H), 4.79 (d, *J* = 8.5 Hz, 1H), 4.92 (d, *J* = 12.2 Hz, 1H), 5.35 (t$_{app}$, *J* = 10.0 Hz, 1H), 7.35 (m, 5H). ^{13}C NMR (100 MHz, CDCl$_3$): δ 18.6, 21.1, 23.1, 23.2, 56.5, 57.3, 72.2, 74.6, 101.6, 129.2, 129.8, 139.4, 172.6, 173.8, 173.8. HRMS calcd for [C$_{19}$H$_{26}$N$_2$O$_6$ + Na]$^+$ requires *m/z* 401.1683. Found 401.1696 (ESI+).

## Synthesis of benzyl 2,4-acetamido-2,4,6-deoxy-β-D-glucopyranoside

12
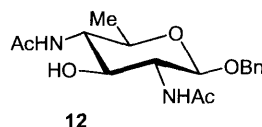
Compound **11** (60 mg, 0.16 mmol, 1 eq) was dissolved in methanol (5 mL) and 20 μL of NaOMe (25 wt% in MeOH) (0.09 mmol, 0.6 eq) was added. The mixture was stirred at room

temperature for 1 hour and the solvent neutralized with Amberlyte IR-120H ion exchange resin. The solvent was removed under vacuum and the product was purified by column chromatography (5% MeOH in CH_2Cl_2) to yield 48 mg of **12** as a white solid (90%) $R_f = 0.18$ (CH_2Cl_2:MeOH 95:5). ^{1}H NMR (400 MHz, CD_3OD): δ 1.26 (d, $J = 6.1$ Hz, 3H), 1.98 (s, 3H), 2.00 (s, 3H), 3.47 (m, 1H), 3.61 (m, 2H), 3.70 (t_{app}, $J = 8.0$ Hz, 1H), 4.53 (d, $J = 8.3$ Hz, 1H), 4.59 (d, $J = 12.1$ Hz, 1H), 4.86 (d, $J = 12.1$ Hz, 1H), 7.30 (m, 5H). ^{13}C NMR (100 MHz, CD_3OD): δ, 17.3, 21.9, 57.3, 58.0, 70.7, 71.3, 100.6, 127.7, 127.8, 128.3, 138.1. HRMS calcd for $[C_{17}H_{24}N_2O_5 + Na]^+$ requires *m/z* 337.1758. Found 337.1755 (ESI+).

Synthesis of benzyl 2,4-acetamido-3-*O*-benzoyl 2,4,6-deoxy-β-D-glucopyranoside

13
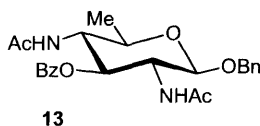
The deprotected compound **12** (36 mg, 0.11 mmol, 1 eq) was dissolved in pyridine (2 mL) and cooled to 0 °C. Benzoyl chloride (26 μL, 0.22 mmol, 2 eq) was added and the mixture was allowed to warm to room temperature overnight. The pyridine was removed under vacuum and the product was purified by column chromatography (5% MeOH in CH_2Cl_2) to yield 29 mg of **13** as a white solid (60% yield). $R_f = 0.28$ (CH_2Cl_2:MeOH 95:5). ^{1}H NMR (400 MHz, $CDCl_3$): δ 1.38 (d, $J = 6.2$ Hz, 3H), 1.76 (s, 1H), 1.83 (s, 3H), 3.62 (m, 1H), 3.95 ($q_{app,}$, $J = 9.0$ Hz, 1H), 4.10 ($q_{app,}J = 9.9$ Hz, 1H), 4.63 (d, $J = 12.2$ Hz, 1H), 4.83 (d, $J = 8.3$ Hz, 1H), 4.94 (d, $J = 12.2$ Hz, 1H), 5.65 (t_{app}, $J = 10.0$ Hz, 1H), 5.93 (d, $J = 8.5$ Hz, 1H), 6.04 (d, $J = 9.3$ Hz, 1H), 7.33 (m, 5H), 7.46 (m, 2H), 7.60 (m, 1H), 8.00 (d, $J = 7.1$ Hz, 2H). ^{13}C NMR (100 MHz, $CDCl_3$): δ18.4, 23.7, 23.7, 56.0, 56.2, 70.9, 71.8, 73.5, 77.1, 77.3, 73.5, 99.4, 128.3, 128.4,

128.9, 129.1, 129.4, 130.0, 134.1, 137.8, 168.0, 170.5, 170.8. HRMS calcd for $[C_{24}H_{28}N_2O_6 +$ Na$]^+$ requires *m/z* 463.1840. Found 463.1852 (ESI+).

Synthesis of 2,4-acetamido-3-*O*-benzoyl-2,4,6-deoxy--D-glucopyranoside

14
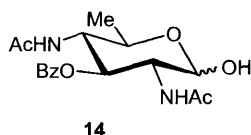
Compound **8** (25 mg, 0.06 mmol, 1 eq) was dissolved in methanol (2 mL) and Pd(OH)$_2$/C (20 wt. %, wet) (5 mg, 0.007 mmol, 0.1 eq) was added. The mixture was stirred under a hydrogen atmosphere at 32 °C for 7 hours. The palladium was filtered off and the product was purified by column chromatography (10% MeOH in CH$_2$Cl$_2$), to yield **14** as a white solid (17 mg, 86%). R$_f$ = 0.52 (CH$_2$Cl$_2$:MeOH 9:1). ^{1}H NMR (400 MHz, CD$_3$OD): δ 1.22 (d, *J* = 6.2 Hz, 3H), 1.82 (s, 3H), 1.86 (s, 3H), 3.34 (m, 1H), 4.00 (m, 1H), 4.19 (m, 1H), 5.14 (d, *J* = 3.5 Hz, 1H), 5.47 (t$_{app}$, *J* = 10.4 Hz, 1H), 7.48 (m, 2H), 7.63 (m, 1H), 7.99 (m, 2H). ^{13}C NMR (100 MHz, CD$_3$OD): δ 21.4, 21.5, 53.3, 53.4, 56.0, 66.0, 72.4, 91.8, 128.6, 129.8, 130.1, 133.4, 133.4, 167.0, 172.3, 172.4. HRMS calcd for $[C_{17}H_{22}N_2O_6 +$ Na$]^+$ requires *m/z* 351.1551. Found 351.1551 (ESI+).

Synthesis of dibenzylphospho-2,4-acetamido-3-*O*-benzoyl-2,4,6-deoxy-α-D-glucopyranoside

15
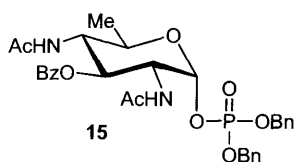
Compound **14** (15 mg, 0.04 mmol, 1 eq) was dissolved in dry THF (2 mL) and the solution cooled to -68 °C. To the cooled solution, lithium hexamethyldisilazide (LiHMDS, 1M

solution in THF) (56 μL, 0.05mmol, 1.3 eq) was added dropwise. After 10 minutes, tetrabenzylpyrophosphate (30 mg, 0.06 mmol, 1.5 eq) was dissolved in dry THF (1 mL) and added dropwise to the reaction mixture. The reaction was stirred at -68 °C for 10 minutes and slowly warmed to 0 °C. The mixture was then diluted with Et_2O (2 mL), washed with chilled saturated $NaHCO_3$ (2mL) and NaCl (2 mL) and dried over anhydrous $MgSO_4$. The solvent was removed under vacuum and the mixture purified by column chromatography (5% MeOH in CH_2Cl_2) to yield the protected phosphate **15** as a colorless oil (16 mg, 50% yield). R_f = 0.27 (CH_2Cl_2:MeOH 95:5). ^{1}H NMR (400 MHz, $CDCl_3$): δ 1.17 (d, J = 6.2 Hz, 3H), 1.75 (s, 3H), 1.79 (s, 3H), 3.95 (m, 1H), 4.23 (q_{app}, J = 10.2 Hz, 1H), 4.61 (m, 1H), 5.08 (m, 4H), 5.29 (t_{app}, J = 10.5 Hz, 1H), 5.73 (dd, J = 5.45, 3.4 Hz, 1H), 6.23 (brs, 1H), 6.30 (brs, 1H), 7.25-7.80 (m, 15H). ^{13}C NMR (100 MHz, $CDCl_3$): δ 17.9, 23.0, 23.3, 54.6, 70.0, 70.0, 128.0, 128.1, 128.2, 128.6, 128.7, 128.9, 129.0, 129.0, 129.1, 130.1, 133.9, 168.0, 170.4, 170.6. ^{31}P NMR (162 MHz, $CDCl_3$): δ -2.07. Mass spectroscopy of this intermediate is complicated by the facile β-elimination of the phosphotriester, which precludes observation of the molecular ion.

Synthesis of phospho-2,4-acetamido-3-*O*-benzoyl-2,4,6-deoxy-α-D-glucopyranoside
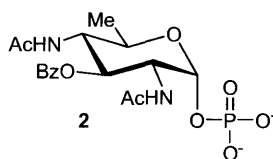
The protected phosphate **15** (16 mg, 0.02 mmol, 1 eq) was dissolved in MeOH (2 mL). Pd/C (10 wt% on activated carbon) (2 mg, 0.002 mmol, 0.1 eq) was added and the mixture was stirred under a hydrogen atmosphere for 3 hours. The palladium was filtered off and the solvent removed to yield 12 mg of **2** as a white solid (> 99%), which was used in the coupling to UMP-morpholidate or Und-P without further purification. ^{1}H NMR (400 MHz, CD_3OD): δ 1.23 (d, J =

6.2 Hz, 3H), 1.81 (s, 3H), 1.85 (s, 3H), 4.04 (t_{app}, J = 10.3 Hz, 1H), 4.19 (m, 1H), 4.41 (m, 1H), 5.43 (t_{app}, J = 10.4 Hz, 1H), 5.75 (dd, J = 6.2, 3.3 Hz, 1H), 7.46 (m, 2H), 7.61 (m, 1H), 7.99 (m, 2H). ^{13}C NMR (100 MHz, CD$_3$OD): δ 16.8, 21.2, 21.4, 52.7, 52.7, 55.2, 68.0, 71.4, 95.3, 128.4, 129.6, 129.6, 133.3, 166.6, 172.1, 172.4. ^{31}P NMR (162 MHz, CD$_3$OD): δ -0.014. HRMS calcd for [C$_{17}$H$_{23}$N$_2$O$_9$P - H]$^-$ requires m/z 429.1068. Found 429.1066 (ESI-).

Synthesis of Uridin-5'-yl (2-4-diacetamido-2,4-6-trideoxy-α-D-glucopyranosyl) diphosphate
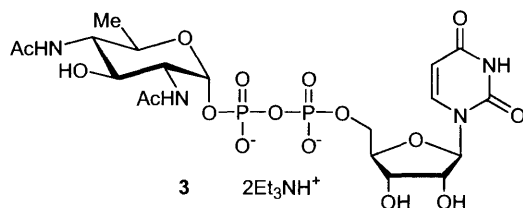
The protected bacillosamine phosphate derivative (**2**) (10 mg, 0.02 mmol, 1 eq) was coevaporated several times with dry pyridine. To this dried compound, 4-morpholine-N,N'-dicyclohexylcarboxamidinium uridine 5'-monophosphormorpholidate (17.9 mg, 0.03 mmol, 1.3 eq) was added and the mixture was coevaporated twice with dry pyridine and dried under vaccum. The solid was then dissolved in 0.5 mL of dry pyridine and 1H-tetrazole (0.04 mmol, 2 eq, 3 mg) was added and the solution stirred at room temperature for 3 days. The solved was removed under vaccum and the residue dissolved in water and purified by RP-HPLC (C18) (0.05 M triethylammonium bicarbonate (TEAB) buffer; gradient 4-18% CH$_3$CN over 30 min; flow rate, 1 mL/min; t_R = 28.7 min). Benzoyl protected UDP-Bac (9 mg, 50%) was obtained as a white solid. ^{1}H NMR (400 MHz, MeOH) δ ^{1}H NMR (400 MHz, CD$_3$OD): δ 1.25 (d, J = 6.8 Hz, 3H), 1.82 (s, 3H), 1.94 (s, 3H), 4.03 (t_{app}, J = 10.3 Hz, 1H), 4.22 (m, 1H), 4.28 (m, 2H), 4.38 (m,

1H), 4.41 (m, 2H), 4.58 (t_{app}, J = 4.2 Hz, 1H), 5.40 (t_{app}, J = 10.4 Hz, 1H), 5.75 (dd, J = 7.3, 3.1

Hz, 1H), 5.85 (d, J = 8.1 Hz, 1H), 6.05 (d, J = 5.3 Hz, 1H), 7.45 (t, J = 7.7 Hz, 2H), 7.58 (t, J =

7.6 Hz, 1H), 7.95 (d, J = 7.1Hz, 2H), 8.18 (d, J = 8.1 Hz, 1H). ^{31}P NMR (162 MHz, CD$_3$OD): δ -

10.8, -13.0. LRMS calcd for [C$_{26}$H$_{33}$N$_4$O$_{17}$P$_2$ - H]$^-$ requires m/z 734.5. Found 734.4 (ESI-).

The protected UDP-Bac analog from above (5 mg, 0.007 mmol, 1 eq) was dissolved in

MeOH (0.5 mL). Sodium methoxide 1.0 µL (25 wt% in MeOH) was added and the mixture

stirred at room temperature for 30 min. The solvent was removed under vacuum, the residue

dissolved in water and purified by RP-HPLC (C18) (0.05 M triethylammonium bicarbonate

(TEAB) buffer; 1 % CH$_3$CN; flow rate, 1 mL/min; t_R = 15.0 min). Compound 1 was obtained as

a white solid (3.0 mg, 68%). LRMS calcd for [C$_{19}$H$_{29}$N$_4$O$_{16}$P$_2$ - H]$^-$ requires m/z 630.4. Found

630.5 (ESI-).

Synthesis of undecaprenyl-pyrophosphate bacillosamine
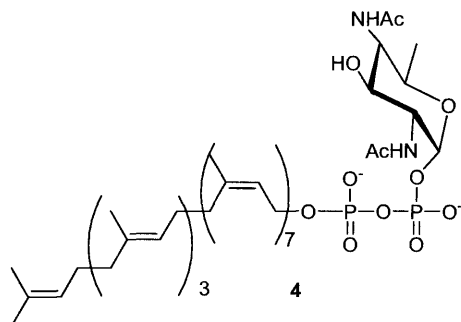
Bacillosamine phosphate **2** (5 mg, 0.008 mmol, 1 eq) was dissolved in DMF (500 µL)

and carbonyl di-imidazole (CDI) (2 mg, 0.013 mmol, 1.6 eq) was added. The mixture was stirred

at room temperature for 6 hours. The excess CDI was quenched with methanol (2.3 µL, 0.072

mmol, 9 eq) and stirred for 30 mins. The undecaprenyl phosphate (7 mg, 0.008 mmol, 1 eq) was dissolved in CH_2Cl_2 (500 µL) and added to the sugar phosphate. The mixture was allowed to stir at room temperature for one week and monitored by ^{31}P NMR. The solvent was removed under vacuum and the mixture purified by silica gel chromatography eluting with $CHCl_3:MeOH:2M$ NH_4OH 85:14:1, to yield 5.0 mg of protected undecaprenyl pyrophosphate bacillosamine as a colorless oil (50%). R_f = 0.48 ($CHCl_3:MeOH:H_2O$ 65:25:4), R_f = 0.82 ($CHCl_3:MeOH:2M$ NH_4OH 9:7:2) ^1H NMR (400 MHz, CD_3OD): see attached spectra. ^{31}P NMR (162 MHz, $CDCl_3:CD_3OD$ (2:1)): δ -4.5, -7.0. LRMS calcd for $[C_{72}H_{112}N_2O_{12}P_2-H]^-$ requires m/z 1257.8. Found 1257.4 (ESI-).
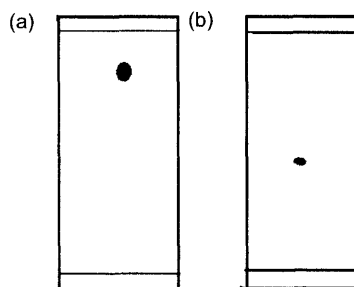
Figure S1: Thin layer chromatography of the protected Und-PP-Bac intermediate. a) $CHCl_3:MeOH:2M$ NH_4OH 9:7:2 b). $CHCl_3:MeOH:H_2O$ 65:25:4

The protected intermediate from above was dissolved in MeOH and 1 µL of NaOMe/MeOH was added and stirred at room temperature for 10 min. The solvent was neutralized with Dowex-50WX8- 200 ion exchange resin (pyridinium form) and removed under vacuum to yield 4.6 mg of **2** (~99 % yield) as a colorless oil. R_f = 0.32 ($CHCl_3:MeOH:H_2O$ 65:25:4). LRMS calcd for $[C_{65}H_{108}N_2O_{11}P_2-H]^-$ requires m/z 1153.7. Found 1153.8 (ESI-).
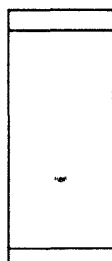
Figure S2: Thin layer chromatography of deprotected Und-PP-Bac in CHCl₃:MeOH:H₂O 65:25:4

Synthesis of benzyl 2,4-acetamido-3,6-*O*-benzoyl-2,4-deoxy-β-D-glucopyranoside
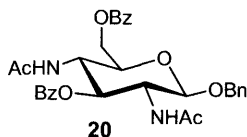
Intermediate **7** in the synthesis of bacillosamine (300 mg, 0.55 mmol, 1 eq) was dissolved in MeOH (10 mL). *N,N*-di-isopropylethylamine (200 μL, 1.1 mmol, 2 eq) was added together with Pd(OH)₂/C (20 wt. %, wet) (110 mg, 0.17 mmol, 0.3 eq). The solution was stirred under a hydrogen atmosphere at 32 °C for 5 hours. The palladium was filtered off and the solvent removed under vacuum to afford a colorless oil. This crude mixture was acetylated without further purification, by dissolving in a 2:1 mixture of pyridine:acetic anhydride (10 mL) and stirred at room temperature overnight. The solvent was removed under vacuum and the resulting residue was purified by column chromatography (10% MeOH in CH₂Cl₂) to yield 200 mg of **20** as a white solid (65% over 2 steps). R_f = 0.29 (CH₂Cl₂:MeOH 95:5). ¹H NMR (400 MHz, CDCl₃): δ 1.75 (s, 3H), 1.76 (s, 3H), 3.94 (m, 1H), 4.07 (m, 1H), 4.42 (t_{app}, J = 10.2 Hz, 1H), 4.46 (dd, J = 12.1, 6.9 Hz, 1H), 4.57 (dd, J = 12.1, 2.4 Hz, 1H), 4.59 (d, J = 10.4 Hz, 1H), 4.76 (d, J = 8.4 Hz, 1H), 4.82 (d, J = 12.0 Hz, 1H), 5.44 (t_{app}, J = 10.3 Hz, 1H), 7.24 (m, 5H), 7.41-8.23 (m, 10H). ¹³C NMR (100 MHz, CDCl₃): δ 23.1, 52.7, 56.5, 65.1, 72.3, 74.2, 75.0, 101.5,

126.0, 128.7, 129.3, 129.4, 129.8, 129.9, 130.1, 130.1, 131.2, 131.2, 131.3, 131.6, 134.9, 135.0, 138.9, 139.1, 150.5, 168.1, 173.9. HRMS calcd for $[C_{31}H_{32}N_2O_8 + Na]^+$ requires m/z 583.2051. Found 583.2027 (ESI+).

Synthesis of 2,4-acetamido-3,6-*O*-acetyl-2,4-deoxy-β-D-glucopyranoside
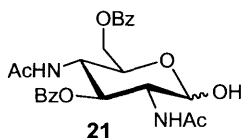
Compound **20** (200 mg, 0.34 mmol, 1 eq) was dissolved in methanol (3 mL) and $Pd(OH)_2/C$ (20 wt. %, wet) (21 mg, 0.03 mmol, 0.1 eq) was added. The mixture was stirred under a hydrogen atmosphere at 32 °C for 7 hours. The palladium was filtered off and the product was purified by column chromatography (10% MeOH in CH_2Cl_2), to yield **21** as a white solid (130 mg, 79%). R_f = 0.18 (CH_2Cl_2:MeOH 95:5). ^{1}H NMR (400 MHz, CD_3OD): δ 1.79 (s, 3H), 1.84 (s, 3H), 4.37-4.53 (m, 5H), 5.21 (d, J = 3.4 Hz, 1H). 5.57 (t_{app}, J = 9.9 Hz, 1H), 7.44-8.10 (m, 10H). ^{13}C NMR (100 MHz, CD_3OD): δ 22.7, 23.0, 52.1, 53.9, 53.9, 64.7, 71.6, 71.6, 71.7, 71.7, 71.9, 72.4, 98.5, 129.2, 129.6, 129.7, 129.9, 130.0, 130.1, 130.2, 130.3, 131.0, 131.2, 131.4, 134.8, 135.1, 137.3, 167.9, 168.1, 173.8, 174.2. HRMS calcd for $[C_{24}H_{26}N_2O_8 + Na]^+$ requires m/z 493.1581. Found 493.1570 (ESI+).

Synthesis of dibenzylphospho-2,4-acetamido-3,6-*O*-benzoyl-2,4-deoxy-α-D-gluco-pyranoside

OBz
AcHN   O
BzO
     AcHN   O
              ‖
          O—P—OBn
22   OBn
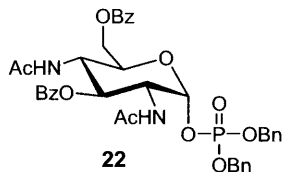
Compound **21** (138 mg, 0.28 mmol, 1 eq) was dissolved in dry THF (3 mL) and the solution cooled to -68 °C. To the cooled solution, lithium hexamethyldisilazide (LiHMDS, 1M solution in THF) (360 μL, 0.36 mmol, 1.3 eq) was added dropwise. After 10 minutes, tetrabenzylpyrophosphate (194 mg, 0.42 mmol, 1.5 eq) was dissolved in dry THF (2 mL) and added dropwise to the reaction mixture. The reaction was stirred at -68 °C for 10 minutes and slowly warmed to 0 °C. The mixture was then diluted with Et_2O (2 mL), washed with chilled saturated $NaHCO_3$ (2mL) and NaCl (2 mL) and dried over anhydrous $MgSO_4$. The solvent was removed under vacuum and the mixture purified by column chromatography (5% MeOH in CH_2Cl_2) to yield the protected phosphate **22** as a colorless oil (100 mg, 50% yield). R_f = 0.35 (CH_2Cl_2:MeOH 95:5). ^{1}H NMR (400 MHz, $CDCl_3$): δ 1.77 (s, 3H), 1.83 (s, 3H), 4.43-4.57 (m, 6H), 5.14, (m, 4H), 5.60 (t_{app}, J = 10.5 Hz, 1H), 5.86 (dd, J = 3.3, 5.9 Hz, 1H), 7.37 (m, 5H), 7.47-8.07 (m, 10H). ^{13}C NMR (100 MHz, $CDCl_3$): δ 22.7, 23.0, 52.1, 53.9, 64.7, 71.6, 71.7, 71.9, 72.4, 98.5, 129.2, 129.6, 129.7, 129.9, 130.0, 130.1, 130.2, 130.3, 131.0, 131.2, 131.4, 134.8, 135.1, 137.3, 167.9, 168.1, 173.8, 174.2. ^{31}P NMR (162 MHz, $CDCl_3$): δ -1.50. Mass spectroscopy of this intermediate is complicated by the facile β-elimination of the phosphotriester, which precludes observation of the molecular ion.

Synthesis of phospho-2,4-acetamido-3,6-*O*-benzoyl-2,4-deoxy-α-D-galacto-pyranoside
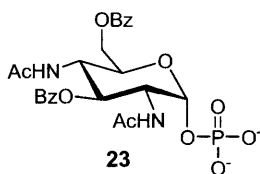
The protected phosphate **22** (100 mg, 0.14 mmol, 1 eq) was dissolved in MeOH (2 mL). Pd/C (10 wt% on activated carbon) (14 mg, 0.014 mmol, 0.1 eq) was added and the mixture was stirred under a hydrogen atmosphere for 3 hours. The palladium was filtered off and the solvent removed to yield 12 mg of **2** as a white solid (> 99%), which was used in the coupling to UMP-morpholidate or Und-P without further purification. ^{1}H NMR (400 MHz, CD$_3$OD): δ 1.82 (s, 3H), 1.93 (s, 3H), 4.41 (m, 2H), 4.51 (m, 1H), 4.58 (m, 2H), 5.61 (m, 2H), 7.46-8.15 (m, 10H). ^{13}C NMR (100 MHz, CD$_3$OD): δ 10.1, 23.1, 47.4, 50.1, 51.8, 54.8, 64.8, 70.6, 74.8, 95.5, 129.9, 130.0, 131.2, 131.6, 131.8, 134.7, 168.2, 168.3, 173.4, 173.9. ^{31}P NMR (162 MHz, CD$_3$OD): δ - 2.56. HRMS calcd for [C$_{24}$H$_{25}$N$_2$O$_{11}$P - H]$^-$ requires *m/z* 549.1269. Found 549.1283 (ESI-).
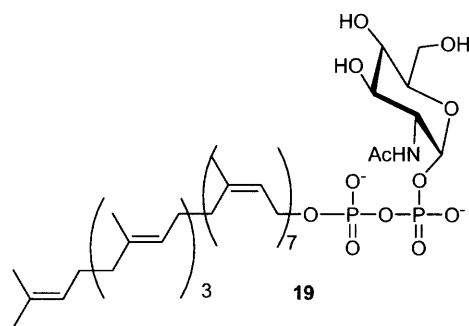

Synthesis of Und-PP-Bac-6-OH
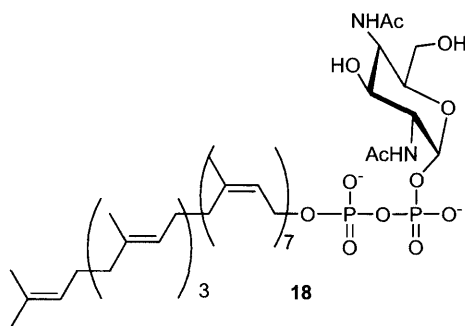
Bac-6-OH phosphate **23** (15 mg, 0.03 mmol, 1 eq) was dissolved in DMF (500 µL) and carbonyl di-imidazole (CDI) (7 mg, 0.04 mmol, 1.6 eq) was added. The mixture was stirred at room temperature for 6 hours. The excess CDI was quenched with methanol (8.6 µL, 0.27 mmol,

9 eq) and stirred for 30 mins. The undecaprenyl phosphate (7 mg, 0.008 mmol, <1 eq) was dissolved in CH_2Cl_2 (500 μL) and added to the sugar phosphate. The mixture was allowed to stir at room temperature for one week and monitored by ^{31}P NMR. The solvent was removed under vacuum and the mixture purified by silica gel chromatography eluting with $CHCl_3$:MeOH:2M NH_4OH 85:14:1, to yield 5.0 mg of protected Und-PP-Bac-6-OH as a colorless oil (50%). R_f = 0.55 ($CHCl_3$:MeOH:H_2O 65:25:4). ^{31}P NMR (162 MHz, $CDCl_3$:CD_3OD (2:1)): δ -5.7, -8.3. The dibenzoyl protection on C-3 and C-6 was deprotected with sodium methoxide as before to yield the deprotected Und-PP-Bac-6-OH as a colorless oil. HRMS calcd for $[C_{65}H_{106}N_2O_{12}P_2$-H]$^-$ requires m/z 1169.7232. Found 1169.7315 (ESI-).

Synthesis of Und-PP-GlcNAc

GlcNAc phosphate (15 mg, 0.03 mmol, 1 eq) was dissolved in DMF (500 μL) and carbonyl di-imidazole (CDI) (7 mg, 0.04 mmol, 1.6 eq) was added. The mixture was stirred at room temperature for 6 hours. The excess CDI was quenched with methanol (8.6 μL, 0.27 mmol, 9 eq) and stirred for 30 mins. The undecaprenyl phosphate (7 mg, 0.008 mmol, <1 eq) was dissolved in CH_2Cl_2 (500 μL) and added to the sugar phosphate. The mixture was allowed to stir at room temperature for one week and monitored by ^{31}P NMR. The solvent was removed under vacuum and the mixture purified by silica gel chromatography eluting with $CHCl_3$:MeOH:2M

NH$_4$OH 85:14:1, to yield 5.0 mg of protected Und-PPGlcNAc as a colorless oil (50%). R$_f$ = 0.40

(CHCl$_3$:MeOH:H$_2$O 65:25:4), ^{31}P NMR (162 MHz, CDCl$_3$:CD$_3$OD (2:1)): δ -5.0, -7.9. LRMS

calcd for [C$_{69}$H$_{111}$NO$_{15}$P$_2$-2H]$^{2-}$ requires m/z 1253.7. Found 1253.5 (ESI-).

The acetyl protecting groups were removed with sodium methoxide as before to yield the

deprotected Und-PP-GlcNAc as a colorless oil.

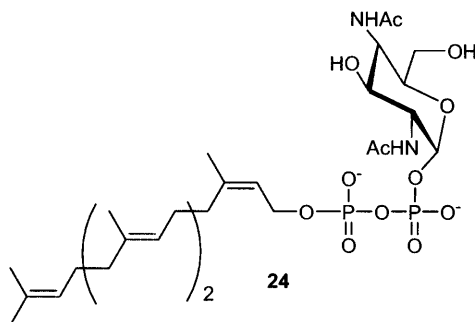**Synthesis of Geranylgeranyl-PP-Bac-6-OH**

Bac-6-OH phosphate **23** (10 mg, 0.018 mmol, 1 eq) was dissolved in DMF (500 µL) and

carbonyl di-imidazole (CDI) (15 mg, 0.09 mmol, 5.0 eq) was added. The mixture was stirred at

room temperature for 6 hours. The excess CDI was quenched with methanol (5 µL, 0.16 mmol, 9

eq) and stirred for 30 mins. Geranylgeranyl phosphate (6.7 mg, 0.018 mmol, 1 eq) was dissolved

in CH$_2$Cl$_2$ (500 µL) and added to the sugar phosphate. The mixture was allowed to stir at room

temperature for one week and monitored by ^{31}P NMR. The solvent was removed under vacuum

and the mixture purified by silica gel chromatography eluting with CHCl$_3$:MeOH:2M NH$_4$OH

85:14:1, to yield 5.0 mg of protected geranylgeranyl-PP-Bac-6-OH as a colorless oil (50%). R$_f$ =

0.25 (CHCl$_3$:MeOH:H$_2$O 65:25:4). ^{31}P NMR (162 MHz, CDCl$_3$:CD$_3$OD (2:1)): δ -5.0, -7.8.

LRMS calculated for [C$_{44}$H$_{60}$N$_2$O$_{14}$P$_2$-2H]$^-$ requires m/z 900.4. Found 900.4 (ESI-). The C-3 and

C-6 benzoyl protection was removed with sodium methoxide as before to yield the fully

deprotected Geranylgeranyl-PP-Bac-6-OH as a colorless oil. LRMS calcd for $[C_{30}H_{52}N_2O_{12}P_2-2H]^{2-}$ requires m/z 692.3. Found 692.5 (ESI-).
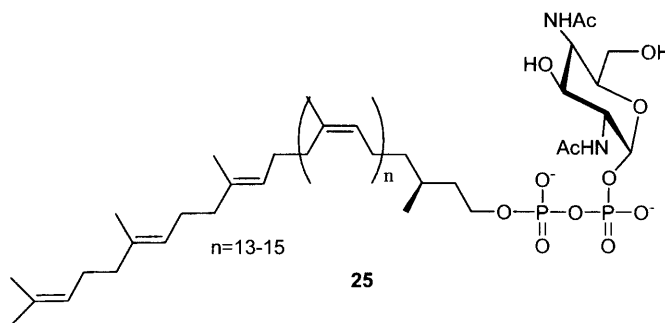
Synthesis of Dolichyl-PP-Bac-6-OH

25

Bac-6-OH phosphate **23** (7 mg, 0.014 mmol, 1 eq) was dissolved in DMF (500 μL) and carbonyl di-imidazole (CDI) (11 mg, 0.07 mmol, 5.0 eq) was added. The mixture was stirred at room temperature for 6 hours. The excess CDI was quenched with methanol (5 μL, 0.12 mmol, 9 eq) and stirred for 30 mins. Dolichyl phosphate (20 mg, 0.014 mmol, 1 eq) was dissolved in CH_2Cl_2 (500 μL) and added to the sugar phosphate. The mixture was allowed to stir at room temperature for one week and monitored by ^{31}P NMR. The solvent was removed under vacuum and the mixture purified by silica gel chromatography eluting with $CHCl_3$:MeOH:2M NH_4OH 85:14:1, to yield 7.0 mg of protected dolichyl-PP-Bac-6-OH as a colorless oil (50%). $R_f = 0.50$ ($CHCl_3$:MeOH:H_2O 65:25:4). ^{31}P NMR (162 MHz, $CDCl_3$:CD_3OD (2:1)): δ -6.3, -8.6. The C-3 and C-6 benzoyl protection was removed with sodium methoxide as before to yield the fully deprotected Geranylgeranyl-PP-Bac-6-OH as a colorless oil. LRMS calcd for $[C_{95-115}H_{158-190}N_2O_{12}P_2-2H]^{2-}$ requires m/z 1579.1, 1647.2, 1715.2, 1783.3, 1851.4. Found 1579.5, 1647.6, 1715.7, 1783.6, 1850.8 (ESI-).

References

1. Burda, P., and Aebi, M. The Dolichol Pathway of N-Linked Glycosylation. *Biochim. Biophys. Acta*, **1999**, *1426*, 239-257.

2. Barnes, G., Hansen, W.J., Holcomb, C.L., and Rine, J. Asparagine-Linked Glycosylation in Saccharomyces-Cerevisiae - Genetic-Analysis of an Early Step. *Mol. Cell. Biol.*, **1984**, *4*, 2381-2388.

3. Imperiali, B., O'Connor, S.E., Hendrickson, T., and Kellenberger, C. Chemistry and Biology of Asparagine-Linked Glycosylation. *Pure Appl. Chem.*, **1999**, *71*, 777-787.

4. Linton, D., Dorrell, N., Hitchen, P.G., Amber, S., Karlyshev, A.V., Morris, H.R., Dell, A., Valvano, M.A., Aebi, M., and Wren, B.W. Functional Analysis of the Campylobacter Jejuni N-Linked Protein Glycosylation Pathway. *Mol. Microbiol.*, **2005**, *55*, 1695-1703.

5. Young, N.M., Brisson, J.R., Kelly, J., Watson, D.C., Tessier, L., Lanthier, P.H., Jarrell, H.C., Cadotte, N., St Michael, F., Aberg, E., and Szymanski, C.M. Structure of the N-Linked Glycan Present on Multiple Glycoproteins in the Gram-Negative Bacterium, Campylobacter Jejuni. *J. Biol. Chem.*, **2002**, *277*, 42530-42539.

6. Wacker, M., Linton, D., Hitchen, P.G., Nita-Lazar, M., Haslam, S.M., North, S.J., Panico, M., Morris, H.R., Dell, A., Wren, B.W., and Aebi, M. N-Linked Glycosylation in Campylobacter Jejuni and Its Functional Transfer into E. Coli. *Science*, **2002**, *298*, 1790-1793.

7. Szymanski, C.M., Yao, R., Ewing, C.P., Trust, T.J., and Guerry, P. Evidence for a System of General Protein Glycosylation in Campylobacter Jejuni. *Mol. Microbiol.*, **1999**, *32*, 1022-1030.

8. Szymanski, C.M., and Wren, B.W. Protein Glycosylation in Bacterial Mucosal Pathogens. *Nat. Rev. Microbiol.*, **2005**, *3*, 225-237.

9. Szymanski, C.M., Logan, S.M., Linton, D., and Wren, B.W. Campylobacter--a Tale of Two Protein Glycosylation Systems. *Trends Microbiol.*, **2003**, *11*, 233-238.

10. Maki, M., and Renkonen, R. Biosynthesis of 6-Deoxyhexose Glycans in Bacteria. *Glycobiology*, **2004**, *14*, 1r-15r.

11. Creuzenet, C. Characterization of Cj1293, a New UDP-GlcNAc C-6 Dehydratase from Campylobacter Jejuni. *FEBS Lett.*, **2004**, *559*, 136-140.

12. Obhi, R.K., and Creuzenet, C. Biochemical Characterization of the Campylobacter Jejuni Cj1294, a Novel UDP-4-Keto-6-Deoxy-GlcNAc Aminotransferase That Generates UDP-4-Amino-4,6-Dideoxy-GalNAc. *J.Biol. Chem.*, **2005**, *280*, 20902-20908.

13. Liav, A., Hildeshe.J, Zehavi, U., and Sharon, N. Synthesis of 2,4-Diacetamido-2,4,6-Trideoxy-D-Glucose and Its Identification with Diacetamido-Sugar of Bacillus-Licheniformis. *J. Chem. Soc.,Chem. Commun.*, **1973**, 668-669.

14. Bundle, D.R., and Josephson, S. Chlorosulfation of Amino-Sugars - Synthesis of Methyl 2-Acetamido-4-Amino-2,4,6-Trideoxy-Beta-D-Glucopyranoside (Bacillosamine) from a D-Glucosamine Derivative. *Can. J. Chem.*,**1980**, *58*, 2679-2685.

15. Banaszek, A., Pakulski, Z., and Zamojski, A. The Synthesis of Derivatives of 2,4-Diamino-2,4,6-Trideoxy-D-Gulo- and L-Altro-Hexopyranoses. *Carbohydr. Res.*, **1995**, *279*, 173-182.

16. Liang, H., and Grindley, T.B. An Efficient Synthesis of Derivatives of 2-Acetamido-4-Amino-2,4,6-Trideoxy-D-Galactopyranose. *J. Carbohydr. Chem.*, **2004**, *23*, 71-82.

17. Liav, A., and Sharon, N. Synthesis of 2,4-Diacetamido-2,4,6-Trideoxy-L-Altrose, -L-Idose, and -L-Talose from Benzyl 6-Deoxy-3,4-O-Isopropylidene-Beta-L-Galactopyranoside. *Carbohydr. Res.*, **1973**, *30*, 109-126.

18. Liav, A., Hildeshe.J, Zehavi, U., and Sharon, N. Synthesis of 2-Acetamido-2,6-Dideoxy-D-Glucose (N-Acetyl-D-Quinovosamine), "2-Acetamido-2,6-Dideoxy-D-Galactose (N-Acetyl-D-Fucosamine), and 2,4-Diacetamido-2,4,6-Trideoxy-D-Glucose from 2-Acetamido-2-Deoxy-D-Glucose. *Carbohydr. Res.*, **1974**, *33*, 217-227.

19. Tsvetkov, Y.E., Shashkov, A.S., Knirel, Y.A., and Zahringer, U. Synthesis and Nmr Spectroscopy of Nine Stereoisomeric 5,7-Diacetamido-3,5,7,9-Tetradeoxynon-2-Ulosonic Acids. *Carbohydr. Res.*, **2001**, *335*, 221-243.

20. Matta, K.L., Johnson, E.A., and Barlow, J.J. Simple Method for Synthesis of 2-Acetamido-2-Deoxy-Beta-D-Galactopyranosides. *Carbohydr. Res.*, **1973**, *26*, 215-218.

21. Zhu, X., Stolz, F., and Schmidt, R.R. Synthesis of Thioglycoside-Based UDP-Sugar Analogues. *J. Org. Chem.*, **2004**, *69*, 7367-7370.

22. Stolz, F., Reiner, M., Blume, A., Reutter, W., and Schmidt, R.R. Novel UDP-Glycal Derivatives as Transition State Analogue Inhibitors of UDP-GlcNAc 2-Epimerase. *J. Org. Chem.*, **2004**, *69*, 665-679.

23. Branch, C.L., Burton, G., and Moss, S.F. An Expedient Synthesis of Allylic Polyprenyl Phosphates. *Synth. Commun.*, **1999**, *29*, 2639-2644.

24. Ye, X.Y., Lo, M.C., Brunner, L., Walker, D., Kahne, D., and Walker, S. Better Substrates for Bacterial Transglycosylases. *J. Am. Chem. Soc.*, **2001**, *123*, 3155-3156.

25. Tai, V.W.F., and Imperiali, B. Substrate Specificity of the Glycosyl Donor for Oligosaccharyl Transferase. *J. Org. Chem.*, **2001**, *66*, 6217-6228.

Chapter 3

Investigating the glycosyltransferases in the Pgl pathway of *Campylobacter jejuni*

A significant portion of the work described in this chapter has been published in:

Dr. K. J. Glover performed all the cloning, protein expression and purification described in this chapter. All the radioactive and HPLC assays were a collaborative effort.

Introduction

The undecaprenyl-pyrophosphate-linked heptasaccharide donor for N-linked glycosylation in *Campylobacter jejuni* is biosynthesized on the cytosolic face of the periplasmic membrane by a series of glycosyltransferases. The enzymes responsible for this process are encoded by a gene cluster dubbed '*pgl*' for protein glycosylation (Figure 3-1).[1,2]
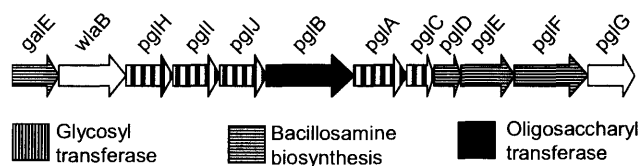

Figure 3-1. The *pgl* gene cluster from *C. jejuni*.

PglC is a glycophosphoryltransferase that transfers the first sugar residue of the N-linked heptasaccharide onto the undecaprenyl-phosphate (Und-P) carrier. The exact identity of the nucleotide-linked glycan substrate for PglC is still in question.[3] Four glycosyltransferases, PglA, PglJ, PglH and PglI are responsible for the completion of the heptasaccharide biosynthesis. The addition of six glycans *via* the action of four glycosyltransferases leads to questions regarding the exact role of each of these enzymes in this biosynthetic process. Our interests lie in validating the activities of the Pgl enzymes *in vitro*, with two major goals in mind:

1. To unambiguously define the nucleotide-linked glycan donor for the glycophosphoryl-transferase, PglC.

2. To determine how the four glycosyltransferases, Pgl A, J, H and I, facilitate the transfer of six glycan units.

90

PglC

PglC is a 23 kD glycophosphoryl-transferase that is postulated to catalyze the transfer of a sugar-phosphate from a UDP-linked donor to Und-P, to form the first, membrane-associated glycan intermediate in the pathway. PglC includes a single N-terminal transmembrane-domain (TMHMM, ExPASY), similar to several bacterial galactose-1-phosphoryl transferases involved in O-antigen biosynthesis, including RfbP (WbaP) in *Salmonella enterica*.[4, 5] Enzymes involved in HexNAc-1-P transfer reactions such as MraY (bacterial cell wall synthesis),[6] WecA (O-antigen biosynthesis),[7] and Alg7 (the eukaryotic enzyme responsible for the first committed step in the dolichol pathway),[8] differ topologically from PglC since they all include 11 or more transmembrane regions.[9]

Currently, there are two accepted models for the role of PglC in the Pgl pathway. In the first model (Figure 3-2 A), UDP-GlcNAc or UDP-GalNAc is converted sequentially by a dehydratase (PglF), an aminotransferase (PglE) and an acetyltransferase (PglD) to afford uridine diphospho-bacillosamine (UDP-Bac). This product then acts as the substrate for PglC, which creates the membrane-associated intermediate, undecaprenyl-pyrophosphate-linked bacillosamine (Und-PP-Bac).[2]

In the second model (Figure 3-2 B), PglC is proposed to utilize UDP-GlcNAc or UDP-GalNAc to create the membrane associated intermediate Und-PP-GlcNAc/GalNAc. This intermediate is then modified sequentially on the Und-P carrier by PglF, PglE and PglD to yield Und-PP-Bac.[1] Our interest lies in defining the precise role of PglC in order to provide vital information regarding the early steps of the Pgl pathway.
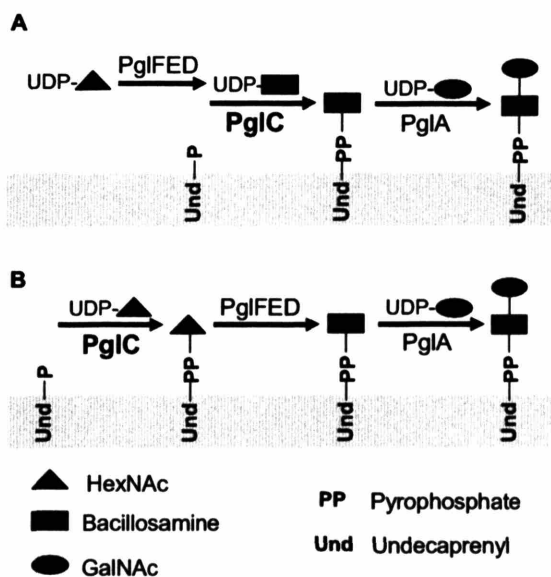
Figure 3-2. The early steps in the Pgl pathway.

Two proposed models: (A) PglC accepts UDP-Bac. (B) PglC accepts UDP-HexNAc.

The Pgl glycosyltransferases

The Und-PP-Bac intermediate is elaborated by PglA, PglJ, PglH and PglI to form the Und-PP-heptasaccharide. The current model postulates that PglA transfers the first GalNAc residue to bacillosamine to form the disaccharide. PglJ catalyzes the subsequent GalNAc transfer to afford the trisaccharide, PglH is responsible for the third GalNAc transfer and PglI catalyzes the final step of the biosynthesis by transferring the branching glucose residue (Figure 3-3).[2] Information on how the trisaccharide is elaborated to the hexasaccharide and the exact role of PglH remains in question. Three probable hypotheses have been presented: PglH transfers a single GalNAc followed by two yet to be identified glycosyltransferases, PglH has polymerase

activity and transfers the remaining three GalNAc residues or PglH and PglJ act alternately to create the hexasaccharide.[1, 2]
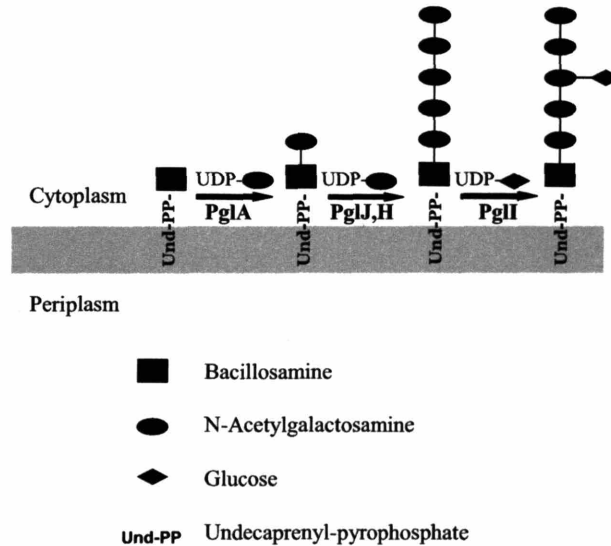

Figure 3-3. The Pgl glycosyltransferases.

Elongation of the Und-PP-Bac intermediate by Pgl A, J, H and I to form the heptasaccharide.

These four glycosyltransferases have very similar sizes and topologies. They have a molecular weight around 40 kD and do not include any predicted transmembrane domains (TMHMM, ExPASY). They bear little homology to the Alg enzymes in the yeast dolichol pathway, which are all integral membrane proteins.[10] However, they do share significant homology with the enzymes involved in the synthesis of the *O*-antigen sugar core.[11]

We are interested in complementing the existing mutational analysis data with *in vitro* biochemical data to define the precise roles of the Pgl glycosyltransferases so as to develop a tractable system for investigating the sequential biosynthetic assembly process that occurs at the membrane interface. Furthermore, the biochemical validation of these transferases will help to

complete the picture of heptasaccharide assembly, setting the stage for using these glycosyltransferases to create native and novel substrates for the bacterial-glycosylation machinery, and providing specific targets for the development of inhibitors that may reduce *C. jejuni* pathogenicity.

Results and Discussion

PglC

PglC was overexpressed, purifed using Ni^{2+}-affinity chromatography and confirmed by SDS-PAGE (Coomassie-stained) and Western-blot analysis (Figure 3-4). The proposed UDP-Bac substrate for PglC was chemically synthesized *via* the coupling of bacillosamine-phosphate to UMP-morpholidate (Chapter 2). Und-P was synthesized from undecaprenol using a previously published procedure, which utilizes phosphoramidite chemistry.[12, 13]
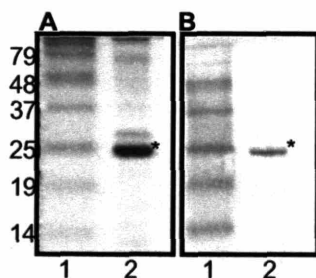

Figure 3-4. Ni-NTA purified PglC.

(A) Coomasie-stained polyacrylamide gel. (B) Anti-T7-Tag western-blot analysis. Molecular weight markers (lane 1). PglC (lane 2).

In one of the postulated models for PglC activity, the enzyme is proposed to facilitate the transfer of bacillosamine-phosphate from UDP-Bac to Und-P to form Und-PP-Bac. This Und-PP-Bac intermediate is then elaborated by PglA, which uses UDP-GalNAc to transfer an α1,3-GalNAc residue to form the GalNAc-α1,3-Bac-α-1-PP-Und disaccharide intermediate (Scheme 3-1).
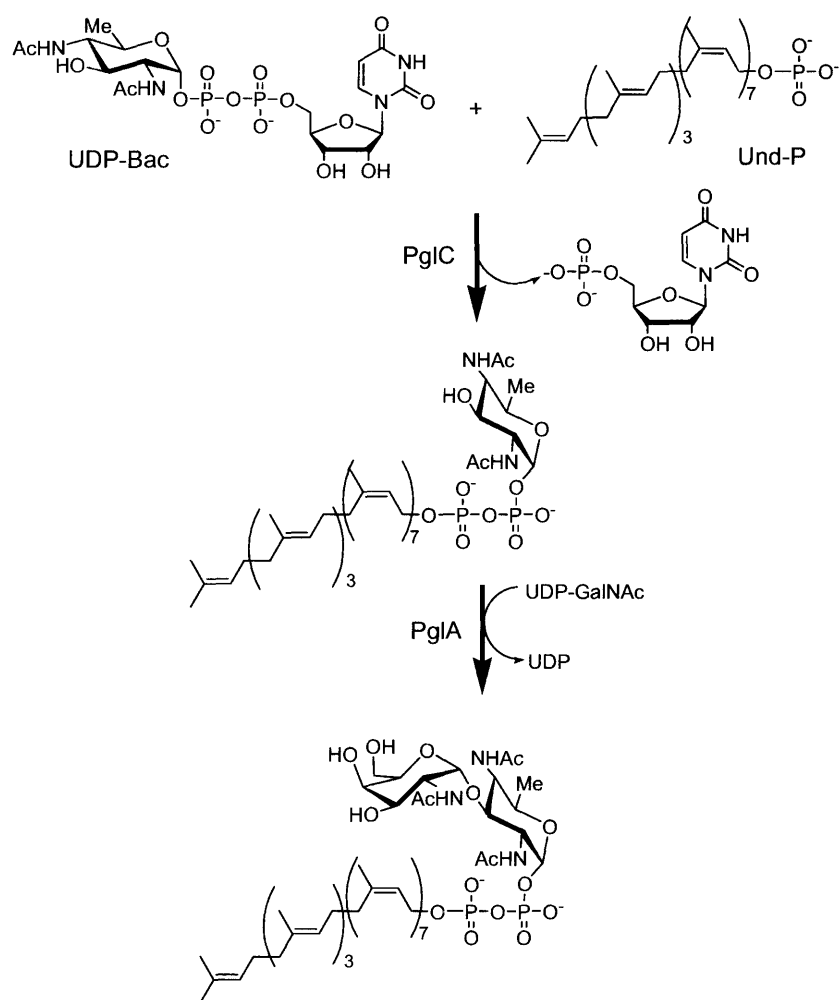
Scheme 3-1. Reactions catalyzed by PglC and PglA.

A typical assay for glycosyltransferase activity involves monitoring the transfer of radiolabeled glycan from the aqueous-soluble UDP-derivative to the organic-soluble Und-PP-derivative. Since radiolabeled UDP-Bac is not readily available through our current synthetic route, the activity of PglC is observed indirectly by coupling the transformation that it effects with that of PglA, which adds a radiolabeled GalNAc moiety to the product of the PglC reaction. UDP-GalNAc is commercially available in radiolabeled form and by coupling the PglC and PglA reactions, the enzyme activity can be monitored by standard radiolabeled assay methods. A schematic of this assay is displayed in Figure 3-5.

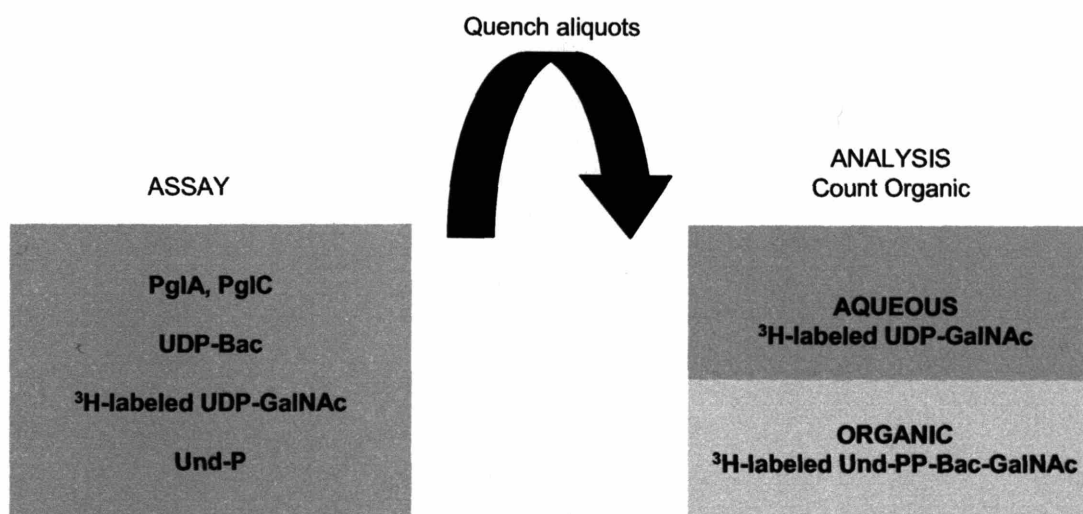**Figure 3-5**. Radiolabeled assay procedure for PglC.

When synthetic UDP-Bac and Und-P are incubated with PglC, PglA and ^{3}H-UDP-GalNAc, a net transfer of radioactivity to the organic layer is observed with time (Figure 3-6). This indicates that both PglC and PglA are acting in concert to produce the radiolabeled GalNAc-α1,3-Bac-α-1-PP-Und disaccharide, which partitions into the organic layer. Control

reactions in which either PglA or PglC were absent, show no net transfer of radioactivity, hence indicating that both enzymes are required for the biosynthesis of the disaccharide product.
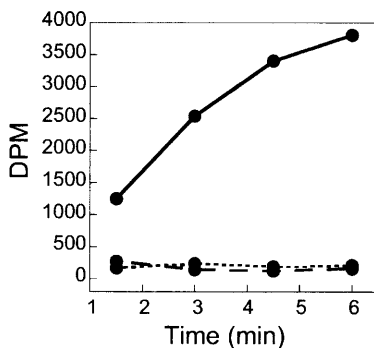

Figure 3-6. PglC accepts UDP-Bac.

Plot of Und-PP-Bac-GalNAc product formation over time with UDP-Bac, UDP-GalNAc and Und-P substrates. Solid line, PglC and PglA. Dotted line, PglA only. Dashed line, PglC only.

In order to further characterize the product from this combined PglC/PglA reaction, the glycan was hydrolyzed from the undecaprenyl-pyrophosphate carrier by acid hydrolysis. The reducing terminus was then labeled with a fluorophore, 2-aminobenzamide (2AB) *via* a reductive amination (Scheme 3-2).
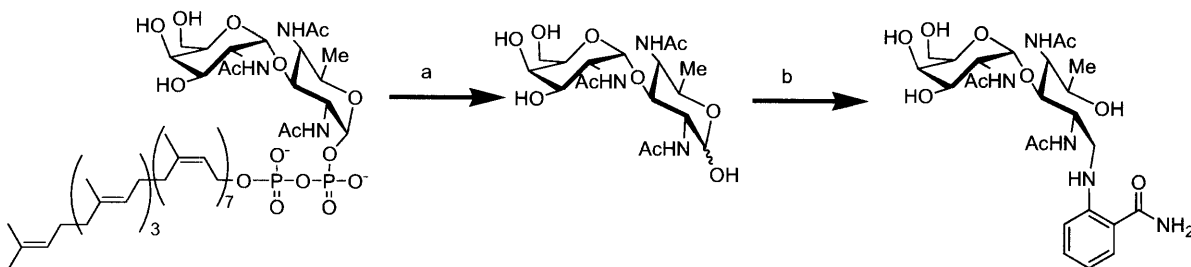
Scheme 3-2. Acid hydrolysis and 2-aminobenzamide-labeling scheme.[a]

[a]Reagents and condition: a) TFA, n-propanol, 50 °C; b) 2-aminobenzamide, sodium cyanoborohydride, DMSO, 60 °C.

97

The resulting 2AB-labeled disaccharide was isolated on a GlykoSepN normal-phase

column and analyzed by MALDI-MS. The HPLC trace shows a peak that elutes at T_R = 22

minutes (Figure 3-7 A) and when subjected to mass-spectral analysis a $[M + Na]^+$ m/z of 592.27^+

is observed, which corresponds to the sodium ion adduct of the disaccharide 2AB-Bac-GalNAc

(where 2AB is 2-aminobenzamide) (Figure 3-7 B).
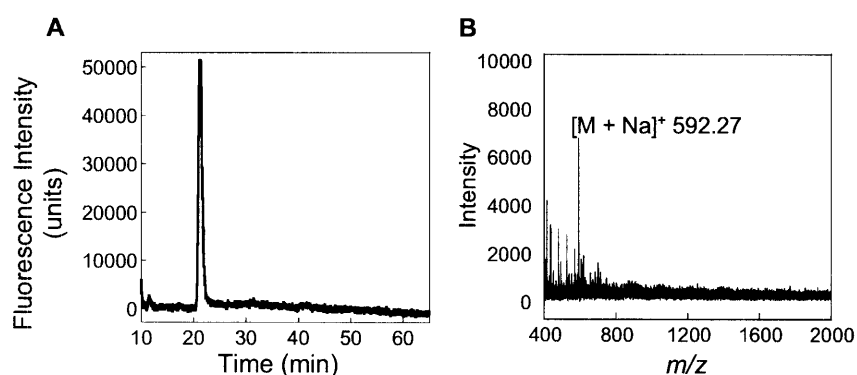
Figure 3-7. HPLC and MALDI-MS analysis of the PglC/PglA reaction.

(A) HPLC analysis of the 2-AB labeled disaccharide. (B) MALDI-MS analysis of the 2-AB

labeled disaccharide. Peak at 592.27 denotes the sodium adduct of the 2AB-labeled disaccharide.

We were interested in clarifying the exact role of PglC in the Pgl pathway. The

radioactive assay results, coupled with the HPLC and MALDI-MS characterization (Figure 3-7)

provide convincing evidence that UDP-Bac is indeed the substrate for PglC. However, it was

necessary to completely rule out the second hypothesis in which UDP-GalNAc or GlcNAc is the

substrate for PglC. Since both these UDP-derivatives are commercially available in radiolabeled

form, they could be directly assayed as substrates for PglC. When compared to the transfer of

radioactivity to the organic phase in the case of UDP-Bac (Figure 3-8 A), there is no radioactivity transferred to the organic phase during the timeframe of the experiment for UDP-GlcNAc and UDP-GalNAc (Figure 3-8 B). This establishes that PglC does not accept UDP-GalNAc or UDP-GlcNAc but does accept UDP-Bac to afford the transfer to Und-P to form Und-PP-Bac. PglC also accepts UDP-6-hydroxybacillosamine (synthesized by M. M. Chen, Imperiali Lab) (Figure 3-8 C) and hence does show some flexibility in the UDP-sugar donor.

Figure 3-8. UDP-sugar substrate specificity of PglA.

(A) Radiolabeled product formation with UDP-Bac. (PglC/PglA coupled assay) (B) No radiolabeled product is formed with UDP-GalNAc or UDP-GlcNAc. (PglC direct assay). (C) Radiolabeled product formation with UDP-Bac-6-OH (PglC/PglA coupled assay).

99

The Pgl glycosyltransferases

PglA, PglH, PglI, and PglJ were overexpressed, purified using Ni^{2+} affinity chromatography, and confirmed by SDS-PAGE (Coomassie-stained) and Western blot analysis (Figure 3-9).
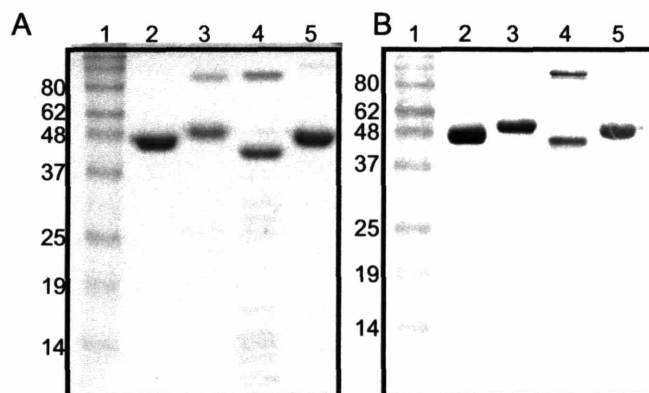

Figure 3-9. Ni-NTA purified glycosyltranferases.

(A) Coomasie stained-polyacrylamide gel. (B) Anti-T7 Western Blot analysis. Molecular weight markers (lane 1). PglA, 43 kD (lane 2). PglH, 41 kD (lane 3). PglI, 36 kD (lane 4). PglJ, 41 kD (lane 5).

Various combinations of the purified enzymes were incubated in the presence of the synthetic Und-PP-Bac (Chapter 2), UDP-Glucose, and UDP-GalNAc. The undecaprenyl-pyrophosphate-linked saccharide products were isolated from the reaction mixture by extraction into organic solvent, and hydrolyzed to remove the Und-PP moiety. The free saccharide was then fluorescently labeled by reductive amination (2-aminobenzamide/sodium cyanoborohydride) for HPLC analysis.

PglA

When PglA alone is reacted, a peak with a retention time of 22 minutes is observed in the
HPLC analysis (Figure 3-10 A). When this peak is collected and subjected to mass spectral
analysis a $[M + Na]^+$ m/z of 592.27^+ is observed, which corresponds to the sodium ion adduct of
the disaccharide 2AB-Bac-GalNAc (where 2AB is 2-aminobenzamide) (Figure 3-10 B). PglH,
PglI, and PglJ all failed to show activity with the undecaprenyl-pyrophosphate-linked
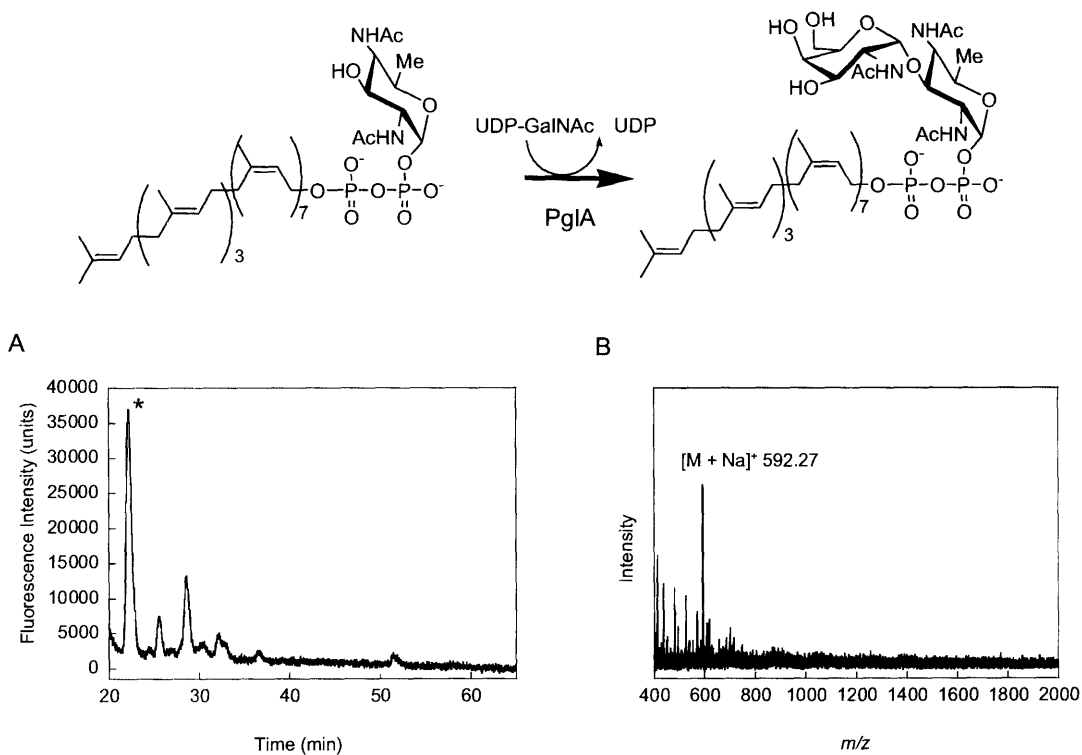monosaccharide (data not shown).

Figure 3-10. HPLC and MALDI-MS analysis of the PglA reaction.

(A) HPLC analysis of the 2-AB labeled disaccharide. (B) MALDI-MS analysis of the 2-AB
labeled disaccharide. Peak at 592.27 denotes the sodium adduct of the 2AB-labeled disaccharide.

To examine the undecaprenyl-pyrophosphate-linked glycan specificity, two non-native

glycosyl acceptors were prepared *via* chemical synthesis: Und-PP-Bac-6-OH and Und-PP-

GlcNAc (Chapter 2). Und-PP-GlcNAc shows product formation that is several-fold less efficient

than the Und-PP-Bac while Und-PP-6-hydroxybacillosamine, which contains a C-6 hydroxyl

group, is accepted significantly more efficiently than Und-PP-GlcNAc (Figure 3-11). These data

are consistent with cellular studies, which show flexibility in PglA activity as GlcNAc is found

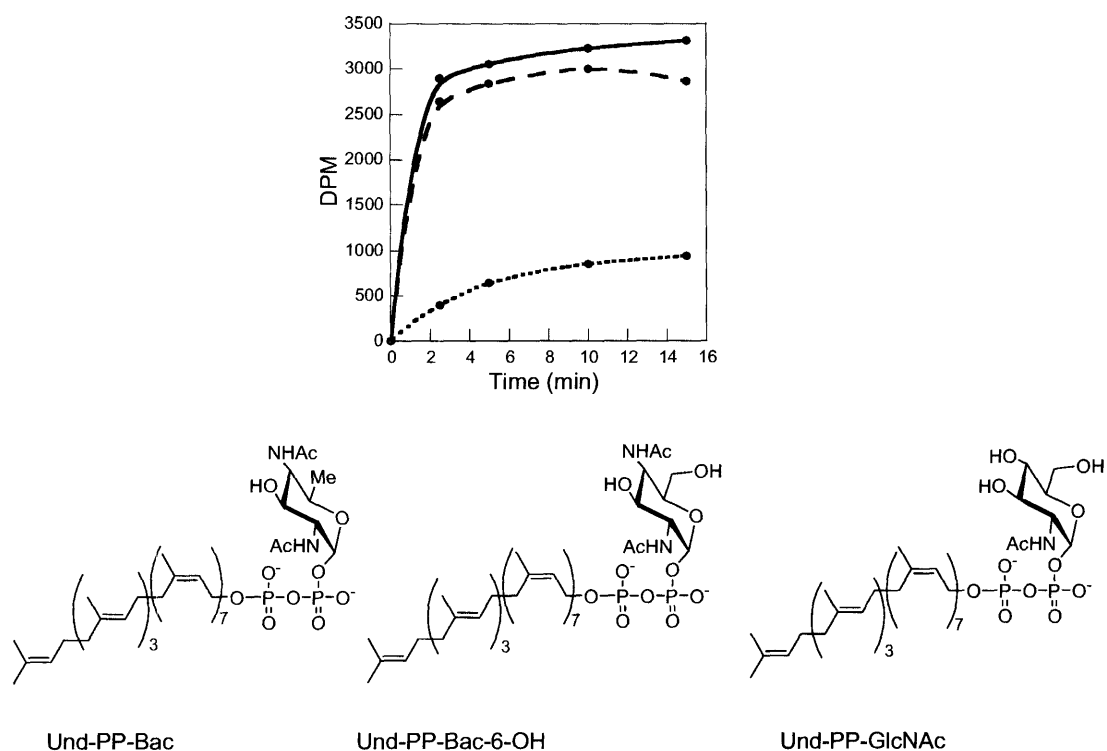in place of bacillosamine when the Pgl cluster is functionally reconstituted in *E. coli*[14].

Figure 3-11. Undecaprenyl-pyrophosphate-linked glycan specificity of PglA.

Plot of Und-PP-disaccharide formation over time using PglA, UDP-GalNAc and the three Und-

PP-monosaccharides shown. Solid line, Und-PP-Bac. Dashed line, Und-PP-Bac-6-OH. Dotted

line, Und-PP-GlcNAc.

We were also interested in exploring the polyisoprene-specificity of PglA. In Chapter 2, the synthesis of two different polyisoprenyl-pyrophosphate-linked compounds, geranylgeranyl-PP-Bac-6-OH and dolichyl-PP-Bac-6-OH was discussed. When PglA was assayed with both of these polyisoprenyl-pyrophosphate-linked compounds, it was found that PglA accepts the dolichyl-linked substrate, but not the shorter geranylgeranyl-linked substrate (Figure 3-12).
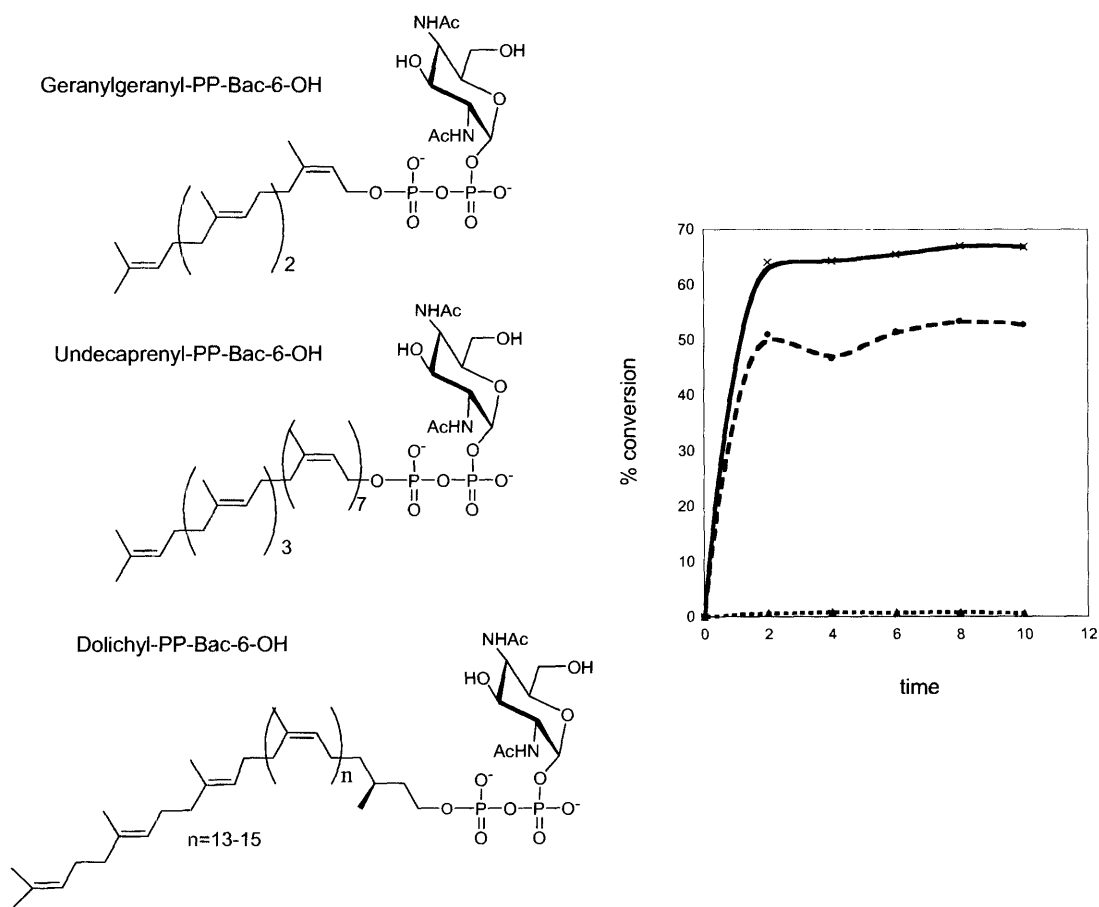

Figure 3-12. Polyisoprene specificity of PglA.

Plot of polyisoprenyl-pyrophosphate-linked disaccharide formation over time using PglA, UDP-GalNAc and the three polyisoprenyl-pyrophosphate-linked monosaccharides shown. Solid line, Undecaprenyl-PP-Bac-6-OH. Dashed line, Dolichyl-PP-Bac-6-OH. Dotted line, Geranylgeranyl-PP-Bac-6-OH.

PglJ

When PglA and PglJ are combined and the products analyzed by HPLC, a peak at 31 minutes is observed, which has a $[M + Na]^+$ m/z of 795.25 (Figure 3-13). This corresponds to the sodium ion adduct of the trisaccharide 2AB-Bac-GalNAc-GalNAc and is consistent with PglJ adding a single GalNAc residue. PglJ is very specific for this step as the combination of PglA and PglH failed to yield anything larger than the disaccharide (data not shown). These results are complementary with the *in vivo* studies, which demonstrated build up of a disaccharide intermediate when PglJ was genetically knocked out. [15]
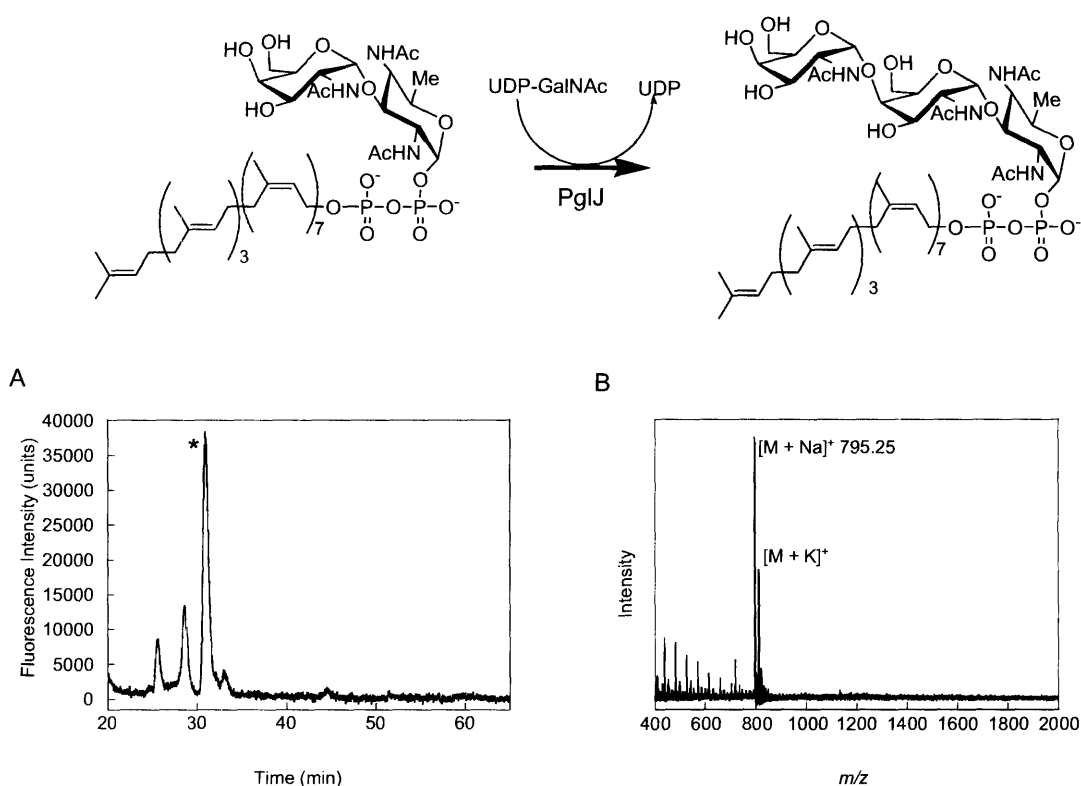

Figure 3-13. HPLC and MALDI-MS analysis of the PglJ reaction.

(A) HPLC analysis of the 2-AB-labeled trisaccharide. (B) MALDI-MS analysis of the 2-AB labeled trisaccharide. Peak at 795.25 denotes the sodium adduct of the 2AB-labeled trisaccharide.

PglH

When PglA, PglH, and PglJ are combined, HPLC analysis reveals a peak with a dramatically increased retention time of 51 minutes (Figure 3-13 A). By mass spectral analysis, this peak has a $[M + Na]^+$ m/z of 1403.26, which corresponds to the hexasaccharide 2AB-Bac-(GalNAc)$_5$ (Figure 3-13 B). In a parallel experiment, when trisaccharide is prepared using PglA and PglJ and then reacted singly with PglH, a hexasaccharide is also observed (data not shown). Taken together, these two studies reveal that PglH is adding the three remaining GalNAc residues.
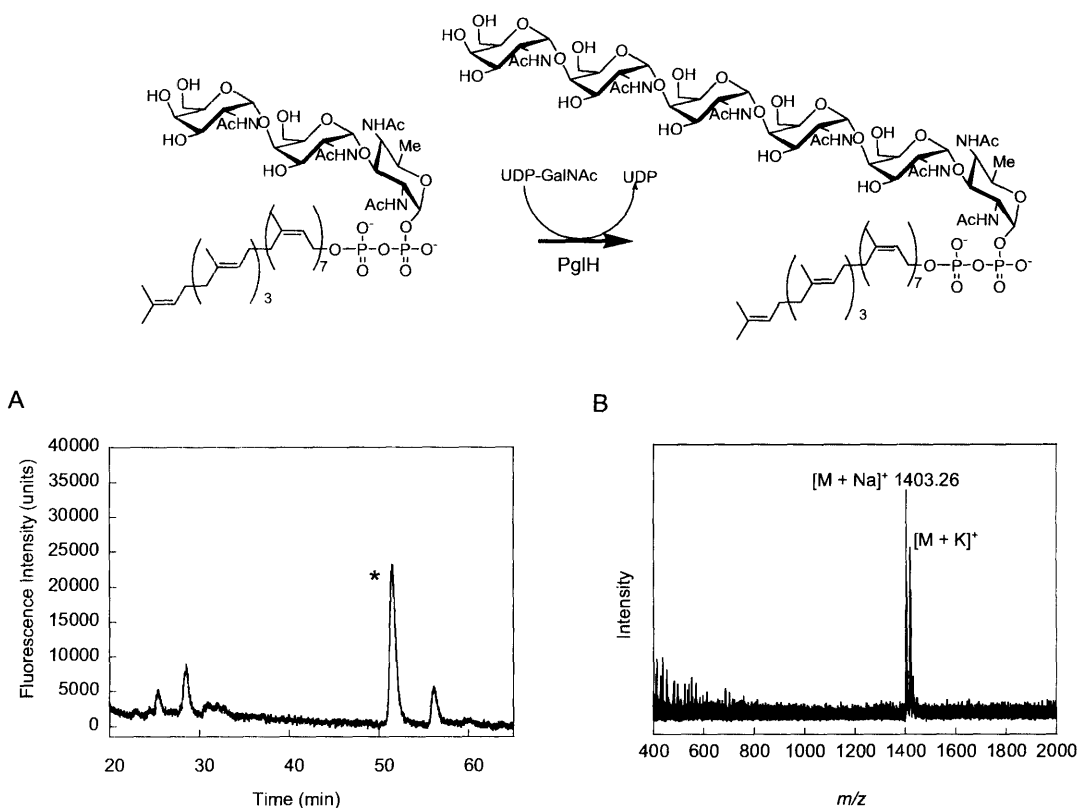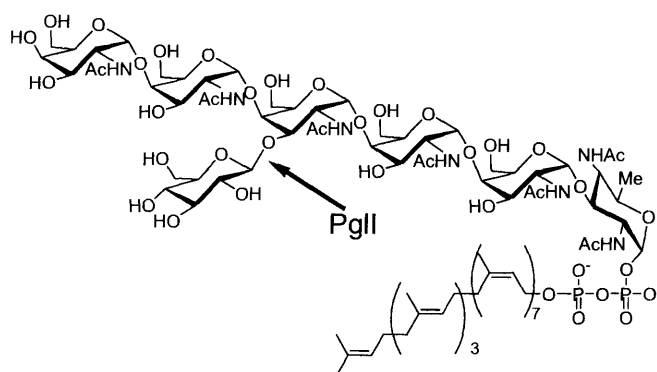

Figure 3-14. HPLC and MALDI-MS analysis of the PglH reaction.

(A) HPLC analysis of the 2-AB labeled hexasaccharide. (B) MALDI-MS analysis of the 2-AB labeled hexasaccharide. Peak at 1403.26 denotes the sodium adduct of the 2AB-labeled hexasaccharide.

PglI

PglI is proposed to add the final branching glucose to complete the heptasaccharide glycosyl donor biosynthesis. When PglI is purified by Ni-NTA chromatography in the presence of Triton X-100, and subjected to SDS-PAGE analysis, two major bands are observed by Coomassie-staining and Western-blot analysis (Figure 3-9). The lower molecular weight band corresponds well to the predicted molecular weight of PglI, while the higher band indicates the presence of protein oligomerization. When added in combination with the other transferases, PglI purified in this manner failed to show any glucosyl transferase activity (data not shown). However, when a bacterial membrane fraction prepared from cells that overexpress PglI is combined with PglA, PglJ and PglH a new peak at 56 minutes is observed (Figure 3-14 A). This peak has a $[M + Na]^+$ m/z of 1567.08 that corresponds to the full heptasaccharide (Figure 3-14 B). Control studies with bacterial membranes prepared using the same over-expression conditions, but using a plasmid that did not have the PglI gene insert, did not show transferase activity confirming that PglI is responsible for the Glc transfer (data not shown).

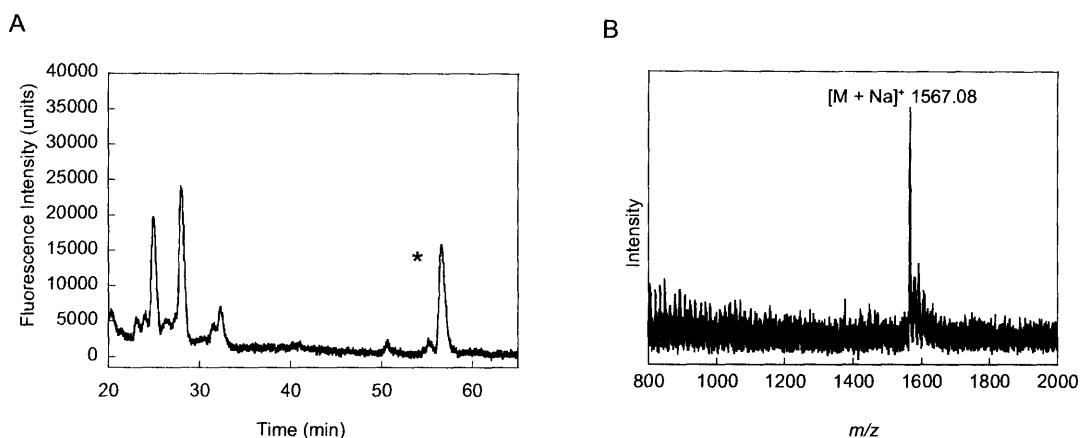A   B

Figure 3-15. HPLC and MALDI-MS analysis of the PglI reaction.

(A) HPLC analysis of the 2-AB-labeled heptasaccharide. (B) MALDI-MS analysis of the 2-AB-labeled heptasaccharide. Peak at 1567.08 denotes the sodium adduct of the 2AB-labeled heptasaccharide.

Conclusion

The processes by which nature assembles complex saccharides is of fundamental biological importance. In the *C. jejuni* system, a heptasaccharide is assembled in a stepwise fashion on an undecaprenyl-pyrophosphate carrier, in a sequential process that involves several glycosyltransferases and a glycophosphoryltransferase.

PglC is a glycophosphoryltransferase that acts in the early steps of the biosynthesis of the undecaprenyl-pyrophosphate-linked heptasaccharide, catalyzing the transfer of bacillosamine-phosphate to undecaprenyl-phosphate to form Und-PP-Bac. In order to study the activity of PglC *in vitro*, significant quantities of UDP-Bac are necessary. Purification from the natural source is not a realistic option due to the low cellular levels of UDP-Bac, and the extreme pathogenicity of

107

C. jejuni, which calls for specialized handling procedures. On the other hand, using the chemical synthesis route presented in Chapter 2, access to the desired compound in milligram quantities is possible.

The *in vitro* biochemical analysis of PglC has revealed that UDP-Bac is the preferred substrate, while UDP-GlcNAc or UDP-GalNAc are not utilized under the conditions examined. This evidence strongly suggests that the later model (Figure 3-2 B) in which a UDP-HexNAc is converted to bacillosamine on the undecaprenyl carrier is not viable. Instead, the enzymes PglF, PglE, and PglD likely convert UDP-HexNAc to UDP-Bac, which is then a substrate for PglC. Favorably, PglC also accepts the UDP-Bac-6-OH derivative, suggesting that the additional N-acetyl group is a major determinant in substrate recognition. Since the synthesis of Bac-6-OH is considerably more facile than that of bacillosamine, this compound represents a suitable alternative for UDP-Bac for future investigations into this process.

The coupling of PglC with PglA shows that PglC works in concert with the other enzymes, and current research focuses on preparation of the undecaprenyl-pyrophosphate-linked heptasaccharide starting from undecaprenyl-phosphate. Furthermore, the confirmation of PglC activity *in vitro* adds crucial clarification to the details of the early steps of heptasaccharide assembly, and is an important milestone in the validation of the *pgl* pathway.

The glycosyltranferases PglA, J, H and I are crucial to the Pgl pathway. PglA is the first glycosyltransferase to act on the Und-PP-Bac substrate to yield the disaccharide. Although *in vivo* reconstitution studies in *E. coli* suggested that PglA had a relaxed substrate specificity, [14] our studies revealed that PglA has strong preference for bacillosamine over GlcNAc *in vitro* consistent with the observation that only bacillosamine is observed in *C. jejuni in vivo*.[3] Interestingly, the intermediate 6-hydroxybacillosamine is accepted almost as well as

bacillosamine indicating that the additional *N*-acetyl group at C-4 is the most important determinant in the specificity of PglA. PglA also accepts a dolichyl-pyrophosphate-linked glycan as the acceptor suggesting that PglA is not hindered by the larger chain length or α-isoprene saturation of the dolichol chain compared to the native undecaprenol. Interestingly, PglA does not accept the geranylgeranyl-pyrophosphate-linked glycan suggesting that although PglA does not include any transmembrane domains, it does interact with the membrane bound regions of the substrate.

PglJ transfers a single sugar while PglH has a polymerase activity adding the three remaining GalNAc residues. Defining the roles of PglJ and PglH was essential for a complete understanding of the glycan assembly process. *In vivo* studies were not able to account for the complete assembly of hexasaccharide, and concluded with three possible scenarios.[15] The studies presented herein clearly indicate that PglH precisely transfers the three terminal GalNAc residues independent of the PglI glucosyl transferase activity, and structural studies are currently underway to investigate this interesting behavior. Finally, PglI is considerably more hydrophobic than the other transferases. In particular, the protein has a modest probability of having a C-terminal transmembrane domain (TMHMM, ExPASY), and when purified in the presence of detergents it shows oligomeric behavior by SDS-PAGE, which is indicative of a poorly stabilized protein. Consequently, when this preparation was used for transferase studies, no activity was observed. However, when a bacterial membrane preparation with PglI was generated it revealed very good activity, highlighting the important role played by native lipids in the stability and activity of the enzyme.

Remarkably, when the various transferases are combined, the intermediate saccharides are not observed; the transfer is complete. This indicates that the sequential activities of the

pathway have been reconstituted and suggests that the enzymes closely interact with each other. Biophysical studies are currently underway to characterize this behavior. Having the exact roles of PglA, PglJ, PglH, and PglI clearly defined helps to complete the picture of glycan assembly in *C. jejuni*, and sets the stage for using this pathway to probe the more fundamental questions surrounding multi-enzyme processes as well as engineering the enzymes to create unnatural saccharides.

Acknowledgements

Experimental

Cloning of the Pgl enzymes

PglC, A, J, H and I were amplified by PCR from *C. jejuni* genomic DNA (ATCC 700819D, designation NCTC 11168) using Pfu Turbo (Stratagene) polymerase, with the oligonucleotides:

Pgl*CBamH I* (CGCGGATCCATGTATGAAAAGTTTTTAAAAGAATTTTTG)

Pgl*CXho I* (CCGCTCGAGGTTCTTGCCATTAAATTTCTCTGTTGTAAC)

Pgl*ABamH I* (CGCGGATCCATGAGAATAGGATTTTTATCACATGCA)

Pgl*AXho I* (TACATTCTTAATTACCCTATCATAAAG)

Pgl*HBamH I* (CGCGGATCCATGATGAAAATAAGCTTTATTATCGCAAC),

Pgl*HXho I* (CCG-CTCGAGGGCATTTTTAACCCTCGGCTATAAGCTTA),

Pgl*IBamH I* (CGCGGATCCATGCCTAAACTTTCTGTTATAGTACCAAC),

Pgl*IXho I* (CCGCTCGAGATTTTTGCATAAAG-CCACCCGAATTTTTG),

Pgl*JBamH I* (CGCGGATCCATGCAAAAATTAGGCATTTTTATTT-ATTC) and Pgl*JXho I* (CCGCTCGAGTCCTAATAAATATTTCAAAGCATCGCGT).

The PCR product was digested with *BamH I* and *Xho I* and cloned into the same sites of vector pET-24a (Novagen). The final gene construct encoded a protein product with a N-terminal T7-Tag and a C-terminal His-Tag.

Protein expression

Starting from a 5 mL overnight culture, PglC, PglA, PglH, PglI, or PglJ were grown at 37°C in LB broth to an OD_{600} of 0.6-0.8. At that point, the temperature was reduced to 16°C and

protein production was induced by the addition of IPTG (1 mM). After 24 hours, the cells were harvested by centrifugation (5000 × g) and frozen at -80°C until needed.

Protein purification

All steps were performed at 4°C. Cell pellets of *E. coli* strains expressing PglC, PglA, PglH, PglI, or PglJ were thawed and resuspended in 5% of the original culture volume in buffer L (50 mM Tris-acetate [pH 8.0], 20 mM imidazole). The cells were then subjected to sonication, followed by the addition of Triton X-100 (1%), and shaking for 10 minutes. Cellular debris and membrane proteins were removed by centrifugation (142,414 × g) for one hour, and the supernatant was loaded on to a column containing Ni-NTA agarose equilibrated with buffer L. After washing with 10 column volumes of buffer W (50 mM Tris-acetate [pH 8.0], 45 mM imidazole), the purified protein was eluted with buffer E (50 mM Tris-acetate [pH 8.0], 250 mM imidazole). Fractions containing at least 500 µg/mL protein were used for enzyme assays, and no further purification of the proteins was undertaken.

Preparation of PglI membrane fraction

All steps were performed at 4°C. The cell pellet of PglI was thawed and resuspended in 5% of the original culture volume in buffer M (50 mM Tris-acetate [pH 8.0], 1 mM EDTA). The cells were then subjected to sonication, unbroken cells were removed by centrifugation at 5,697 × g for 15 minutes, and the membrane fraction was collected by centrifugation at 142,414 × g for 60 minutes. The pellet was washed once with buffer M, centrifuged again, and resuspended in 0.25% of the original culture volume in buffer M. The final suspension was aliquoted and stored at -80°C.

PglC radioactive enzyme assay using UDP-Bac and UDP-6-hydroxybacillosamine

To a tube containing 0.06 mg of dried Und-P, 3 μL of DMSO and 7 μL of 14.3% (v/v) Triton X-100 were added. After vortexing and sonication (water bath), 70 μL of H_2O, 4.5 μL of 1 M Tris-acetate pH 8.0, 1 μL 1 M $MgCl_2$, 5 μL PglC, and 5 μL PglA were added. For the two control reactions, the PglA or PglC samples were replaced with buffer E. The reaction was initiated by the addition of 4.5 μL of a 53 μM solution of UDP-Bac or UDP-6-hydroxy bacillosamine, and 22 μM UDP-GalNAc (2.02 nCi/nmol). Reactions were run at room temperature and 20 μL aliquots were taken at 1.5, 3, 4.5, and 6 minutes. Reactions were quenched by addition to a tube containing 400 μL of 2:1 chloroform : methanol and 200 μL of pure solvent upper phase (PSUP, [15 mL chloroform/240 mL methanol/1.83 g potassium chloride in 235 mL water]). After vortexing briefly the layers were allowed to separate and the aqueous layer removed. The organic layer was washed two times with 400 uL of PSUP and dried under a stream of nitrogen. The residue was redissolved in 200 μL of Solvable (Perkin-Elmer) by vigorous vortexing followed by the addition of 5 mL of Formula 989 scitillation fluid (Perkin-Elmer). The tubes were allowed to rest for 1 hour and counted in a scintillation counter (5 minutes per sample).

PglC Enzyme assay for HPLC and MALDI-MS analysis

To a tube containing 0.06 mg of dried Und-P, 3 μL of DMSO and 7 μL of 14.3% (v/v) Triton X-100 were added. After vortexing and sonication (water bath), 15 μL H_2O, 4 μL 1 M Tris-Acetate pH 8.0, 1 μL 1 M $MgCl_2$, and 5 μL of PglC were added. The reaction was initiated by the addition of 4.5 μL of a 53 μM solution of UDP-Bac or UDP-6-hydroxy bacillosamine, and

22 µM UDP-GalNAc. The reaction was run at room temperature for 120 minutes and quenched by addition to a tube containing 800 µL of 2:1 chloroform : methanol and 160 µL of PSUP. After vortexing for 20 seconds, the tubes were centrifuged briefly and the organic layer (bottom) containing product was removed and dried.

PglA, J, H and I enzyme assays for HPLC and MALDI-MS analysis

80 µL of buffer A (50 mM Tris-Acetate pH 7.0, 3 mM dithiothreitol, 5 mM MgCl$_2$, 1% (v/v) NP-40, 250 mM sucrose) was added to a tube containing 0.06 mg of dried Und-PP-Bac, UDP-GalNAc (4 mM final concentration), and/or UDP-Glucose (0.4 mM final concentration). The mixture was vortexed vigorously and sonicated (water bath). Next 5 µL each of the desired enzymes were added and if necessary, buffer E was added to bring the total volume to 100 µL. Reactions were run at room temperature for 120 minutes and quenched by addition to a tube containing 800 µL of 2:1 chloroform : methanol and 160 µL of pure solvent upper phase. After vortexing for 20 seconds, the tubes were centrifuged briefly and the organic layer (bottom) containing product was removed and dried.

PglA radioactive enzyme assay

To a tube containing 0.06 mg of dried Und-PP-Bac, Und-PP-6-hydroxybacillosamine or Und-PP-GlcNAc, 3 µL of DMSO, and 7 µL of 14.3% (v/v) Triton X-100 were added. After vortexing and sonication (water bath), 77 µL of H$_2$O, 5 µL of 1 M Tris-Acetate pH 8.0, 1 µL 1 M MgCl$_2$, and 1 µL PglA (660 µg/mL) were added. The reaction was initiated by the addition of 6 µL of UDP-GalNAc (2.25 µCi/nmol). 15 µL aliquots were taken at 2.5, 5, 10, and 15 minutes. Und-PP-6-hydroxybacillosamine and Und-PP-GlcNAc were prepared via the coupling of Und-P

with the corresponding sugar phosphate similar to the procedure used in the synthesis of Und-PP-Bac.[‡] 6-hydroxy bacillosamine is an intermediate in the synthetic route to bacillosamine[‡] and GlcNAc-phosphate was synthesized according to a previously reported procedure.[16]

HPLC and MALDI-MS analysis of saccharides

The saccharide sample was hydrolyzed by dissolving the dried sample in *n*-propanol:2M TFA (1:1), heating to 50°C for 15 minutes, and then evaporating to dryness. The hydrolyzed sugars were then labeled with 2-aminobenzamide. To prepare the labeling reagent, a solution of 2-aminobenzamide (5mg) in 100 µL acetic acid:DMSO (1:2.3) was prepared. This dye solution was added to 6 mg of sodium cyanoborohydride and aliquots of 5 µL of this reagent were added to dried samples of hydrolyzed, desalted glycans, and heated to 60°C for 2-4 hours. Post-labeling clean-up was accomplished using GlykoClean S cartridges (ProZyme Inc.) according to the manufacturer's instructions. The labeled glycans were separated on a normal phase analytical HPLC column (GlykoSepN, ProZyme, Inc.) using 50 mM ammonium formate, pH 4.4 (Solvent A) and acetonitrile (Solvent B) as eluents. A gradient of 20-52% solvent A over 80 minutes was used at a flow rate of 0.4 mL/min. The peaks were detected using a fluorescence detector (λ_{ex} = 330 nm, λ_{em} = 420 nm) and collected and characterized with MALDI-MS using a matrix composed of DHB (2,5-dihydrobenzoic acid) in acetonitrile/water with Nafion perfluorinated resin (Aldrich) and trifluoroacetic acid as additives.

References

1. Szymanski, C.M., Logan, S.M., Linton, D., and Wren, B.W. Campylobacter--a Tale of Two Protein Glycosylation Systems. *Trends Microbiol.*, **2003**, *11*, 233-238.

2. Szymanski, C.M., and Wren, B.W. Protein Glycosylation in Bacterial Mucosal Pathogens. *Nat. Rev. Microbiol.*, **2005**, *3*, 225-237.

3. Young, N.M., Brisson, J.R., Kelly, J., Watson, D.C., Tessier, L., Lanthier, P.H., Jarrell, H.C., Cadotte, N., St Michael, F., Aberg, E., and Szymanski, C.M. Structure of the N-Linked Glycan Present on Multiple Glycoproteins in the Gram-Negative Bacterium, Campylobacter Jejuni. *J. Biol. Chem.*, **2002**, *277*, 42530-42539.

4. Wang, L., and Reeves, P.R. Involvement of the Galactosyl-1-Phosphate Transferase Encoded by the Salmonella-Enterica RfbP Gene in O-Antigen Subunit Processing. *J. Bacteriol.*, **1994**, *176*, 4348-4356.

5. Wang, L., Liu, D., and Reeves, P.R. C-Terminal Half of Salmonella Enterica WbaP (RfbP) Is the Galactosyl-1-Phosphate Transferase Domain Catalyzing the First Step of O-Antigen Synthesis. *J. Bacteriol.*, **1996**, *178*, 2598-2604.

6. Bouhss, A., Crouvoisier, M., Blanot, D., and Mengin-Lecreulx, D. Purification and Characterization of the Bacterial MraY Translocase Catalyzing the First Membrane Step of Peptidoglycan Biosynthesis. *J. Biol. Chem.*, **2004**, *279*, 29974-29980.

7. Amer, A.O., and Valvano, M.A. Conserved Aspartic Acids Are Essential for the Enzymic Activity of the WecA Protein Initiating the Biosynthesis of O-Specific Lipopolysaccharide and Enterobacterial Common Antigen in Escherichia Coli. *Microbiol.*, **2002**, *148*, 571-582.

8. Kukuruzinska, M.A., and Robbins, P.W. Protein Glycosylation in Yeast - Transcript Heterogeneity of the Alg7 Gene. *Proc. Natl. Acad. Sci. USA*, **1987**, *84*, 2145-2149.

9. Price, N.P., and Momany, F.A. Modeling Bacterial UDP-HexNAc: Polyprenol-P HexNAc-1-P Transferases. *Glycobiology*, **2005**, *15*, 29r-42r.

10. Burda, P., and Aebi, M. The Dolichol Pathway of N-Linked Glycosylation. *Biochim. Biophys. Acta.*, **1999**, *1426*, 239-257.

11. Samuel, G., and Reeves, P. Biosynthesis of O-Antigens: Genes and Pathways Involved in Nucleotide Sugar Precursor Synthesis and O-Antigen Assembly. *Carbohydr. Res.*, **2003**, *338*, 2503-2519.

12. Ye, X.Y., Lo, M.C., Brunner, L., Walker, D., Kahne, D., and Walker, S. Better Substrates for Bacterial Transglycosylases. *J. Am. Chem. Soc.*, **2001**, *123*, 3155-3156.

13.   Branch, C.L., Burton, G., and Moss, S.F. An Expedient Synthesis of Allylic Polyprenyl Phosphates. *Synth. Comm.*, **1999**, *29*, 2639-2644.

14.   Wacker, M., Linton, D., Hitchen, P.G., Nita-Lazar, M., Haslam, S.M., North, S.J., Panico, M., Morris, H.R., Dell, A., Wren, B.W., and Aebi, M. N-Linked Glycosylation in Campylobacter Jejuni and Its Functional Transfer into E. Coli. *Science*, **2002**, *298*, 1790-1793.

15.   Linton, D., Dorrell, N., Hitchen, P.G., Amber, S., Karlyshev, A.V., Morris, H.R., Dell, A., Valvano, M.A., Aebi, M., and Wren, B.W. Functional Analysis of the Campylobacter Jejuni N-Linked Protein Glycosylation Pathway. *Mol. Microbiol.*, **2005**, *55*, 1695-1703.

16.   Tai, V.W.F., and Imperiali, B. Substrate Specificity of the Glycosyl Donor for Oligosaccharyl Transferase. *J. Org. Chem.*, **2001**, *66*, 6217-6228.

Chapter 4:

Investigating the activity of PglB, the oligosaccharyl transferase of *Campylobacter jejuni*

Introduction

Oligosaccharyl transferase (OT) is the enzyme responsible for the transfer of an oligosaccharide from a polyisoprenyl-pyrophosphate-linked glycan donor to the asparagine side chain within the Asn-Xaa-Ser/Thr sequon of nascent proteins.[1] In eukaryotes, the OT complex comprises 8 membrane-bound protein subunits.[2,3] The complexity of this system has hindered in-depth investigation into the catalytic mechanism of this enzyme. Recent genetic and biochemical studies in yeast have demonstrated that the Stt3p protein is an essential component of the OT complex.[4,5] Homologs of *STT3* are found in eukaryotic and archaeal systems (Table 4-1), but until recently, no homologs were found in bacteria. The recent discovery of a *STT3* homolog, *pglB*, in the genome of *C. jejuni* suggested the presence of an *N*-linked protein glycosylation system in this bacterium.[6]

Domain	Species	Conserved sequence
Archaeal	*Pyrococcus abyssi (976 aa)*	ATATS**WWD**YGYWIE
Archaeal	*Pyrococcus horikoshii (976 aa)*	ATATS**WWD**YGYWIE
Archaeal	*Archaeoglobus fulgidus (593 aa)*	YAVLS**WWD**YGYWIL
Eukaryal	*Saccharomyces cerevisiae (718 aa)*	SKVAA**WWD**YGYQIG
Eukaryal	*Arabidopsis thaliana (779 aa)*	DKVAS**WWD**YGYQTT
Eukaryal	*Mus musculus (823 aa)*	ARVMS**WWD**YGYQIA
Eukaryal	*Drosophila melanogaster (774 aa)*	ARVMS**WWD**YGYQIA
Eukaryal	*Anopheles gambiae (806 aa)*	ARVMS**WWD**YGYQIA
Eukaryal	*Caenorhabditis elegans (757 aa)*	ARVMS**WWD**YGYQIA
Eukaryal	*Schizosaccharomyces pombe (752 aa)*	TKVMS**WWD**YGYQIA
Eukaryal	*Toxoplasma gondii (723 aa)*	ARIMS**WWD**YGYQAT
Eukaryal	*Leishmania major (833 aa)*	ARVLA**WWD**YGYQIT
Bacterial	***Campylobacter jejuni (713 aa)***	DYVVT**WWD**YGYPVR
Archaeal	*Methanobacterium thermoautotrophicum (845aa)*	TVVMS**WWD**FGHLFA
Archaeal	*Pyrococcus furiosus (743 aa)*	DIVLT**WWD**WGHFVT
Archaeal	*Pyrococcus horikoshii (758 aa)*	DVILA**WWD**WGHFIT
Archaeal	*Methanococcus jannaschii (933 aa)*	SVITC**WWD**NGHIYT

Table 4-1. Alignment of conserved residues from putative oligosaccharyl transferase homologs from archaeal, bacterial and eukaryal domains. (Figure adapted from Wacker et al.)[7]

Figure 4-1 shows a hydropathy plot of PglB and Stt3p (TMHMM, ExPasy), which clearly demonstrates that the predicted membrane topology is highly conserved. PglB comprises a large hydrophobic domain located at the N-terminus that accounts for approximately 2/3 of the protein. This region is proposed to include 10 to 12 transmembrane domains. In contrast, the C-terminal domain is highly hydrophilic and contains the WWDYGY motif (Table 4-1), which is highly conserved among the Stt3 protein family.
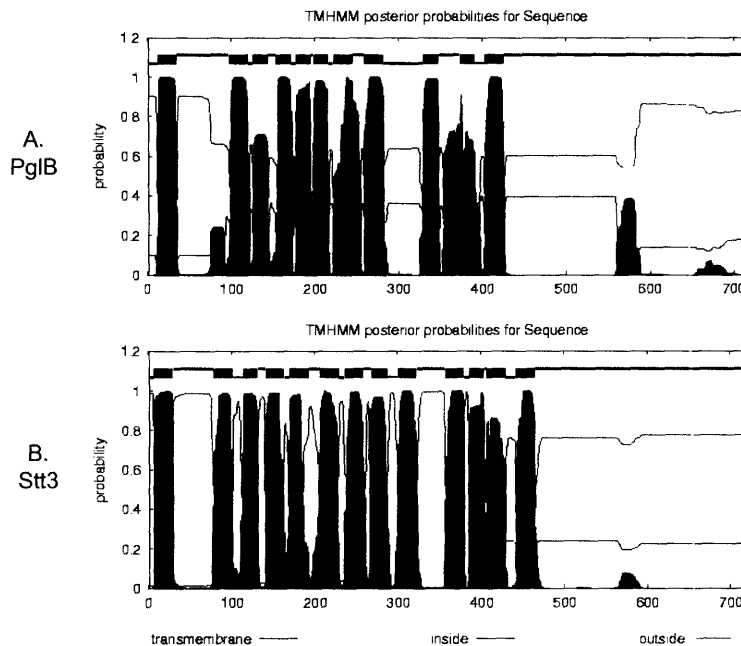

Figure 4-1. Membrane topology of PglB and Stt3p.

Hydropathy plot of (A) PglB and (B) Stt3p, showing that the membrane topology is highly conserved (TMHMM, ExPASY).

The substrate for the PglB-catalyzed oligosaccharide transfer is the undecaprenyl-pyrophosphate-linked heptasaccharide that is biosynthesized on the cytosolic side of the

periplasmic membrane (Chapter 3). A *C. jejuni, pglB* mutant strain shows changes in immunoreactivity that is in agreement with the phenotype expected for reduced glycosylation.[7] When the Pgl cluster is reconstituted in *E.coli,* two *Campylobacter*-derived proteins AcrA and PEB3 are shown to be glycosylated with the heptasaccharide structure. However, if the highly conserved WWDYGY sequence in PglB is mutated, the proteins are produced in unglycosylated form.[7] These data suggest a direct involvement of PglB in the generation of *N*-linked glycoproteins in *C. jejuni*.
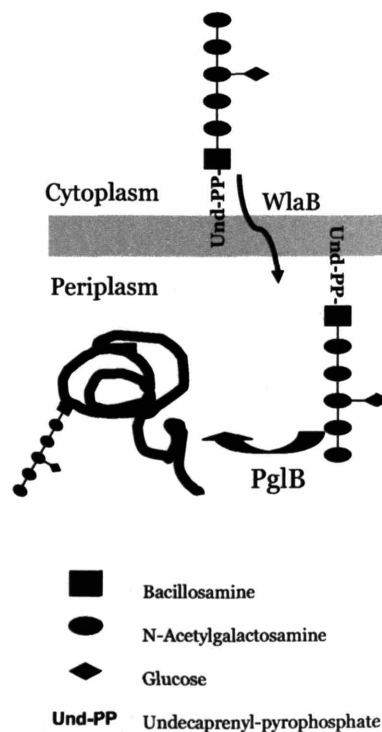
Figure 4-2. Glycoprotein biosynthesis by PglB.

C. jejuni glycosylation occurs at the Asn-Xaa-Ser/Thr acceptor sequence.[8] Both serine and threonine can serve as the hydroxyl amino acid in the acceptor sequence and proline as

121

residue X prevents glycosylation. Not all potential sites are glycosylated indicating that the N-X-S/T sequon is necessary but not sufficient for glycosylation to occur.[8] These observations are in agreement with the parallel process of eukaryotic glycosylation.[1]

The native substrate for PglB is the undecaprenyl-pyrophosphate-linked heptasaccharide containing bacillosamine as the first sugar. However, if *E. coli* cells are engineered so that the *N*-glycosylation pathway and the lipopolysaccharide biosynthesis pathway converge, PglB is shown to transfer O-polysaccharide from the undecaprenyl-pyrophosphate carrier to an acceptor protein.[9] A variety of O-antigen glycan structures are transferred, demonstrating the substrate flexibility of PglB. When individual enzymes in the Pgl pathway are genetically knocked-out, truncated saccharides are transferred to protein.[10] This suggests that PglB also displays flexible specificity with respect to the length of the glycan it transfers.

The requirement for the canonical consensus peptide sequence and the polyisoprenyl-pyrophosphate-linked glycan substrate in bacterial glycosylation suggests that the mechanism of oligosaccharide transfer in *C. jejuni* and eukaryotes is very similar. It is proposed that PglB fulfills the oligosaccharyltransferase-function in the prokaryote and Stt3 represents the catalytic subunit of the eukaryotic enzyme complex. Since the mechanism of action of the eukaryotic OT complex is poorly understood due to its significant complexity, PglB provides a simpler model system for studying this intriguing reaction in more depth.

Results and Discussion

In this chapter, work done towards establishing a system for demonstrating and studying PglB activity *in vitro* is described. This *in vitro* system utilizes an *E. coli* membrane-fraction

overexpressing PglB in the presence of a truncated polyisoprenyl-pyrophosphate-linked glycan donor and a peptide acceptor derived from known *C. jejuni* glycoproteins.

Substrates for *N*-linked glycosylation *in vitro*

In order to define the optimal peptide substrate, sequence alignments were performed with two known periplasmic proteins, PEB3 and AcrA, which have been established to be glycosylated in *C. jejuni*.[7] AcrA is glycosylated at Asn123, while PEB3 is glycosylated at Asn90.[10, 11] When those regions are aligned, the first three amino acids preceding the Asn are strictly conserved. However the residue after the Asn appears to be variable. The next two residues, Ser and Lys are also highly conserved in these two proteins. In order to create the consensus sequence for glycosylation, all conserved residues were retained to create the KDFNVSKA-octapeptide sequence, which was prepared with a free N-terminus and a C-terminal amide by standard solid-phase peptide synthesis (Figure 4-3).

PEB3 (N90)   K D F N V S K I K
AcrA (N123)   K D F N R S K A L
Consensus   K D F N V S K A

Figure 4-3. A peptide substrate for PglB.

Sequence alignment of residues flanking the glycosylation sites of PEB3 and AcrA and the resulting consensus peptide.

The polyisoprenyl-pyrophosphate-linked sugar substrate for PglB was obtained chemo-enzymatically, using the glycosyltransferase PglA, which adds a GalNAc to the synthetically-

obtained Und-PP-Bac in high yield (also accepts Und-PP-GlcNAc and Und-PP-6-hydroxybacillosamine to a lesser extent) (Chapter 3). Und-PP-Bac is obtained *via* chemical synthesis (Chapter 2) and radiolabeled-disaccharide substrate can be readily prepared by using commercially available UDP-^{3}H-GalNAc as the substrate for the PglA reaction (Scheme 4-1 A). The disaccharide was selected as the minimal glycan donor for PglB based upon precedent from the yeast OT, which readily accepts a truncated disaccharide substrate *in vitro*.[1, 3] If the PglB reaction proceeds as anticipated, a glycosylated peptide containing the tritium-labeled disaccharide will be formed (Figure 4-1 B).
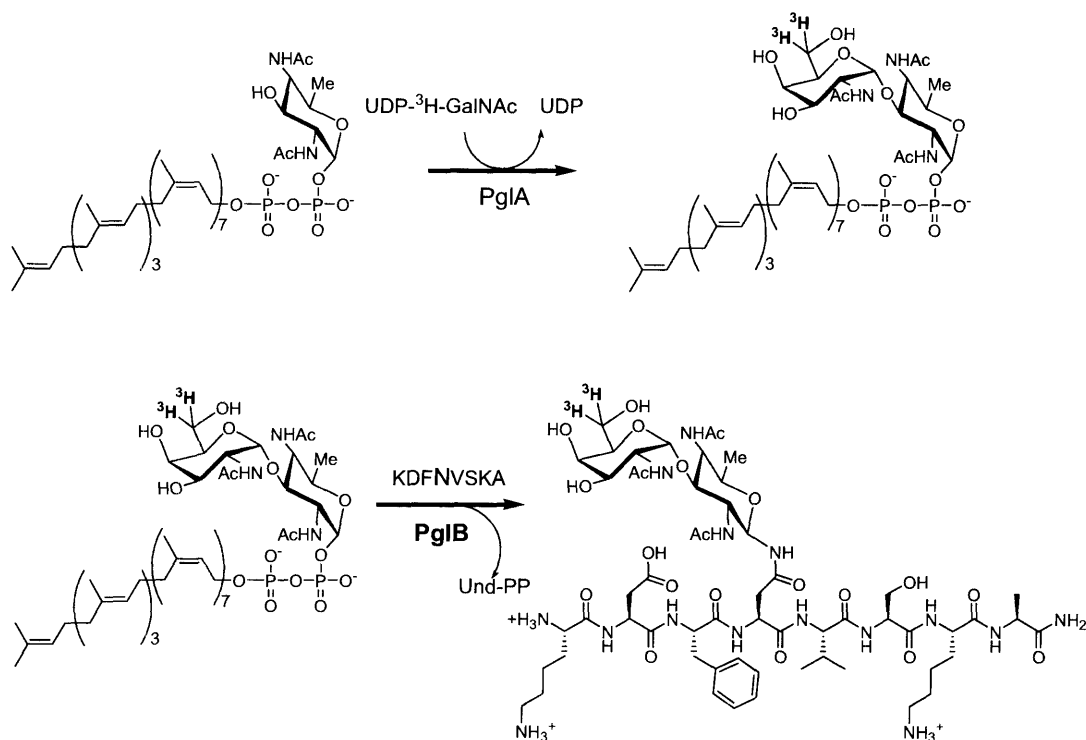
Scheme 4-1. Overview of PglA and PglB reactions.

(A) Synthesis of radiolabeled undecaprenyl-linked disaccharide using PglA, synthetic undecaprenyl-pyrophosphate bacillosamine, and tritiated UDP-GalNAc.

(B) Overview of desired PglB reaction using the polyisoprenyl-pyrophosphate-linked disaccharide and the peptide based on the consensus sequence.

124

Preparation of PglB

Two PglB constructs were prepared. The native, 83 kD PglB protein was cloned into a vector such that the final construct was a T7·Tag-PglB-(His)$_6$ (T7·Tag is for antibody recognition). In addition, since the bacterial, fungal, and yeast OTs all contain a highly conserved, 6-residue signature sequence WWDYGY (Figure 4-4 A) a second homologous construct, in which two of these residues were mutated, was also prepared. According to the work of Aebi and co-workers, if the WWDYGY signature sequence is mutated to WAAYGY, PglB is no longer active *in vivo*.[7] Both proteins were over-expressed in *E. coli* and a bacterial membrane preparation of each was used for activity assays. The anti-T7 western-blot analysis shown in Figure 4-4 B confirms that both the wild type and mutant are expressed in *E. coli*.

A

Wild Type   **W   W   D   Y   G   Y**
Mutant   **W   A   A   Y   G   Y**

B

1   2
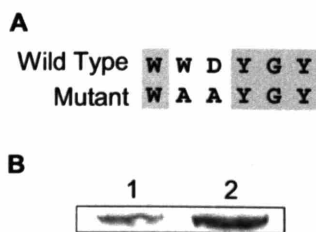
Figure 4-4. Overexpression of PglB, wild-type and mutant.

(A) Sequence alignment of the highly conserved OT hexa-peptide motif. The PglB mutant has two alanine substitutions within this motif.

(B) Anti-T7·Tag Western blot of bacterial membranes overexpressing PglB and the mutant counterpart. Lane 1, Wild type; Lane 2, Mutant.

Both the mutant and wild-type PglB proteins were assayed for oligosaccharyl transferase activity. The assay used was one that was already established for assaying yeast microsomes for

OT activity.[3, 12] Briefly, PglB is added to a solution containing the peptide and glycan substrates. An aqueous/organic phase separation partitions the peptide into the aqueous phase, while the tritiated undecaprenyl-pyrophosphate-linked saccharide starting material remains in the organic phase. If oligosaccharyl transferase activity is observed there will be a net transfer of radioactivity from the organic phase into the aqueous phase (Figure 4-5).

Quench aliquots

ASSAY

PglB

Peptide substrate

^{3}H-GalNAc-Bac-PP-Und

ANALYSIS
Count Aqueous

AQUEOUS
^{3}H-labeled glycopeptide

ORGANIC
^{3}H-GalNAc-Bac-PP-Und
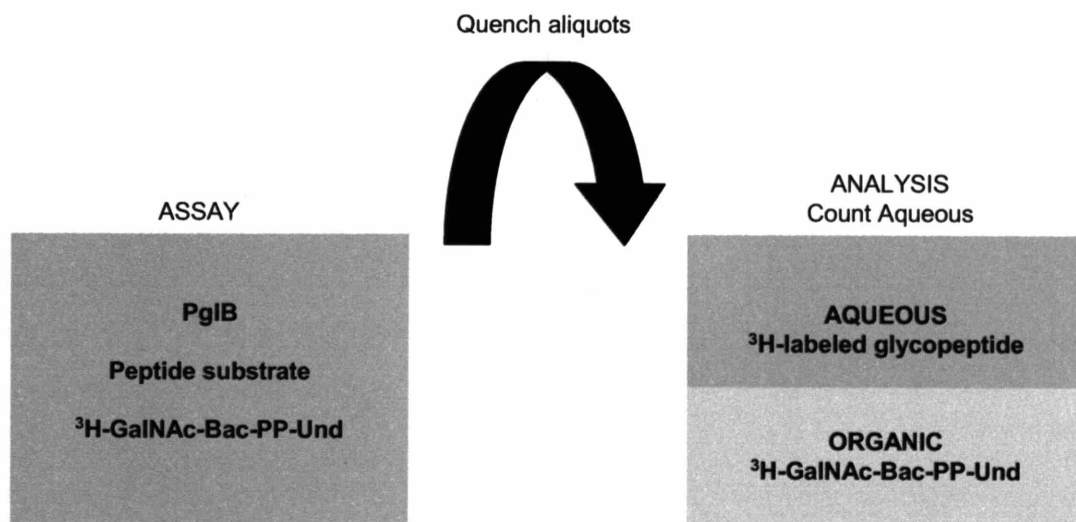
Figure 4-5. Radiolabeled assay procedure for PglB.

The results from this assay are shown in Figure 4-6. These data clearly show that the wild type PglB readily transfers radioactive carbohydrate to an aqueous soluble fraction whereas the mutant shows no comparable activity. Furthermore, the constant slope of the mutant-PglB rate profile confirms that there is negligible hydrolysis of the Und-PP-Bac pyrophosphate within the timeframe of the experiment.
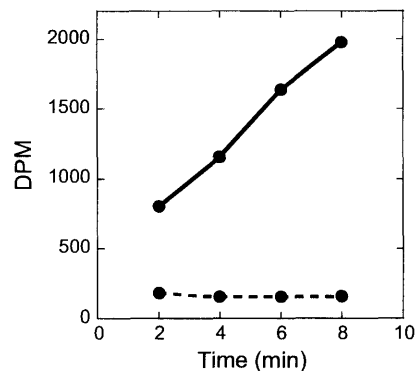
Figure 4-6. Radiolabeled assay data for PglB.

Plot of glycopeptide product formation as a function of time. Solid line, Wild type. Dashed line, Mutant.

In order to further confirm the presence of glycopeptide, the peptide products of reactions run for extended periods (12 hours) were subjected to HPLC analysis. With wild-type PglB, the peptide product elutes from the reverse-phase column at $T_R = 17.5$ minutes (Figure 4-7 A), while the peptide product from a similar incubation using the mutant PglB elutes at $T_R = 18.0$ minutes (Figure 4-7 C). When these peaks are collected and subjected to mass spectral analysis, a mass corresponding to the glycopeptide is observed for the wild type (Figure 4-7 B) while the mutant analysis shows only the unglycosylated peptide (Figure 4-7 D). The presence of single peaks in the mass spectra shows that the glycopeptide product is not modified by host transferases or proteases. Interestingly, the presence of only a single product in the mass spectrum coupled with the appearance of a single HPLC peak suggests that there is complete conversion to glycopeptide.
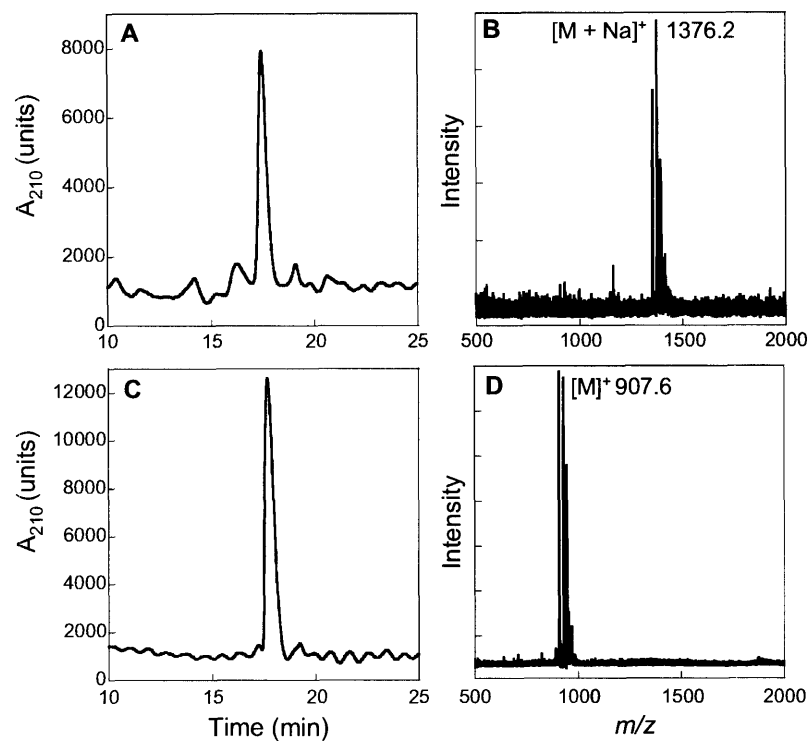
Figure 4-7. HPLC analysis of the PglB reaction.

Reverse-phase HPLC traces of peptide and glycopeptide products. (These experiments were performed with the more synthetically accessible 6-hydroxybacillosamine derivative).

(A) HPLC trace of peptide after PglB wild type reaction.

(B) MALDI-MS of glycopeptide product.

(C) HPLC trace of peptide after PglB mutant reaction.

(D) MALDI-MS of unglycosylated peptide product.

Furthermore if radiolabeled glycan substrate is used, an abundant radioactive peak at a similar retention time as the glycopeptide (Figure 4-8 A), confirms that the radioactivity in the

aqueous layer is a result of radiolabeled glycopeptide and not a hydrolysis product, which would elute very early (T_R < 4 minutes) from the reverse-phase column (Figure 4-8 A arrow). In addition, the assay including the mutant PglB, does not show an abundant radioactive peak (Figure 4-8 B) further confirming that it is deficient in transferase activity.
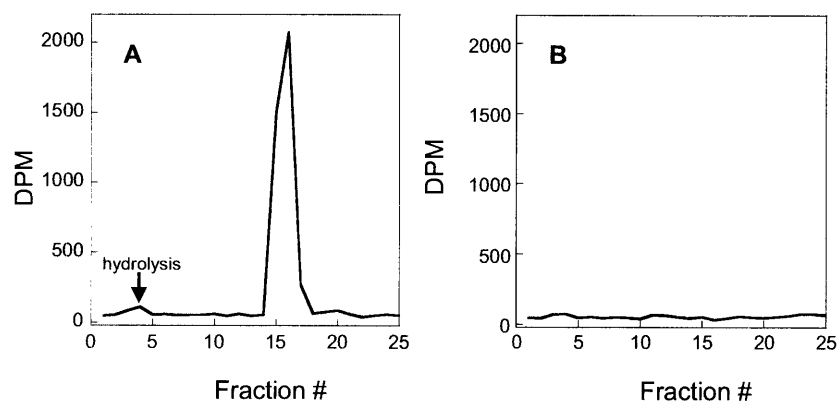
Figure 4-8. Radiolabeled HPLC analysis of the PglB reaction.

Radioactive HPLC traces of peptide and glycopeptide products. (These experiments were performed with the more synthetically accessible 6-hydroxybacillosamine derivative).

(A) Radioactive HPLC trace of wild type PglB reaction.

(B) Radioactive HPLC trace of mutant PglB reaction.

Utilization of diverse undecaprenyl-pyrophosphate-linked glycans by PglB

Three disaccharide substrates were prepared using a chemo-enzymatic approach. As presented previously, synthetic Und-PP-bacillosamine, Und-PP-6-hydroxybacillosamine and Und-PP-GlcNAc were reacted with PglA and UDP-GalNAc to form the corresponding radiolabeled disaccharides (Chapter 3). These polyisoprene-linked disaccharides were then

129

assayed with PglB to determine the specificity of PglB for the saccharide proximal to the pyrophosphate moiety. The results from this assay are shown in Figure 4-9. In these studies, it appears that PglB will accept the unnatural 6-hydroxybacillosamine and GlcNAc analogs, although the bacillosamine substrate appears to be the most efficient of the three. These data agree with *in vivo* studies, which have shown that PglB can transfer several structurally different O-antigen saccharides to protein. [9]
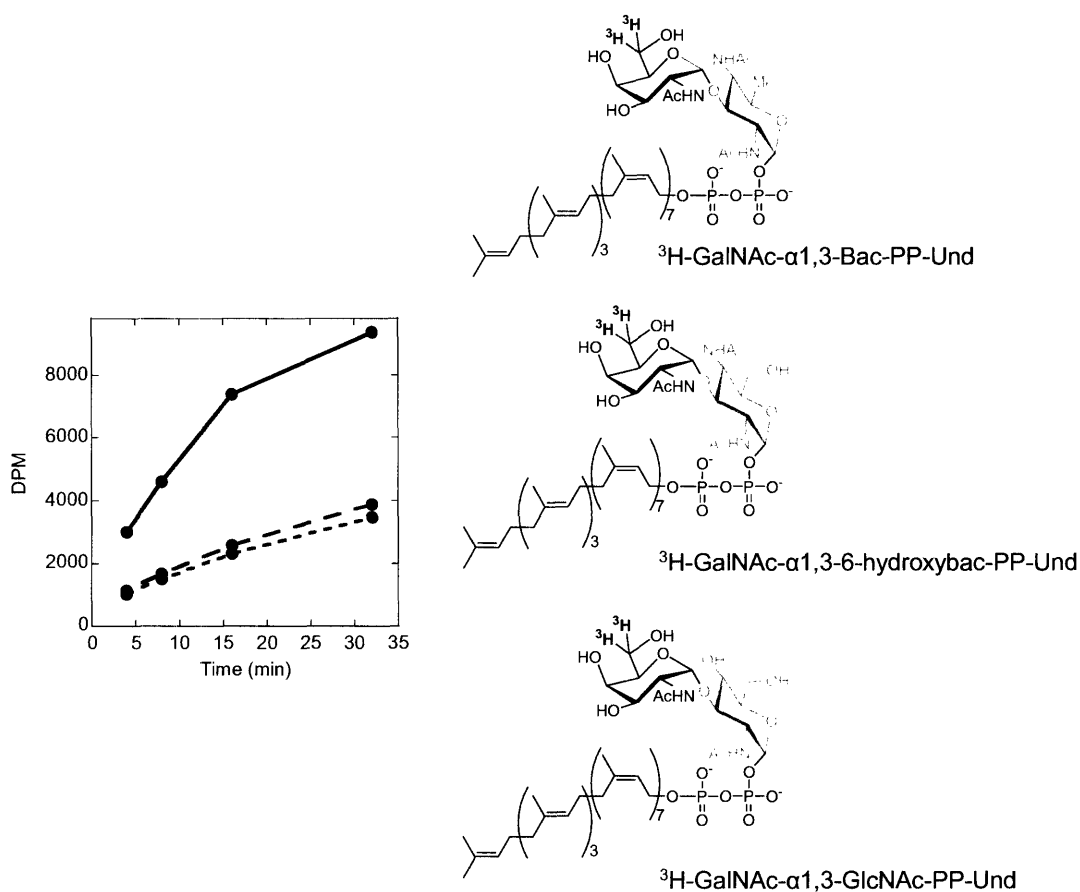
Figure 4-9. Undecaprenyl-pyrophosphate linked disaccharide (^{3}H-GalNAc-X-PP-Und) specificity of PglB. Solid line, X= bacillosamine; Dashed line, X= GlcNAc; Dotted line, X= 6-hydroxybacillosamine.

The native substrate for PglB is a heptasaccharide, but we have shown that PglB readily accepts a disaccharide *in vitro*. Mutagenesis studies of proteins in the Pgl pathway have resulted in the transfer of truncated saccharides to protein, strongly suggesting that PglB has low specificity for the substrate length when functionally reconstituted in *E. coli*.[10] *In vitro* we can access truncated substrates using various combinations of glycosyltransferases in the *pgl* pathway. With PglA we can access the disaccharide (GalNAc-Bac), with PglA and J, we form the trisaccharide (GalNAc$_2$-Bac), with PglA, J and H, we form hexasaccharide (GalNAc$_5$-Bac) and finally, with PglA, J, H and I we form the heptasaccharide (GalNAc$_2$(Glc)GalNAc$_3$-Bac) (Chapter 3). Studies with these intermediate sugars *in vitro*, demonstrate that PglB readily accepts these substrates (Figure 4-10), further reinforcing the observations made by Linton *et al.*[10] in the *in vivo* system. Since radioactive bacillosamine is not currently available to us using our synthetic scheme, a similar study on the transfer of monosaccharide could not be performed. However, using HPLC analysis and mass spectral analysis, it was shown that PglB also accepts the Und-PP-linked monosaccharide, bacillosamine, in addition to the di-, tri-, hexa- and heptasaccharide substrates (data not shown).
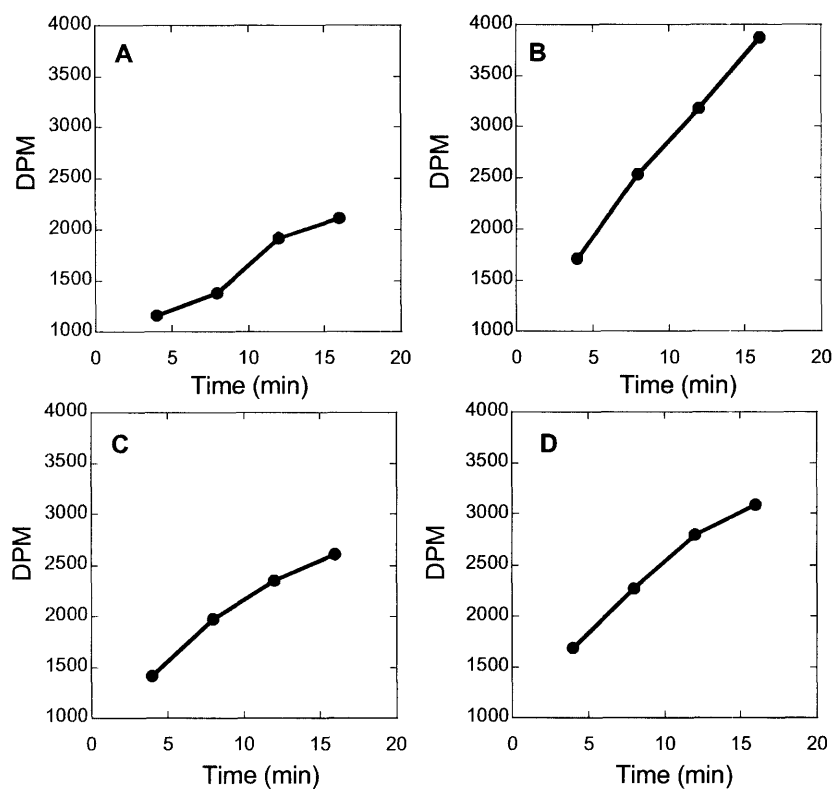
Figure 4-10. PglB accepts glycans of varying length.

Utilization of various undecaprenyl-pyrophosphate linked saccharide intermediates. The more readily available 6-hydroxybacillosamine substrate was used for this study in place of the native bacillosamine. The variation in the levels of DPM incorporation reflects the different amounts of polyisoprenyl-phosphate-linked sugar substrate in each reaction; conclusions about the relative reaction rates cannot be made.

(A) ^{3}H-GalNAc-Bac-PP-Und.

(B) GalNAc-^{3}H-GalNAc-Bac-PP-Und.

(C) (GalNAc)$_4$-^{3}H-GalNAc-Bac-PP-Und.

(D) (Glc)-(GalNAc)$_4$-^{3}H-GalNAc-Bac-PP-Und.

Polyisoprene specificity of PglB

In order to study the specificity of PglB towards the polyisoprene that acts as the carrier for the glycan, two substrates, dolichyl-PP-Bac-6-OH and geranylgeranyl-PP-Bac-6-OH were synthesized (Chapter 2). Dolichol is of longer chain length (15-20 isoprene units) compared to the native undecaprenyl (11 isoprene units) and also contains a saturated α-isoprene unit. Geranylgeraniol is of shorter chain length (4 isoprene units) and different geometry compared to undecaprenol. Using PglA, the radiolabeled dolichyl-PP-Bac-6-OH-GalNAc substrate was accessed. Using the standard radioactive assay, it appears that PglB accepts this dolichol-linked substrate, albeit at lower efficiency than the native undecaprenyl-linked substrate (Figure 4-11).
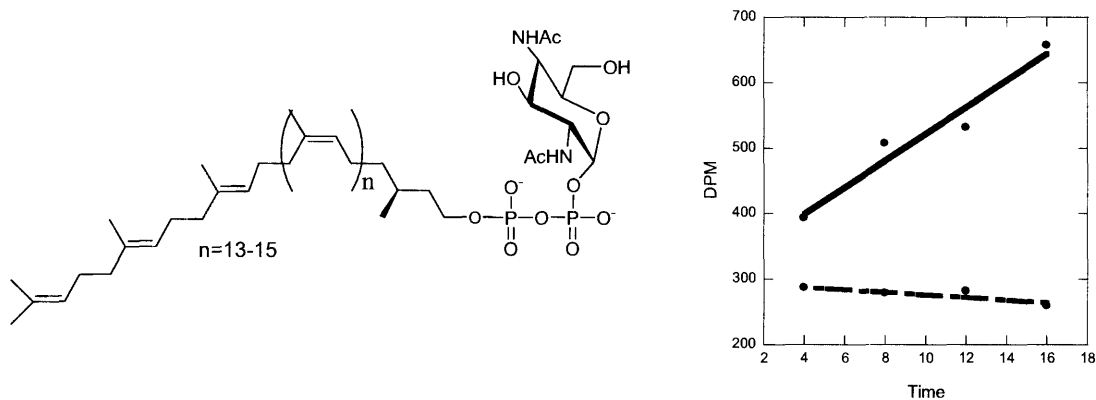
Figure 4-11. PglB accepts the dolichyl-pyrophosphate-linked disaccharide. Radiolabeled glycopeptide formation using dolichyl-PP-Bac-6-OH-GalNAc as the substrate. Solid line, Dol-PP-Bac-6-OH with PglB. Dashed line, Dol-PP-Bac-6-OH, no PglB.

PglA does not accept the shorter chain geranylgeraniol, hence it was not possible to chemo-enzymatically access the radiolabeled disaccharide substrate for the PglB assay. However, since PglB accepts the polyisoprenyl-pyrophosphate-linked monosaccharide substrate,

the geranylgeranyl-pyrophosphate-Bac-6-OH was incubated with PglB and the product analyzed by HPLC. Mass spectral analyses of the resulting HPLC product peaks show that the Bac-6-OH monosaccharide is transferred to protein (Figure 4-12). Hence it appears that PglB accepts the geranylgeranyl-bound glycan substrate as well as the longer chain, α-saturated dolichyl-pyrophosphate-linked substrate.
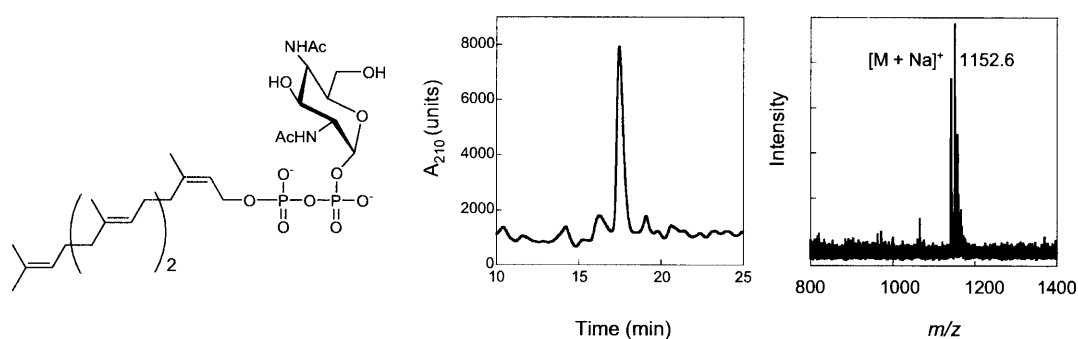
Figure 4-12. PglB accepts geranylgeranyl-PP-Bac-6-OH.

HPLC and MALDI-MS analysis of the geranylgeranyl-PP-Bac-6-OH and PglB reaction. Peak at m/z 1152.6 corresponds to sodium adduct of glycopeptide.

Peptide specificity of PglB

Next, the oligosaccharyl transferase activity of PglB on the octapeptide was compared with the known tripeptide acceptor (Bz-NLT-NHMe), commonly used for the yeast OT system.[1,12] As the results from the radioactive assay demonstrate, clearly, Bz-NLT-NHMe is a poor acceptor in the bacterial system (Figure 4-13 A). The Bz-NLT-NHMe peptide was subjected to HPLC analysis similar to the consensus peptide discussed above. The HPLC trace (Figure 4-13 B) shows the presence of 2 peaks, which were confirmed by mass spectral analysis to be the glycopeptide (T_R = 27.0 min, Figure 4-13 C) and the unglycosylated counterpart (T_R = 29.0 min,

data not shown). The presence of a significant amount of unglycosylated peptide, when subjected to the exact reaction conditions as the consensus peptide, together with the low rate of transfer observed in Figure 4-13A, strongly suggests that PglB has additional determinants for the peptide substrate beyond the minimal tripeptide consensus sequence that is well recognized in eukaryotes.[8] Further studies into the peptide specificity of PglB are currently in progress.
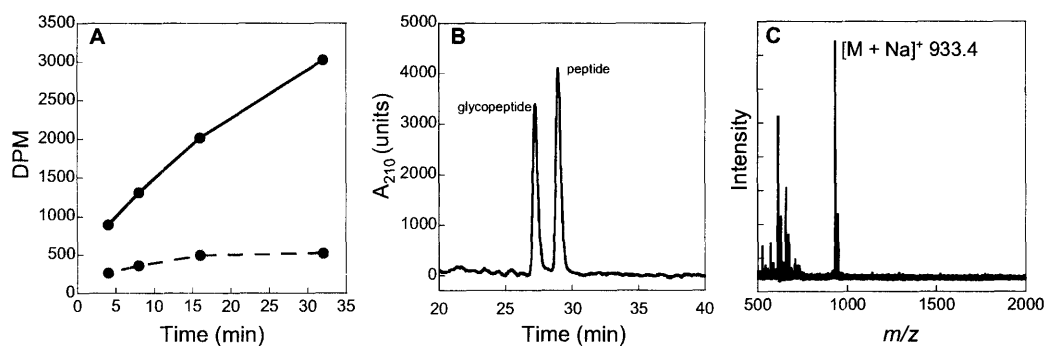
Figure 4-13. Peptide substrate specificity of PglB.

(A) Radioactive Assay of product formation. Solid line, octapeptide consensus sequence (KDFNVSKA); Dashed line, tripeptide consensus for yeast OT (Bz-NLT-NHMe). 6-hydroxybacillosamine was used in this study in place of the native bacillosamine substrate. (B) HPLC trace of Bz-NLT-NHMe reaction. (C) MALDI-MS of Bz-NLT-NHMe glycopeptide (peak at $T_R = 27.0$ min).

Conclusion

The experiments described here afford the first *in vitro* observation of oligosaccharyl transferase activity of PglB. By using a combination of chemical synthesis and enzymology we were able to prepare various glycan and peptide substrates for this remarkable enzyme. The results demonstrate that PglB has a relaxed substrate specificity accepting peptide substrates in place of full-length proteins, making it potentially useful for the preparation of artificial glycopeptides. Furthermore, the observation that PglB can transfer undecaprenyl-pyrophosphate-linked saccharides of various lengths (2-7 saccharides) adds to the promise of using PglB in the synthesis of diverse glycopeptide products. However, PglB does require determinants in the peptide sequence beyond the canonical N-X-S/T tripeptide and with this *in vitro* assay in place, further studies to determine the role of the amino acid binding determinants can be readily undertaken. Although the experiments described in this work strongly suggest that PglB is solely responsible for oligosaccharyl transferase activity, the presence of an accessory protein(s) in *E. coli* cannot be ruled out due to the use of a bacterial membrane fraction. Currently, work is underway to purify PglB to homogeneity, which will unambiguously demonstrate whether it alone is responsible for the oligosaccharyl transferase activity. Furthermore, detailed investigations into the poorly understood mechanism of oligosaccharyl transferases are more feasible using PglB as opposed to the multi-subunit eukaryotic complexes. Lastly, the development of powerful inhibitors of this enzyme would be valuable in the quest for antibiotics for *C. jejuni*-induced gastrointestinal disorders.

Acknowledgements

For the work presented in this chapter, I have had the pleasure of collaborating with Dr. Jebrell Glover. The preparation of the PglB membrane fraction was performed by Jebrell and all the HPLC and radioactive assay data represent a collaborative effort. The help of Langdon Martin and Matthieu Sainlos with obtaining the MALDI-MS data is greatly appreciated.

Experimental Procedures

Expression of PglB

Starting from a 5 mL overnight culture, *E. coli* strains expressing PglB were grown at 37°C in LB broth to an OD_{600} of 0.6-0.8. At that point, the temperature was reduced to 16°C and protein production was induced by the addition of IPTG (1 mM). After 24 hours, the cells were harvested by centrifugation (5000 × *g*) for 30 minutes, washed once with 0.9% NaCl solution, recentrifuged (5000 x g) for 30 minutes and frozen at -80°C until needed.

Preparation of PglB membrane fraction

All steps were performed at 4°C. The *E.coli* cell pellets expressing PglB (wild type and mutant) were thawed and resuspended in 5% of the original culture volume in buffer M (50 mM Tris-acetate [pH 8.0], 1 mM EDTA). The cells were then subjected to sonication (3 × 15 s), unbroken cells were removed by centrifugation at 5,697 × *g* for 15 minutes, and the membrane fraction was collected by centrifugation at 142,414 × *g* for 60 minutes. The pellet was washed once with

137

buffer M, centrifuged again, and resuspended in 0.25% of the original culture volume in buffer M. The final suspension was aliquoted and stored at -80°C.

Preparation of undecaprenyl-pyrophosphate-linked disaccharides using PglA

To a tube containing 0.06 mg of dried Und-PP-Bac, Und-PP-6-hydroxybacillosamine or Und-PP-GlcNAc, 3 μL of DMSO, and 7 μL of 14.3% (v/v) Triton X-100 were added. After vortexing and sonication (water bath), 58 μL of H_2O, 4 μL of 1 M Tris-Acetate pH 8.5, 1 μL 1 M $MgCl_2$, and 20 μL PglA (660 μg/mL) were added. The reaction was initiated by the addition of 7.5 μL of UDP-GalNAc (55 nCi/nmol). After 120 minutes, the reaction was quenched in 1.6 mL of 2:1 chloroform:methanol and extracted three times with 320 μL of pure solvent upper phase (15 mL chloroform: 240 mL methanol: 235 mL water: 1.83 g KCl). The organic layer was aliquoted and dried under a stream of nitrogen (~20,000-30,000 DPM/tube).

Peptide Synthesis

Peptides were synthesized by automated peptide synthesis (Applied Biosystems ABI 431A peptide synthesizer) using standard Fmoc peptide synthesis conditions on PAL-PEG-PS resin. The resulting peptides were cleaved from the resin using trifluoroacetic acid and purified by preparative reverse phase high-pressure liquid chromatography using a standard acetonitrile/water gradient.

PglB Assay

To a tube of the labeled disaccharide, 10 μL DMSO, 100 μL of 2× assay Buffer (100mM HEPES, pH 7.5, 280 mM sucrose, 2.4% v/v Triton-X100), 2 μL of 1M $MnCl_2$ and 28 μL of

water was added. Reactions were initiated by the addition of 10 μL of a 2 mM stock of the peptide in DMSO. 40 μL of the reaction mixture was removed at various time points and quenched into 1 mL of 3:2 chloroform: methanol + 200 μL of 4 mM $MgCl_2$. The aqueous layer was extracted and the organic layer was washed twice with 600 μL of pure solvent upper phase. The aqueous layers were combined, mixed with 5 mL of scintillation fluid (EcoLite, MP Biomedicals) and subjected to scintillation counting (2 minutes per tube).

HPLC analysis of glycopeptides

A 40 μL reaction aliquot was quenched as above and the aqueous layer was dried under vacuum. The residue was resuspended in 50 μL of water and injected on a reverse-phase C18 column and eluted under a standard water/acetonitrile gradient. Fractions were collected every minute, mixed with scintillation fluid (EcoLite, MP Biomedicals) and subjected to scintillation counting (2 minutes per tube).

References

1.   Imperiali, B., and Hendrickson, T.L. Asparagine-Linked Glycosylation: Specificity and Function of Oligosaccharyl Transferase. *Bioorg. Med. Chem.*, **1995**, *3*, 1565-1578.

2.   Kaplan, H.A., Welply, J.K., and Lennarz, W.J. Oligosaccharyl Transferase: The Central Enzyme in the Pathway of Glycoprotein Assembly. *Biochim. Biophys. Acta*, **1987**, *906*, 161-173.

3.   Sharma, C.B., Lehle, L., and Tanner, W. N-Glycosylation of Yeast Proteins. Characterization of the Solubilized Oligosaccharyl Transferase. *Eur. J. Biochem.*, **1981**, *116*, 101-108.

4.   Yan, Q., and Lennarz, W.J. Studies on the Function of Oligosaccharyl Transferase Subunits. Stt3p Is Directly Involved in the Glycosylation Process. *J. Biol. Chem.*, **2002**, *277*, 47692-47700.

5.   Chavan, M., Rekowicz, M., and Lennarz, W. Insight into Functional Aspects of Stt3p, a Subunit of the Oligosaccharyl Transferase. Evidence for Interaction of the N-Terminal Domain of Stt3p with the Protein Kinase C Cascade. *J. Biol. Chem.*, **2003**, *278*, 51441-51447.

6.   Szymanski, C.M., Yao, R., Ewing, C.P., Trust, T.J., and Guerry, P. Evidence for a System of General Protein Glycosylation in Campylobacter Jejuni. *Mol. Microbiol.*, **1999**, *32*, 1022-1030.

7.   Wacker, M., Linton, D., Hitchen, P.G., Nita-Lazar, M., Haslam, S.M., North, S.J., Panico, M., Morris, H.R., Dell, A., Wren, B.W., and Aebi, M. N-Linked Glycosylation in Campylobacter Jejuni and Its Functional Transfer into E. Coli. *Science*, **2002**, *298*, 1790-1793.

8.   Nita-Lazar, M., Wacker, M., Schegg, B., Amber, S., and Aebi, M. The N-X-S/T Consensus Sequence Is Required but Not Sufficient for Bacterial N-Linked Protein Glycosylation. *Glycobiology*, **2005**, *15*, 361-367.

9.   Feldman, M.F., Wacker, M., Hernandez, M., Hitchen, P.G., Marolda, C.L., Kowarik, M., Morris, H.R., Dell, A., Valvano, M.A., and Aebi, M. Engineering N-Linked Protein Glycosylation with Diverse O Antigen Lipopolysaccharide Structures in Escherichia Coli. *Proc. Natl. Acad. Sci. U. S. A.*, **2005**, *102*, 3016-3021.

10.   Linton, D., Dorrell, N., Hitchen, P.G., Amber, S., Karlyshev, A.V., Morris, H.R., Dell, A., Valvano, M.A., Aebi, M., and Wren, B.W. Functional Analysis of the Campylobacter Jejuni N-Linked Protein Glycosylation Pathway. *Mol. Microbiol.*, **2005**, *55*, 1695-1703.

11.   Linton, D., Allan, E., Karlyshev, A.V., Cronshaw, A.D., and Wren, B.W. Identification
of N-Acetylgalactosamine-Containing Glycoproteins PEB3 and CgpA in Campylobacter
Jejuni. *Mol Microbiol*, **2002**, *43*, 497-508.

12.   Imperiali, B., and Shannon, K.L. Differences between Asn-Xaa-Thr-Containing Peptides:
A Comparison of Solution Conformation and Substrate Behavior with
Oligosaccharyltransferase. *Biochemistry*, **1991**, *30*, 4374-4380.

Chapter 5

Design and synthesis of peptidomimetic inhibitors of oligosaccharyl transferase

Introduction

Protein glycosylation plays a key role in numerous cellular processes including immune response, intracellular targeting, intercellular recognition and protein folding and stability.[1] Asparagine-linked glycosylation in eukaryotes is a major class of protein modification mediated by a single enzyme, oligosaccharyl transferase (OT) that is localized in the lumen of the endoplasmic reticulum (ER).[2]
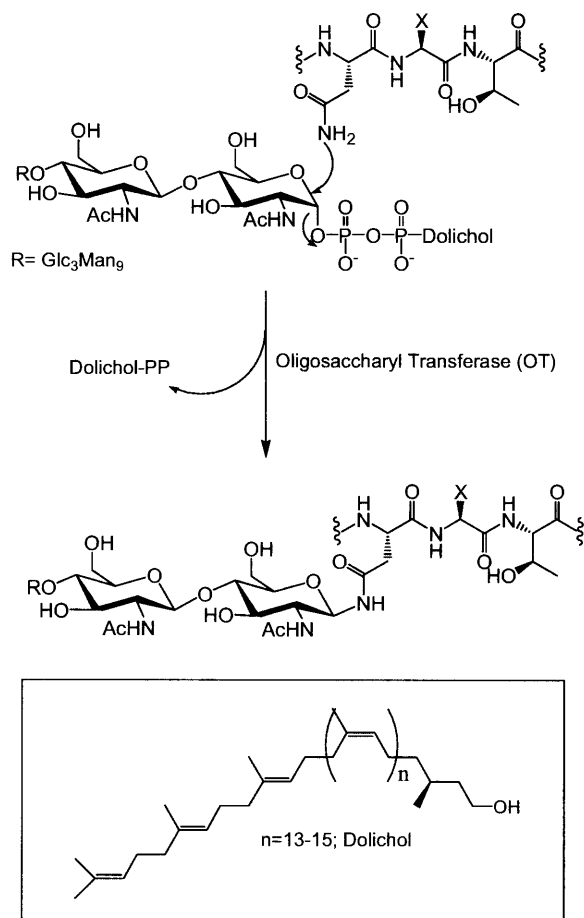
Scheme 5-1. The reaction catalyzed by oligosaccharyl transferase.

The multimeric membrane-bound OT enzyme facilitates the co-translational transfer of a dolichol-pyrophosphate-linked tetradecasaccharide ($GlcNAc_2Man_9Glc_3$) to an asparagine side chain in the Asn-Xaa-Thr/Ser consensus sequence within a nascent polypeptide chain as shown in Scheme 5-1.[3] This step constitutes the first committed step in N-linked glycosylation and the resulting glycoprotein is further processed by various glycosidases and glycosyltransferases in the ER and Golgi to create a diverse array of oligosaccharide units.[4] The availability of OT-specific inhibitors that function *in vivo* may ultimately provide insight into this vital protein modification reaction.

Previous studies in the Imperiali group have yielded inhibitors with low nanomolar affinity for OT (Figure 5-1). These inhibitors are based on the Asn-Xaa-Thr/Ser (**1**) consensus sequence of the natural substrate. To ensure that a short peptide could be used as a substrate for OT, the capped tripeptide, Bz-Asn-Leu-Thr-NHMe (**2**) was synthesized and determined to be a substrate for the enzyme (K_M = 20 µM).[3] The first generation of OT inhibitors was based on this truncated substrate and was discovered by systematically replacing the asparagine residue in the tripeptide with several asparagine mimics. This study resulted in a substrate mimic Bz-Dab-Leu-Thr-NHMe (diaminobutyric acid, Dab) (**3**) that is a weak competitive inhibitor of OT.[5] Therefore replacing the amide of the asparagine with an isosteric amine containing side chain (**3** vs. **4**) results in a compound that binds to the enzyme but is not turned over. Studies extending from this initial discovery exploited binding interactions between OT and residues beyond the consensus sequence through the synthesis of a small library of peptides that included different combinations of amino acids C-terminal to the tripeptide sequence. This study yielded a hexapeptide Bz-Dab-Ala-Thr-Val-Thr-Nph-NH_2 (Nph = *p*-nitrophenylalanine) (**4**) that is a potent *in vitro* inhibitor of yeast (*Saccharomyces cerevisiae*) OT (K_i = 69 nM).[6] Additionally,

peptides that are structurally constrained to an Asx-turn conformation show enhanced affinity for the enzyme.[3] For example, the linear hexapeptide inhibitor (4) demonstrates a two-fold increase in potency when constrained *via* a side-chain to backbone macrocyclization (5).
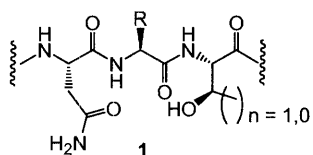
Asn-Xaa-Ser/Thr
Consensus sequence for *N*-linked glycosylation
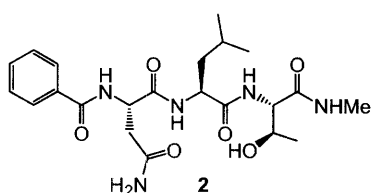
Bz-Asn-Leu-Thr-NH_2
Substrate
$K_M = 20$ μM
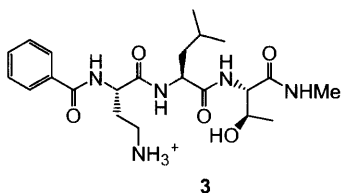
Bz-Dab-Leu-Thr-NH_2
Weak competitive inhibitor
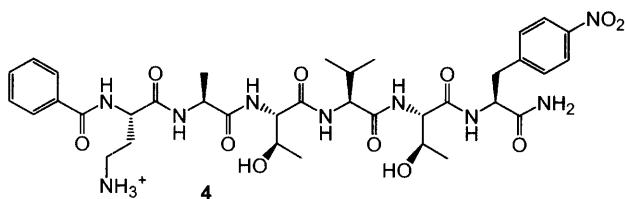
Bz-Dab-Ala-Thr-Val-Thr-Nph-NH_2
Potent Inhibitor
$K_i = 62$ nM
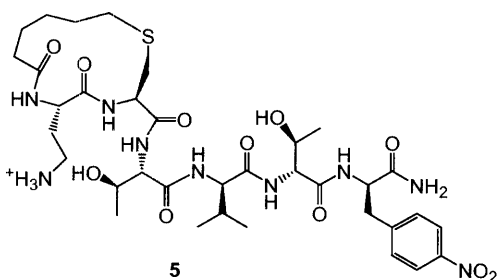
c[Hex-Dab-Cys]-Thr-Val-Thr-Nph-NH_2
Potent Inhibitor
$K_i = 37$ nM

Figure 5-1. Previous peptide-based inhibitors of OT

Inhibitors based upon these initial discoveries yielded potent OT inhibitors, yet none displayed any activity *in vivo*. This is due to the fact that peptides exhibit poor membrane permeation and limited proteolytic stability and are not optimal for studies under cell-based conditions. Hence it is desirable to obtain non-peptidic, bioavailable inhibitors of OT that would enable the study of N-linked glycosylation and its role in cellular processes *in vivo*.

Currently, the only inhibitor of N-linked protein glycosylation to function within a cellular environment is the microbial product tunicamycin (Figure 5-2). Tunicamycin is a bisubstrate analog that inhibits the GlcNAc-1-P glycophosphoryltransferase, Alg 7, which is responsible for the first committed step in the assembly of the dolichol-pyrophosphate-linked oligosaccharide donor.[7] Since this transformation occurs numerous steps prior to the reaction catalyzed by OT, the effect of the inhibitor on the actual glycosylation step is not immediate, nor does it have the potential to reveal the specific consequences of blocking N-linked glycosylation. The focus of this study is the generation of rationally designed, non-peptidic inhibitors that specifically target OT *in vivo*.
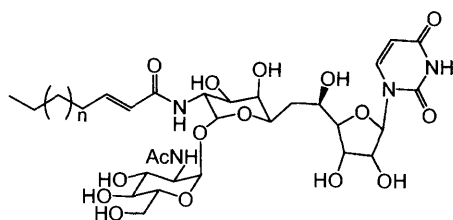
Figure 5-2. Tunicamycin: A microbial natural product that inhibits *N*-linked glycosylation.

146

The linear hexapeptide Bz-Dab-Ala-Thr-Val-Thr-Nph-NH$_2$ (**4**) was used as the prototype for designing an inhibitor of OT that would function *in vivo*. The readily modifiable and modular peptide platform provided a foundation for the rational design of a peptidomimetic inhibitor. Our aim was to design and synthesize a potent inhibitor of OT that includes reduced amide bond content to minimize proteolysis and increase hydrophobicity to enhance cell permeability. Toward this goal, we have modified the inhibitor structure *via* isosteric replacement of the peptide backbone to afford a non-peptidic entity that demonstrates higher proteolytic stability and increased lipophilicity, which would facilitate passive permeation of the cellular and endoplasmic reticulum membranes. There has been no previous record of peptidomimetics or pseudopeptides that act as potent inhibitors of OT. In this study we report the systematic introduction of non-peptidic character to our previous inhibitors to yield several compounds with nanomolar inhibition potency for OT.

Results and Discussion

Dipeptide Isosteres

We initiated this project with the goal of identifying a suitable dipeptide isostere to replace the Val-Thr unit within our prototype hexapeptide, Bz-Dab-Ala-Thr-Val-Thr-Nph-NH$_2$ (**4**). Modifications to the consensus inhibitory sequence (Dab-Xaa-Thr) results in a dramatic decrease in affinity, hence these residues were left unaltered in this study. In early investigations into inhibitor design, the Val-Thr dipeptide unit was installed based on statistical studies on glycosylation sites found in native proteins.[8] Further studies showed that hydrophobic residues are well tolerated in the positions occupied by the Val-Thr unit,[9] hence this would be the ideal

location for a lipophilic isostere. It is also believed that these residues can be altered without deleterious effects to binding. In this study we evaluated the ability of a non-peptide based scaffold to replace the Val-Thr dipeptide unit and orient the crucial residues for optimal interaction with the OT exosites. The third residue in the C-terminal extension sequence, *p*-nitrophenylalanine was introduced to allow facile concentration determination by absorption spectroscopy.

The class of dipeptide isosteres that was the focus of this study is based on an aminobenzoic acid framework. Similar scaffolds have been successfully incorporated within the structures of Src SH2 domain antagonists,[10] growth hormone-releasing peptide receptor agonists[11] and Ras farnesyl protein transferase inhibitors.[12] These isosteres introduce rigidity and hydrophobicity to the system as well as reducing the total amide bond content. We replaced the Val-Thr unit with a variety of spacers that incorporated 3-or 4-aminobenzoic acid and 3- or 4-aminomethylbenzoic acid.

All the unprotected aminobenzoic acids (**9, 11** and **13**, Scheme 5-2) were commercially available except for 3-amino(methyl)benzoic acid (**7**), which was synthesized from 3-cyanobenzoic acid (**6**) by catalytic hydrogenation. The amine or aniline groups of the building blocks were protected as the Alloc (allyloxy carbonyl) derivative using allyl chloroformate, to afford the corresponding protected spacer units (**8, 10, 12** and **14**).
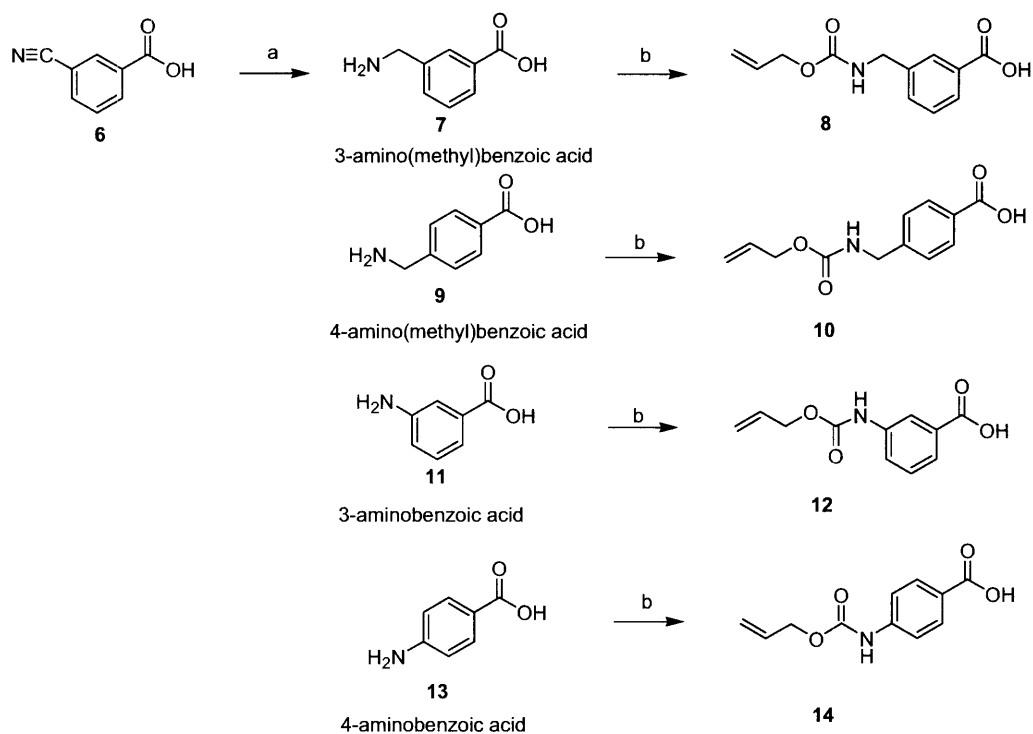
Scheme 5-2. Synthesis of alloc-protected aminobenzoic acid spacers for solid phase synthesis.[a]

[a]Reagents and conditions: (a) H_2, Pd/C, MeOH, 70%; (b) allyl chloroformate, 60-80% yield.

These spacer units were then incorporated into the peptide backbone using standard solid phase coupling conditions (Scheme 5-3). The synthesis of these peptidomimetics involves the solid phase peptide coupling of the corresponding Alloc protected aminobenzoic acid to the *p*-nitrophenylalanine residue using *O*-(7-azabenzotriazol-1-yl)-*N,N,N',N'*-tetramethyluronium hexafluorophosphate (HATU) as a coupling agent. The coupling step was followed by alloc deprotection using tetrakis(triphenylphosphine)palladium(0) and phenylsilane and subsequent coupling of the threonine, alanine and diaminobutyric acid residues using standard peptide synthesis protocols. In the final step, the entire construct was cleaved from the resin using 95% trifluoroacetic acid (TFA).
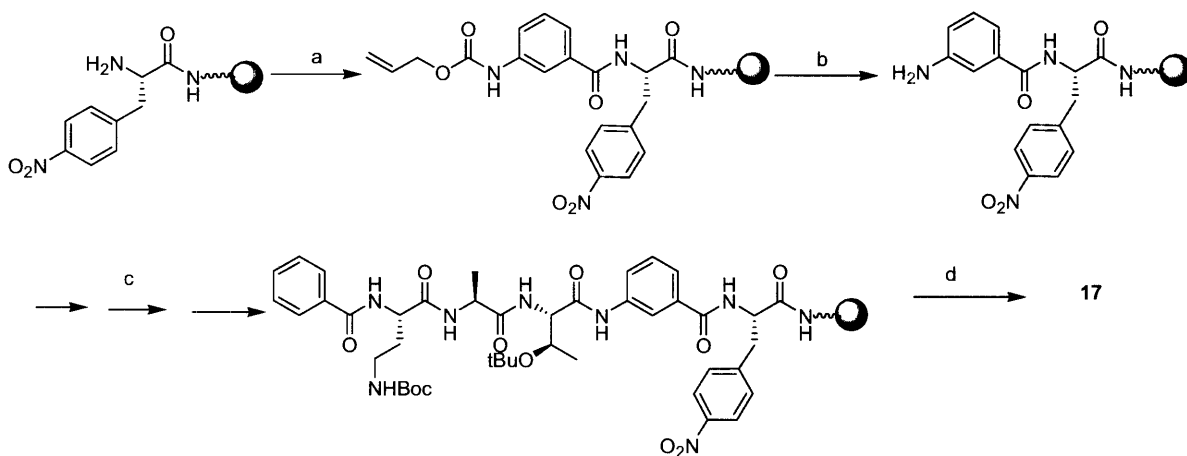
Scheme 5-3. Solid phase incorporation of aminobenzoic-acid spacers.[a]

[a]Reagents and conditions: (a) alloc-protected aminobenzoic acid, HATU, DIPEA; (b) Pd(PPh₃)₄, phenylsilane; (c) standard Fmoc-based solid-phase coupling of Fmoc-Thr(tBu)-OH, Fmoc-Ala-OH, Fmoc-Dab(Boc)-OH with HATU and DIPEA followed by capping with benzoic anhydride, pyridine; (d) cleavage from resin (95% TFA).

The four inhibitors synthesized via this method are illustrated in Figure 5-3. These compounds were assayed for their *in vitro* potency as OT-inhibitors using solubilized yeast (*S. cerevisiae*) microsomes. The prototype peptide inhibitor (**4**) has low nanomolar affinity for OT (K_i = 62 nM). When the Val-Thr dipeptide of **4** was replaced with each of the four spacers, it was found that the longer spacers, the 3-amino(methyl) (**15**) and 4-amino(methyl)-benzoic acid (**16**) groups caused a significant drop in enzyme affinity. In contrast, the shorter spacers, the 3-amino (**17**) and 4-aminobenzoic acid (**18**) showed a less dramatic effect on enzyme binding. In fact the

3-aminobenzoic acid dipeptide isostere (**17**) (K_i = 84 nM) was the most effective mimic of the

Val-Thr unit with inhibitory activity similar to the parent peptidic inhibitor.
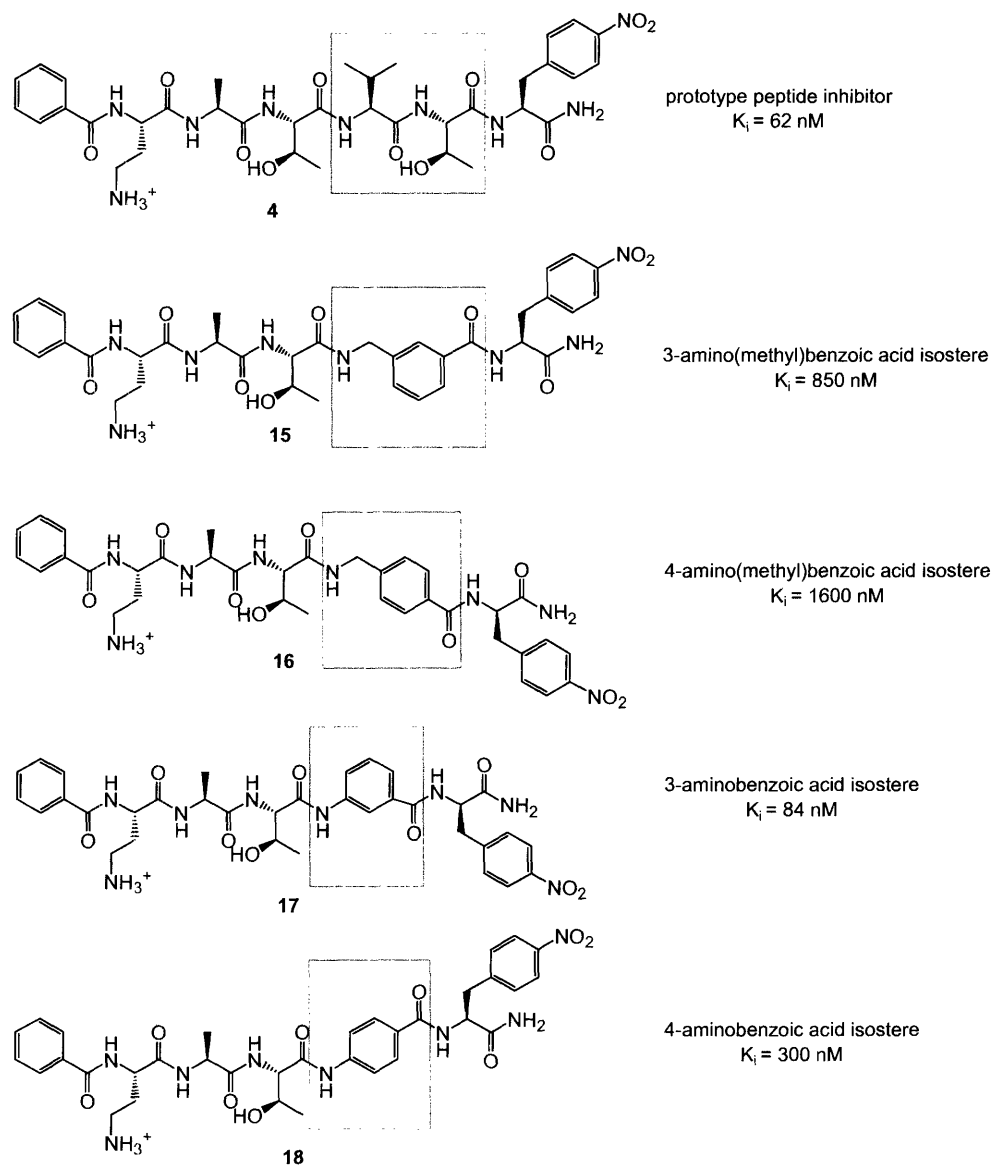
Figure 5-3. Inhibitors incorporating dipeptide isosteres in place of the Val-Thr dipeptide.

The reduced amide bond content and increased rigidity of the 3-aminobenzoic acid isostere has minimal effects on affinity for OT, confirming that the amide bond between the Val-Thr unit can be replaced without deleterious consequences for inhibitor binding. Due to the high affinity of compound **17** for OT, it was considered to be an excellent candidate for further modification toward the goal of generating a more hydrophobic non-peptide inhibitor.

Further Modifications

Since the dipeptide isostere studies yielded a promising candidate for OT inhibition, subsequent attempts were made to systematically reduce the hydrophilicity of compound **17**. The first site explored was the central residue within the Dab-Xaa-Thr sequence. Previous studies have shown that a higher binding affinity is observed in the truncated tripeptide substrate when this site is occupied by hydrophobic residues (Leu, Ala) whereas charged residues (Asp) are disfavored.[3] In order to confirm that this effect also applies to the interaction of the peptide isostere with OT, the alanine was substituted with norvaline, an unnatural amino acid that contains a longer hydrophobic side chain. This modification resulted in compound **19** that displays a slightly improved affinity for OT (Figure 5-4).
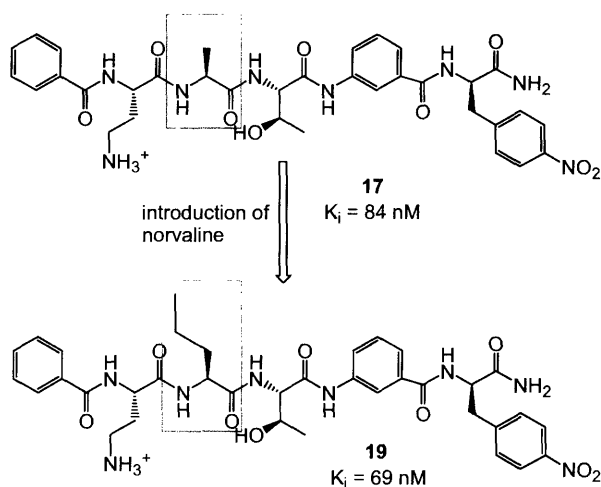
Figure 5-4. Replacement of Ala with Nva (norvaline).

In general, a minimal number of amide bonds are desirable so as to confer proteolytic stability as well as lipophilicity.[13] In accordance with this hypothesis, further modifications were investigated to systematically reduce the number of amide bonds within the inhibitor structure and determine which of these bonds are important for interaction with the enzyme. The initial dipeptide isostere studies removed the amide between the Val-Thr unit with minimal effect on binding. The next amide bond targeted was the amide C-terminal to the p-nitrophenylalanine residue. Replacement of this amide was achieved by using a 4-nitrobenzylamine capping group instead of the intact p-nitrophenylalanine residue. The resulting inhibitor 20 (Figure 5-5) shows a 3-fold loss in affinity, suggesting that there is a weak interaction between this terminal amide and residues in the OT exosites.
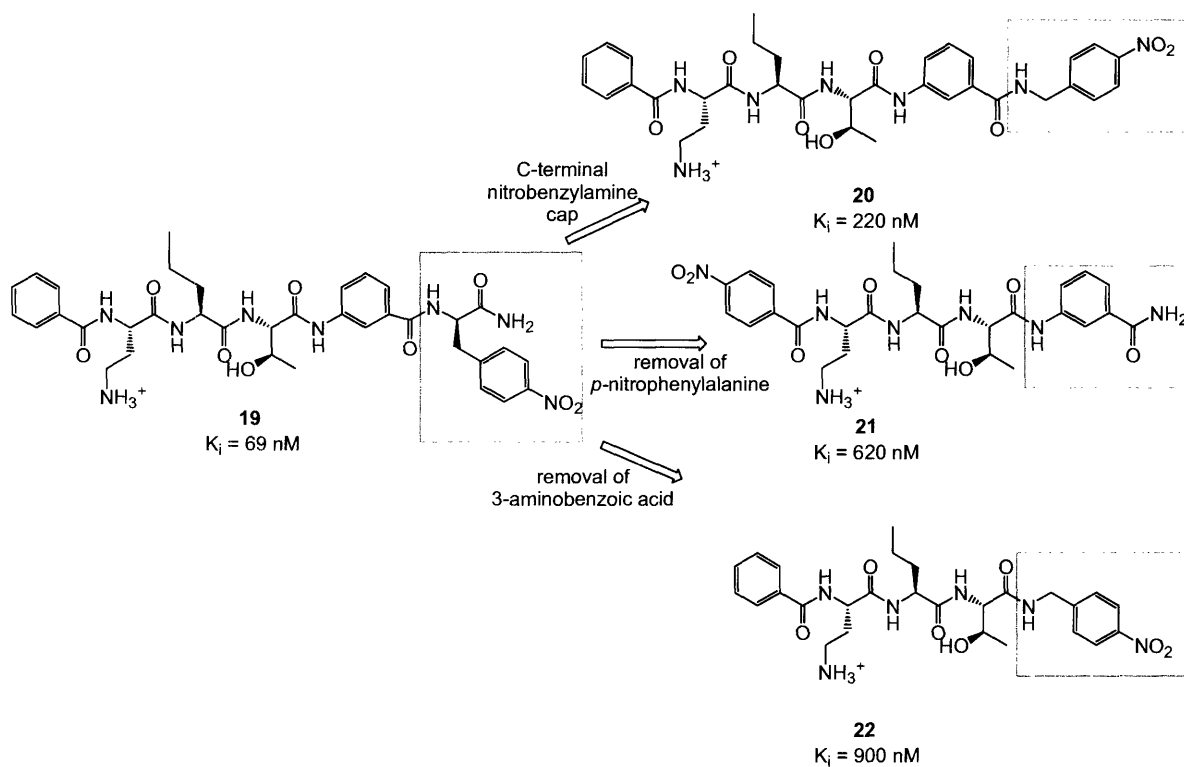
Figure 5-5. Further modifications to reduce the amide-bond character and the overall size of the inhibitors.

A second factor that promotes cell permeability is the overall size of the molecule and studies show that higher molecular weight compounds are less likely to be orally active and therefore cell permeable relative to smaller entities.[13] In accordance with this hypothesis, two inhibitors, **21** and **22** (Figure 5-5) were synthesized with molecular weights in the 500 Da range were synthesized. Inhibitor **21** eliminates the p-nitrophenylalanine residue that is used for quantification purposes, but installs an N-terminal nitrobenzyl cap instead, to enable accurate quantitation as described previously. Previous studies in the group have shown that substitution of nitrobenzyl for benzyl at the N-terminal cap causes minimal effects in enzyme affinity.[6] These

154

two changes to the inhibitor structure resulted in compound **21** which exhibited a K_i of 620 nM, a seven-fold decrease in enzyme affinity. This result indicates that the C-terminal *p*-nitrophenylalanine provides valuable interactions with residues at the enzyme exosite, in addition to supplying a quantification tool. Inhibitor **22** installs the nitrophenyl group adjacent to the consensus sequence and excludes the Val-Thr dipeptide isostere. The resulting molecule is small and contains only 4 amide bonds, factors that enhance its cell permeability and proteolytic stability. The higher K_i value (900 nM) of this compound confirms that the Val-Thr dipeptide and hence the aminobenzoic acid unit is essential for enzyme binding. It is believed that the rigid structure of the aminobenzoic spacer orients the nitrophenyl group at the optimal distance for favorable interactions as well as providing additional hydrophobic contacts.

The synthesis of inhibitors **20** and **22,** containing the C-terminal nitrobenzyl group, involved the use of an aldehyde-functionalized resin, to which the nitrobenzyl functionality was introduced *via* a reductive amination using 4-nitrobenzylamine. A schematic of the synthetic scheme used to access the C-terminal nitrobenzyl cap is shown in Scheme 5-4 for the synthesis of inhibitor **20.** After the nitrobenzyl group was installed, the alloc protected 3-aminobenzoic acid spacer was coupled to the resulting resin-bound secondary amine using HATU as a coupling agent. The peptide was then extended after deprotection of the alloc group using standard peptide synthesis protocols and cleaved from the resin with TFA to yield **20.**
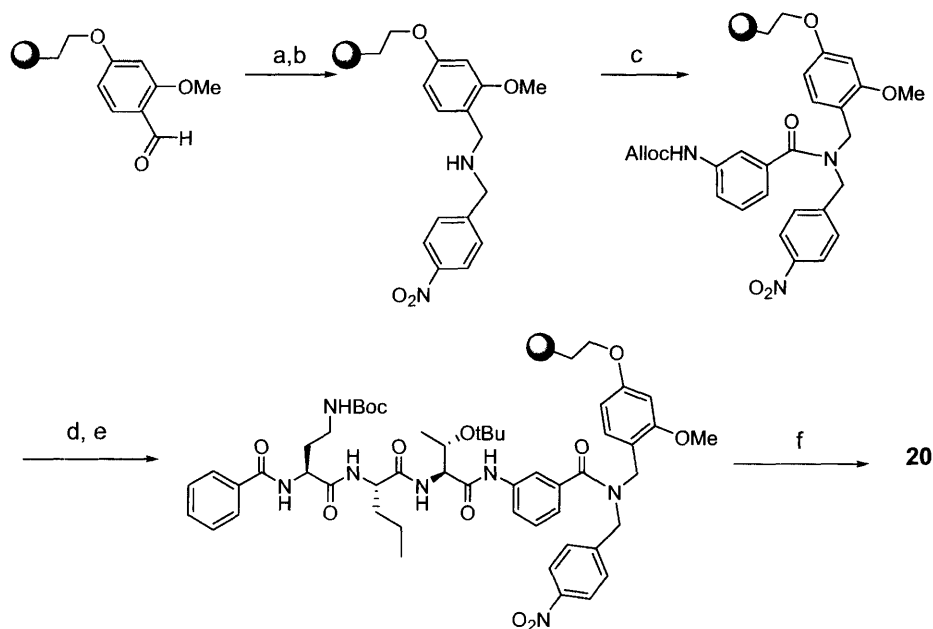
Scheme 5-4. Synthesis of C-terminal nitrobenzyl-capped inhibitor **20.**[a]

[a]Reagents and conditions: a) nitrobenzylamine, trimethyl orthoformate, dichloroethane; b) sodium triacetoxyborohydride; c) Alloc-3-aminobenzoic acid, HATU, N,N-diisopropyltheylamine (DIPEA); d) Pd(PPh₃)₄, phenylsilane; e) standard solid phase peptide coupling of Fmoc-Thr(tBu)-OH, Fmoc-Nva-OH, Fmoc-Dab(Boc)-OH, followed by capping with benzoic anhydride and pyridine; f) cleavage from resin (95% TFA).

Of the changes that were made to increase the bioavailability of the initial lead compound (**19**), the installation of the nitrobenzyl cap (**20**) had the least effect on enzyme affinity. This compound displays significantly decreased amide bond character relative to the first generation inhibitors. Further modifications were then performed on this inhibitor skeleton, to add more 'drug-like' character and increase enzyme affinity. To do this, we focused our attention on hydrophobicity, which is an important aspect of bioavailability.[13] Throughout this study,

theoretical water-octanol partition coefficients were calculated for each of the inhibitors using the ACD/LogP software that calculates log P values based on an algorithm using a database of over 14,600 compounds.[14] These computed values of log P were used to obtain a rough estimate of the relative hydrophobicity of the synthesized compounds. The values of log P for the inhibitors discussed in this paper ranged from 1.19 to 5.04 (Table 5-1) and even the most hydrophobic inhibitors were well behaved under the aqueous assay conditions at concentrations up to 10 μM.

The most hydrophobic of the inhibitors synthesized was **23** (Figure 5-6), which incorporated a nitrobenzyl group at the amine terminus of the Dab side chain. This modification was chosen due to information suggesting the presence of aromatic amino acids at the site of interaction of carbohydrate binding proteins.[15, 16] It was hypothesized that the sugar-binding site of OT comprises similar aromatic side chains and previous studies in the group demonstrated that incorporating a naphthyl group at this position enhanced inhibition.[17] Since p-nitrobenzyl provided enhanced binding when incorporated at the C-terminus of the inhibitor, this same group was chosen to modify the amine side chain of the diaminobutyric acid. We proposed that this would place the p-nitrobenzyl group within the same binding pocket as the carbohydrate moiety and should hypothetically exploit the aromatic π-stacking interactions at this site to enhance binding affinity.
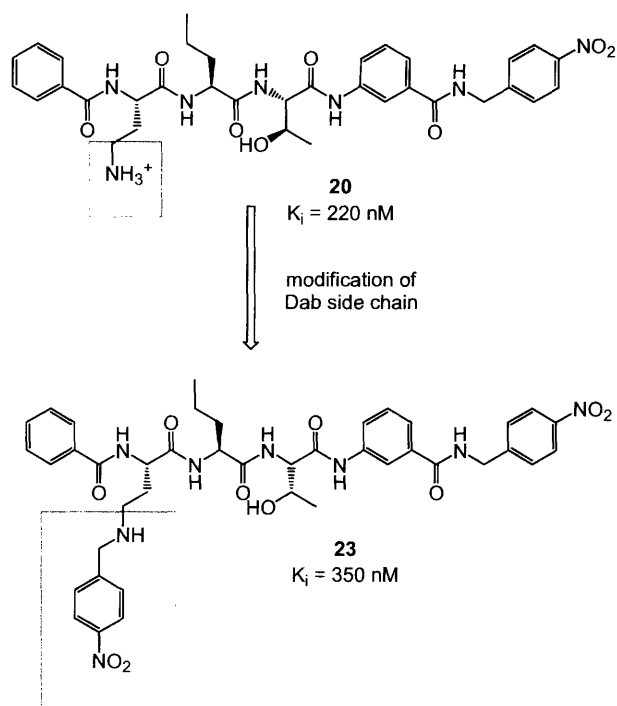
Figure 5-6. Modification of the Dab side chain.

The synthesis of compound **23** used Fmoc-Dab(Alloc)-OH instead of the Boc-protected unit used in the previous syntheses. This orthogonal protection scheme enabled the selective deprotection of the Dab side chain using tetrakis(triphenylphosphine)palladium(0) and phenylsilane. The nitrobenzyl moiety was installed via a reductive amination with nitrobenzaldehyde and sodium triacetoxyborohydride (Scheme 5-5). The resulting inhibitor, **23**, shows a K_i value of 350 nM with no solubility problems under the aqueous assay conditions. The added aromatic group does not increase affinity as desired, suggesting that the nitrobenzyl group is not oriented optimally for π-stacking interactions at the active site.
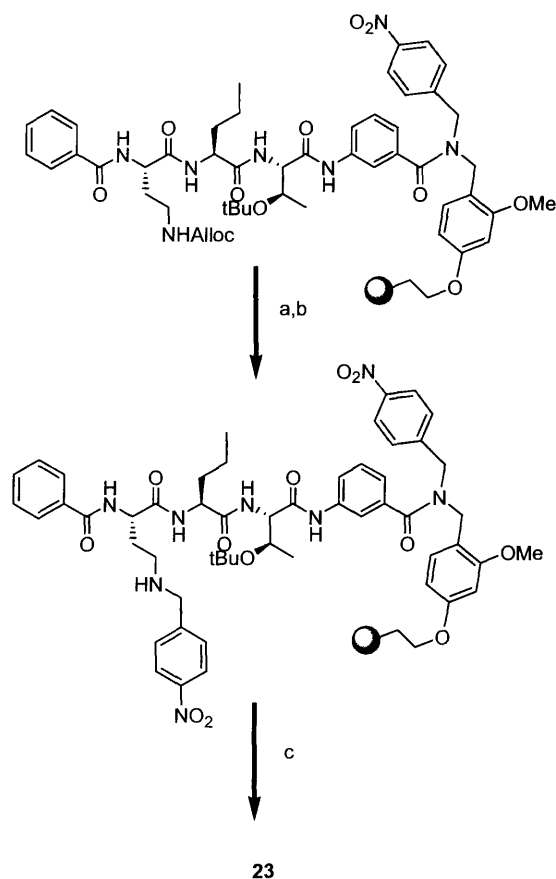
Scheme 5-5. Reductive amination with *p*-nitrobenzaldehyde.[a]

[a]Reagents and conditions: (a) Pd(PPh₃)₄, phenylsilane; (b) *p*-nitrobenzaldehyde, sodium triacetoxyborohydride; (c) cleavage from resin (95% TFA).

Cyclization into an Asx-turn

Molecular rigidity is proposed to play a role in cell permeation by locking out access to degradative enzymes while retaining inhibitory potency against the target.[18] Conformational studies on different OT substrates have indicated that peptides constrained to an Asx-turn motif, through side-chain to main-chain macrocyclization, are more competent substrates for OT as compared to the corresponding linear analogs.[5] The Asx-turn motif displays a "more open"

peptide backbone when contrasted with the more common β-turn structure. These turns are characterized by an extensive hydrogen bonding network between the side chain of the asparagine and the backbone amides (Figure 5-7). Since the peptide substrate for OT is hypothesized to bind in an Asx-turn conformation, introducing this turn to the inhibitor structure could potentially enhance binding affinity as well as increase the molecular rigidity of the compound.

A   B
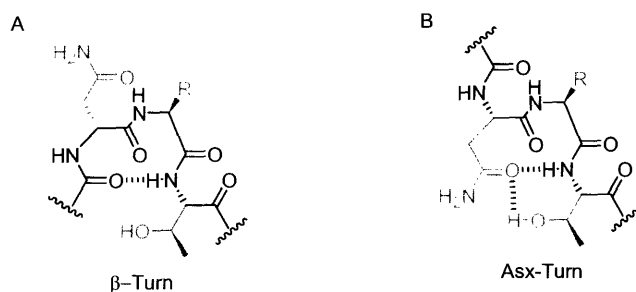
β-Turn   Asx-Turn

Figure 5-7. Illustration of a typical Asx-turn (**A**) and a β-turn (**B**).

Since the cyclized hexapeptide, cyclo(Hex-Dab-Cys)-Thr-Val-Thr-Nph-NH$_2$ (**5**) was found to be twice as potent as its linear counterpart,[3] it was decided that the macrocycle motif be incorporated into the non-peptidic inhibitor that is the focus of this study. This cyclization, together with a modified C-terminal cap, 4-nitrophenethylamine instead of nitrobenzylamine, resulted in a 2-fold increase in enzyme affinity. This resulted in inhibitor **24** (Figure 5-8) that possesses very little peptidic character, yet reasonable affinity for OT (K$_i$ = 130nM), and is one of the most hydrophobic and conformationally constrained inhibitors to demonstrate activity against this enzyme. This inhibitor fulfils the requirements for bioavailability such as reduced

amide bond character, increased hydrophobicity and rigidity as well as being smaller in size relative to the parent peptidic compound (4).

24
$K_i = 130$ nM
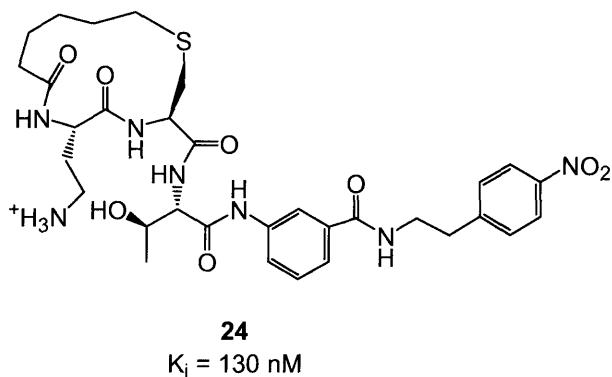
Figure 5-8. Structure of the cyclized peptidomimetic inhibitor (**24**).

The synthesis of inhibitor **24** utilized an aldehyde-functionalized resin (2-(4-formyl-3-methoxyphenoxy)ethyl polystyrene) that enables installation of the C-terminal 4-nitrophenethylamine cap via a reductive amination as illustrated in Scheme 5-6. The Alloc-protected 3-aminobenzoic acid moiety was coupled to the resulting secondary amine using HATU as a coupling reagent. Subsequent Alloc deprotection was then followed by coupling of threonine, 4-methoxytrityl (Mmt) protected cysteine, and diaminobutyric acid residues using standard peptide synthesis procedure. The peptide was capped with 6-bromohexanoic acid, the Mmt group was deprotected with 1% TFA and the cyclization between the resultant thiolate and the 6-bromohexanoyl group was achieved in degassed DMF, using an excess of 1,1,3,3-tetramethylguanidine as base.[19] The inhibitor was then cleaved from resin using 95% TFA and purified by HPLC.
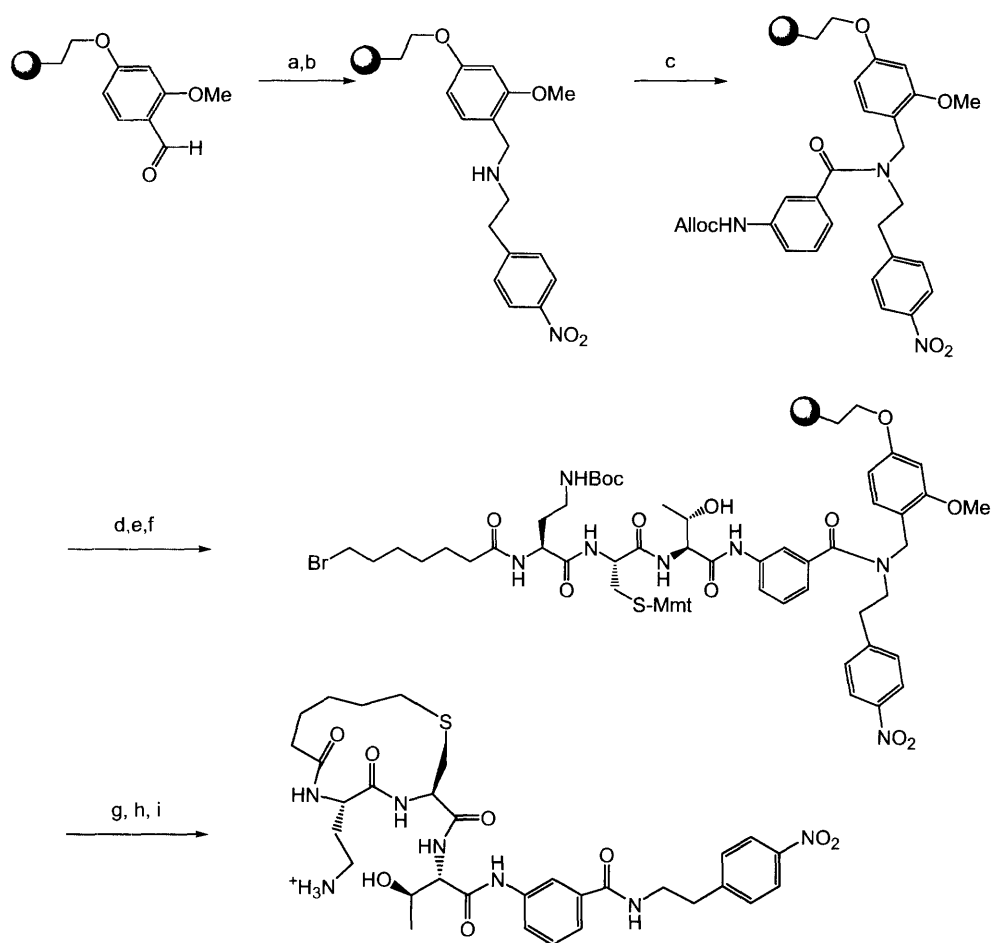
Scheme 5-6. Synthesis of the cyclized inhibitor (**11**).[a]

[a]Reagents and conditions: a) nitrophenethylamine, trimethyl orthoformate, dichloroethane; b) sodium triacetoxyborohydride; c) Alloc-3-aminobenzoic acid, HATU, *N,N*-diisopropylethylamine (DIPEA); d) Pd(PPh₃)₄, phenylsilane; e) standard solid phase peptide coupling of Fmoc-Thr(tBu)-OH, Fmoc-Cys(Mmt)-OH, Fmoc-Dab(Boc)-OH; f) 6-bromohexanoic acid, benzotriazol-1-yloxytrispyrrolidinophosphonium hexafluorophosphate (PyBOP), DIPEA; g) 1% TFA; h) 1,1,3,3-tetramethylguanidine; i) cleavage from resin (95% TFA).

Since these compounds would eventually be tested as potential inhibitors of OT within mammalian cells, it was necessary to confirm that they inhibit mammalian as well as the yeast (*S. cerevisiae*) OT. Hence assays were carried out as before, but using pig liver microsomes instead of yeast, to determine the K_i for a representative mammalian OT. These studies show that there is no significant difference in K_i values between the yeast and mammalian enzymes; hence none of the inhibitors are selective for either species, although inhibition in yeast is generally more potent. The results of these enzyme assays are summarized in Table 1.

Table 5-1. K_i values for yeast and porcine liver oligosaccharyl transferase inhibition.

Inhibitor	K_i (nM) yeast OT	K_i (nM) porcine OT	log P^*	Molecular Weight
17	84	110	1.48 +/- 0.88	705
19	69	180	2.55 +/- 0.88	733
20	220	710	3.13 +/- 0.87	676
21	620	780	1.19 +/- 0.83	586
22	900	600	1.79 +/- 0.85	557
23	350	760	5.04 +/- 0.88	810
24	130	180	2.96 +/- 0.84	686

*LogP values are crude estimates calculated using the ACD/LogP software[15]

The OT inhibitors developed in these studies will be implemented into a cellular assay for OT function to determine if they inhibit efficiently within a cellular environment. Their potency *in vivo* will be determined with the use of a high throughput assay system for analyzing OT inhibition in a stable mammalian cell line (Chapter 6).

Conclusion

Protein N-glycosylation is a crucial post-translational modification that is mediated by the enzyme oligosaccharyl transferase (OT). Due to its multimeric, membrane bound nature, there is no structural information available on the catalytic domain, and any studies into its inhibition could provide valuable information regarding the structure and function of this elusive enzyme. Moreover, inhibition of this enzyme within a cellular environment could be the means of obtaining knowledge on the immediate biological implications of N-linked glycosylation. Although potent inhibitors of OT exist, they all display peptidic character with poor proteolytic stability and membrane permeability, which render them unsuitable for *in vivo* studies. For this reason, this study describes efforts made towards the design and synthesis of a non-peptidic, hydrophobic OT inhibitor that demonstrates affinity for the enzyme in the nanomolar range.

In order to follow a rational approach to peptidomimetic inhibitor design, the hexapeptide Bz-Dab-Ala-Thr-Val-Thr-Nph-NH$_2$ (4) was used as the prototype. Studies on various Val-Thr dipeptide mimetics established 3-aminobenzoic acid to be the most suitable isostere for this unit with a K$_i$ value that is almost identical to the parent peptide. Further modifications conferred structural features on the inhibitor to improve its bioavailability. The amide bond character of the inhibitor was significantly reduced, thus increasing the proteolytic stability of the molecules. Since smaller size is favored for cell permeation, systematic deletion of units from the prototype inhibitor yielded compounds with low molecular weight. Hydrophobicity as estimated from log *P* values is also an important factor in cell permeation. Removal of hydrogen-bond donor groups and the introduction of aromatic groups assist in increasing the log *P* values whilst maintaining

solubility under the assay conditions. Finally, rigidity was introduced to the inhibitor by an aminobenzoic acid scaffold as well as by cyclization into an Asx turn conformation.

This study yielded a family of non-peptidic inhibitors that display varying degrees of hydrophobicity, size and rigidity as viable candidates for the *in vivo* inhibition of OT. In addition to providing potential bioavailable inhibitors of OT, the structure-activity relationships that have been defined in this study provide valuable clues to the nature of the interactions occurring at the OT exosites. The prime candidate for cellular studies is inhibitor **24**, which displays all the properties that are desired of *in vivo* inhibitors whilst maintaining a very high affinity for the enzyme (130 nM). The success of such a bioavailable inhibitor on a cellular level would enable studies related to understanding the effects of depleting N-linked glycoproteins on downstream biological processes.

Experimental Procedures

General procedure for solid-phase peptide synthesis

All peptides unless otherwise noted, were synthesized by manual solid phase methods using Fmoc-PAL PEG resin and using Fmoc (9-fluorenylmethoxycarbonyl) as the protecting group for the α-amino functionalities. Amino acids were coupled using either 1-benzotriazolyloxytris(pyrrolidino)phosphonium hexafluorophosphate (PyBOP) or *O*-(7-azabenzotriazol-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate (HATU) and DIPEA to generate the activated ester. The resin was swelled in dichloromethane (5 min.) and DMF (5 min.) prior to use. Couplings to the resin were performed in the following order: removal of the Fmoc group (20% piperidine solution in DMF, 3 × 5 min.), wash (DMF, 5 × 1 min.), coupling (amino acid/PyBOP/DIPEA, 1:2:2 in DMF, 1 hour) and rinse (DMF 2 × 1 min., CH_2Cl_2 2 × 1 min.). All amino acids were obtained commercially as the N-α-Fmoc-protected derivatives with the following side-chain protection: Dab(Boc), Dab(Alloc), Cys(Mmt), Thr(t-Bu). At the conclusion of the peptide synthesis, the peptides were capped with a large excess (10 eq) of benzoic anhydride and pyridine (10 eq). Cleavage from resin was performed with trifluoroacetic acid (TFA): CH_2Cl_2: triisopropylsilane: H_2O (90:5:2.5:2.5). The mixture was then filtered to remove the resin and concentrated under a stream of nitrogen. The resulting pellet was triturated with cold ether and redissolved in DMSO. The DMSO solution was then diluted into a solution of 1:1 water:acetonitrile and the peptides were purified by preparative HPLC (C_{18}) with a gradient of increasing acetonitrile/0.1%TFA (solvent A) in water/0.1%TFA (solvent B): the standard gradient used in this study was 93:7 A:B to 0:100 A:B in 35 mins. Elution from the column was monitored by UV absorbance at 280 nm and 228 nm and the product peak was

collected and lyophilized. All peptide identities were confirmed by electrospray mass spectrometry (ESI-MS) and the peptide solutions were quantified by UV absorbance at 280 nm (ε_{280} = 12,800 M^{-1} cm^{-1} for the *p*-nitrophenyl group).

Bz-Dab-Ala-Thr-(3-amino(methyl)benzoic acid)-Nph-NH$_2$ (15)
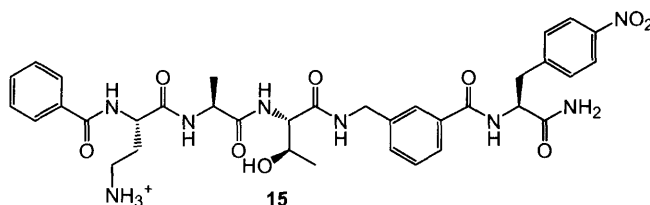
Fmoc-PAL-PEG PS resin was used and the nitrophenylalanine residue was coupled using standard Fmoc-based peptide synthesis methods. After Fmoc-deprotection, alloc protected 3-amino(methyl)benzoic acid was added in excess (4 eq) with HATU (4 eq) and diisopropylethylamine (DIPEA) (8 eq). The coupling was allowed to proceed overnight, after which alloc deprotection was performed with tetrakis(triphenylphosphine)Palladium (0) (0.2 eq) and phenylsilane (25 eq) in CHCl$_2$. Three 20-minute deprotections were performed under an N$_2$ atmosphere. Fmoc-Thr(tBu)-OH coupling was performed using HATU as above, and the subsequent alanine and diaminobutyric acid residues were coupled using standard procedure. The peptide was capped with benzoic anhydride, cleaved from resin and purified by HPLC.

HPLC t_R= 21.60 min (C$_{18}$, 7-100% B in 28 min)

LRMS calcd for **15** (C$_{35}$H$_{43}$N$_8$O$_9$$^+$) requires *m/z* 719.76. Found 719.5 (ESI+)

Bz-Dab-Ala-Thr-(4-amino(methyl)benzoic acid)-Nph-NH₂ (16)

The standard procedure outlined above was used, where alloc protected 4-amino(methyl)benzoic acid was coupled to the nitrophenylalanine on PAL-PEG PS resin.

HPLC t_R= 22.60 min (C_{18}, 7-100% B in 28 min)

LRMS calcd for **16** ($C_{35}H_{43}N_8O_9^+$) requires *m/z* 719.76. Found 719.5

Bz-Dab-Ala-Thr-(3-aminobenzoic acid)-Nph-NH₂ (17)

The standard procedure outlined above was used, where alloc protected 3-aminobenzoic acid was coupled to the nitrophenylalanine on PAL-PEG PS resin.

HPLC t_R= 21.43 min (C_{18}, 7-100% B in 28 min)

LRMS calcd for **17** ($C_{34}H_{41}N_8O_9^+$) requires *m/z* 705.74. Found 705.5

Bz-Dab-Ala-Thr-(4-aminobenzoic acid)-Nph-NH₂ (18)

The standard procedure outlined above was used, where alloc protected 4-aminobenzoic

acid was coupled to the nitrophenylalanine on PAL-PEG PS resin.

HPLC t_R= 19.22 min (C$_{18}$, 7-100% B in 28 min)

LRMS calcd for **18** (C$_{34}$H$_{41}$N$_8$O$_9$$^+$) requires m/z 705.74. Found 705.1
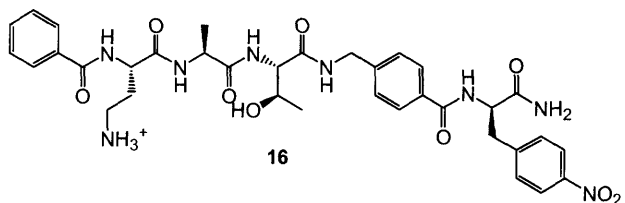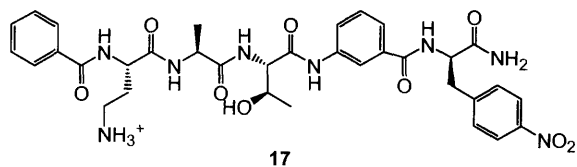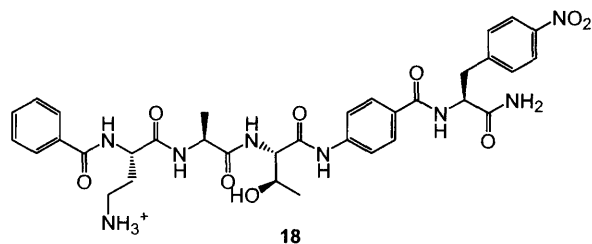
Synthesis of Bz-Dab-Nva-Thr-(3-aminobenzoic acid)-Nph-NH₂ (19)
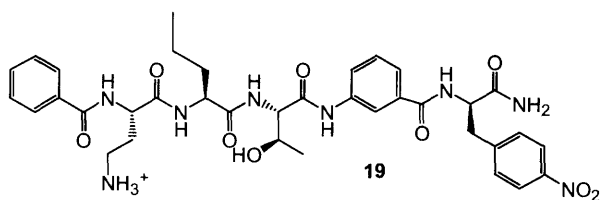
The standard procedure for the synthesis of inhibitor **15** was applied here, except that

Fmoc-Ala-OH was replaced with the commercially available unnatural amino acid Fmoc-Nva-

OH (norvaline).

HPLC t_R= 20.26 min (C$_{18}$, 7-100% B in 28 min)

LRMS calcd for **19** (C$_{36}$H$_{45}$N$_8$O$_9$$^+$) requires m/z 733.79. Found 733.3

Synthesis of Bz-Dab-Nva-Thr-(3-aminobenzoic acid)-NH-CH₂-Ph-NO₂ (20)

20

This peptide was synthesized on 2-(4-formyl-3-methoxyphenoxy)ethyl polystyrene (FMPE) resin available from Novabiochem, which is an aldehyde-functionalized resin. The resin was swollen in trimethylorthoformate (TMOF) and dichloroethane (DCE) 3:2 for 20mins. 4-nitrobenzylamine (10 eq) was added to the resin and stirred under N_2 for 3 hours, then 10 eq of sodium triacetoxyborohydride was added and the mixture shaken overnight. Alloc-protected 3-aminobenzoic acid (4 eq) was coupled to the resulting secondary amine using HATU and DIPEA. Fmoc-Thr(t-Bu)-OH, Fmoc-Nva-OH and Fmoc-Dab(Boc)-OH were coupled using standard procedure. The peptide was capped with benzoic anhydride and purified by HPLC as before.

HPLC t_R= 23.85 min (C_{18}, 7-100% B in 28 min)

LRMS calcd for **20** ($C_{34}H_{41}N_7O_8{}^+$) requires *m/z* 676.73. Found 676.4

170

Synthesis of Bz(NO₂)-Dab-Nva-Thr-(3-aminobenzoic acid)-NH₂ (21)

21

Alloc-protected 3-aminobenzoic acid was coupled to PAL-PEG-PS resin, followed by the coupling of Fmoc-Thr(t-Bu)-OH, Fmoc-Nva-OH and Fmoc-Dab(Boc)-OH using standard peptide synthesis procedure. The peptide was capped with 4-nitrobenzoic (2 eq) acid,using PyBOP (2 eq) as the coupling reagent and DIPEA (4eq).

HPLC t_R= 23.02 min (C_{18}, 7-100% B in 28 min)

LRMS calcd for **21** ($C_{27}H_{36}N_7O_8^+$) requires m/z 586.62. Found 586.4

Synthesis of Bz-Dab-Nva-Thr-NH-CH₂-Ph-NO₂ (22)

22

The procedure was the same as for inhibitor **20** above, except that Fmoc-Thr(tBu)-OH was coupled to the resin after the reductive amination, instead of 3-aminobenzoic acid. The peptide was cleaved and purified as before.

HPLC t_R = 23.20 min (C_{18}, 7-100% B in 28 min)

LRMS calcd for **22** ($C_{27}H_{36}N_6O_7^+$) requires *m/z* 557.62. Found 557.3

Synthesis of Bz-Dab(4-nitrobenzyl)-Nva-Thr-(3-aminobenzoic acid)-NH-CH$_2$-CH$_2$-Ph-NO$_2$ (23)

23

FMPE resin was used to couple the 4-nitrobenzylamine as before. The residues, alloc-3-aminobenzoic acid, Fmoc-Thr(tBu)-OH, Fmoc-Nva-OH and Fmoc-Dab(alloc)-OH were coupled using standard procedure outlined previously. The peptide was capped with benzoic anhydride and pyridine as before. The orthogonal alloc protection on the Dab side chain was removed with Pd(PPh$_3$)$_4$ and phenylsilane as before and a mixture of DCE:TMOF (3:2) with 10% acetic acid was added to the resin. 4-nitrobenzaldehyde (10 eq) was added to the resin and shaken for 5 hours. The reagent mixture was drained and the resin washed with TMOF. NaBH(OAc)$_3$ (10eq) was added and the resin stirred under N$_2$ overnight. The resin was washed with TMOF, DCE and dichloromethane and cleaved and purified as before.

HPLC t_R = 27.70 min (C_{18}, 7-100% B in 28 min)

LRMS calcd for **23** ($C_{41}H_{46}N_8O_{10}$) requires *m/z* 810.85. Found 811.3

Synthesis of cyc[Hex-Dab-Cys]-Thr-(3-aminobenzoic acid)-NH-CH₂-CH₂-Ph-NO₂ (24)

24

This peptide was synthesized on FMPE resin swollen with TMOF and DCE. 4-nitrophenethylamine (10 eq) was added to the resin and stirred under N_2 for 3 hours, then 10 eq of sodium triacetoxyborohydride was added and the mixture shaken overnight. Alloc-protected 3-aminobenzoic acid (4 eq) was coupled to the resulting secondary amine using HATU and DIPEA. After alloc deprotection as before, Fmoc-Cys(Mmt)-OH and Fmoc-Dab(Boc)-OH were coupled using standard peptide coupling procedure described above. The peptide was capped with 6-bromohexanoic acid and the Mmt protecting group was removed using 1% TFA in CH_2Cl_2. Cyclization was effected between the thiolate of the cysteine and the 6-bromohexanoyl group in degassed DMF using a large excess of 1,1,3,3-tetramethylguanidine as a base (24 hours). The peptide was cleaved and purified as before.

HPLC t_R= 23.90 min (C_{18}, 7-100% B in 28 min)

LRMS calcd for **24** ($C_{41}H_{46}N_8O_{10}^+$) requires *m/z* 686.80. Found 686.3

173

Determination of K$_i$ values

In a typical yeast OT assay, [^{3}H]DPPC (dolichol pyrophosphate chitobiose) (50,000 dpm, 60 Ci/mmol) was aliquoted from a chloroform/methanol stock solution into a microcentrifuge tube, and residual solvent removed under a gentle stream of nitrogen. Mixtures of increasing amounts of inhibitor (in 10 µl of DMSO) and a constant amount of enzyme in buffer were incubated on ice for 30 min. This incubation time was required to ensure pre-equilibration of the inhibitor with the enzyme. The reaction was initiated by adding 10 µL of a solution of Bz-Asn-Leu-Thr-NHMe (4 K$_m$) (2 mM for yeast and 10 mM for pig liver microsomes) substrate in DMSO. Aliquots (40 µL) of the reaction mixture were quenched after 2, 4, 6 and 8 min, and the aqueous phase containing the radioactive N-glycosylated peptide was extracted and quantified using a scintillation counter. Preliminary experiments employed a broad range of inhibitor concentrations for each peptide to afford a rough estimate of the IC$_{50}$. Three concentrations were then selected to give between 30-70% inhibition. All experiments were run in duplicate and in each case the K$_i$ was determined using the following equation:

$$K_i = \frac{[I] \times (1-i)}{i + \left(\dfrac{[S]}{K_m} \times i\right)}$$

where i represents the fraction inhibition, [I] is the concentration of inhibitor, and [S] is the concentration of Bz-Asn-Leu-Thr-NHMe. This assessment of K$_i$ assumes competitive inhibition for all the peptides under investigation.

References

1. Varki, A. Biological Roles of Oligosaccharides - All of the Theories Are Correct. *Glycobiology*, **1993**, *3*, 97-130.

2. Dempski, R.E., and Imperiali, B. Oligosaccharyl Transferase: Gatekeeper to the Secretory Pathway. *Curr. Opin. Chem. Biol.*, **2002**, *6*, 844-850.

3. Imperiali, B., and Hendrickson, T.L. Asparagine-Linked Glycosylation: Specificity and Function of Oligosaccharyl Transferase. *Bioorg. Med. Chem. Lett.*, **1995**, *3*, 1565-1578.

4. Trombetta, E.S., and Parodi, A.J. N-Glycan Processing and Glycoprotein Folding. *Adv Protein Chem.*, **2002**, *59*, 303-344.

5. Imperiali, B., Shannon, K.L., and Rickert, K.W. Role of Peptide Conformation in Asparagine-Linked Glycosylation. *J. Am. Chem. Soc.*, **1992**, *114*, 7942-7944.

6. Peluso, S., Ufret, M. de L., Imperiali, B. Unpublished Results.

7. Heifetz, A., Keenan, R.W., and Elbein, A.D. Mechanism of Action of Tunicamycin on the UDP-GlcNAc-Dolichyl-Phosphate GlcNAc-1-Phosphate Transferase. *Biochemistry*, **1979**, *18*, 2186-2192.

8. Gavel, Y., and Von Heijne, G. Sequence Differences between Glycosylated and Non-Glycosylated Asn-X-Thr/Ser Acceptor Sites: Implications for Protein Engineering. *Protein Eng.*, **1990**, *3*, 433-442.

9. Kellenberger, C., Hendrickson, T.L., and Imperiali, B. Structural and Functional Analysis of Peptidyl Oligosaccharyl Transferase Inhibitors. *Biochemistry*, **1997**, *36*, 12554-12559.

10. Lunney, E.A., Para, K.S., Rubin, J.R., Humblet, C., Fergus, J.H., Marks, J.S., and Sawyer, T.K. Structure-Based Design of a Novel Series of Nonpeptide Ligands That Bind to the PP60Src SH2 Domain. *J. Am. Chem. Soc.*, **1997**, *119*, 12471-12476.

11. Peschke, B., Madsen, K., Hansen, B.S., and Johansen, N.L. Aminomethylthiophene-2-Carboxylic Acids as Dipeptide Mimetic in New Growth Hormone Secretagogues. *Bioorg. Med. Chem. Lett.*, **1997**, *7*, 1969-1972.

12. Qian, Y., Blaskovich, M.A., Saleem, M., Seong, C.M., Wathen, S.P., Hamilton, A.D., and Sebti, S.M. Design and Structural Requirements of Potent Peptidomimetic Inhibitors of P21ras Farnesyltransferase. *J. Biol. Chem.*, **1994**, *269*, 12410-12413.

13. Lipinski, C.A., Lombardo, F., Dominy, B.W., and Feeney, P.J. Experimental and Computational Approaches to Estimate Solubility and Permeability in Drug Discovery and Development Settings. *Adv. Drug Delivery Rev.*, **1997**, *23*, 3-25.

14. Advanced Chemistry Development Inc, Toront, Ontario, Canada. *http:www.acdlabs.com*, **2001**.

15. Johnson, P.E., Tomme, P., Joshi, M.D., and McIntosh, L.P. Interaction of Soluble Cellooligosaccharides with the N-Terminal Cellulose-Binding Domain of Cellulomonas Fimi Cenc. 2. NMR and Ultraviolet Absorption Spectroscopy. *Biochemistry*, **1996**, *35*, 13895-13906.

16. Wiesmann, C., Hengstenberg, W., and Schulz, G.E. Crystal Structures and Mechanism of 6-Phospho-B-Galactosidase from Lactococcus Lactis. *J. Mol. Biol.*, **1997**, *269*, 851-860.

17. Ufret, M.de.L., and Imperiali, B. Probing the Extended Binding Determinants of Oligosaccharyl Transferase with Synthetic Inhibitors of Asparagine-Linked Glycosylation. *Bioorg. Med. Chem. Lett.*, **2000**, *10*, 281-284.

18. Veber, D.F., Johnson, S.R., Cheng, H.-Y., Smith, B.R., Ward, K.W., and Kopple, K.D. Molecular Properties That Influence the Oral Bioavailability of Drug Candidates. *J. Med. Chem.*, **2002**, *45*, 2615-2623.

19. Virgilio, A.A., and Ellman, J.A. Simultaneous Solid-Phase Synthesis of B-Turn Mimetics Incorporating Side-Chain Functionality. *J. Am. Chem. Soc.*, **1994**, *116*, 11580-11581.

Chapter 6

Evaluation of peptidomimetic inhibitors of oligosaccharyl transferase in cell-based systems

Introduction

Oligosaccharyl transferase (OT) is a multimeric membrane bound protein that facilitates the co-translational transfer of a dolichol-pyrophosphate-linked tetradecasaccharide to the asparagine side chain of a nascent protein within the Asn-Xaa-Thr/Ser consensus sequence.[1] This intriguing enzyme is located in the lumen of the endoplasmic reticulum (Figure 6-1).
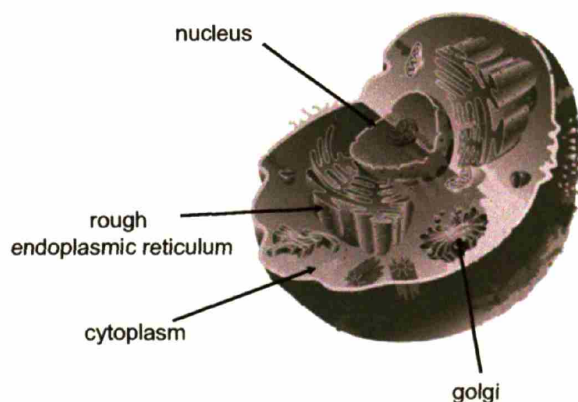
Figure 6-1. The oligosaccharyl transferase complex is located in the lumen of the rough endoplasmic reticulum. (Figure taken from http://202.195.138.253/ycx/chapter2/2-2.htm).

An inhibitor that targets OT within a cellular context has the potential to provide valuable information on the effect of glycosylation on downstream biological processes. N-linked glycosylation is one of the key post-translational protein modification processes that plays a role in immune response, intracellular targeting, intercellular recognition and protein folding and stability.[2] Currently the only inhibitor of N-linked glycosylation that functions *in vivo* is tunicamycin.[3] This microbial product is a bisubstrate-analog that inhibits the first

committed step in the assembly of the dolichol-pyrophosphate-linked oligosaccharide donor. Since this transformation occurs numerous steps prior to the reaction catalyzed by OT, the effect of the inhibitor on the actual protein-glycosylation step is not immediate, nor does it have the potential to reveal the specific consequences of blocking N-linked glycosylation. For these reasons, a specific inhibitor of OT that functions *in vivo* has several important applications.
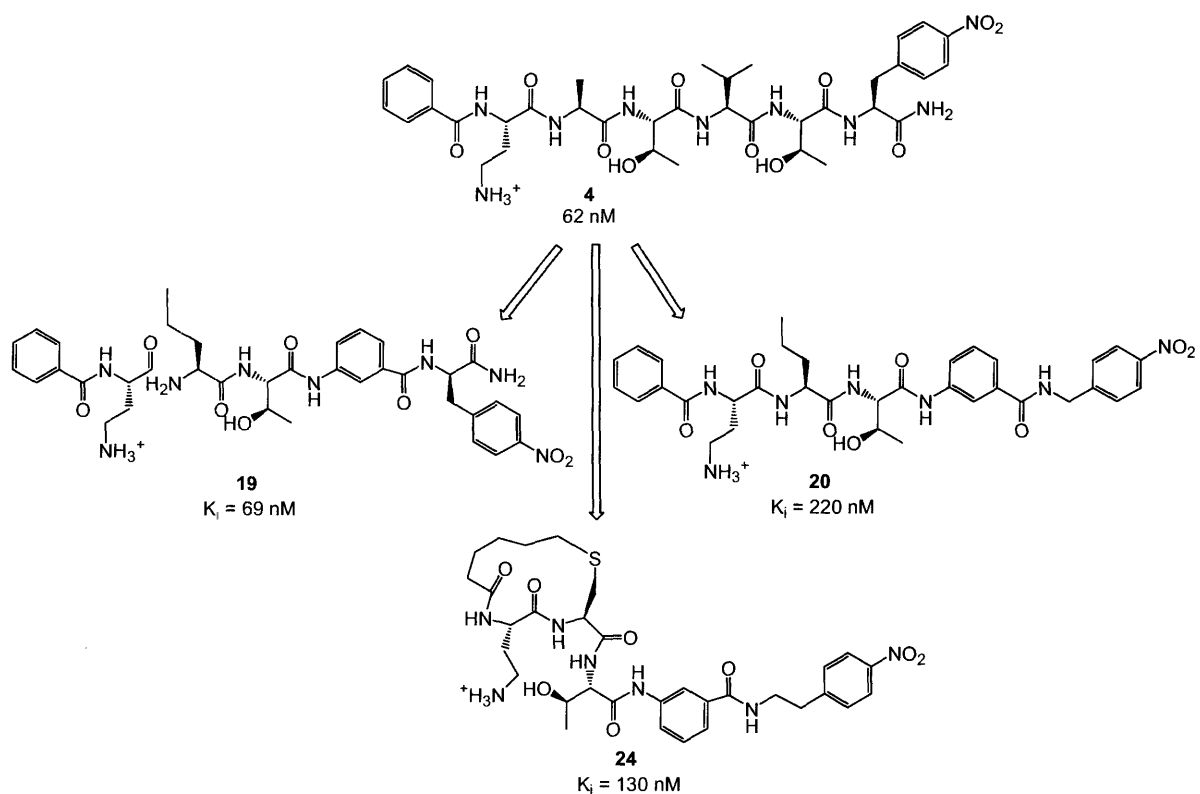
Figure 6-2. Peptidomimetic inhibitors of oligosaccharyl transferase.[4]

An inhibitor that targets OT *in vivo* needs to permeate both the plasma membrane and the endoplasmic reticulum membrane. The inhibitor also has to show a high degree of

179

stability to proteolytic degradation by cytoplasmic proteases. In Chapter 5, research towards increasing the non-peptide like character of previously synthesized peptide-based inhibitors is discussed. In those studies, the peptide backbone was modified through the incorporation of dipeptide isosteres to yield a small library of peptidomimetic structures with high affinity for OT (representatives of this library are shown in Figure 6-2).[4] In this chapter, these peptidomimetic inhibitors are investigated for their cell permeability and cell lysate stability properties, as well as their ability to inhibit N-linked glycosylation *in vivo* using a novel high-throughput assay.

Results and Discussion

Cell lysate stability

One of the major advantages of using a peptidomimetic inhibitor rather than a peptide-based inhibitor is the reduced amide-bond character that affords greater proteolytic stability.[5, 6] We were interested in comparing the cell lysate stability of our peptidomimetics with the parent peptide-based inhibitor (4) to verify this prediction.

In order to explore the cell lysate stability, solutions of 20 µM of either the peptide inhibitor (4) or the peptidomimetic inhibitor (20) were incubated in Chinese Hamster Ovary (CHO) cell lysate for 24 hours and the resulting mixture analyzed by HPLC (Figure 6-3). After exposure to proteases in cell lysate, several degradation peaks were apparent on the HPLC trace of the hexapeptide 4. The peptidomimetic 20, however, was present as mainly a single peak with negligible degradation products. Hence it is apparent that the reduced amide-bond character of the peptidomimetic, as well as the incorporation of non-native amino acid

scaffolds within the structure, renders the resulting compound considerably less susceptible to protease degradation.
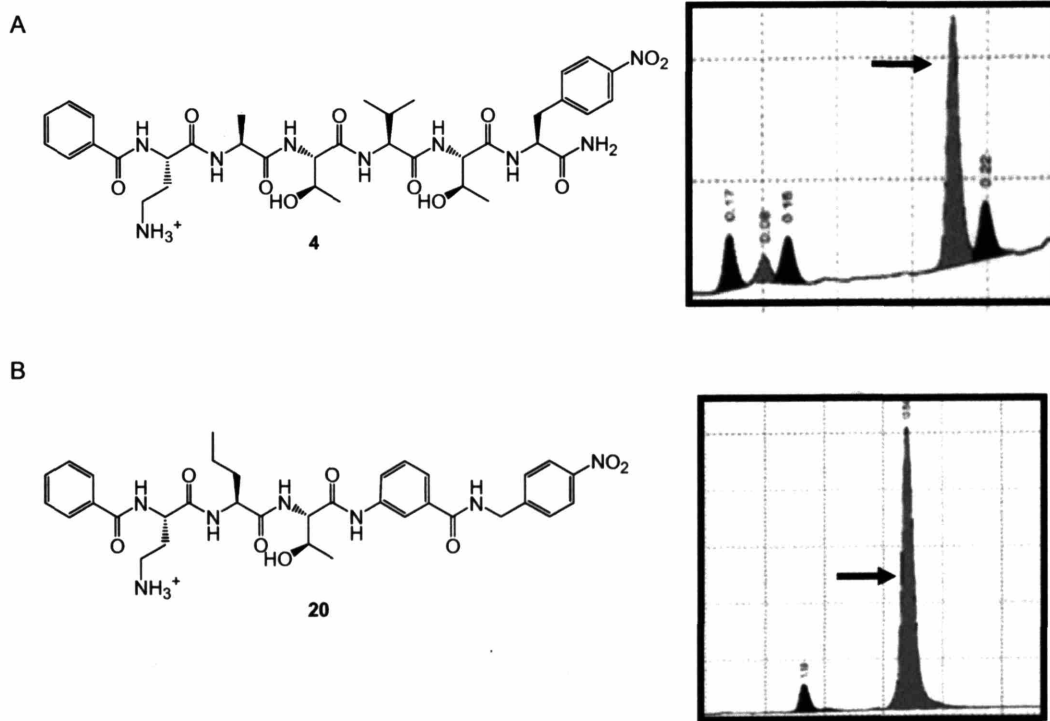
Figure 6-3. Cell lysate stability.

HPLC traces demonstrating the cell lysate stability of the peptide and peptidomimetic inhibitors. (A) Peptide (**4**). (B) Peptidomimetic (**20**). Arrows designate peak corresponding to non-degraded peptide or peptidomimetic.

Cell permeability

In order to explore the cell permeability of the inhibitors, analogs of the peptide and peptidomimetic structure were synthesized with a fluorophore label. In order to minimize the effect of the fluorophore on the properties of the inhibitors, a nitrobenzoxadiazolyl (NBD) moiety was chosen. NBD has a relatively small size, a polar character, and a small dipole moment and as such would not be expected to interfere with peptide-membrane interactions.[7, 8] Two NBD-labeled compounds were synthesized, utilizing an N-α-Fmoc-Dap(NBD)-OH amino acid building block (Dap = diaminopropionic acid) that has been previously synthesized.[9] This fluorescent amino acid was incorporated into the peptide and peptidomimetic structures using standard Fmoc-based solid-phase peptide synthesis methods to afford the two compounds 25 and 26 (Figure 6-4).

NBD-peptide
25

NBD-peptidomimetic
26
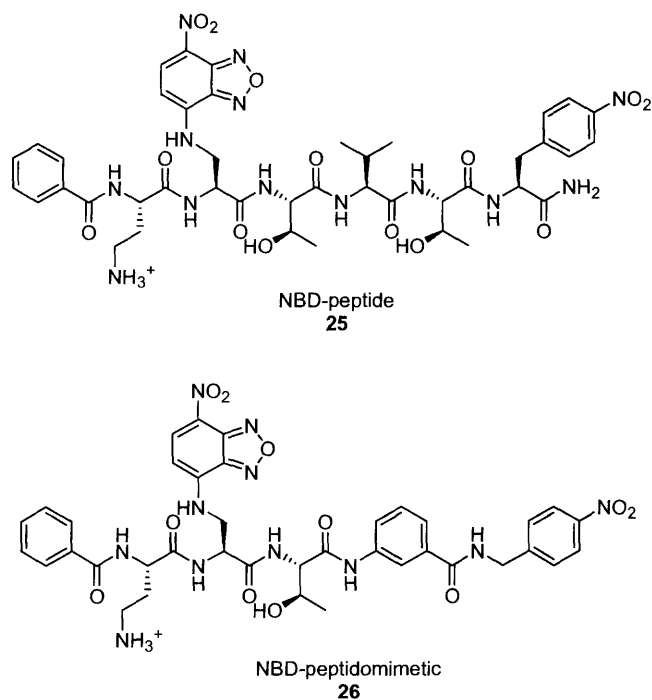
Figure 6-4. NBD labeled structures. (A) Peptide (**25**). (B) Peptidomimetic (**26**).

These NBD-labeled compounds were incubated with CHO cells at 20 μM for 2 hours and visualized under a fluorescence microscope to determine if the peptide and peptidomimetic are able to permeate the cell membrane. These fluorescence studies showed no fluorescence inside the cells for both the NBD-labeled peptide and peptidomimetic compounds (25 and 26) indicating that the peptidomimetic, although more hydrophobic than the peptide, was still unable to permeate the lipophilic cell membrane.

The peptidomimetic inhibitor (26) incorporates several aromatic moieties in its backbone and has reduced amide-bond character that enhances its hydrophobicity. However, the presence of the charged amine on the diaminobutyric acid (Dab) side chain may encumber the ability of the peptidomimetic to interact with the lipophilic cell membrane.[6] To investigate if the charge on the Dab side chain was hampering cell permeability, a further NBD-tagged construct was synthesized where the free amine on the Dab side chain was capped with the NBD fluorophore (27). The resulting aryl amine has a lower pK_a relative to the primary amine and is hence neutral at physiological pH (Figure 6-5).

27

Figure 6-5. Peptidomimetic construct (27) with the NBD-capped Dab side chain.

This compound was synthesized on solid-phase using an Alloc-protected Dab amino acid derivative, which was orthogonally deprotected using tetrakistriphenylphosphine palladium (0) and elaborated with NBD-chloride to yield the NBD-capped inhibitor **27** (Scheme 6-1).[10]
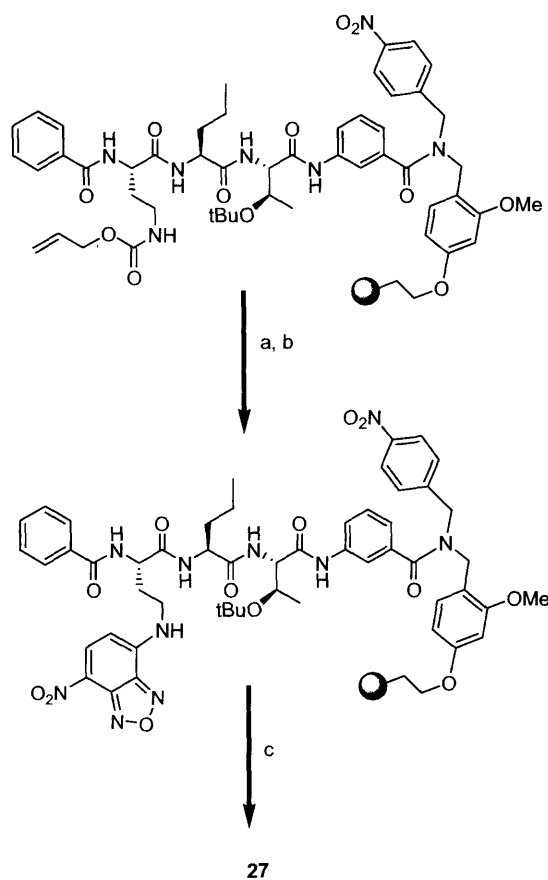
Scheme 6-1. Synthesis of the NBD-labeled peptidomimetic **27**.[a]

[a]Reagents and conditions: (a) Pd(PPh₃)₄, phenylsilane, CH₂Cl₂; (b) NBD-Cl (nitrobenzoxadiazolyl chloride), DIPEA, DMF; (c) cleavage from resin (95% TFA).

When compound **27** was incubated with CHO cells as before and visualized under the fluorescence microscope, the cells were brightly fluorescent (Figure 6-6). This compound permeates the cell membrane very efficiently, suggesting that the charged amine on the Dab side chain was indeed hindering the ability of the inhibitor to permeate the cell membrane.

Figure 6-6. Cellular uptake studies with the NBD-labeled peptidomimetic **27**.

Coumarin-based prodrug approach

The fluorescence uptake experiments suggest that the charged amine on the Dab side chain prevents cell permeation. This charged amine is, however, crucial for interaction with the OT active site, as inhibitors without this charge show no activity *in vitro*. Hence, it is necessary to transiently mask the side chain using a prodrug approach to inhibitor design. The prodrug strategy involves derivatizing certain polar functional groups transiently and bioreversibly to 'mask' the undesirable physicochemical characteristics of these groups without permanently altering the pharmacological properties of the molecule.[11] This strategy has been very successful in cases where a carboxyl or a hydroxyl group is derivatized into an ester group that is readily hydrolyzed *in vivo* by esterases.[12, 13] In the case of amine

functionalities, the task becomes more complicated as there are few chemical methods available to derivatize an amino group bioreversibly due to the inherent chemical stability of an amide linkage relative to that of an ester.[14]

The coumarin-based prodrug system takes advantage of the facile lactonization of *cis*-coumarinic acid and its derivatives. In such a strategy, a latent nucleophile can be unmasked using an esterase triggering mechanism that, in turn, initiates the cyclization reaction to release the parent drug (Figure 6-7).[15-17] Similar coumarinic derivatives have been successfully utilized in the synthesis of opioid peptides,[18] and peptidomimetic glycoprotein IIb/IIIa antagonists,[17, 19] and have shown increased membrane permeability.
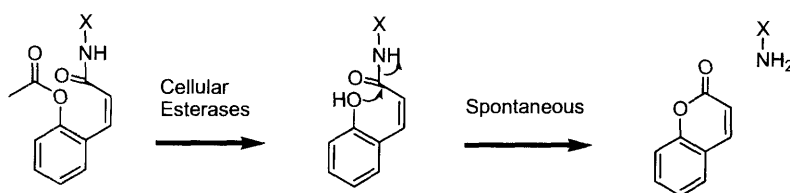
Figure 6-7. Mechanism of release of the coumarin-based prodrug, X = inhibitor.
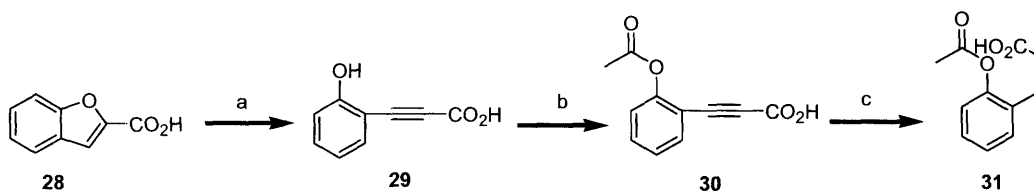
Scheme 6-2. Synthesis of the Coumarin-based prodrug precursor (**31**).[a]

[a]Reagents and conditions: (a) LDA/THF, -78 °C; (b) Ac₂O, Et₃N; (c) H₂, Lindlar's catalyst, EtOH.

The synthesis of the coumarin precursor was previously described in the literature[17] and involves a 4-step procedure starting with benzofuran-2-carboxylic acid (28) (Scheme 6-2). Lithium diisopropylamine (LDA) treatment produces 3-(2-hydroxyphenyl)propynoic acid (29). Esterification of the phenol with acetic anhydride yields the phenolic ester (30). The final step involves a catalytic hydrogenation of the alkyne with Lindlar's catalyst to yield the *cis* double bond in the coumarin precursor (31).
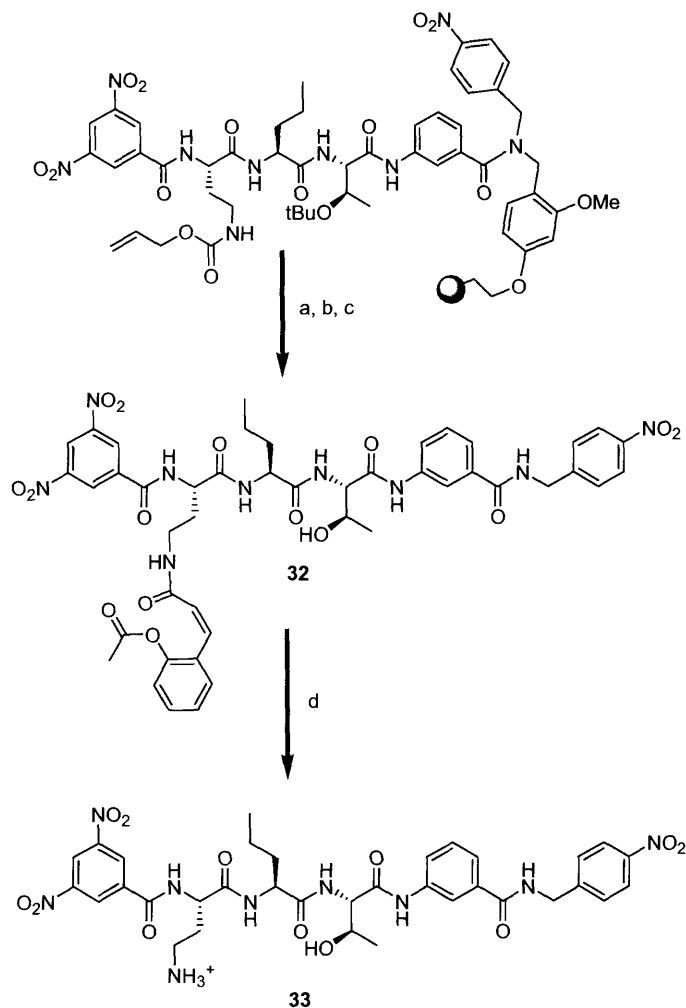
Scheme 6-3. Synthesis and esterase-cleavage of the coumarin-based prodrug.[a]

[a]Reagents and conditions: (a) Pd(PPh3)4, phenylsilane; (b) 31, HATU, DIPEA, DMF; (c) cleavage from resin (95% TFA); (d) cellular esterases.

This precursor is then coupled to the free amine on the Dab side chain on solid phase to afford the final prodrug **32** after cleavage with 95% TFA (Scheme 6-3). A dinitrobenzoyl cap was installed at the *N*-terminus to enable visualization in cells using a commercial antibody that specifically recognizes dinitrobenzene.[20] Upon exposure to cellular esterases, the coumarin masking group is released to reveal the active inhibitor **33**.

Incubation of the coumarin-based prodrug **32** in CHO-cell lysate showed minimal cleavage of the coumarin after 12 hours by HPLC analysis. Since the rate limiting step is the esterase-catalyzed cleavage of the acetate that triggers cyclization, it appears that the coumarin-bound inhibitor **32** is a poor substrate for these cellular esterases. Since there is negligible active inhibitor **33** formed after 12 hours, it is unlikely that this prodrug approach will be successful in targeting OT inhibitors into cells and hence this approach was abandoned.

Caged inhibitor

Another method, similar to the coumarin-based prodrug approach, to transiently mask a functionality, is the use of a photo-labile protecting group. Similar methods have been successful in the investigation of nitric oxide synthase, carbonic anhydrase and other signaling pathways.[21-23] Upon incubation with the protected form for several hours, the cells would be illuminated to cleave the photo-labile group and release free inhibitor within the cell. A nitroveratryloxycarbonyl (NVoc) group was chosen as the 'cage' and incorporated into the inhibitor framework on solid phase to afford compound **34** (Scheme 6-4).[24, 25]
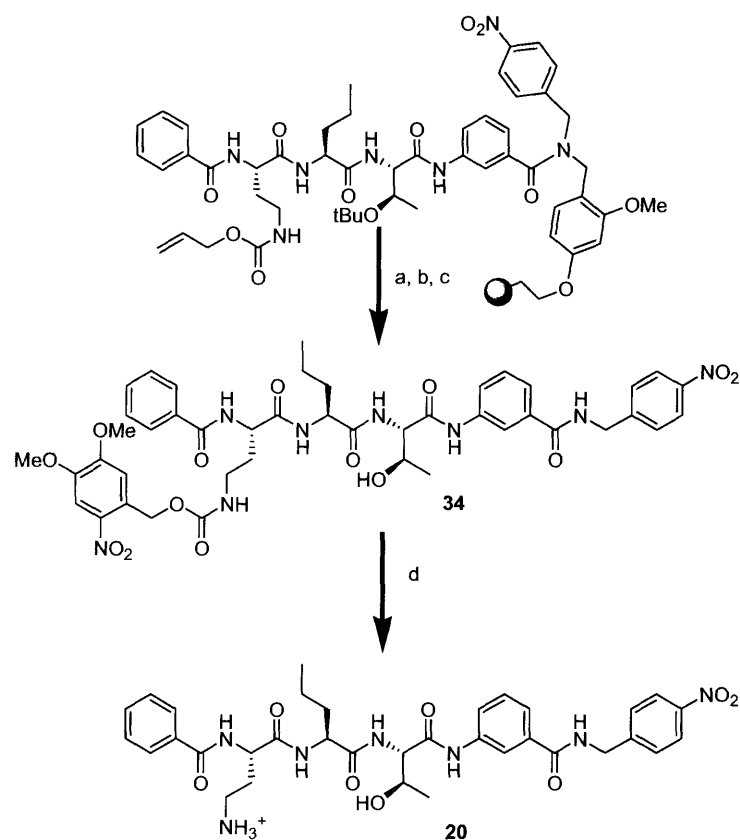
Scheme 6-4. Synthesis and photolysis of the NVoc-labeled inhibitor.[a]

[a]Reagents and conditions: (a) Pd(PPh₃)₄, phenylsilane; (b) NVoc-Cl, DIPEA, DMF; (c) cleavage from resin (95% TFA); (d) illuminate at λ = 365 nm.

Cleavage of the NVoc protecting group was tested *in vitro*, by illuminating a solution of the caged inhibitor (**34**) using a transilluminator and monitoring the release of the inhibitor (**20**) using HPLC analysis. Within 5 minutes, almost 70% cleavage of the NVoc group had occurred (Figure 6-8). This compound could be used as a potential inhibitor by incubating with cells, illuminating the cells at 365 nm to remove the NVoc protection and monitoring the activity of the inhibitor.

0 min.   2 min.   5 min.
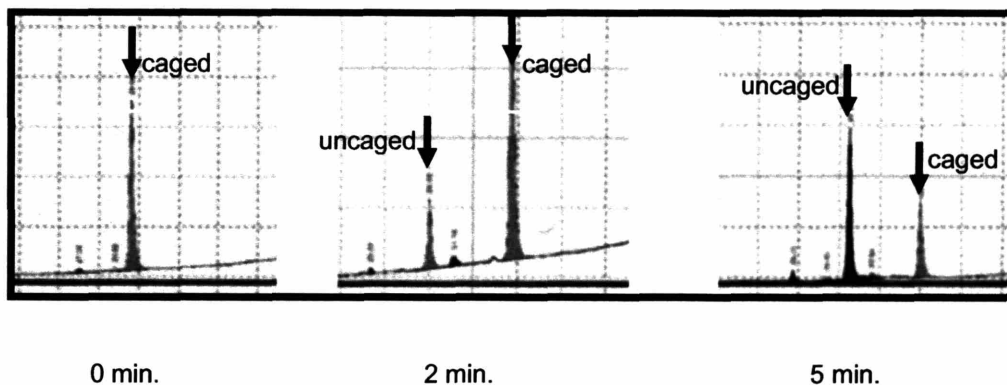
Figure 6-8. Photolysis of the NVoc group (progress monitored by HPLC).

Secreted Alkaline Phosphatase (SeAP) Assay

In order to study the effectiveness of OT inhibitors *in vivo*, it was necessary to develop an assay to monitor *N*-linked glycosylation within cells. Since we now have a series of potential candidates for inhibition *in vivo*, it was desirable to design an assay system that is high-throughput and amenable to screening several compounds concurrently.

The assay that we developed is based on a stably-transfected CHO cell line expressing a reporter-recombinant glycoprotein, SeAP (Secreted Alkaline Phosphatase). SeAP is a glycoprotein that contains several *N*-linked glycosylation sites,[26, 27] and previous studies on insect cells have shown that pre-treating cells with tunicamycin results in a dramatic decrease in the amount of SeAP secreted into the media.[27] This sparked the idea of a possible use for SeAP to monitor levels of glycosylation in cells. Inhibiting *N*-linked glycosylation would result in a decrease in SeAP activity and this activity can be easily detected in the extracellular media through a fluorescence, chemiluminescence or UV-based absorbance assay.

The Kompala group at UC Boulder (Department of Chemical Engineering) has developed a stably transfected CHO cell line that produces SeAP when activated with a hormone, dexamethasone.[28-30] An MMTV-promoter replaces the SV40-promoter found in the commercially available plasmid, resulting in 10-times greater protein production. The ability to induce protein production is also an advantage as the cells can be incubated with inhibitors prior to initiating protein production.
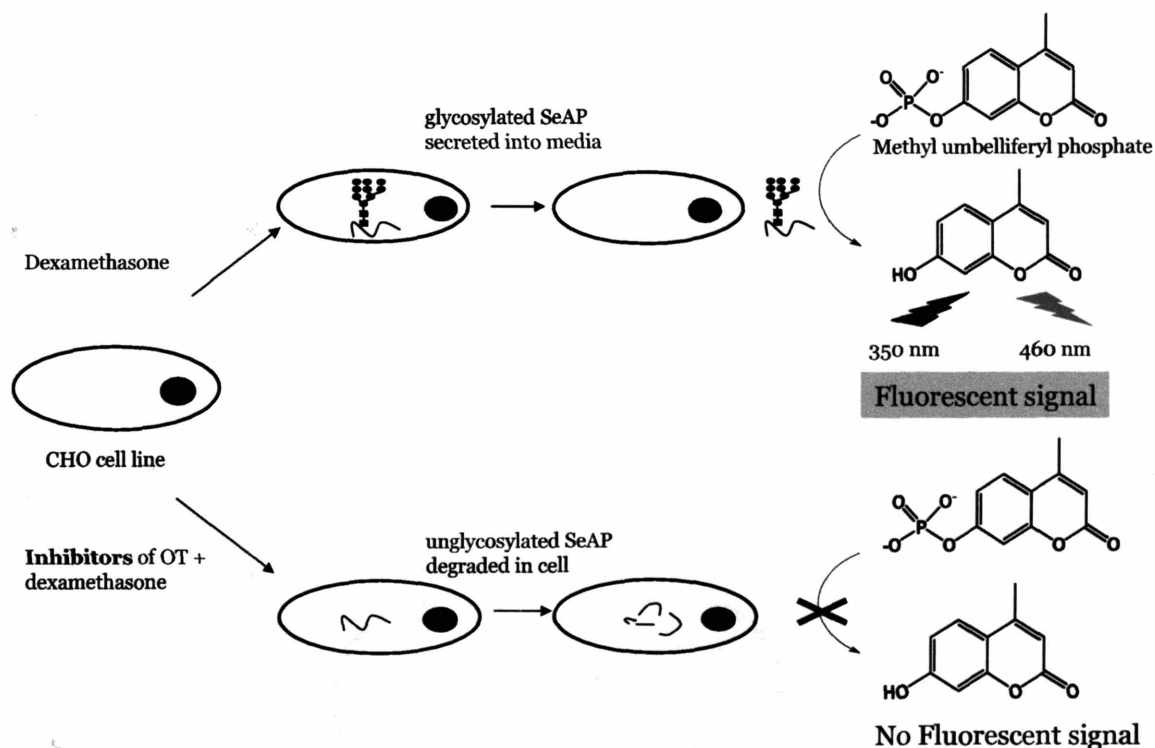
Figure 6-9. Schematic of the SeAP assay with and without inhibitors of *N*-linked glycosylation.

We obtained this cell line from the Kompala Group and started cell growth in α-MEM supplemented with Geneticin to select for cells containing the SeAP plasmid. The cells are then grown in a 96-well plate and incubated overnight with inhibitors after which SeAP

production is induced with the addition of dexamethasone. After 24 hours, a sample of the media is removed from each well and subjected to the addition of methylumbelliferyl phosphate (MUP). MUP is a coumarin derivative that is not fluorescent, but upon the action of the phosphatase, the phosphate is cleaved to yield a fluorescent coumarin derivative.[31] Hence, control cells that do not contain inhibitor will produce fully glycosylated SeAP that is released into the media resulting in a high fluorescence reading. On the other hand, cells that are exposed to an inhibitor of N-linked glycosylation will produce unglycosylated SeAP that is not properly folded and hence not secreted. Thus there will be lower levels of SeAP in the extracellular milieu in the presence of inhibitor, resulting in a lower fluorescence reading (Figure 6-9).

In order to validate the principal behind the SeAP assay, tunicamycin was used as a positive control. Cells exposed to tunicamycin show reduced protein glycosylation, hence less SeAP is secreted into the media, resulting in a lower fluorescence signal. This observation was verified by incubating the cells with increasing amounts of tunicamycin and monitoring SeAP production. As shown in Figure 6-10, as the amount of tunicamycin is increased, there is a stepwise decrease in the fluorescence signal, indicating that there is reduced SeAP in the media.
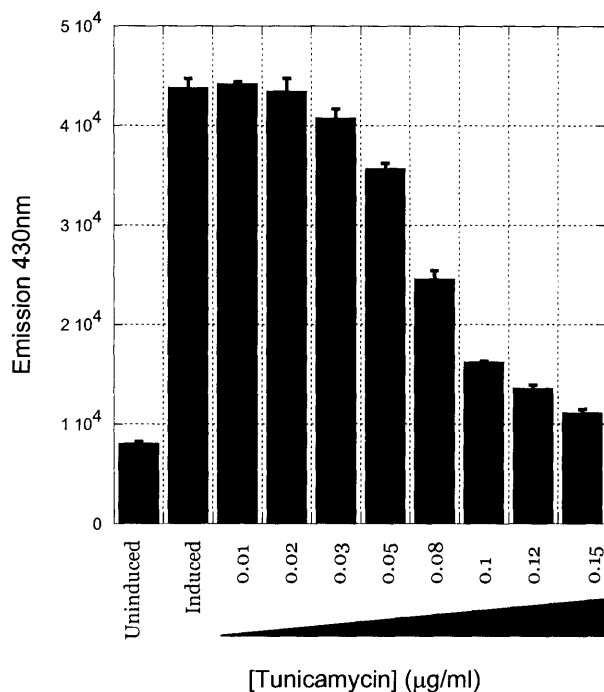
Figure 6-10. SeAP assay with increasing concentrations of tunicamycin.

Inhibitor library.

All of the inhibitors synthesized were tested using the SeAP assay. The inhibitors shown in Figure 6-11 are all *in vitro* inhibitors of OT with K_i values in the low to high-nanomolar range. These inhibitors display a variety of physical characteristics differing in size, hydrophobicity and amide-bond content. Several of the inhibitors have been verified to permeate the cell membrane and are shown to be stable to proteolytic digestion. It is as yet unknown if any or all of these compounds are able to permeate the membrane of the endoplasmic reticulum and bind to the peptide-binding site of the membrane-bound OT complex.
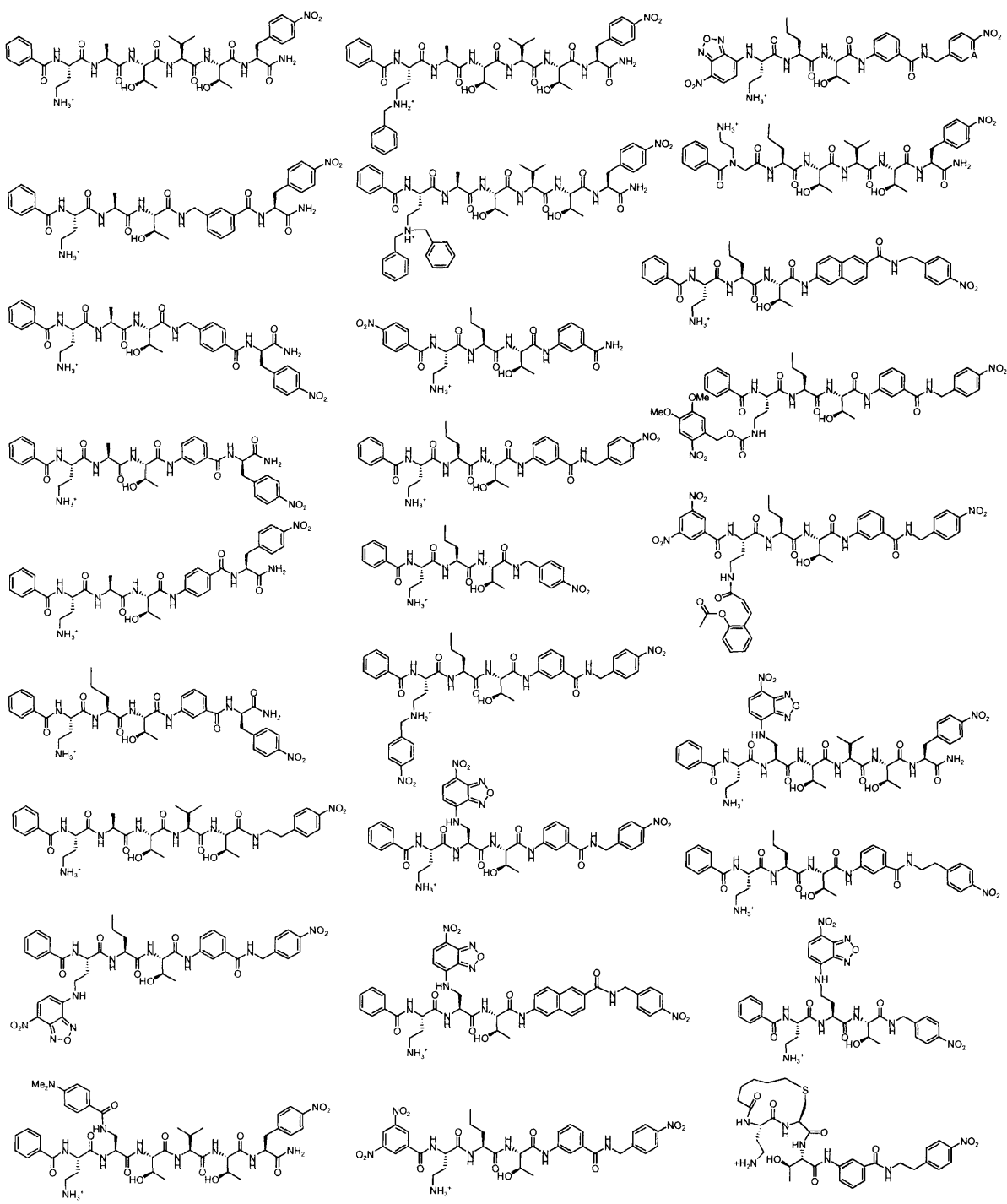
Figure 6-11. Library of peptide and peptidomimetic inhibitors tested in the SeAP assay.

194

Naphthyl based inhibitor

All of the inhibitors illustrated in Figure 6-11 were assayed in the SeAP system at 50 μM concentration. Only one inhibitor yielded a positive result and this was compound **35**, which bears a naphthyl group in the backbone. This compound was synthesized on solid phase, by the coupling of Alloc-protected 6-amino-2-naphthoic acid to the resin. Alloc deprotection was then followed by standard peptide-coupling chemistry to yield **35** (Scheme 6-5).
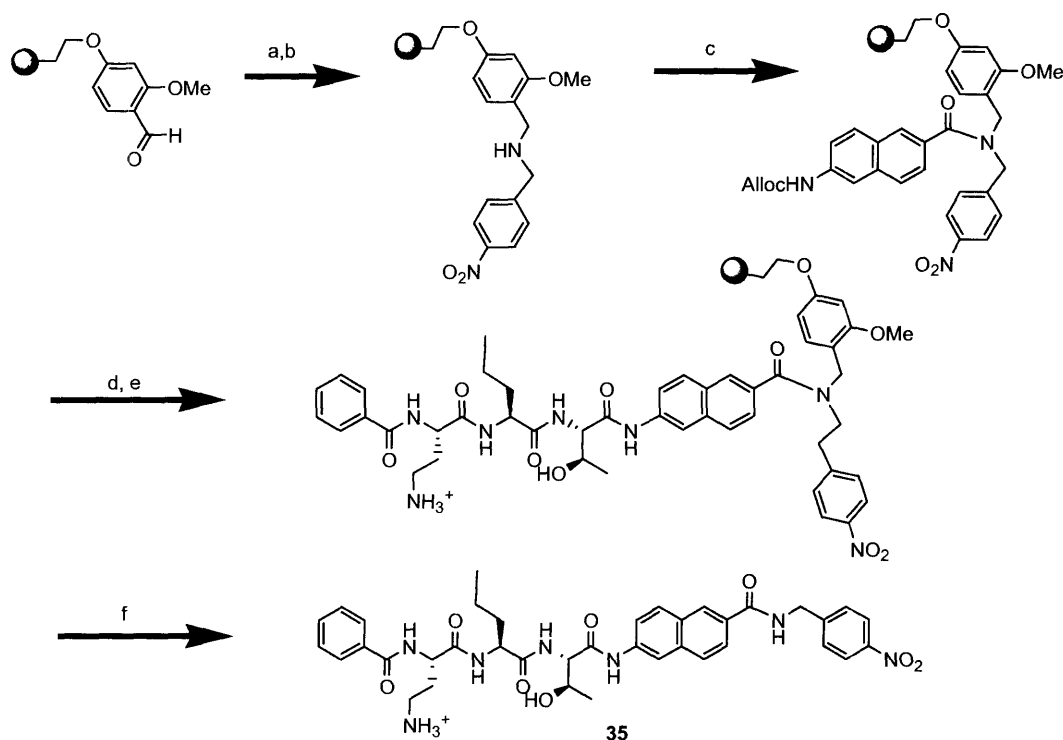
Scheme 6-5. Solid-phase synthesis of the naphthyl-based inhibitor (**36**).[a]

[a]Reagents and conditions: a) nitrobenzylamine, trimethyl orthoformate, dichloroethane; b) sodium triacetoxyborohydride; c) Alloc-6-amino-3-naphthoic acid, HATU, DIPEA; d) Pd(PPh₃)₄, phenylsilane; e) standard solid phase peptide coupling of Fmoc-Thr(tBu)-OH, Fmoc-Nva-OH, Fmoc-Dab(Boc)-OH, followed by capping with benzoic anhydride and pyridine; f) cleavage from resin (95% TFA).

Similar to the SeAP assay with tunicamycin, the fluorescence data for the naphthyl-based inhibitor displayed a stepwise decrease in fluorescence upon addition of increasing inhibitor (Figure 6-12). Even at levels as low as 8 μM of inhibitor, there is a significant decrease in the amount of SeAP in the media, which is a very promising indication of an inhibitor that functions *in vivo*. This naphthyl-based inhibitor has a K_i of 900 μM *in vitro*, which is relatively poor compared to some of the other inhibitors in the library. However, this compound has a log *P* value that is 4.01 +/- 1.16, which is higher than any other member of the library.[32] These data suggests that this naphthyl-based inhibitor functions *in vivo* due to its extreme hydrophobicity and its ability to cross the lipophilic cell membrane more efficiently than the other members of the library.
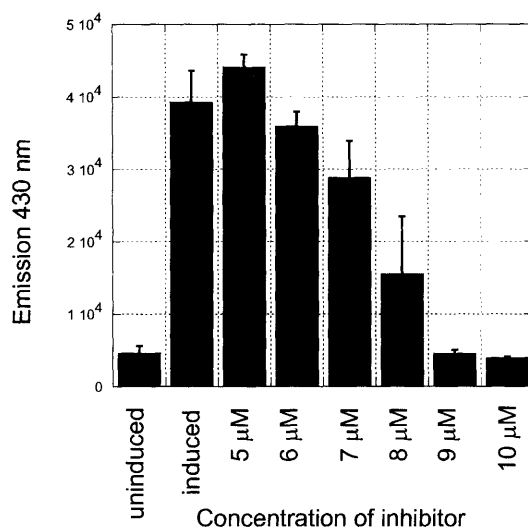
Figure 6-12. SeAP data for the naphthyl-based inhibitor **35**.

Metabolic Labeling Assay.

Although the SeAP assay functions well as a high-throughput assay to screen a library of compounds concurrently, it is a "negative read-out" assay, which is not optimal. Hence we propose using the SeAP assay as a first-round high-throughput screen for inhibitors of glycosylation, then following up with a more labor intensive secondary assay on the "hits" from the SeAP assay in order to further confirm the data.

The secondary assay that was used relies on the same cell line as before. Instead of simply monitoring the formation of SeAP by fluorescence analysis of the media, we probed for both glycosylated and unglycosylated forms of SeAP using a metabolic labeling/immunoprecipitation approach. The cells were plated and incubated with inhibitor as before and SeAP production was induced with dexamethasone. The media from the wells was collected and the cells lysed to release all proteins sequestered inside. The combined cell media and lysate mixtures were treated with an antibody specific to SeAP. This antibody binds both glycosylated and unglycosylated SeAP in the solution, and protein G beads are used to pull down this antibody-protein complex (Figure 6-13). The proteins released from the G-beads by boiling are separated by SDS-page electrophoresis, which should show glycosylated SeAP in the control lanes with no inhibitor, but unglycosylated SeAP in those cells treated with inhibitor. The unglycosylated band should run noticeably faster than the glycosylated band.
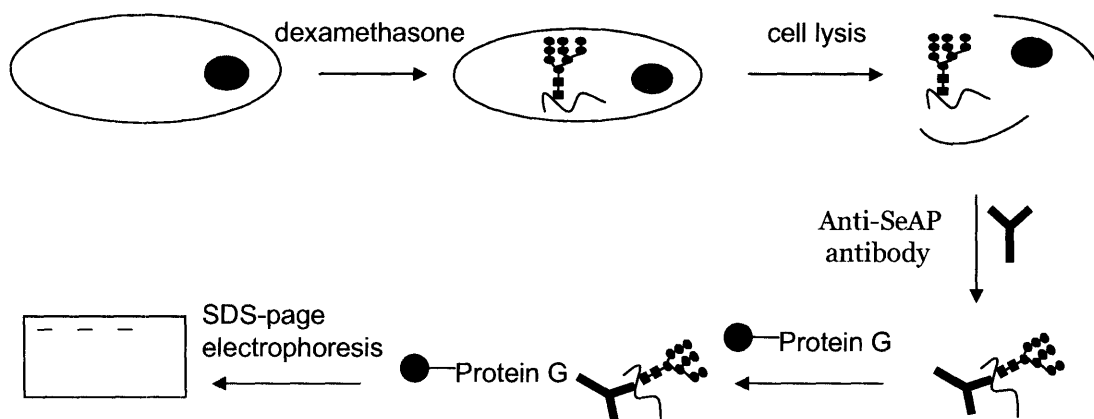
Figure 6-13. Schematic of the metabolic labeling/immunoprecipation assay.

Our first few attempts at this assay used a commercially available anti-human placental alkaline phosphatase antibody. However, we discovered that this antibody and all others that are commercially available are conformational antibodies that recognize only the fully-folded protein. Since the unglycosylated SeAP that we are monitoring is unfolded, we were unable to visualize a band corresponding to unglycosylated protein. In order to circumvent this problem, we custom synthesized an antibody to SeAP based on a 15 amino acid sequence that is present in an exposed-loop region of SeAP. With this antibody in hand, attempts to definitively prove that inhibitor **35** is affecting N-linked glycosylation can be undertaken.

Conclusion

In order to investigate the downstream effects of glycosylation on protein function, it is vital to obtain an inhibitor of *N*-linked glycosylation that functions in a cellular milieu. Although tunicamycin, a microbial natural product, does act *in vivo*, it acts several metabolic steps prior to the actual glycosylation step, hence its effect is not immediate. For these reasons, it would be ideal to obtain an inhibitor that directly targets OT *in vivo*.

As discussed in Chapter 5, peptide-based inhibitors of OT were modified to yield several peptidomimetic structures that demonstrated low nanomolar affinity for the enzyme. Through cell lysate incubations, we demonstrate that a representative member of the peptidomimetic library (compound **20**) displays greater proteolytic stability compared to the peptide counterpart (**4**). Several prodrug and caging strategies are used to transiently mask the charge on the amine of the Dab side chain, which hinders cell permeability. These studies afforded several inhibitors as potential candidates for *in vivo* inhibition of OT, and also provided insight into the properties that make compounds more cell-permeable and stable to proteases.

In order to monitor glycosylation in cells, a novel, high-throughput assay based on the activity of a secreted glycoprotein, SeAP, was developed. This assay allows inhibitor screening in a 96-well format, providing a fast read-out that reveals the cellular activity of the inhibitors. Using this SeAP assay, a hydrophobic inhibitor, based on a naphthyl scaffold (**35**), emerged as a successful candidate. Ongoing and future work involves the development of a secondary assay to directly visualize the production of unglycosylated protein in the presence of this inhibitor, so as to unambiguously prove that it is targeting *N*-linked glycosylation.

Experimental

General procedure for solid-phase peptide synthesis

All peptides, unless otherwise noted, were synthesized by manual solid-phase methods using Fmoc-PAL PEG resin and using Fmoc (9-fluorenylmethoxycarbonyl) as the protecting group for the α-amino functionalities. Amino acids were coupled using either 1-benzotriazolyloxytris(pyrrolidino)phosphonium hexafluorophosphate (PyBOP) or O-(7-azabenzotriazol-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate (HATU) and DIPEA to generate the activated ester. The resin was swelled in dichloromethane (5 min.) and DMF (5 min.) prior to use. Couplings to the resin were performed in the following order: removal of the Fmoc group (20% piperidine solution in DMF, 3 × 5 min.), wash (DMF, 5 × 1 min.), coupling (amino acid/PyBOP/DIPEA, 1:2:2 in DMF, 1 hour) and rinse (DMF, 2 × 1 min., CH_2Cl_2, 2 × 1 min.). All amino acids were obtained commercially as the N-α-Fmoc-protected derivatives with the following side-chain protection: Dab(Boc), Dab(Alloc), Thr(t-Bu). At the conclusion of the peptide synthesis, the peptides were capped with a large excess (10 eq) of benzoic anhydride and pyridine (10 eq). Cleavage from resin was performed with trifluoroacetic acid (TFA): CH_2Cl_2: triisopropylsilane: H_2O (90:5:2.5:2.5). The mixture was then filtered to remove the resin and concentrated under a stream of nitrogen. The resulting pellet was triturated with ice-cold ether and redissolved in DMSO. The DMSO solution was then diluted into a solution of 1:1 water:acetonitrile and the peptides were purified by preparative HPLC (C_{18}) using a gradient of increasing acetonitrile/0.1%TFA (solvent A) in water/0.1%TFA (solvent B): the standard gradient used in this study was 93:7 A:B to 0:100 A:B in 35 mins. Elution from the column was monitored by UV absorbance at 280 nm and

228 nm, and the product peak was collected and lyophilized. All peptide identities were confirmed by electrospray mass spectrometry (ESI-MS) and the peptide solutions were quantified by UV absorbance at 280 nm (ε_{280} = 12,800 M^{-1} cm^{-1} for the p-nitrophenyl group).

Synthesis of Bz-Dab-Dap(NBD)-Thr-Val-Thr-Nph-NH$_2$ (25)
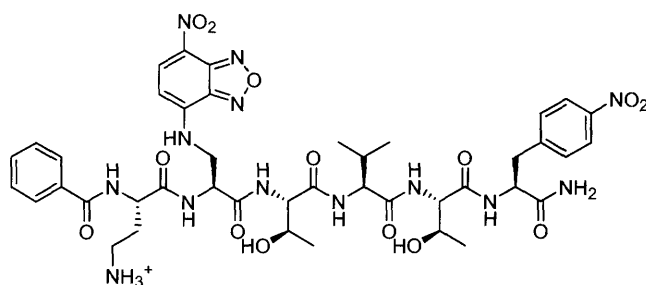
Fmoc-PAL-PEG PS resin was used and Fmoc-Nph-OH, Fmoc-Thr(tBu)-OH, Fmoc-Val-OH, and Fmoc-Dab(Boc)-OH were coupled using standard Fmoc-based peptide synthesis. A Fmoc-Dap(NBD)-OH amino acid that was previously synthesized in the group by Dr. Eugenio Vazquez was used to install the NBD fluorophore. This coupling was performed using standard procedure. The peptide was capped with benzoic anhydride, cleaved from resin and purified by HPLC.

HPLC t_R= 24.11 min (C$_{18}$, 7-100% B in 28 min)

LRMS calcd for 25 (C$_{42}$H$_{54}$N$_{13}$O$_{14}^{+}$) requires m/z 964.9. Found 964.3 (ESI+)

Synthesis of Bz-Dab-Dap(NBD)-Thr-(3-aminobenzoic acid)-NH-CH₂-Ph-NO₂ (26)
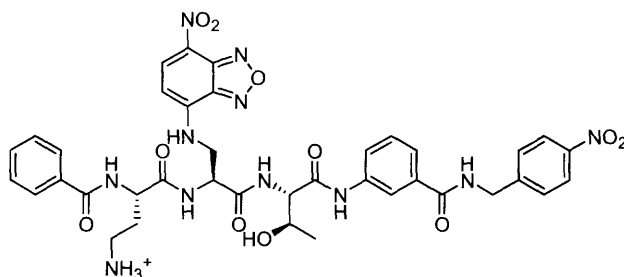
This peptide was synthesized on 3-(4-formyl-3-methoxyphenoxy)ethyl polystyrene (FMPE) resin available from Novabiochem. The resin was swelled in trimethylorthoformate (TMOF) and dichloroethane (DCE) 3:2 for 20 minutes. 4-nitrobenzylamine (10 eq) was added to the resin and stirred under N_2 for 3 hours, then 10 eq of sodium triacetoxyborohydride was added and the mixture shaken overnight. Alloc-protected 3-aminobenzoic acid was added in excess (4 eq) with HATU (4 eq) and diisopropylethylamine (DIPEA) (8 eq). The coupling was allowed to proceed overnight, after which alloc deprotection was performed with tetrakis(triphenylphosphine) palladium (0) (0.2 eq) and phenylsilane (25 eq) in $CHCl_2$. Three 20-minute deprotections were performed under an N_2 atmosphere. Fmoc-Thr(tBu)-OH coupling was performed using HATU as above, and the subsequent Fmoc-Dap(NBD)-OH and Fmoc-Dab(Boc)-OH residues were coupled using standard procedure.

HPLC t_R= 26.29 min (C_{18}, 7-100% B in 28 min)

LRMS calcd for 26 ($C_{38}H_{40}N_{11}O_{11}^+$) requires *m/z* 826.8. Found 826.2 (ESI+)

Synthesis of Bz-Dab(NBD)-Nva-Thr-(3-aminobenzoic acid)-NH-CH$_2$-Ph-NO$_2$ (27)
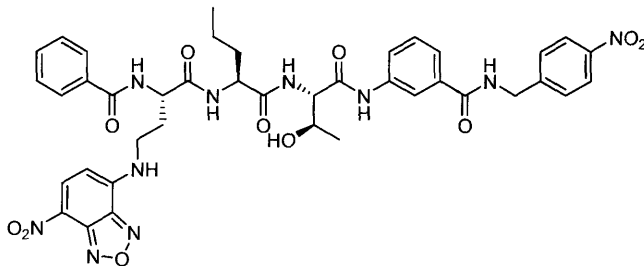
The procedure for the synthesis of compound **26** was followed except that after the coupling of Fmoc-Thr(tBu)-OH, Fmoc-Nva-OH was coupled. This was followed by the coupling of Fmoc-Dab(Alloc)-OH. The peptide was capped with benzoic anhydride as before. The alloc side chain on the Dab residue was removed with tetrakis(triphenylphosphine) palladium (0) (0.2 eq) and phenylsilane (25 eq) in CHCl$_2$. Three 20-minute deprotections were performed under a N$_2$ atmosphere. The side chain on the Dab was capped with 4 eq. of nitrobenzoxadiazolyl chloride (NBD-Cl) and 4 eq. of diisopropylethylamine (DIPEA) in DMF. The reaction was allowed to proceed for 2 hours and the compound was cleaved from the resin and purified by HPLC as before.

HPLC t$_R$= 29.02 min (C$_{18}$, 7-100% B in 28 min)

LRMS calcd for 27 (C$_{40}$H$_{42}$N$_{10}$O$_{11}$$^+$) requires *m/z* 838.3. Found 839.6 (ESI+)

Synthesis of the coumarin-based prodrug (3,5-DNB-Dab(Coum)-Nva-Thr-(3-aminobenzoic acid)-NH-CH₂-Ph-NO₂) (32)
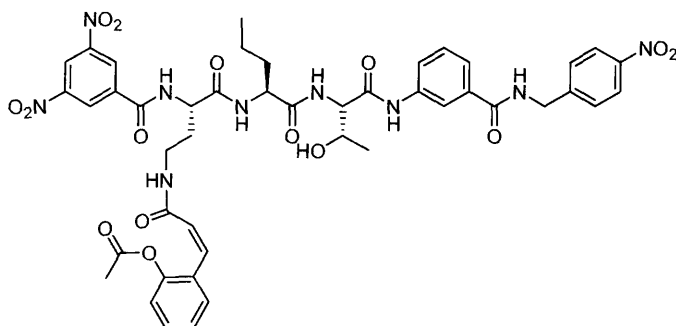
The procedure for the synthesis of compound **27** was followed. After the coupling of Fmoc-Dab(alloc)-OH, the peptide was capped with 3,5-dinitrobenzoyl chloride (4 eq.) and DIPEA (4 eq.) in DMF. The alloc side chain on the Dab residue was removed as described before with tetrakis(triphenylphosphine) palladium (0) (0.2 eq) and phenylsilane (25 eq) in CH_2Cl_2. The coumarinic acid precursor was synthesized as described in the literature[17] and coupled to the Dab side chain using HATU and DIPEA. The compound was then cleaved from the resin and purified by HPLC as before.

HPLC t_R= 29.51 min (C_{18}, 7-100% B in 28 min)

LRMS calcd for 32 ($C_{45}H_{47}N_9O_{15}^+$) requires *m/z* 953.3. Found 954.2 (ESI+)

Synthesis of the caged inhibitor (Bz-Dab(NVoc)-Nva-Thr-(3-aminobenzoic acid)-NH-CH₂-Ph-NO₂) (34)
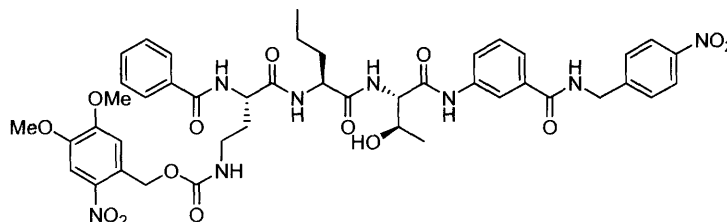
The procedure for the synthesis of compound **27** was followed. After deprotection of the alloc protection on the Dab side chain, the nitroveratryloxycarbonyl (NVoc) group was installed by coupling the free amine to NVoc-chloride (2 eq.) with DIPEA (4 eq.). The peptide was cleaved from resin and purified as before.

HPLC t_R= 30.41 min (C_{18}, 7-100% B in 28 min)

LRMS calcd for 34 ($C_{48}H_{48}N_{12}O_{17}^{+}$) requires m/z 1065.0. Found 1065.3 (ESI+)

Synthesis of Naphthyl-based inhibitor (Bz-Dab-Nva-Thr-(6-amino-2-naphthoic acid)-NH-CH₂-Ph-NO₂) (36)
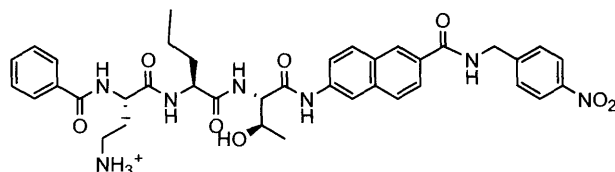
This inhibitor was synthesized by performing a reductive amination on FMPE resin with 4-nitrobenzylamine as described previously. This was followed by the coupling of alloc-

protected 6-amino-2-naphthoic acid to the resulting resin-bound secondary amine using HATU (4 eq.) and DIPEA (4 eq.). The alloc protection is removed with tetrakis(triphenylphosphine) palladium (0) (0.2 eq) and phenylsilane (25 eq) in CH_2Cl_2. The remaining amino acids are then coupled using standard procedure.

HPLC t_R= 30.41 min (C_{18}, 7-100% B in 28 min)

LRMS calcd for 34 ($C_{38}H_{44}N_7O_8^+$) requires m/z 726.8. Found 726.9 (ESI+)

Cell lysate stability assay.

CHO K1 cells were grown on a 6-well plate until they reached 80% confluency. The media was removed from the wells and the cells lysed with 1 mL of PBS containing 1% Triton X-100. After agitating the cells in the lysis buffer for 15 minutes, the lysate was transferred to an eppendorf tube and centrifuged to remove cell debris. Solutions of the peptide (**4**) and peptidomimetic (**20**) in DMSO were added to 1 mL of lysate to a final concentration of 20 μM (1% final volume of DMSO). The lysates were incubated at 35 °C for 24 hours, filtered to remove any precipitate and analyzed by analytical HPLC (C_{18}) using a gradient of increasing acetonitrile/0.1%TFA in water/0.1%TFA (93% to 0% water in 35 mins.). Elution of the peptide was detected by UV absorbance at both 228 nm and 280 nm. A blank run of lysate with no inhibitor was used as a control to determine which of the peaks were due to the exogenous compounds added to the lysate.

Cell permeability assays

CHO K1 cells were plated on glass slides (2E6 cells/well) and allowed to grow for 4-6 hours. The inhibitors (in DMSO) were then added with complete media (200 μL with 1% final

volume of DMSO) and the cells incubated overnight. The media was removed and the cells washed thoroughly with PBS + 0.5% FBS and observed under the fluorescent microscope with a filter set corresponding to the absorption and emission wavelengths of NBD (λ_{ex} = 481 nm, λ_{em} = 541 nm).

Cell lysate incubation of the coumarin-based prodrug

CHO K1 cell lysate was prepared as before using PBS + 1% Triton X-100. 50 μM of the Coumarin-based prodrug was incubated in the cell lysate and cleavage of the prodrug moiety was monitored by HPLC (C_{18}). Aliquots of the lysate mixture were removed at 1 hour, 2 hour, 5 hour and 15 hour time points and injected on the HPLC. Cleavage of the coumarin was monitored by observing the peaks at λ = 228 nm and λ = 280 nm.

Photolysis of NVoc-protected inhibitor

The concentrated inhibitor stock (in DMSO) was diluted into PBS buffer to a final concentration of 50 μM. The samples were then irradiated with a UVP UV Transilluminator, 365 nm, 7330 μW/cm² at the surface, in vessels with a path length of 1 mm. The samples were irradiated for 2 min. and 5 min. and the extent of photolysis monitored by injecting the samples on the HPLC.

SeAP Assay

MMapG CHO cells were grown in α-MEM supplemented with 10 % FBS, 4 mg/ml Geneticin and Penn/Strep. To run the assay, the cells were plated in a 96-well plate at 1E6 cells/ml in 200 μL of CD-CHO serum-free media (Gibco/Invitrogen) supplemented with 1% FBS. The

cells were allowed to settle for 4 hours and inhibitors (in DMSO) were added (2.5% final DMSO concentration). The cells were incubated overnight in the presence of inhibitors. The media was then removed and replaced with fresh media containing fresh inhibitors and 1 mM dexamethasone. The cells were then incubated for 24 hours, after which they were assayed for SeAP activity. 5 µL from each well was transferred to a black well plate, and 95 µL of methyl umbelliferyl phosphate (MUP) solution was added. The MUP solution is premixed and is a 1 µM solution of MUP in 2 mM diethanolamine. The plate was then incubated in the dark at 35-40 °C and the fluorescence recorded using a HTS 7000 BioAssay Reader (Perkin-Elmer). λ_{ex} = 360 nm, λ_{em} = 430 nm.

Metabolic Labeling Assay:

This assay is performed in a 24-well plate with 500 µL of media. The night before, MMapG CHO cells were plated at 3E5 cells/well in 500 µL of media (α-MEM, 5% dialyzed FBS, penn/strep, 100 µM MSX). Inhibitors were added at this time. 24 hours later, the media was removed from cells and washed with PBS. 500 µL per well of Met-/Cys- DMEM (+5% FBS, Penn/Strep, MSX) was added. Inhibitor was added along with 5 µL of Trans S-35 label. The cells were incubated for 5 hours at 37 °C in the CO_2 incubator. The media was transferred to an eppendorf tube, spun to remove cell debris and the supernatant transferred to a fresh tube. The cells were lysed with 300 µL PBS + 1% Triton X and added to the cell media. A 4-uL portion of antibody (anti-human placental alkaline phosphatase, 3 µg/ml) was added together with 50 µL of 1 M Tris-Cl pH 7.5. The media was incubated at 4 °C for 1 hour, then 30 µL of well-resuspended protein-G beads were added and incubated for another hour. The beads were pelleted and washed 3 times with 500 µL wash buffer (1X PBS with 0.5% Triton-

X) and after the final wash the beads were taken up in 15 μL Laemmeli running buffer. The samples were boiled and the supernatant loaded on a 15% SDS-PA gel using pre-stained markers and C-14 labeled markers. After running the gel, it was dried on a gel-dryer and read on a PhosphoImager after exposing for 24 hours.

References

1.	Dempski, R.E., and Imperiali, B. Oligosaccharyl Transferase: Gatekeeper to the Secretory Pathway. *Curr. Opin. Chem. Biol.*, **2002**, *6*, 844-850.

2.	Varki, A. Biological Roles of Oligosaccharides - All of the Theories Are Correct. *Glycobiology*, **1993**, *3*, 97-130.

3.	Heifetz, A., Keenan, R.W., and Elbein, A.D. Mechanism of Action of Tunicamycin on the Udp-GlcNAc-Dolichyl-Phosphate GlcNAc-1-Phosphate Transferase. *Biochemistry*, **1979**, *18*, 2186-2192.

4.	Weerapana, E., and Imperiali, B. Peptides to Peptidomimetics: Towards the Design and Synthesis of Bioavailable Inhibitors of Oligosaccharyl Transferase. *Org. Biomol. Chem.*, **2003**, *1*, 93-99.

5.	Veber, D.F., Johnson, S.R., Cheng, H.-Y., Smith, B.R., Ward, K.W., and Kopple, K.D. Molecular Properties That Influence the Oral Bioavailability of Drug Candidates. *J. Med. Chem.*, **2002**, *45*, 2615-2623.

6.	Lipinski, C.A., Lombardo, F., Dominy, B.W., and Feeney, P.J. Experimental and Computational Approaches to Estimate Solubility and Permeability in Drug Discovery and Development Settings. *Adv. Drug Delivery Rev.*, **1997**, *23*, 3-25.

7.	Uchiyama, S., Santa, T., and Imai, K. Fluorescence Characteristics of Six 4,7-Disubstituted Benzofurazan Compounds: An Experimental and Semi-Empirical Mo Study. *J. Chem. Soc. Perk. T. 2*, **1999**, 2525-2532.

8.	Bertorelle, F., Dondon, W., and Fery-Forgues, S. Compared Behavior of Hydrophobic Fluorescent Nbd Probes in Micelles and in Cyclodextrins. *J. Fluoresc.*, **2002**, *12*, 205-207.

9.	Dufau, I., and Mazarguil, H. Design of a Fluorescent Amino Acid Derivative Usable in Peptide Synthesis. *Tetrahedron Lett.*, **2000**, *41*, 6063-6066.

10.	Rapaport, D., and Shai, Y. Interaction of Fluorescently Labeled Pardaxin and Its Analogs with Lipid Bilayers. *J. Biol. Chem.*, **1991**, *266*, 23769-23775.

11.	Pauletti, G.M., Gangwar, S., Siahaan, T.J., Aube, J., and Borchardt, R.T. Improvement of Oral Peptide Bioavailability: Peptidomimetics and Prodrug Strategies. *Adv. Drug Delivery Rev.*, **1997**, *27*, 235-256.

12.	Wang, B.H., Gangwar, S., Pauletti, G.M., Siahaan, T.J., and Borchardt, R.T. Synthesis of a Novel Esterase-Sensitive Cyclic Prodrug System for Peptides That Utilizes a "Trimethyl Lock"-Facilitated Lactonization Reaction. *J. Org. Chem.*, **1997**, *62*, 1363-1367.

13. Gangwar, S., Pauletti, G.M., Siahaan, T.J., Stella, V.J., and Borchardt, R.T. Synthesis of a Novel Esterase-Sensitive Cyclic Prodrug of a Hexapeptide Using an (Acyloxy)Alkoxy Promoiety. *J. Org. Chem.*, **1997**, *62*, 1356-1362.

14. Pitman, I.H. Prodrugs of Amides, Imides, and Amines. *Med. Res. Rev.*, **1981**, *1*, 189-214.

15. Wang, B.H., Zhang, H.J., and Wang, W. Chemical Feasibility Studies of a Potential Coumarin-Based Prodrug System. *Bioorg. Med. Chem. Lett.*, **1996**, *6*, 945-950.

16. Wang, B.H., Zhang, H.J., Zheng, A.L., and Wang, W. Coumarin-Based Prodrugs. Part 3: Structural Effects on the Release Kinetics of Esterase-Sensitive Prodrugs of Amines. *Bioorg. Med. Chem.*, **1998**, *6*, 417-426.

17. Zheng, A.L., Wang, W., Zhang, H.J., and Wang, B. Two New Improved Approaches to the Synthesis of Coumarin-Based Prodrugs. *Tetrahedron*, **1999**, *55*, 4237-4254.

18. Wang, B.H., Wang, W., Zhang, H.J., Shan, D.X., and Smith, T.D. Coumarin-Based Prodrugs .2. Synthesis and Bioreversibility Studies of an Esterase-Sensitive Cyclic Prodrug of DADLE, an Opioid Peptide. *Bioorg. Med. Chem. Lett.*, **1996**, *6*, 2823-2826.

19. Wang, W., Borchardt, R.T., and Wang, B. Orally Active Peptidomimetic RGD Analogs That Are Glycoprotein IIb/IIIa Antagonists. *Curr. Med. Chem.*, **2000**, *7*, 437-453.

20. Bromberg, R., Kessler, N., and Addadi, L. Antibody Recognition of Specific Crystal Faces of 1,4-Dinitrobenzene. *J. Cryst. Growth*, **1998**, *193*, 656-664.

21. Kehayova, P.D., Bokinsky, G.E., Huber, J.D., and Jain, A. A Caged Hydrophobic Inhibitor of Carbonic Anhydrase II. *Org. Lett.*, **1999**, *1*, 187-188.

22. Curley, K., and Lawrence, D.S. Caged Regulators of Signaling Pathways. *Pharmacol. Ther.*, **1999**, *82*, 347-354.

23. Montgomery, H.J., Perdicakis, B., Fishlock, D., Lajoie, G.A., Jervis, E., and Guillemette, J.G. Photo-Control of Nitric Oxide Synthase Activity Using a Caged Isoform Specific Inhibitor. *Bioorg. Med. Chem.*, **2002**, *10*, 1919-1927.

24. Rusiecki, V.K., and Warne, S.A. Synthesis of N-Alpha-Fmoc-N-Epsilon-NVoc-Lysine and Use in the Preparation of Selectively Functionalized Peptides. *Bioorg. Med. Chem. Lett.*, **1993**, *3*, 707-710.

25. Liu, Z.C., Shin, D.S., Lee, K.T., Jun, B.H., Kim, Y.K., and Lee, Y.S. Synthesis of Photolabile O-Nitroveratryloxycarbonyl (NVoc) Protected Peptide Nucleic Acid Monomers. *Tetrahedron*, **2005**, *61*, 7967-7973.

26.   Zhang, F.M., Murhammer, D.W., and Linhardt, R.J. Enzyme Kinetics and Glycan
       Structural Characterization of Secreted Alkaline Phosphatase Prepared Using the
       Baculovirus Expression Vector System. *Appl. Biochem. Biotechnol.*, **2002**, *101*, 197-210.

27.   Davis, T.R., Shuler, M.L., Granados, R.R., and Wood, H.A. Comparison of
       Oligosaccharide Processing among Various Insect-Cell Lines Expressing a Secreted
       Glycoprotein. *In Vitro Cell Dev.*, **1993**, *29A*, 842-846.

28.   Lipscomb, M.L., Palomares, L.A., Hernandez, V., Ramirez, O.T., and Kompala, D.S.
       Effect of Production Method and Gene Amplification on the Glycosylation Pattern of a
       Secreted Reporter Protein in CHO Cells. *Biotechnol. Progr.*, **2005**, *21*, 40-49.

29.   Mangalampalli, V.R.M., Mowry, M.C., Lipscomb, M.L., James, R.I., Johnson, A.K., and
       Kompala, D.S. Increased Production of a Secreted Glycoprotein in Engineered CHO
       Cells through Amplification of a Transcription Factor. *Cytotechnology*, **2002**, *38*, 23-35.

30.   James, R.I., Elton, J.P., Todd, P., and Kompala, D.S. Engineering CHO Cells to
       Overexpress a Secreted Reporter Protein Upon Induction from Mouse Mammary Tumor
       Virus Promoter. *Biotechnol. Bioeng.*, **2000**, *67*, 134-140.

31.   Yang, T.T., Sinai, P., Kitts, P.A., and Kain, S.R. Quantification of Gene Expression with
       a Secreted Alkaline Phosphatase Reporter System. *BioTechniques*, **1997**, *23*, 1110-1114.

32.   Advanced Chemistry Development Inc, Toront, Ontario, Canada. *http:www.acdlabs.com*,
       2001.

Appendix

NMR data

Benzyl 2-acetamido-3,6-di-*O*-benzoyl-2-deoxy-β-D-galactopyranoside (6)
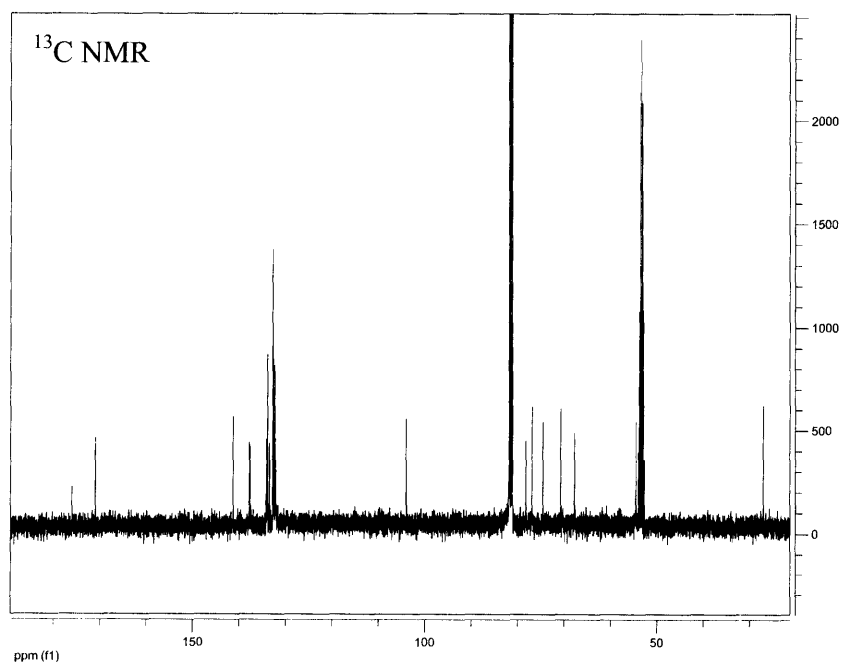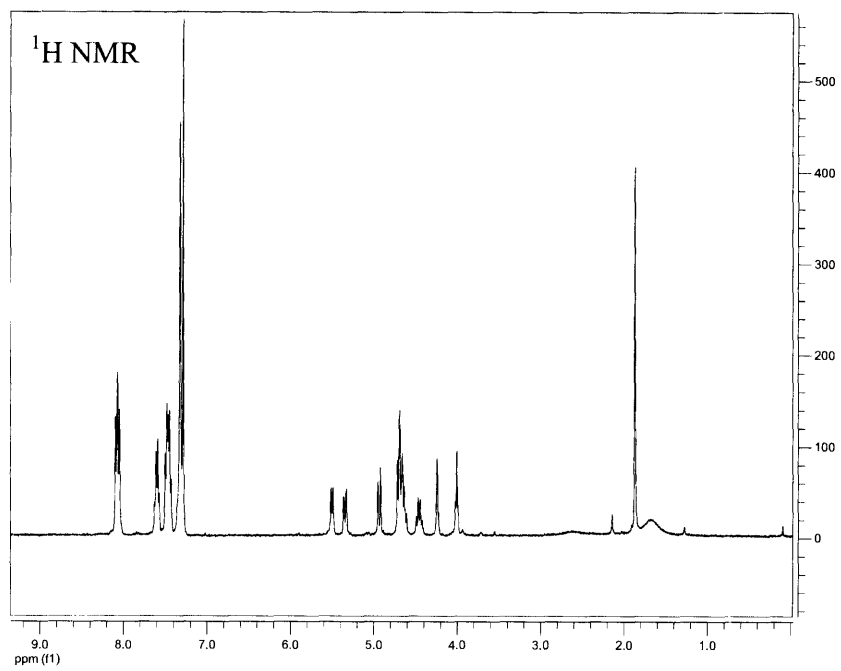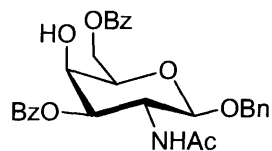
Benzyl 2-acetamido-4-azido-3,6-di-*O*-benzoyl-2,4-deoxy-β-D-glucopyranoside (7)

¹H NMR

¹³C NMR
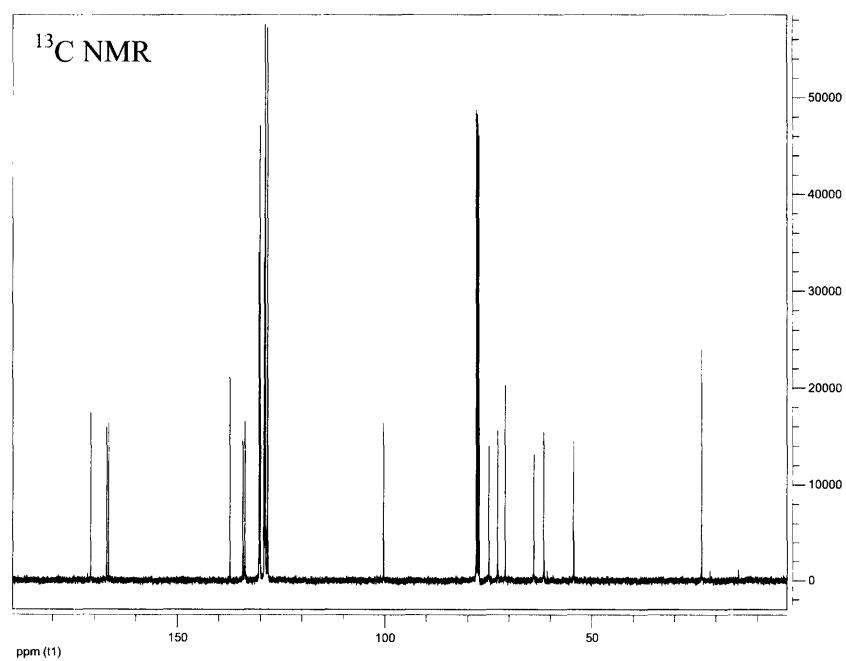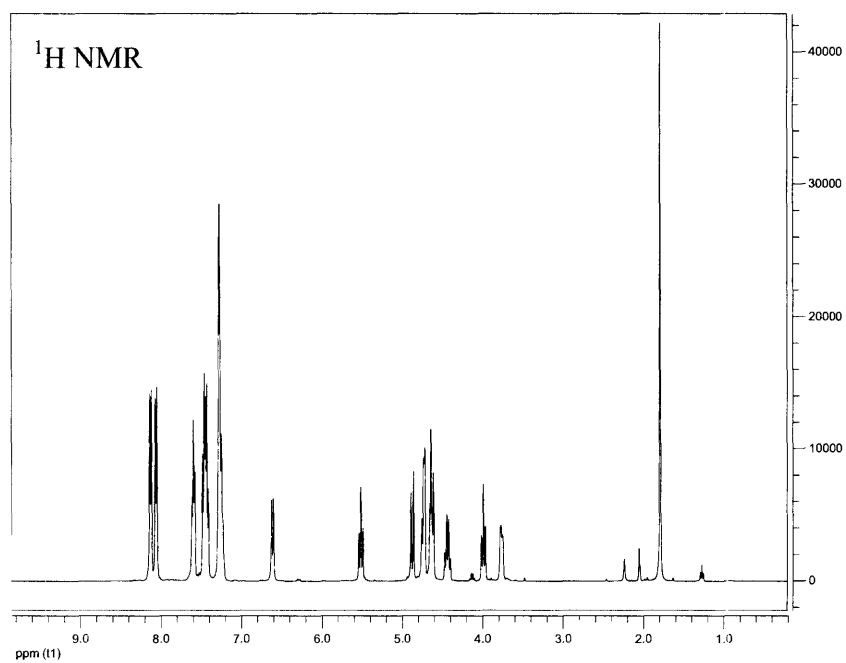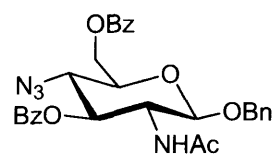
Benzyl 2-acetamido-4-azido-2,4-deoxy-β-D-glucopyranoside (8)
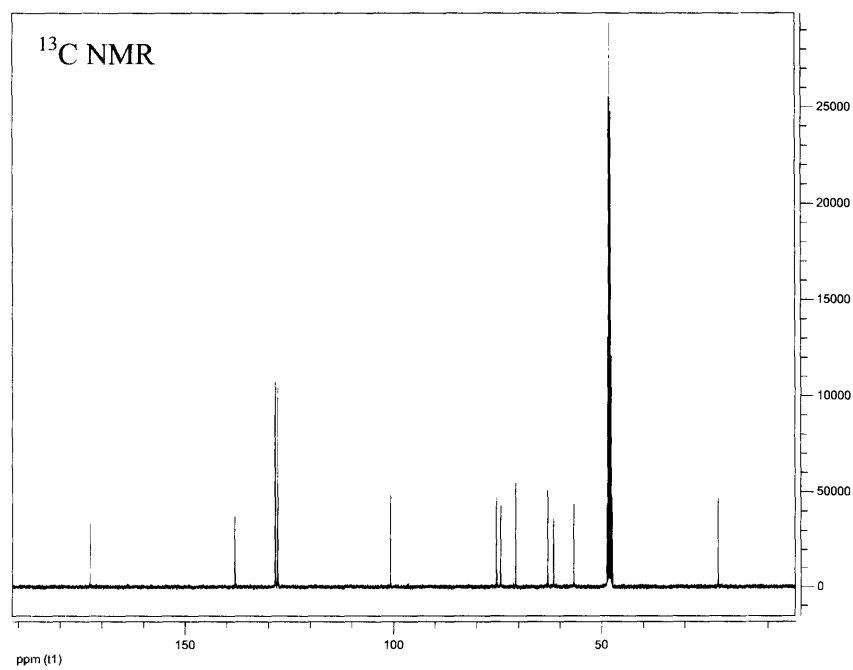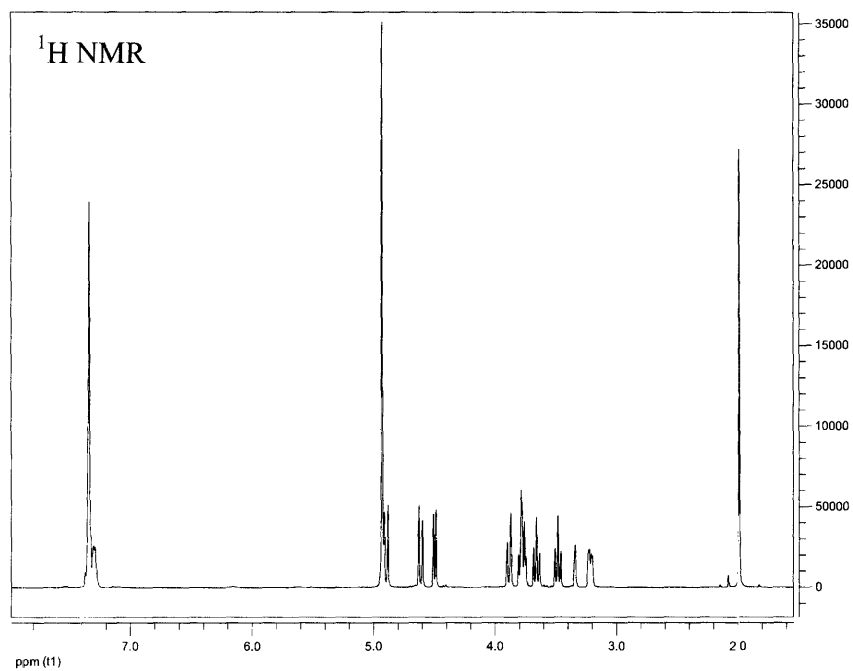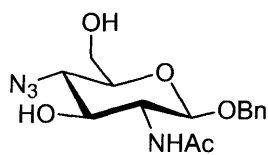
Benzyl 2-acetamido-4-azido-6-tosyloxy-2,4-deoxy-β-D-glucopyranoside (9)

217
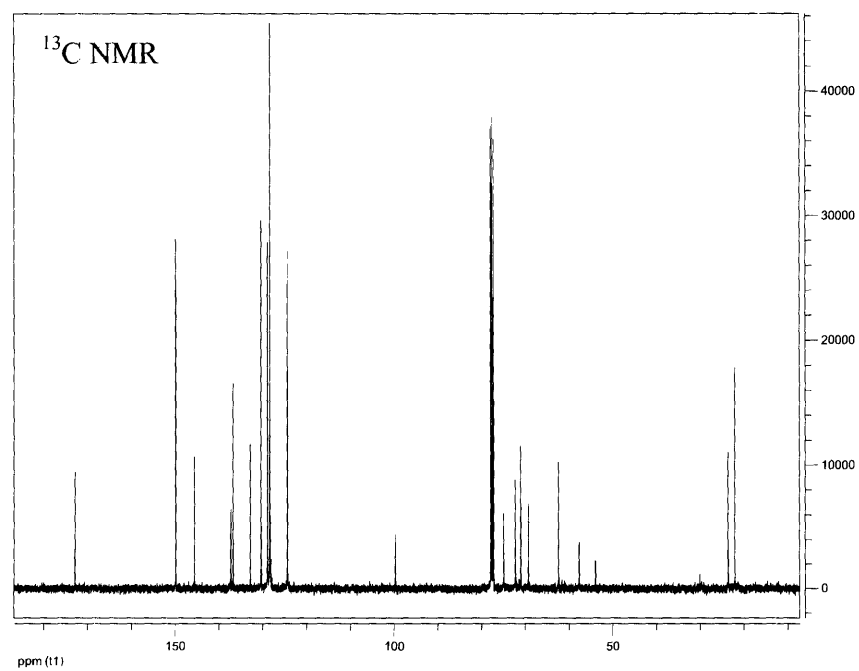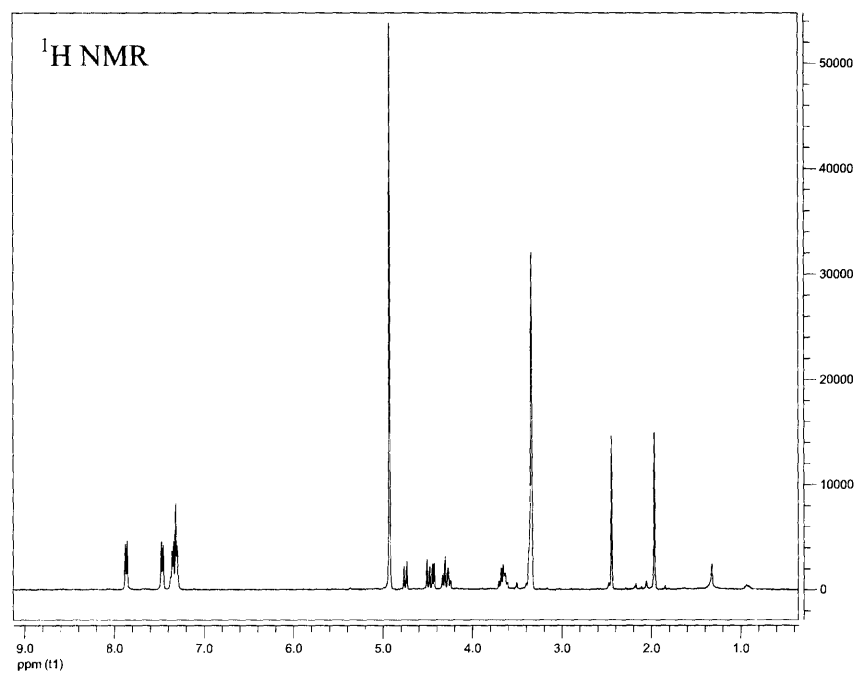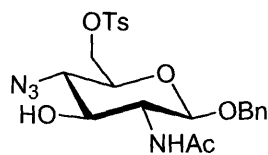
# Benyzl 2-acetamido-4-azido-6-iodo-2,4,6-deoxy-β-D-glucopyranoside (10)
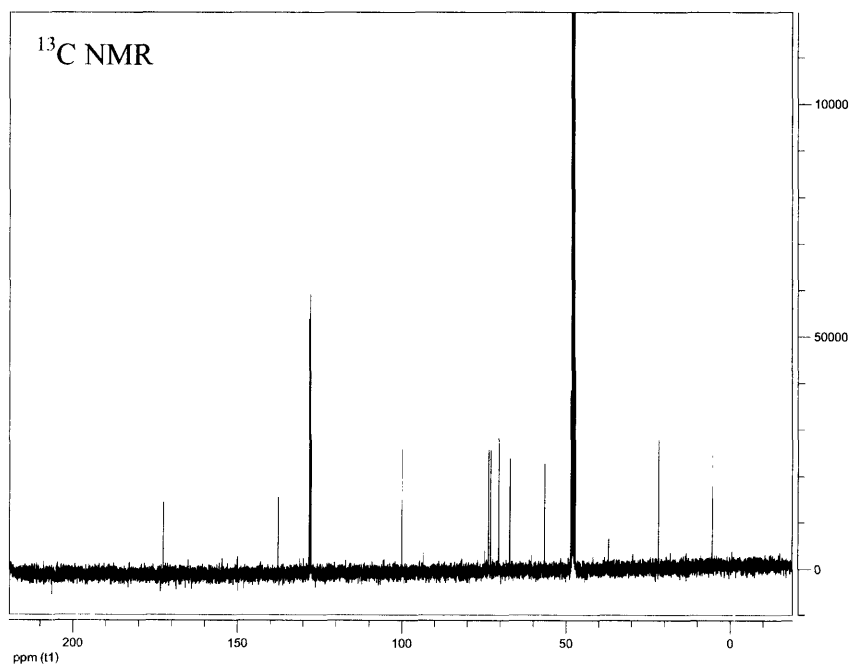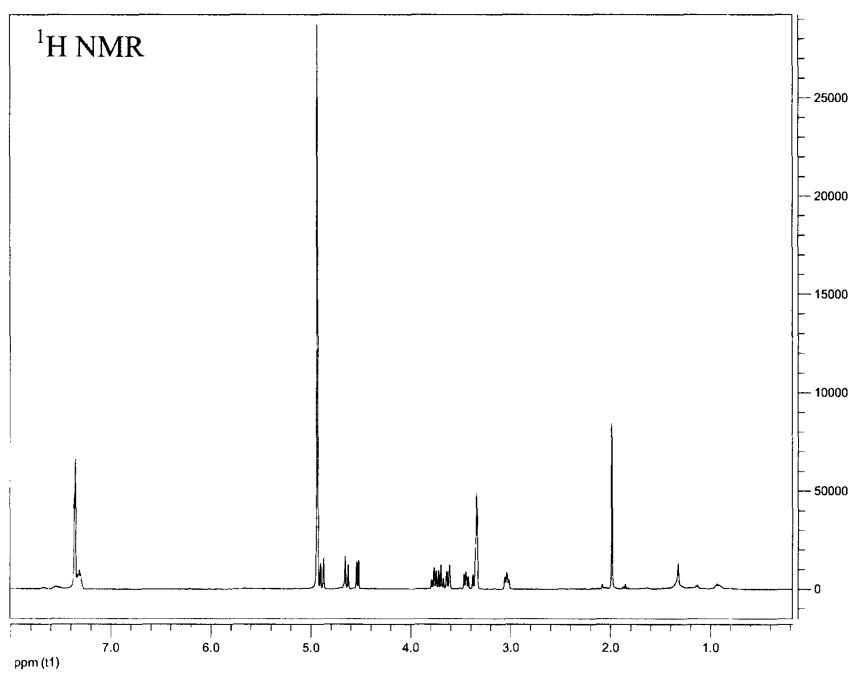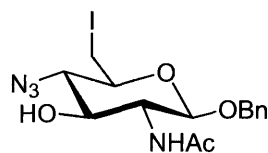
Benzyl 2,4-acetamido-3-*O*-acetyl-2,4,6-deoxy-β-D-glucopyranoside (11)
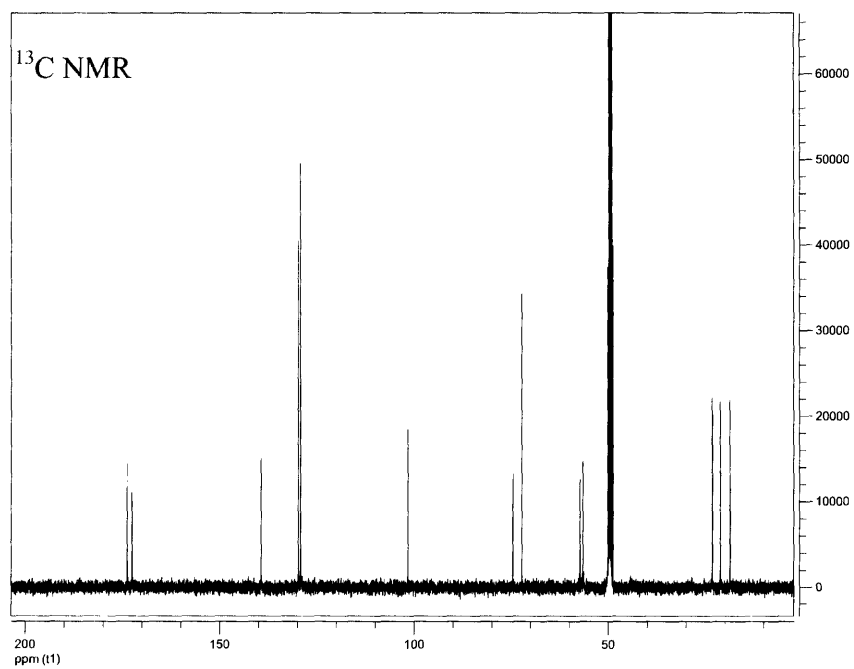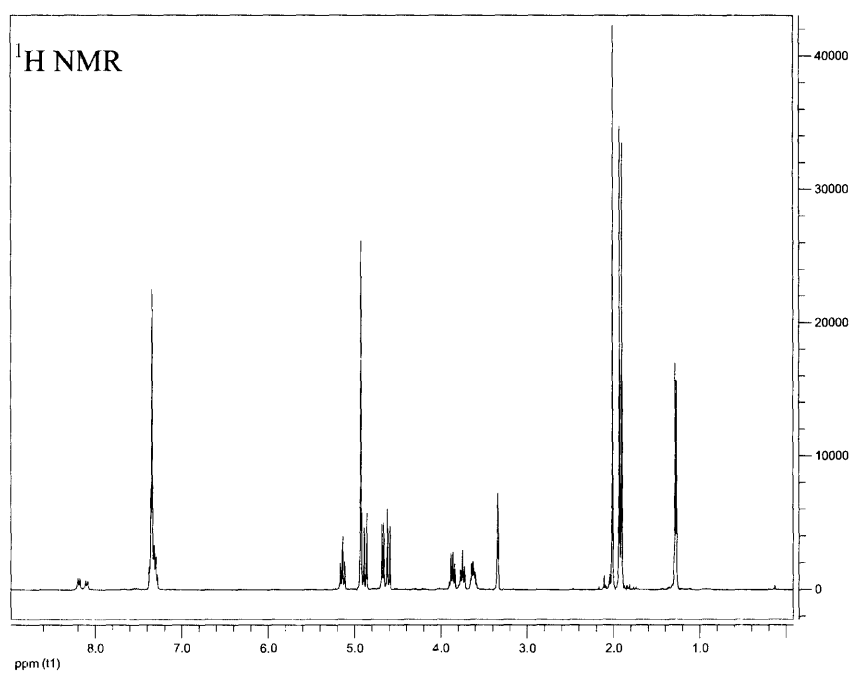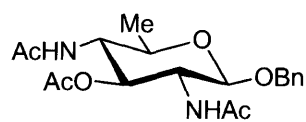
Benzyl 2,4-acetamido-2,4,6-deoxy-β-D-glucopyranoside (12)
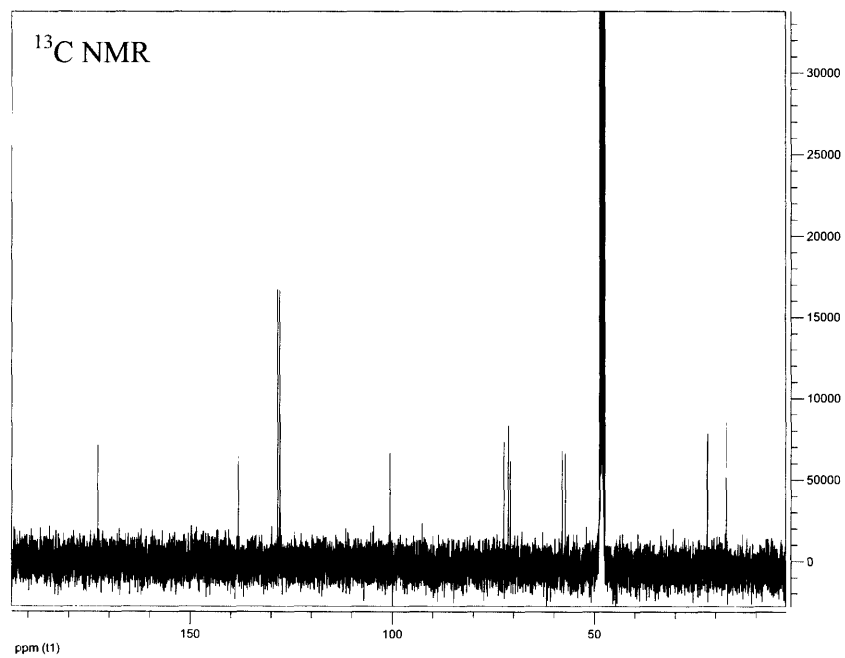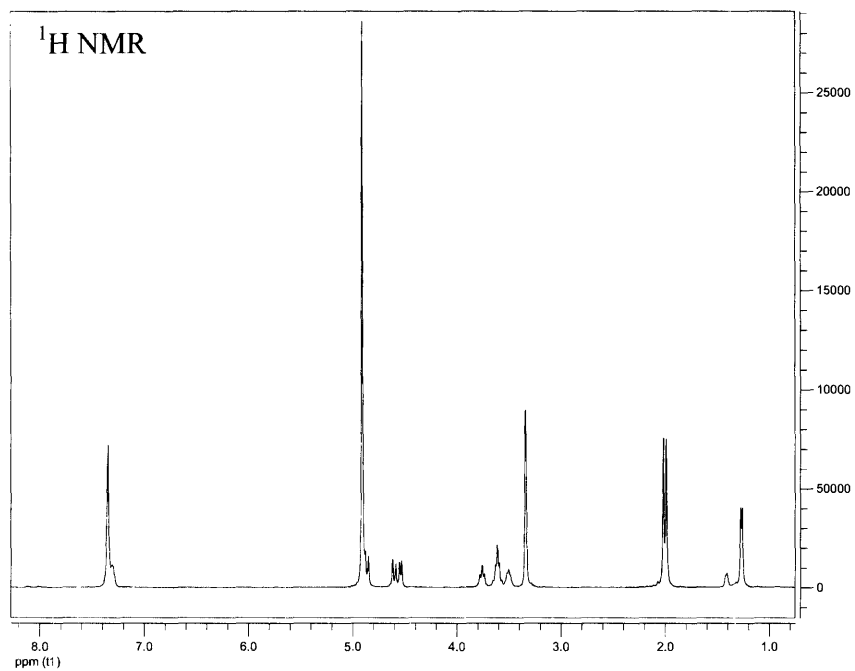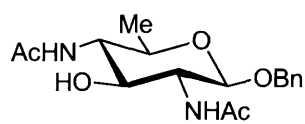
Benzyl 2,4-acetamido-3-*O*-benzoyl-2,4,6-deoxy-β-D-glucopyranoside (13)
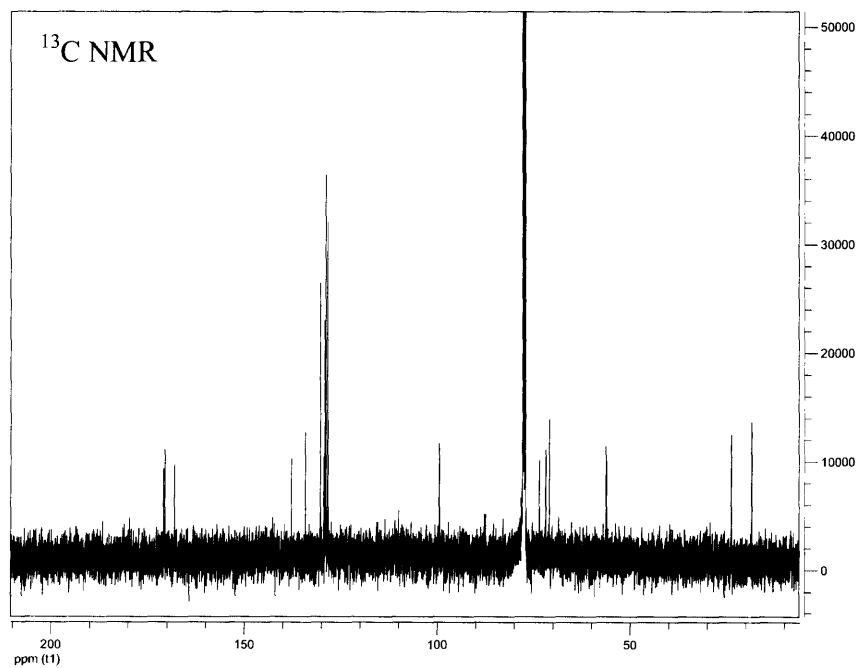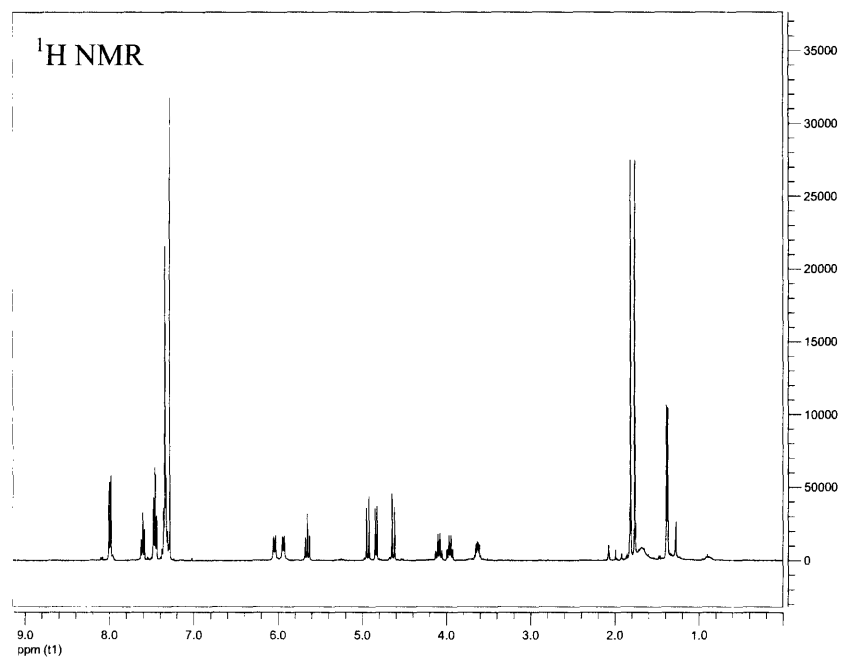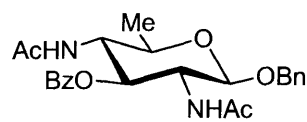
2-4-acetamido-3-*O*-benzoyl-2,4,6-deoxy-D-glucopyranoside (14)
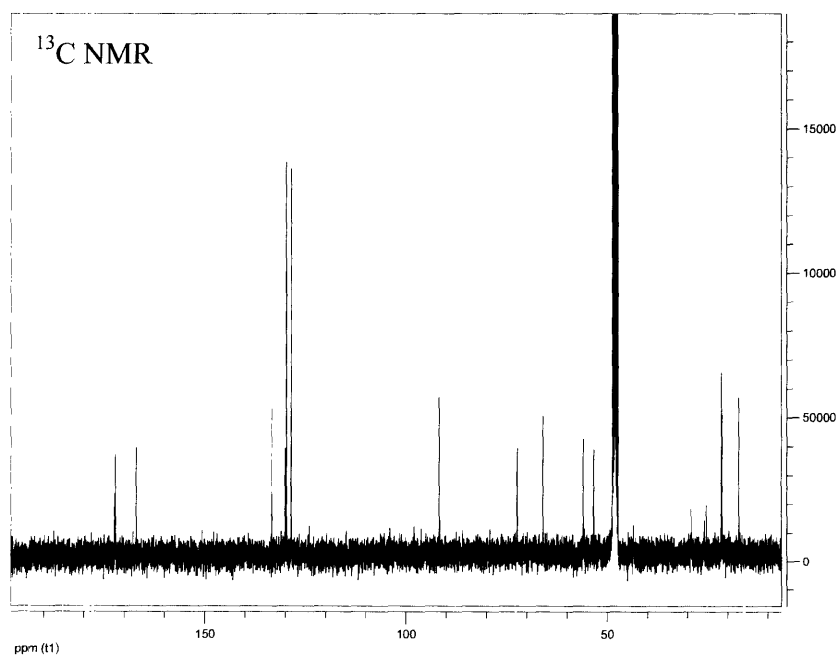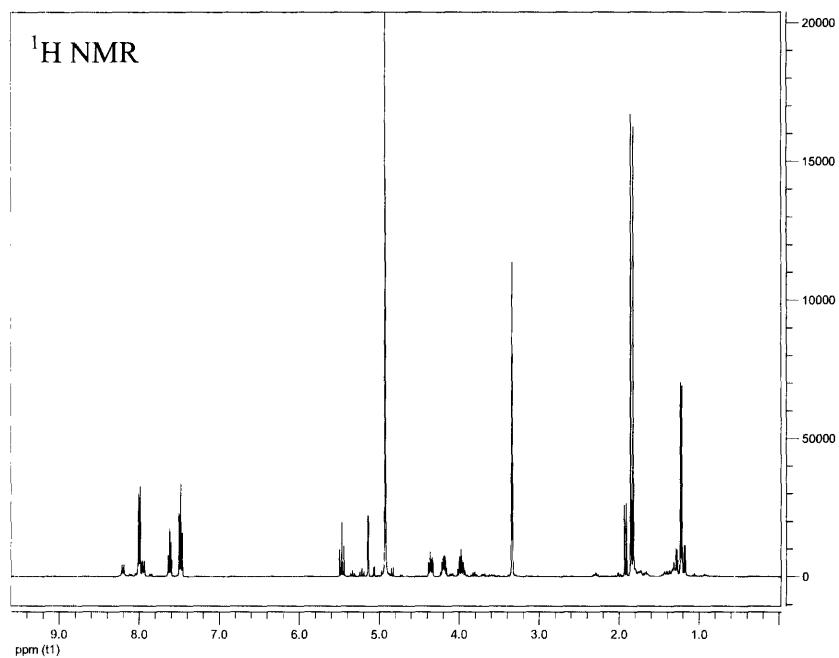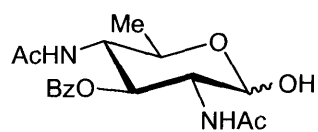
Dibenzylphospho-2,4-acetamido-3-*O*-benzoyl-2,4,6-deoxy-α-D-gluocopyranoside (15)
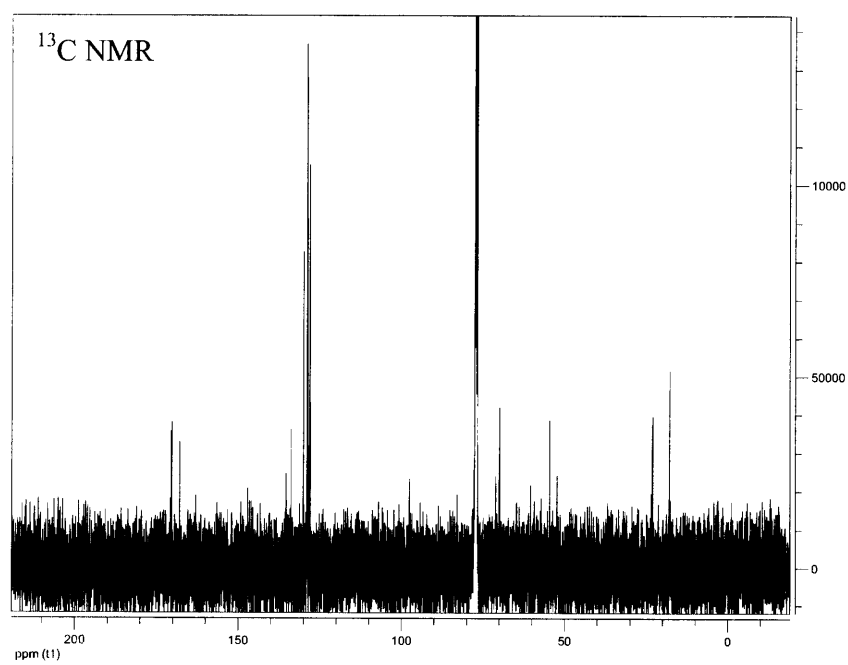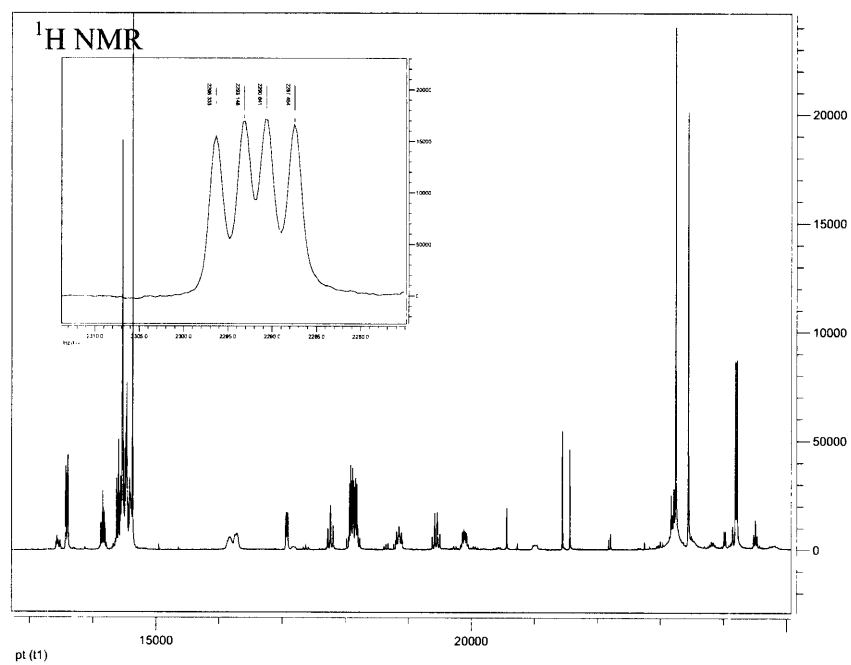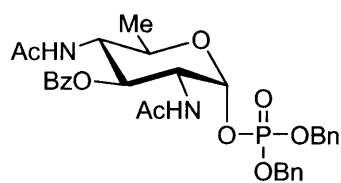
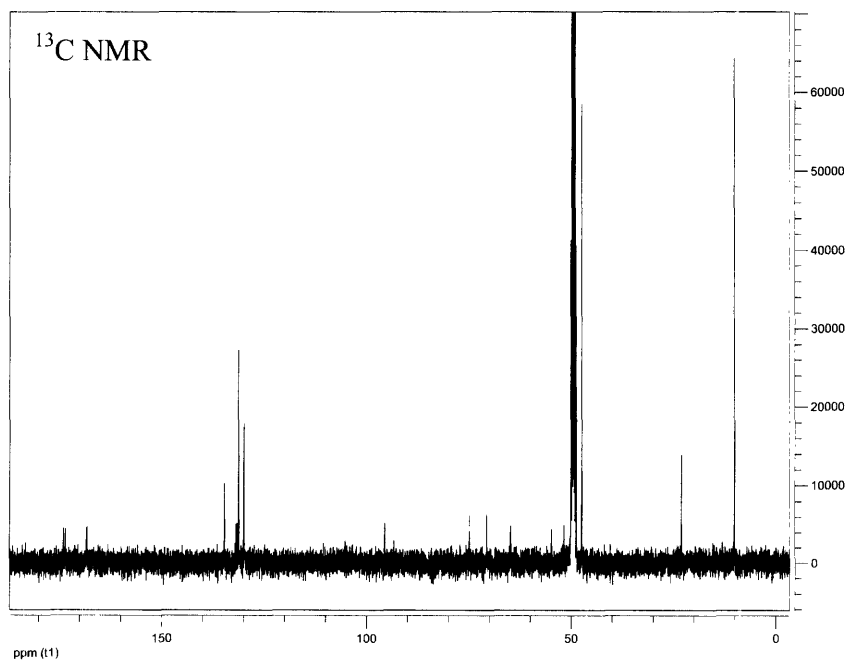Dibenzylphospho-2,4-acetamido-3-*O*-benzoyl-2,4,6-deoxy-α-D-gluocopyranoside (15) (continued)
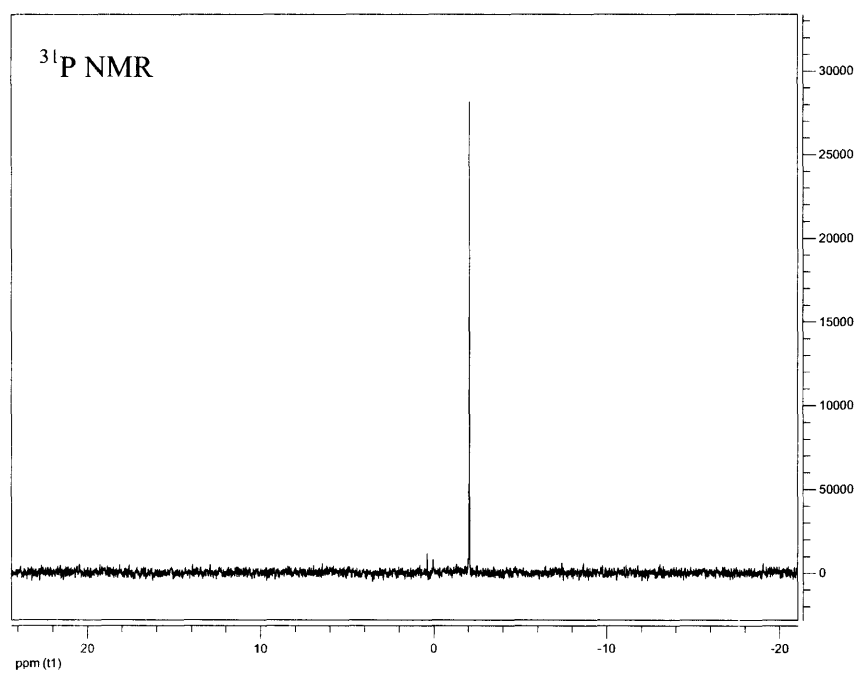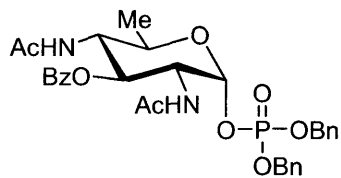
Dibenzylphospho-2,4-acetamido-3-*O*-benzoyl-2,4,6-deoxy-α-D-glucopyranoside (2)
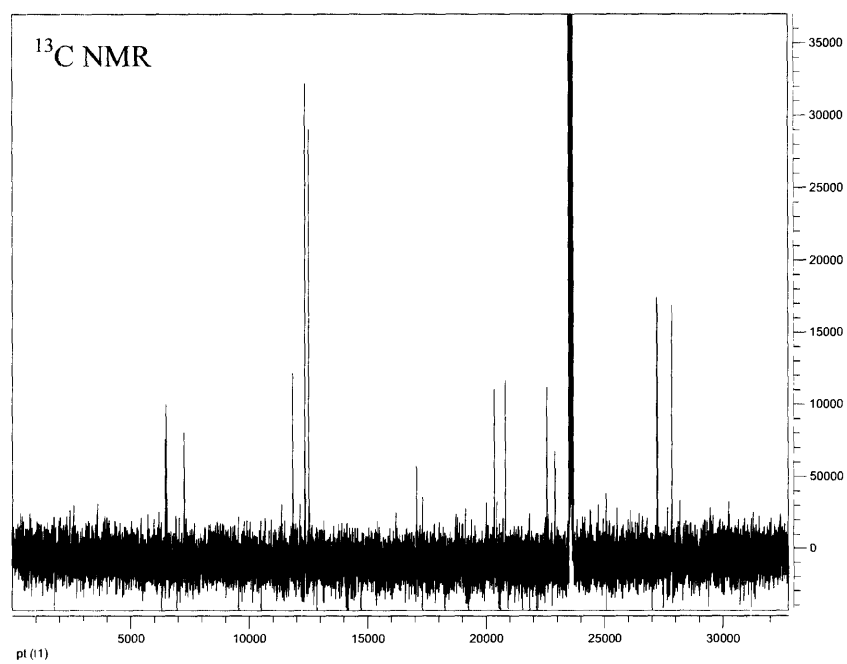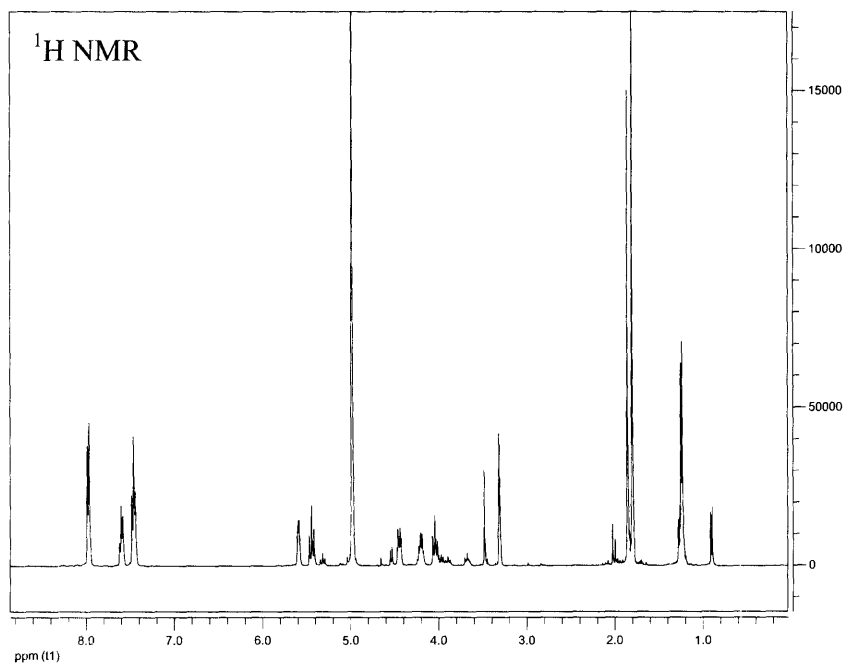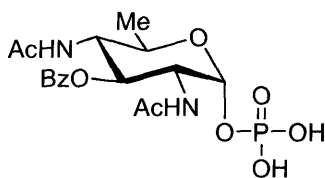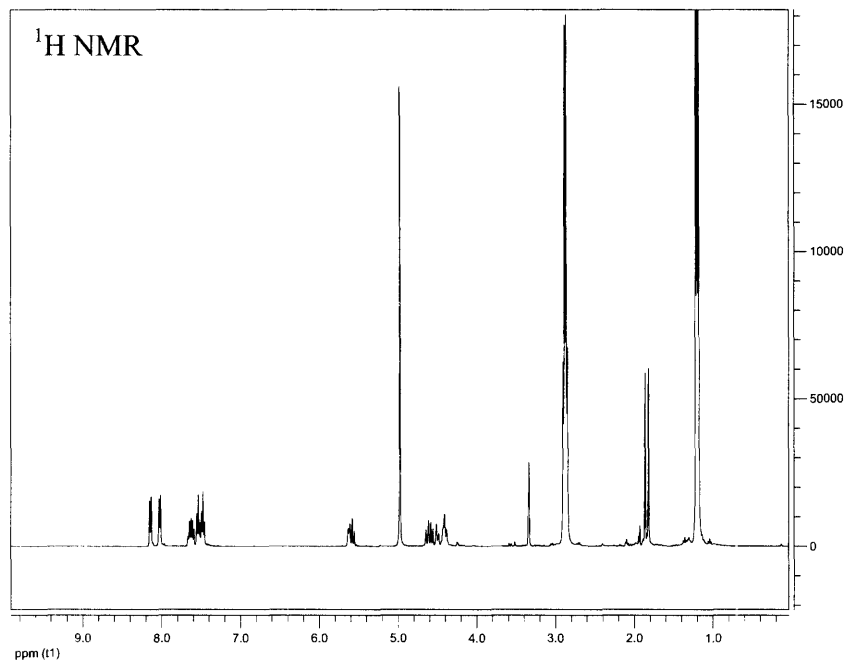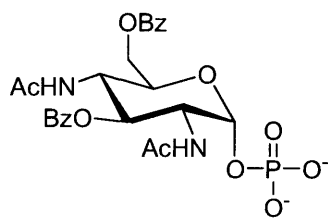
Dibenzylphospho-2,4-acetamido-3-*O*-benzoyl-2,4,6-deoxy-α-D-glucopyranoside (2) (continued)
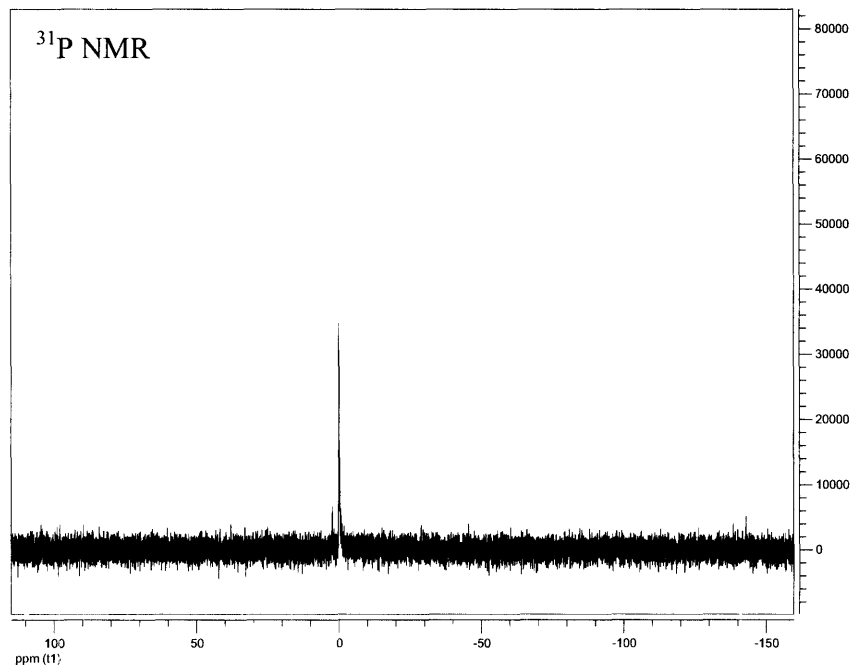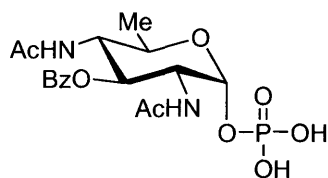
Uridin-5'-yl (2,4-diacetamido-2,4,6-trideoxy-α-D-glycopyranosyl) diphosphate (3)
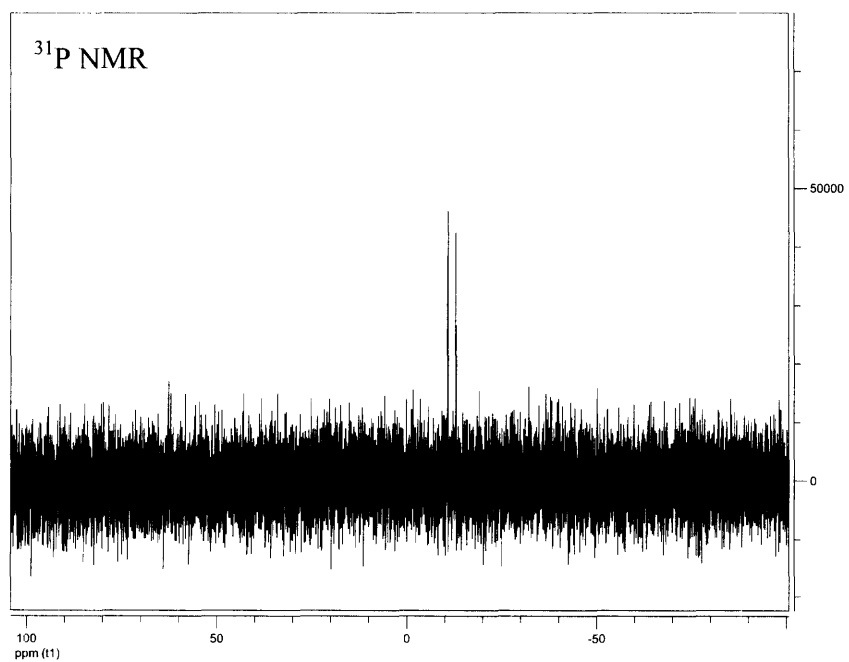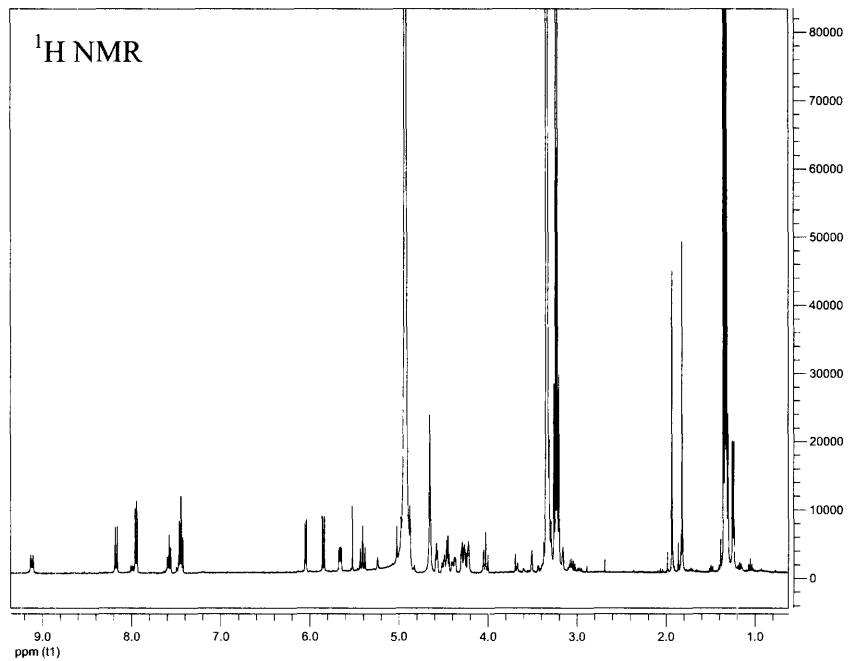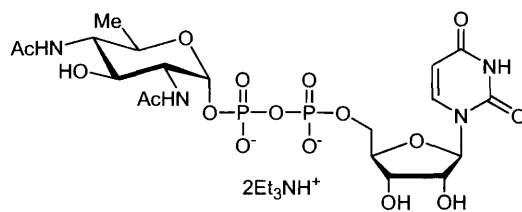
Undecaprenyl-pyrophosphate bacillosamine (4)

^{1}H NMR
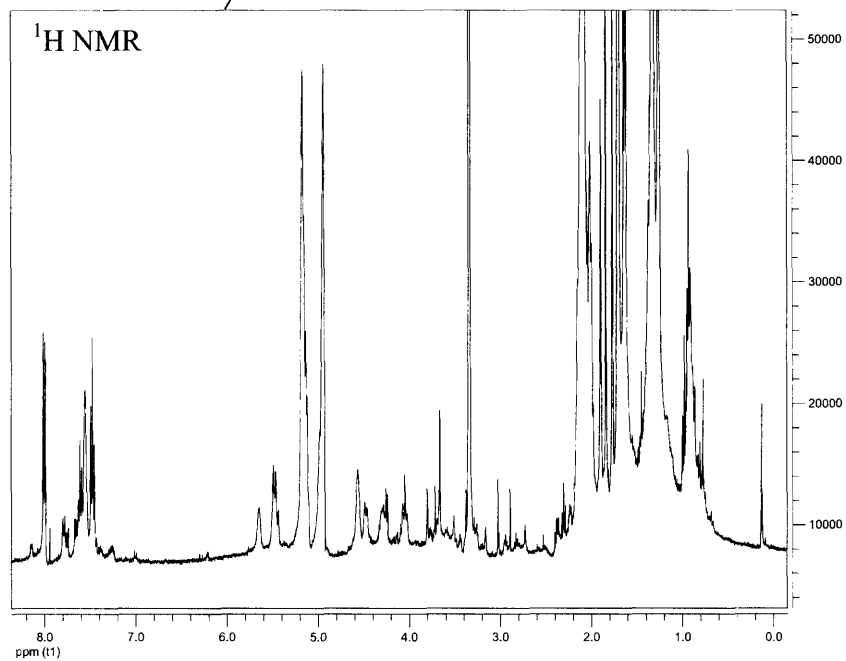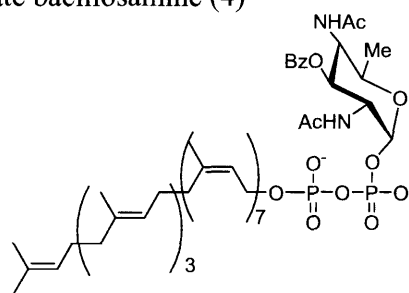
(Water suppression performed on 1H spectra by irradiating at 4.9 ppm)
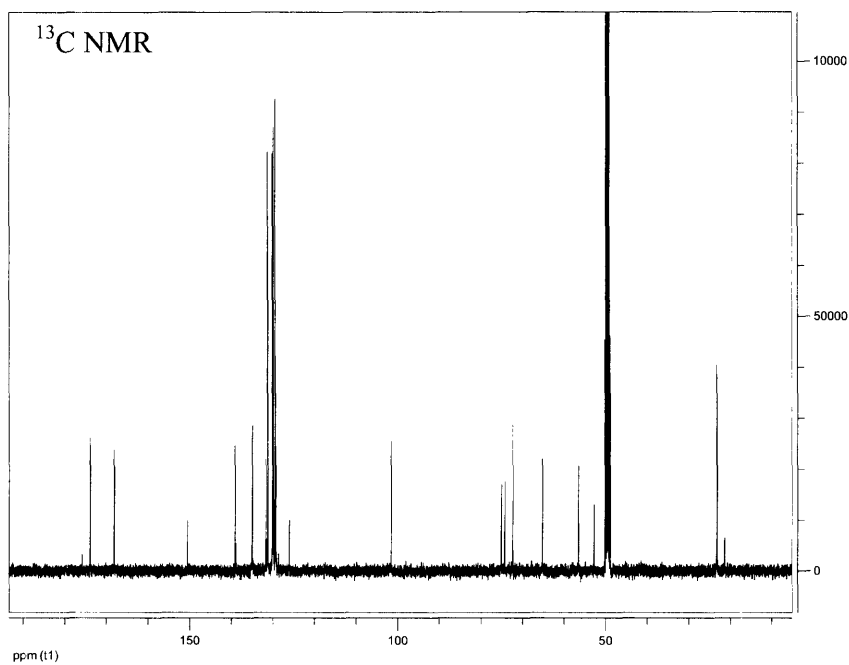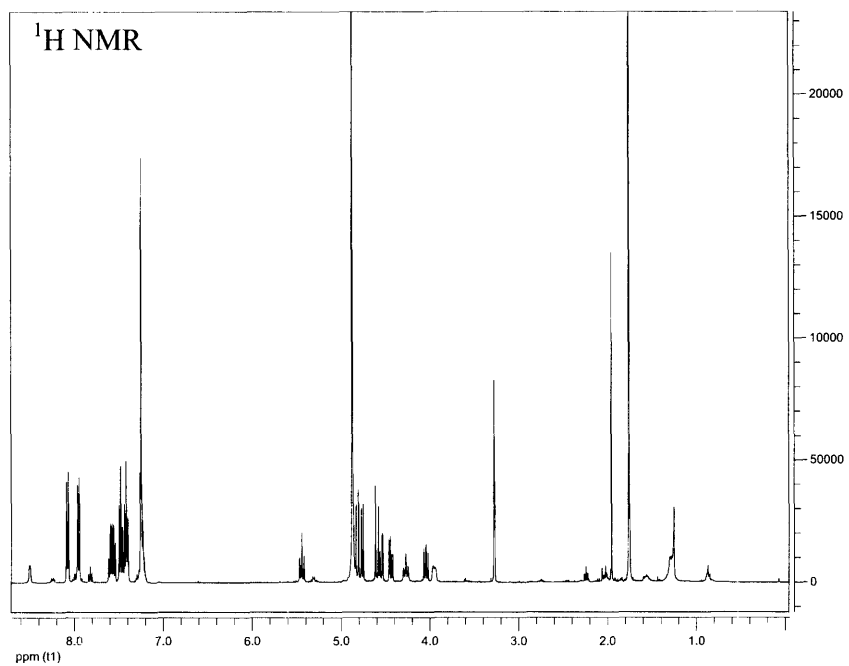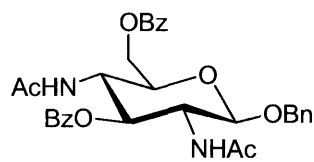
^{31}P NMR

Benzyl 2,4-acetamido-3,6-*O*-benzoyl-2,4-deoxy-β-D-gluocpyranoside (20)

2,4-acetamido-3,6-*O*-acetyl-2,4-deoxy-β-D-glucopyranoside (21)
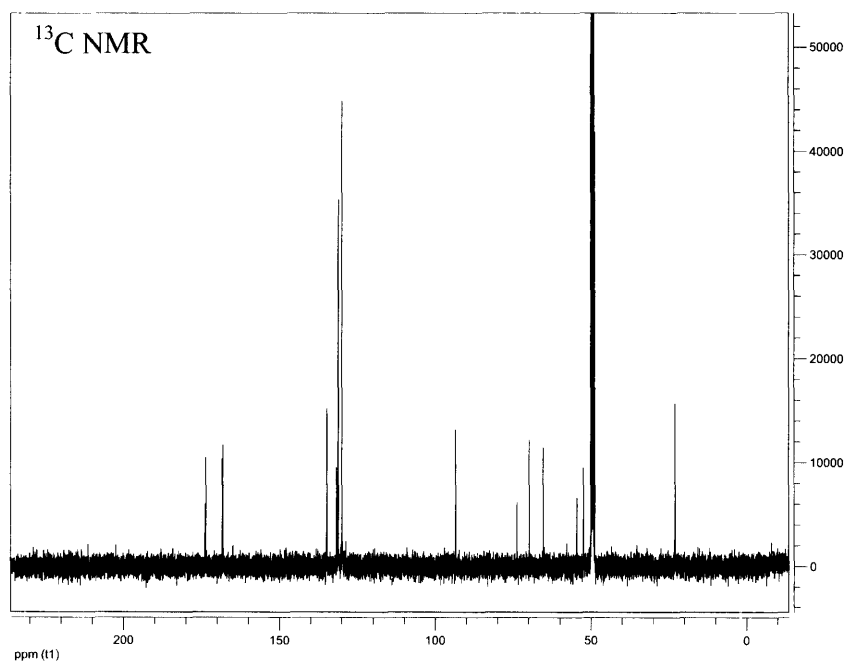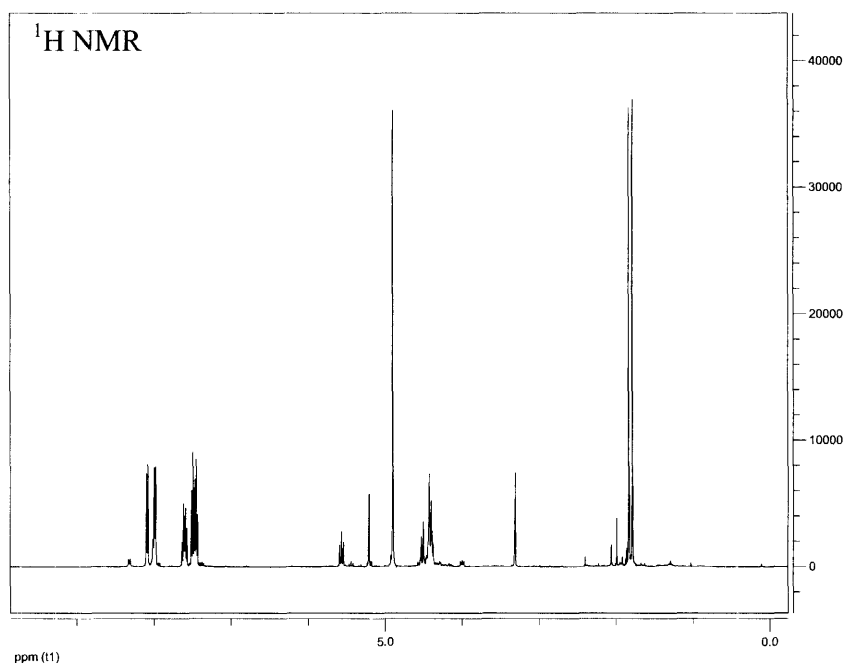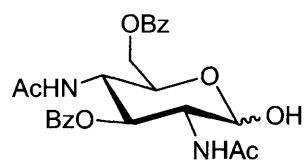
¹H NMR

¹³C NMR

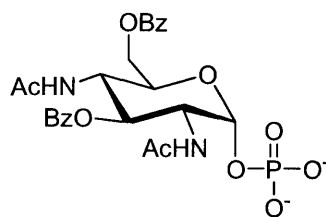Dibenzylphospho-2,4-acetamido-3,6-*O*-benzoyl-2,4-deoxy-α-D-glucopyranoside (22)

¹H NMR

¹³C NMR
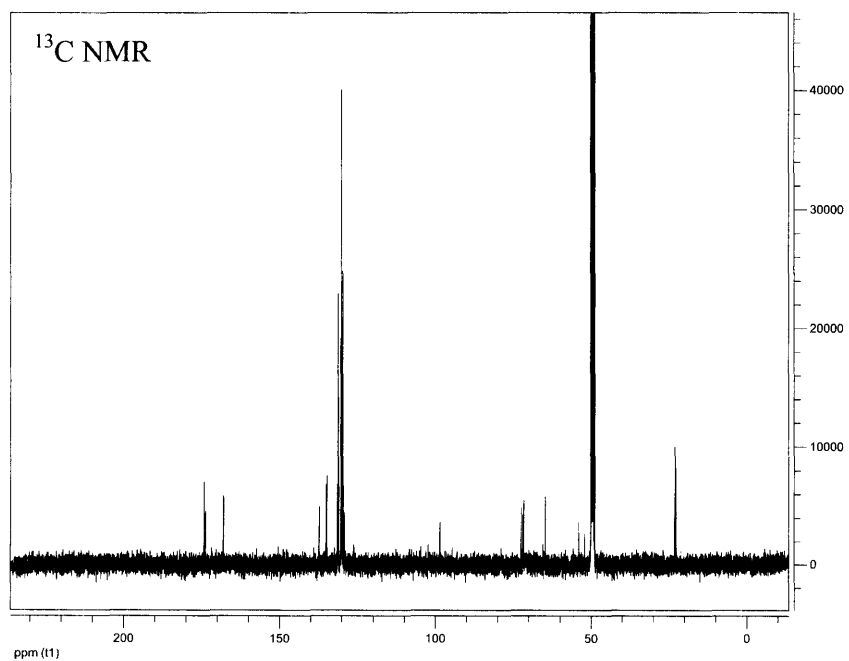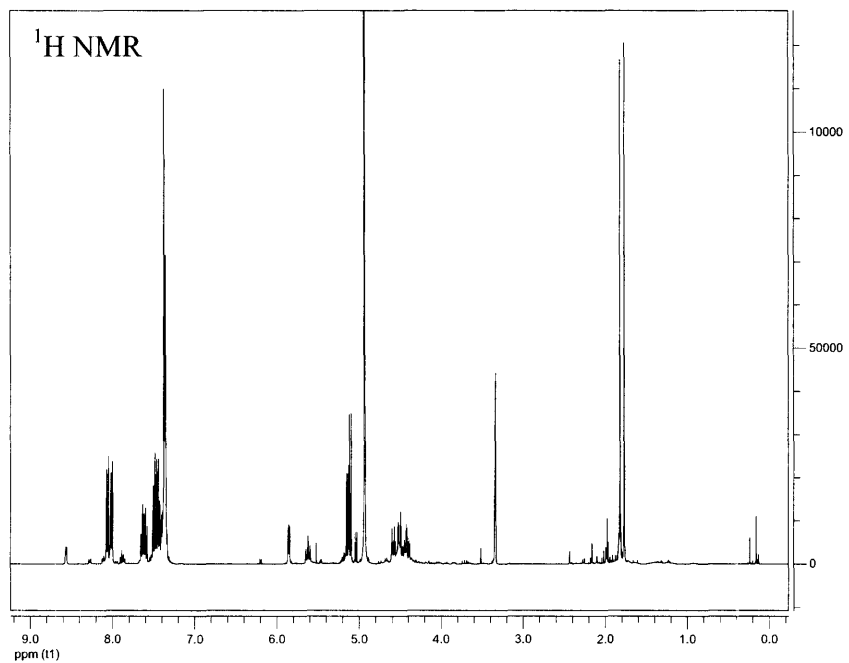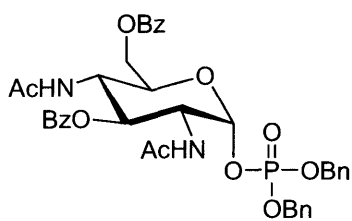
Dibenzylphospho-2,4-acetamido-3,6-*O*-benzoyl-2,4-deoxy-α-D-glucopyranoside (22)
(continued)
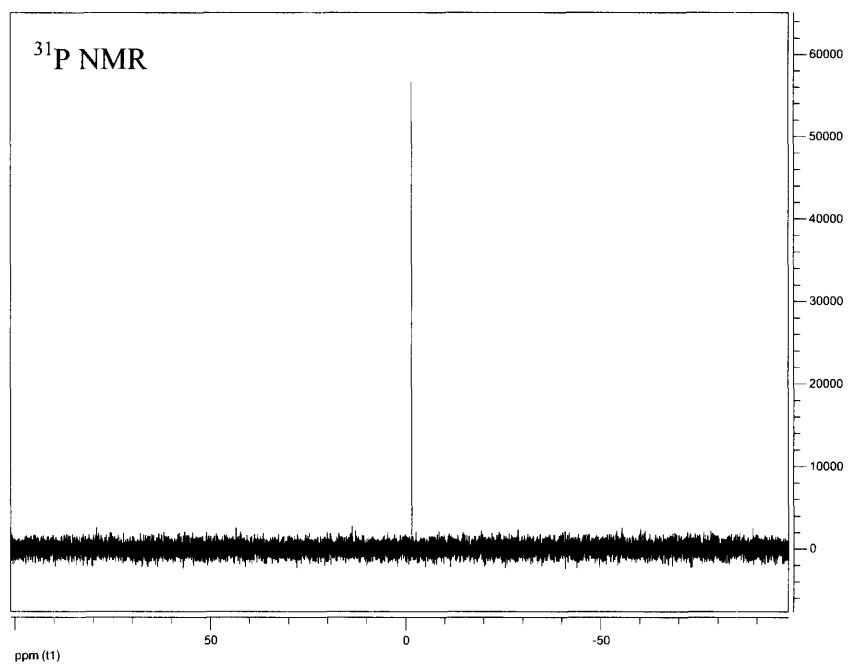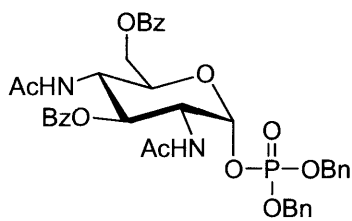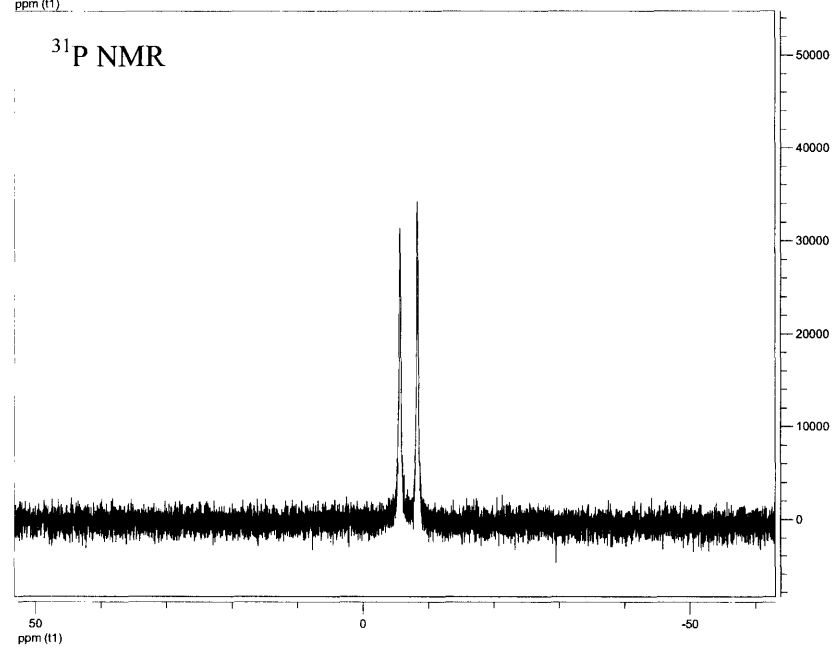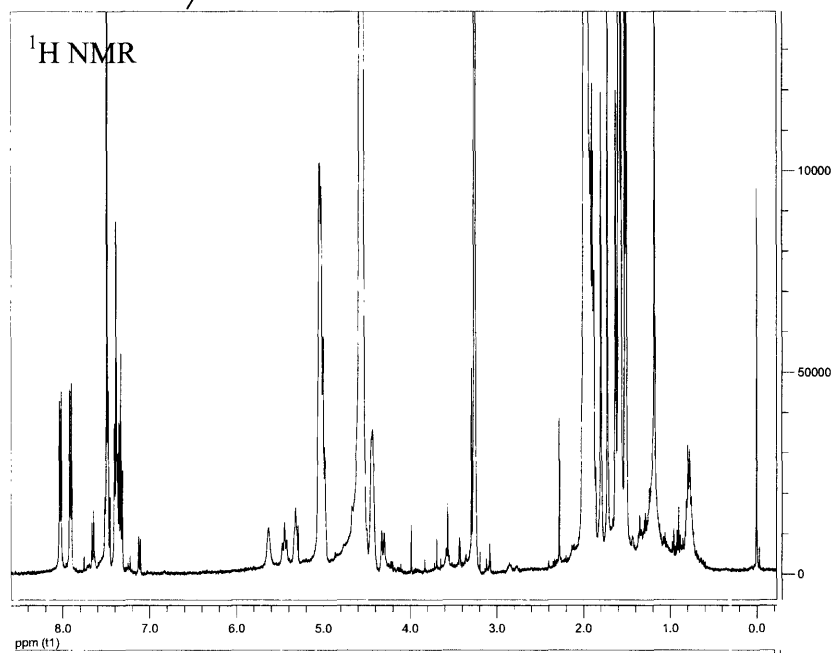
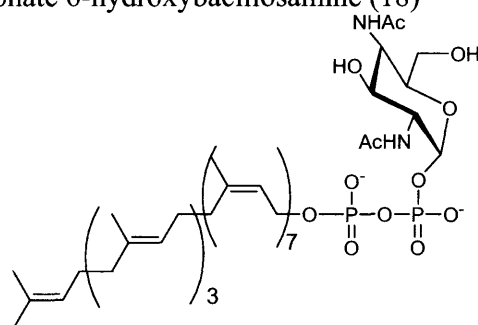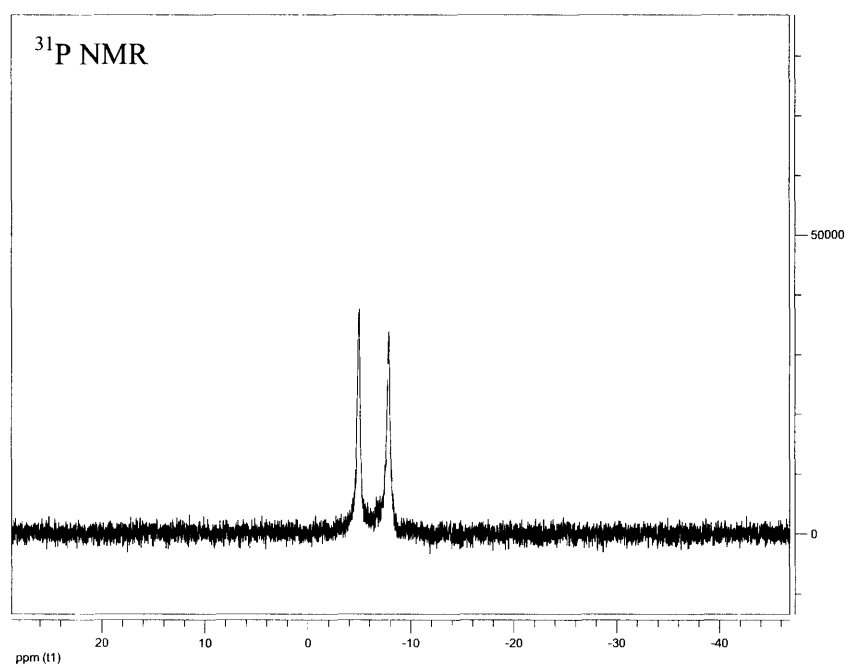Phospho-2,4-acetamido-3,6-*O*-benzoyl-2,4-deoxy-α-D-glucopyranoside (23)
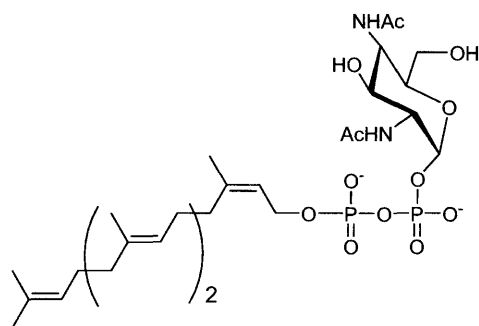
Phospho-2,4-acetamido-3,6-*O*-benzoyl-2,4-deoxy-α-D-glucopyranoside (23)
(continued)
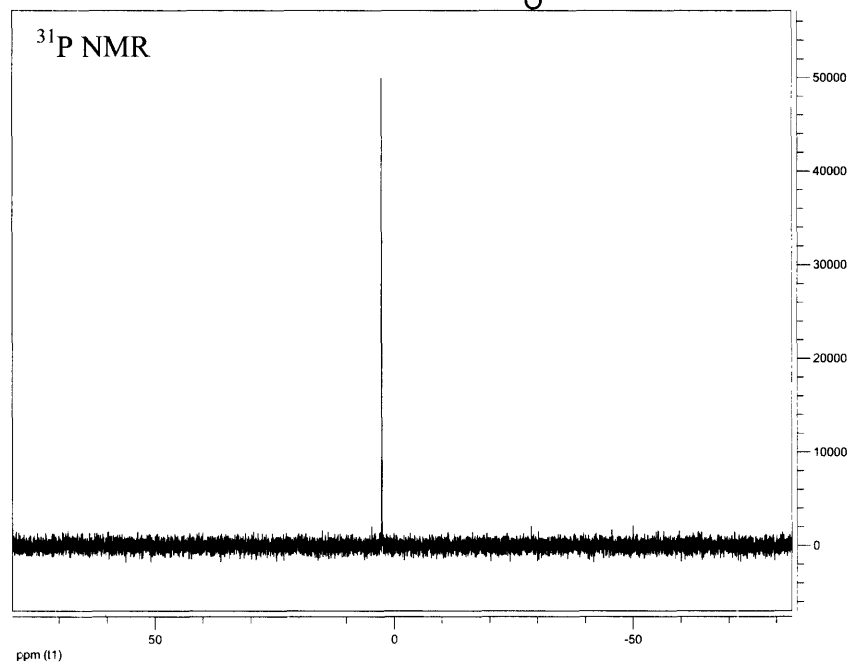
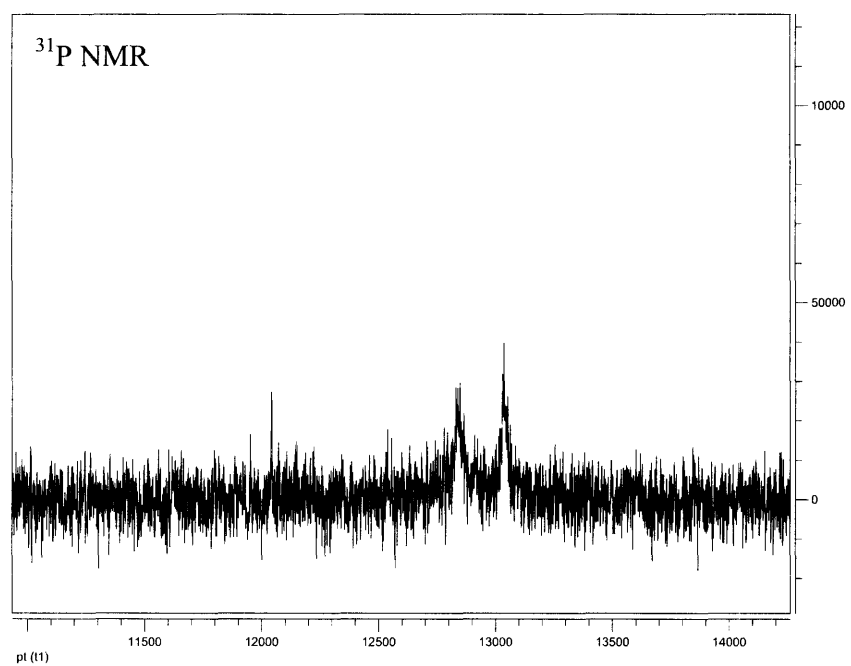# Undecaprenyl-pyrophosphate 6-hydroxybacillosamine (18)

Undecaprenyl-pyrophosphate N-acetylglucosamine (19)

Geranylgeranyl-PP-Bac-6-OH (24)

Geranylgeranyl-PP-Bac-6-OH (24) (continued)

^{31}P NMR

Dolichol-pyrophosphate Bac-6-OH (25)

31P NMR
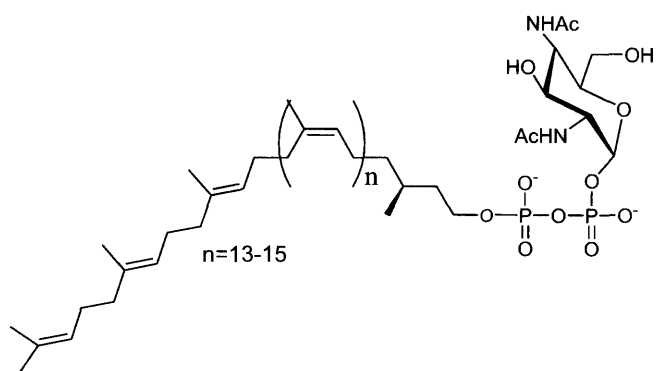
ERANTHIE WEERAPANA
Department of Chemistry
Massachusetts Institute of Technology
Building 18-563
Cambridge, MA 02139
Tel: 617 596 3197
eranthie@mit.edu
Citizenship: Sri Lanka

Education:

2005   Ph.D. Organic Chemistry
            Massachusetts Institute of Technology, Cambridge, MA

2000   B.S. Chemistry, *Cum Laude* with Honors in Chemistry
            Yale University, New Haven, CT

Research Experience:

2001-2005   **Massachusetts Institute of Technology**, Cambridge, MA
                   Graduate Research Assistant
                   Advisor: Professor Barbara Imperiali
                   Investigating asparagine-linked protein glycosylation in prokaryotic and
                   eukaryotic systems

1999-2000   **Yale University**, New Haven, CT
                   Undergraduate Research Assistant
                   Advisor: Professor David J. Austin
                   Methodology development for the synthesis of substituted pyrroles as
                   potential protein kinase inhibitors

Summer 1998   **Pharmacia and Upjohn**, Kalamazoo, MI
                   Summer Intern
                   Advisor: Linda Maggiora, Medicinal Chemistry Department
                   Design and synthesis of small molecule inhibitors of Hepatitis-C viral
                   protease

Summer 1997   **Yale University**, New Haven, CT
                   Summer Research Fellow
                   Advisor: Professor Andrew D. Hamilton
                   Synthesis of phosphotyrosine-based peptidomimetic inhibitors of STAT
                   dimerization

Publications

Glover, K. J.*; Weerapana, E.*, Chen, M. M., Imperiali, B. UDP-Bacillosamine is the substrate for the PglC glycophosphoryltransferase involved in C. jejuni N-linked glycosylation. *Submitted.*
*These authors contributed equally to this work.

Glover, K. J.*; Weerapana, E.*, Numao, S., Imperiali, B. Chemo-enzymatic synthesis of glycopeptides using PglB, a bacterial oligosaccharyl-transferase from *Campylobacter jejuni.*
Chem. Biol. **2005**
*These authors contributed equally to this work.

Glover, K.J., Weerapana, E., and Imperiali, B. In vitro assembly of the undecaprenylpyrophosphate-linked, heptasaccharide for prokaryotic N-linked glycosylation. *Proc. Nat.l Acad. Sci. USA*, **2005**, *102*, 14255-14259.

Weerapana, E., Glover, K.J., Chen, M.M., and Imperiali, B. Investigating bacterial N-linked glycosylation: Synthesis and glycosyl acceptor activity of the undecaprenyl pyrophosphate-linked bacillosamine. *J. Am. Chem. Soc.*, **2005**, *127*, 13766-13767.

Weerapana, E., and Imperiali, B. Peptides to peptidomimetics: towards the design and synthesis of bioavailable inhibitors of oligosaccharyl transferase. *Org. Biomol. Chem.*, **2003**, *1*, 93-99.

Presentations

"Investigating asparagine-linked protein glycosylation in Campylobacter jejuni" Weerapana, E., Glover, K. J., Imperiali, B., Gordon Research Conference on Carbohydrates, Tilton, NH, July 2005.

"Inhibiting *N*-linked glycosylation" Weerapana, E., Imperiali, B., 228th National Meeting of the American Chemical Society, Philadelphia, PA, August 2004 (Poster Session).

Teaching Experience, Chemistry Department, MIT.

2000   Teaching Assistant, Laboratory Chemistry

2001   Teaching Assistant, Introductory Chemical Experimentation

2002   Teaching Assistant, Graduate Level Bioorganic Chemistry