HHS Public Access

Author manuscript

Nat Chem Biol. Author manuscript; available in PMC 2013 December 01.

Published in final edited form as:

Nat Chem Biol. 2013 June ; 9(6): 367–373. doi:10.1038/nchembio.1249.

Biochemical evidence for an alternate pathway in N-linked glycoprotein biosynthesis

Angelyn Larkin[1,2,3], **Michelle M. Chang**[1,3], **Garrett E. Whitworth**[1], and **Barbara Imperiali**[1,*]

[1]Department of Chemistry and Department of Biology, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA 02139

Abstract

Asparagine-linked glycosylation is a complex protein modification conserved among all three domains of life. Herein we report the *in vitro* analysis of N-linked glycosylation from the methanogenic archaeon *Methanococcus voltae*. Using a suite of synthetic and semisynthetic substrates, we show that AglK initiates N-linked glycosylation in *M. voltae* through the formation of α-linked dolichyl monophosphate N-acetylglucosamine (Dol-P-GlcNAc), which contrasts with the polyprenyl-diphosphate intermediates that feature in both eukaryotes and bacteria. Intriguingly, AglK exhibits high sequence homology to dolichyl-phosphate β-glucosyltransferases, including Alg5 in eukaryotes, suggesting a common evolutionary origin. The combined action of the first two enzymes, AglK and AglC, afforded an α-linked Dol-P-glycan that serves as a competent substrate for the archaeal oligosaccharyl transferase AglB. These studies provide the first biochemical evidence revealing that despite the apparent similarity of the overall pathways, there are actually two general strategies to achieve N-linked glycoproteins across the domains of life.

INTRODUCTION

The diverse and essential functions of N-linked glycosylation make it a process of major significance in human health and disease and great relevance in medicine and biotechnology, as protein therapeutics have become important elements in the modern pharmacopoeia. It is now known that proteins decorated with asparagine-linked glycans are pervasive in all domains of life.[1–3] The relatively recent discovery of N-linked glycosylation in selected bacteria reveals common themes in the molecular logic of the process relative to the well-studied and highly conserved eukaryotic system.[4] In particular, both pathways feature the stepwise assembly of a glycan onto a polyprenyl-diphosphate (Pren-PP) carrier followed by glycan transfer to an asparagine residue within an AsnXaaSer/Thr sequence.

[*]Correspondence should be addressed to: imper@mit.edu.

[2]Present address: Department of Cell Biology, Harvard Medical School, 240 Longwood Avenue, Boston, MA 02115.

[3]These authors contributed equally to this work.

AUTHOR CONTRIBUTIONS B.I., A.L. M.M.C. and G.E.W. designed the experiments. A.L. M.M.C. and G.E.W. performed the experiments. The manuscript was written through contributions of all authors.

COMPETING FINANCIAL INTERESTS The authors declare no competing financial interests.

The earliest membrane-committed step in the process is noteworthy, as each pathway invokes a phosphoglycosyl transferase that catalyzes the reaction between a nucleotide-diphosphate-activated carbohydrate and a polyprenyl-phosphate (Pren-P) to form the corresponding α-linked polyprenyl-diphosphate monosaccharide (Pren-PP-monosaccharide, Fig. 1). These amphiphilic products are then extensively elaborated by various glycosyltransferases to ultimately afford glycosyl donors for the oligosaccharyl transferases (OTases), which generate a β-linked glycosyl bond to asparagine. Certainly there are also differences between the bacterial and eukaryotic pathways,[2,3] but these essential elements, together with the observed conservation of the Stt3 catalytic subunit of the OTases, have suggested a basic framework of events which for many years was presumed to be conserved in archaea as well. Unlike in bacteria, N-linked glycosylation in archaea is believed to be widespread, as a recent bioinformatics study revealed that 54 out of 56 sequenced archaeal genomes contain at least one archaeal OTase, termed AglB.[5] AglB is a homolog of both the bacterial PglB and the Stt3 subunit of the eukaryotic OTase complex, and thus has been presumed to be uniquely responsible for this function in archaea.[6,7]

Although the key enzymes and specific transformations involved in eukaryotic and bacterial N-linked glycosylation are now well understood, much less is known about the analogous pathways in archaea despite the fact that the first archaeal glycoproteins were observed nearly 40 years ago.[8] Since then, it has been suggested that both dolichyl-monophosphate (Dol-P) and dolichyl-diphosphate (Dol-PP) linked glycosyl donor substrates are implicated in archaeal glycan assembly.[9–13] For example, recent studies using *Haloferax volcanii* and *Sulfolobus acidocaldarius* identify Dol-P-glycans as intermediates in the assembly of N-linked glycoproteins; however the details of the particular enzymes, the stereochemical outcome of the transformations, as well as overall architecture of the archaeal pathways in general remain elusive.[14,15] This observation raises important questions concerning the molecular features of the assembly process, particularly the implication of Dol-P as an alternate glycan carrier, as this finding sharply diverges from the characterized pathways that use Dol-PP-linked glycans, which are well-established in bacteria and eukaryotes and have been proposed for selected methanogens including *Methanococcus voltae* and *Methanococcus maripaludis*.[16] In addition, the use of Dol-P in place of Dol-PP as a glycan carrier is intriguing with respect to the function of the OTase, since the metal-ion coordinated diphosphate-linked glycosyl donor would be anticipated to be a considerably more activated substrate than the corresponding monophosphate derivative. As such, it would be very valuable to be able to reliably predict which pathways are either diphosphate- or monophosphate-dependent as a foundation for understanding the divergence of mechanisms and the evolutionary origin of the complex present in the widespread phenomenon of N-linked glycosylation. Furthermore, a deeper understanding of the component steps of archaeal protein glycosylation pathways will also reveal opportunities for exploiting the robust properties of archaeal enzymes as tools for engineering complex glycoconjugates.

In this study we present the *in vitro* biochemical analysis of key enzymes in the N-linked pathway of the marine methanogen *M. voltae*, which represents one of the three most intensively studied archaeal pathways, together with *M. maripaludis*[17] and *H. volcanii*.[7] *M.*

voltae is known to generate N-linked glycoproteins with a unique trisaccharide composed of ManNAc(6Thr)A-β1,4-Glc-2,3-diNAcA-β1,3-GlcNAc (Fig. 2a), in which threonine is linked via an amide to the terminal mannuronic acid.[18] Several *M. voltae* proteins modified with this trisaccharide are defined, most notably the proteins that comprise the archaeal flagellum and the S-layer protein.[18] Genetic analyses have established a number of candidates involved in glycoprotein assembly in *M. voltae*, including the putative glycosyltransferases AglH, AglK, AglC, and AglA, as well as the OTase AglB.[6,19,20] Deletion of several of these genes result in profound defects on flagellar structure and severe impairment of cell motility.[6]

Herein we applied purified enzymes and synthetic and semisynthetic substrates to establish the biochemical framework of the *M. voltae* N-linked glycosylation pathway (Fig. 2b). We demonstrated that the first key step involves formation of α-linked Dol-P-GlcNAc (Fig. 3a), rather than the proposed Dol-PP-GlcNAc.[19,20] Surprisingly, this reaction was catalyzed by AglK, an enzyme previously assigned to the second step in the pathway. Dol-P-GlcNAc from the AglK reaction was further elaborated by the glycosyltransferase AglC, which catalyzed transfer from the highly modified glycosyl donor UDP-Glc-2,3-diNAcA to afford Dol-P-GlcNAc-Glc-2,3-diNAcA. With this disaccharide substrate in hand, we then showed that the OTase AglB efficiently transferred disaccharide to an acceptor peptide, indicating that the Dol-P-charged glycan represents a truncated form of the oligosaccharide donor for archaeal N-linked glycosylation. These results clearly indicate that the N-linked protein glycosylation pathway in selected archaea can be distinguished from the known N-linked glycosylation pathways in both eukaryotes and bacteria, and presents a foundation for further investigation of the important mechanistic consequences resulting from assembly using a monophosphate donor.

RESULTS

AglK is a specific Dol-P-GlcNAc synthase

In order to investigate the biosynthesis of the *M. voltae* N-linked glycan donor, we first focused our attention on AglH. Previous studies report that the *aglH* gene complements a Δ*alg7* mutation in *Saccharomyces cerevisiae*, implying that AglH is responsible for the formation of Dol-PP-GlcNAc.[20] In this case, Dol-PP-GlcNAc would then be the first polyprenyl-linked intermediate in the *M. voltae* pathway by analogy with the known eukaryotic and bacterial systems. Since the polyprenol for glycan assembly in archaea is known to be of the dolichol family, which features an α-saturated isoprene unit as in eukaryotes, we investigated the activity of polyprenyl-phosphate substrates featuring both short and long dolichols, which are reported to be characteristic of archaea and eukaryotes (C55–60 vs. C85–105).[21] Archaeal dolichols, in particular those from selected halophiles, also include an additional element of saturation at the ω-isoprene unit.[13,14] The enzyme which reduces this ω-position isoprene in the short (C55 and C60) dolichyl-phosphates from *H. volcanii* has recently been identified[22] and mass spectrometry analyses reveal that both the α-saturated and α-, ω-saturated short dolichols are competent intermediates for the assembly of the Dol-P-glycans. Therefore, we were confident that the implementation of α-saturated polyprenols was suitable for the current studies, although in the future it will be

important to confirm whether methanogens such as *M. voltae* also feature the additional polyprenol ω-saturation. To test the activity of AglH, the enzyme was overexpressed in *Escherichia coli*, isolated as a membrane fraction, and incubated with either short (C55–60) or long (C85–105) (*S*)-Dol-P and UDP-[^{3}H]GlcNAc. However, even after an exhaustive screen of reaction conditions, we did not observe formation of the Dol-PP-GlcNAc product (Fig. 3b), which was predicted based on the AglH complementation studies and the pathway proposals reported in the literature.

We then turned our attention to AglK, a 28.2 kDa protein proposed to function downstream of AglH.[19] AglK was overexpressed in *E. coli* with N-terminal T7 and C-terminal His_6 tags, purified from bacterial membranes in high yield, and confirmed by SDS-PAGE and Western blot analyses (Supplementary Results, Supplementary Fig. 1). Although this enzyme has been previously described to function in tandem with a second gene product, AglC, to produce the downstream Dol-PP-disaccharide from the Dol-PP-monosaccharide,[19] a closer examination of the AglK sequence revealed that it exhibits a strikingly high sequence similarity (> 50%) to Alg5, the Dol-P-Glc (DPG) synthase from *S. cerevisiae*.[23] Interestingly however AglK was also homologous to AglJ, the enzyme proposed to carry out the first step in the *H. volcanii* pathway,[24] and an unassigned ORF (MMP1170) from *M. maripaludis* (Supplementary Fig. 2).[17] While AglK was not predicted to include a true transmembrane domain (TMHMM server[25]), the protein sedimented with the membrane fraction and required detergent for solubility, suggesting it may contain hydrophobic regions that associate with the membrane. Upon incubation of AglK with short (C55–60) (*S*)-Dol-P and UDP-[^{3}H]GlcNAc, we observed rapid formation of a radiolabeled organic-soluble product, characteristic of a dolichyl-linked compound. The reaction was specific for UDP-GlcNAc, as we detected no turnover with radiolabeled UDP-Glc, UDP-GalNAc, UDP-Gal, or GDP-Man (Fig. 3b). Additionally, enzymatic activity required divalent metal cations (Ca^{2+}, Mn^{2+}, or Mg^{2+}, Fig. 3c). We purified the resulting (C55–60) Dol-P-linked product using normal phase HPLC and identified it as Dol-P-GlcNAc by MS and NMR (Supplementary Figs. 3 and 4, Supplementary Table 1). Analysis of the reaction products by capillary electrophoresis also revealed that the nucleotide product of this reaction was UDP rather than UMP, providing further evidence that AglK acts as a Dol-P-GlcNAc synthase and transfers GlcNAc to the Dol-P acceptor, rather than GlcNAc-1-P (Supplementary Fig. 5). Finally, AglK showed much higher activity with the shorter, native-like (C55–60) Dol-P substrate compared with the longer (C85–105) Dol-P, indicating a strong preference of AglK for the shorter dolichols found in archaea (Fig. 3d).[14,26] To provide further evidence for our findings, we carried out mass spectrometry analysis of a cellular lipid extract from the closely related methanogen *M. maripaludis*, which has been previously reported to generate a very similar N-linked glycan to that of *M. voltae*.[27] The MS data indicated the presence of a component ($[M-2H+Na]^-$ = 1692.0, Supplementary Fig. 6), which corresponded to a short (C55) Dol-P-trisaccharide bearing two sites of unsaturation, presumably the α- and ω-isoprene units predicted for archaeal dolichols and a trisaccharide, with the mass calculated for the first three sugars of the asparagine-linked glycan previously identified in *M. maripaludis*.[27]

Dol-P-GlcNAc exhibits an α-glycosidic linkage

Due to the high sequence similarity of AglK with DPG and DPM synthases, it was predicted that AglK would afford β-linked Dol-P-GlcNAc, as previous studies using a variety of methods including acid/base hydrolysis, enzymatic digestion, and MS suggest that both Dol-P-Man and Dol-P-Glc produced by Dpm1 and Alg5 show this anomeric configuration.[28–30] However, it was essential to unambiguously determine the anomeric configuration of the product since this stereocenter is critical in the ultimate OTase reaction catalyzed by AglB. We investigated the stereochemistry of the GlcNAc moiety bound to Dol-P using phosphorus-decoupled ^{1}H NMR, and determined the $J_{1,2}$ value of the anomeric proton to be 3.4 Hz (Supplementary Fig. 4, Supplementary Table 1). This value strongly indicates that the Dol-P-GlcNAc formed by AglK is the α-anomer and that the AglK reaction proceeds with retention of stereochemistry (Fig. 3a).[13,31]

AglC is a UDP-Glc-2,3-diNAcA glycosyltransferase

In order to examine the downstream steps of glycan assembly, we focused our efforts on the characterization of AglC. AglC had previously been implicated in the assembly of the Dol-linked disaccharide in conjunction with AglK based on genetic studies;[19] however, the role of this enzyme in the pathway was now unclear in light of our characterization of AglK. We envisioned two possible functions for AglC. In the first scenario, AglC would behave as a DPG homolog similar to AglK, utilizing UDP-Glc-2,3-diNAcA and Dol-P to generate Dol-P-Glc-2,3-diNAcA. Based on sequence analysis AglC was also annotated as a GT2-family glycosyltransferase (GT) along with DPG and AglK; however, the sequence similarity of AglC to DPG was much lower (21%) compared with AglK to DPG (53%). Alternatively, AglC could directly act on UDP-Glc-2,3-diNAcA, transferring the highly modified carbohydrate to the Dol-P-GlcNAc product of AglK. To test these hypotheses, we first overexpressed AglC in *E. coli* with a N-terminal T7 and C-terminal His_6 tag and purified it from the membrane fraction (Supplementary Fig. 1). Sequence analysis predicted that AglC likely contains 1–2 transmembrane domains at the C-terminus,[25] and thus required detergent to maintain solubility. We also prepared UDP-[^{3}H]Glc-2,3-diNAcA using the enzymes WbpB, WbpE, and WbpD from *Pseudomonas aeruginosa* as previously described.[32] We incubated AglC with UDP-[^{3}H]Glc-2,3-diNAcA and either Dol-P (for scenario 1) or Dol-P-GlcNAc (scenario 2, Fig. 4a) as a glycosyl acceptor. We found that AglC readily converted the aqueous-soluble starting material UDP-[^{3}H]Glc-2,3-diNAcA to the organic-soluble Dol-P-disaccharide when Dol-P-GlcNAc was used as the glycosyl acceptor. We detected no turnover with radiolabeled UDP-Glc, UDP-GalNAc, or UDP-Gal, and observed only minimal turnover for UDP-GlcNAc (Fig. 4b). The enzyme was specific for the monophosphate-linked Dol-P-GlcNAc, as no conversion was observed with Dol-P or Dol-PP-GlcNAc (Fig. 4c). MS analysis confirmed the resulting product as Dol-P-GlcNAc-Glc-2,3-diNAcA (Supplementary Fig. 7), indicating that AglC is a Glc-2,3-diNAcA glycosyltransferase (Fig. 4a).

Dol-P-GlcNAc-Glc-2,3-diNAcA is a glycan donor for the AglB

With the Dol-P-disaccharide in hand, we then investigated its competence as a substrate for the predicted OTase AglB. Although the full length *M. voltae* N-linked glycan is a

trisaccharide, previous *in vitro* studies demonstrate that disaccharide-linked substrates are generally sufficient for OTase activity.[33–36] Sequence analysis predicted that AglB contains 11–13 transmembrane domains located at the N-terminus of the protein and a large (39.1 kDa) C-terminal soluble domain, which is similar to the overall topology observed for both Stt3 and PglB.[3,36,37] We overexpressed AglB in *E. coli* with N-terminal T7 and C-terminal His_6 tags and purified it from the membrane fraction in good yield (2–5 mg/L *E. coli* culture, Supplementary Fig. 1). The OTase was then incubated with Dol-P-GlcNAc-[^{3}H]Glc-2,3-diNAcA and the acceptor peptide Ac(YKYNESSYK*p*NF)NH_2, which was based on the natively glycosylated FlaB2 protein sequence. Remarkably, we observed efficient formation of a radiolabeled glycopeptide product (Fig. 5). AglB was specific for the Dol-P-disaccharide donor and showed minimal activity with Dol-P-GlcNAc (Fig. 5b). The activity of AglB, like the other OTases characterized to date, required the addition of divalent metal cations such as Mn^{2+} (Fig. 5c). We readily separated the glycopeptide product from the peptide starting material using reverse phase HPLC (Fig. 5d), and confirmed the product by MS analysis (Supplementary Fig. 8). Finally, no glycosylation activity was detected with the control peptide Ac(YKYQESSYK*p*NF)NH_2, indicating that AglB activity is dependent on the canonical AsnXaaSer/Thr sequon (Figs. 5a,b).

DISCUSSION

Complex glycoconjugates including glycoproteins play central roles in biology, yet the biochemical details of the assembly and recognition of these structures are largely incomplete due to the technical challenges associated with preparing homogenous substrates for unambiguous characterization. In this study, we have utilized purified substrates and enzymes along with detailed product characterization to define the three key processes of N-linked glycosylation in *M. voltae*. These are (1) the initial reaction between polyprenyl-phosphate and a UDP-activated glycosyl donor, which affords the first amphiphilic membrane-associated substrate; (2) the second glycosyltransferase step, which yields the simplest competent substrate for the OTase; and (3) the culminating OTase reaction to form the N-linked glycan. These findings have enabled us to parse out key differences among N-linked protein glycosylation pathways across the three domains of life.

In this work, we demonstrated that AglK is a Dol-P-GlcNAc synthase that utilizes UDP-GlcNAc and (C55–60) Dol-P to produce α-linked Dol-P-GlcNAc. This initial step immediately deviates from the biochemically characterized bacterial and eukaryotic pathways, which feature polyprenyl-diphosphate-linked (Pren-PP) products (Fig. 1), as well as the proposed pathways in selected methanogens.[16] Interestingly, AglK showed high sequence similarity to *S. cerevisiae* Alg5, the Dol-P-Glc synthase, which is believed to afford β-linked Dol-P-GlcNAc (Supplementary Fig. 2a).[28,30] This suggests a common evolutionary origin for both enzymes, and may shed light on the intriguing observation that in eukaryotes Dol-P-monosaccharides serve as glycosyl donors for the glycosyltransferases within the ER lumen in the second phase of the dolichol pathway, compared with the nucleotide sugars that are utilized in the first phase.

An earlier study suggested another protein, AglH (Mv1751), is responsible for the first glycosyltransferase step based on a genetic study in which overexpression of the *M. voltae*

aglH gene is able to rescue growth of an *alg7* deletion in *S. cerevisiae*.[20] From this work, the authors extrapolated that Dol-PP-GlcNAc must be the initial polyprenyl-linked monosaccharide produced in the *M. voltae* N-linked glycosylation pathway. However, using recombinant protein along with pure Dol-P and UDP-GlcNAc, we did not observe the proposed transferase activity of AglH *in vitro* (Fig. 3b), and instead demonstrated that AglK carries out the first glycosyltransferase step. It is noteworthy that the authors of the complementation study suggest that AglH may also be involved in different and essential lipid-glycoconjugate biosynthetic pathways in *M. voltae*, including assembly of the unique GlcNAc-1-P diether glycolipid found in the plasma membrane.[20,38]

One of the most intriguing aspects of this study is the finding that AglK produced Dol-P-GlcNAc bearing an anomeric α-linkage (Fig. 3a). While unexpected, the α-configuration of the AglK Dol-P-GlcNAc product would be consistent with the inversion of stereochemistry catalyzed by known OTases, which generate a β-glycosyl amide linkage in the ultimate glycoprotein product.[18] The N-linked glycosylation pathway glycosyltransferases in archaea were originally proposed to belong principally to either the CAZy GT2 or GT4 families.[5] Based on sequence similarity, AglK was assigned to the family-GT2 inverting GTs, which includes the biochemically and structurally characterized NDS transferase SpsA from *Bacillus subtilis*.[39,40] Comparison of the AglK and SpsA sequences suggested that both proteins may share a similar nucleotide and metal binding fold. However, since it is now clear that the GT fold identity does not necessarily dictate stereochemical outcome,[41] and because we confirmed that AglK acts with retention of stereochemistry, the family annotation will need to be revised through detailed structural and mechanistic analysis. In this context, studies on mannosylglycerate synthase (MGS), which based on sequence is most similar to the family-GT2 inverting GTs, are instructive since in this case MGS was also originally assigned to the GT2 family.[42] However, in depth analysis revealed that only minor changes in structure could accommodate a change in mechanism and stereochemical outcome, prompting reassignment of MGS to the family-GT78 retaining GTs.

Since the first membrane-committed step in the *M. voltae* pathway involves an enzyme that catalyzes a rather unique transformation, we questioned whether it might represent a hallmark for the dolichyl-monophosphate-dependent pathways. In this context, examination of enzyme sequences and resulting pathway intermediates from other archaeal species can be informative, even if the molecular details remain unclear. In the case of *H. volcanii*, AglJ is proposed to carry out the first glycosyl transfer step, and exhibits 43% similarity to AglK (Supplementary Fig. 2b.[24] Previous analysis of *H. volcanii* lipid extracts confirms the presence of Dol-P-linked glycans, suggesting that this glycosylation pathway may be more similar to that of *M. voltae* than originally proposed, although the identity of the carbohydrates is different and the stereochemistry at the glycosyl phosphate linkage is not defined.[24] Similarly, while the enzyme that catalyzes the first step of the *M. maripaludis* N-linked glycosylation pathway is not yet identified, a gene that is proposed to be involved in the process, based on its location in a gene cluster, (MMP1170) encodes a protein that shows extremely high homology to AglK (79% similarity; Supplementary Fig. 2c).[17] Furthermore, mass spectrometry analysis of an *M. maripaludis* lipid extract contained a significant component with a mass corresponding to a Dol-P-trisaccharide (Supplementary Fig. 6a–c).[27]

This finding strongly supports the notion that like halophiles, methanogens such as *M. maripaludis* and the closely related *M. voltae* rely on dolichol as the polyprenol anchor in N-linked glycan biosynthesis, and additionally require the monophosphate linkage rather than the diphosphate as previously proposed. In addition, the observation of this Dol-P-trisaccharide in *M. maripaludis* is compelling evidence that the intermediates defined in this study are consistent with those produced *in vivo*. Therefore, taken together, it is likely that the N-linked glycosylation pathways in both *H. volcanii* and *M. maripaludis* may also follow a similar monophosphate-dependent pathway as that characterized herein for *M. voltae*.

The second step of the *M. voltae* pathway involves the glycosyltransferase AglC, which converts Dol-P-GlcNAc to Dol-P-GlcNAc-Glc-2,3-diNAcA (Fig. 4a). The study of this enzyme has been challenging due to the unique nature of the UDP-sugar substrate. Indeed, analysis of glycan structures across both the archaeal and bacterial domains reveals highly modified carbohydrates, which frequently make the unambiguous biochemical annotation of target glycosyltransferase enzymes challenging. In the case of *M. voltae*, while the identity of the second carbohydrate in the glycan had been known for some time through analysis of the glycoprotein products,[18] the enzymes responsible for its biosynthesis had not been elucidated. Therefore, in order to access the essential glycosyl donor substrate for AglC, we took advantage of a known enzymatic pathway in *P. aeruginosa* lipopolysaccharide biosynthesis and developed an efficient chemoenzymatic synthesis of UDP-Glc-2,3-diNAcA.[32] As more and more archaeal and bacterial genomes are sequenced and characterized, it is likely that the compendium of unusual carbohydrate building blocks that becomes available will further enable new discoveries and insight into complex glycoconjugate assembly pathways.

The final step of N-linked glycosylation involves oligosaccharide transfer from the Dol-P donor to asparagine in acceptor proteins by the OTase AglB (Fig. 5a). It has previously been reported that the catalytic subunit of OTases from bacteria, archaea and eukaryotes share common topologies and conserved sequences.[36,37] Thus, it is intriguing that despite this homology, AglB transfers carbohydrates to asparagine in the same conserved AsnXaaSer/Thr sequon from a monophosphate- rather than a diphosphate-linked donor, as the diphosphate moiety has been typically assigned important roles in metal binding and catalysis.[36,37] In particular, the metal-coordinated diphosphate-linked glycosyl donor would be considerably more activated than the corresponding monophosphate derivative due to the higher affinity of the diphosphate for divalent cations, which would in turn render the metal-coordinated complex a better leaving group as a result of the more acidic pK_a (e.g. 5.3 for $ADPH^{2-}$-Mg^{2+} versus ~7.0 for typical monophosphate esters) of the departing phosphate species.[43] The structural and mechanistic consequences of this significant change in the glycosyl donor substrate will be extremely interesting to pursue. To date, the only structure of a full-length, catalytically active OTase is that described for PglB from *Campylobacter lari*, which uses an undecaprenyl-diphosphate linked (Und-PP) glycosyl donor substrate.[37] Additionally, structures of the large, but catalytically inactive, C-terminal soluble domains of OTases from *Pyrococcus furiosus*[44] and *Archeoglobus fulgidus*[45] have been reported; however, it is not yet known whether these employ polyprenyl-monophosphate (Pren-P) or

polyprenyl-diphosphate (Pren-PP) substrates because activity assays on the full length proteins were carried out using crude membrane fractions as a source of the glycosyl donor.

In conclusion, the *in vitro* biochemical analysis of three key enzymes in the N-linked glycosylation pathway of the archaeon *M. voltae* using purified components establishes the biochemical details of a second general strategy for N-linked glycosylation. While it is not yet known how common this strategy is amongst archaea, certainly, the comparative analysis of the structural and mechanistic aspects of the two strategies will provide new information on this complex post-translational modification throughout evolution.

METHODS

Cloning

The *aglK* and *aglC* gene sequences were optimized for expression in *E. coli* and synthesized by Genscript (Piscataway, NJ). The *aglH* and *aglB* genes were amplified by PCR from *M. voltae* PS genomic DNA using Pfu Turbo polymerase (Agilent). Both the commercial genes and PCR products were treated with the restriction endonucleases BamHI and XhoI and cloned into the same sites in the pET24a vector (Novagen) to give encoded proteins with N-terminal T7 and C-terminal His_6 tags.

Protein expression and purification

Plasmids were transformed into *E. coli* BL21-CodonPlus (DE3)-RIL cells (Agilent) using kanamycin and chloramphenicol for selection. LB media (1 L) supplemented with antibiotics was inoculated with a 5 mL starter culture and incubated at 37 °C with shaking until an optical density (600 nm) of 0.8 AU was obtained. The cultures were then cooled to 16 °C and protein expression was induced with 1 mM IPTG. After 16 h, the cells were harvested by centrifugation (4000 × *g*). For protein purification, cell pellets were resuspended in 5% of the original culture volume in buffer A (50 mM HEPES, pH 7.5/300 mM NaCl) plus 10 mM imidazole and a protease inhibitor cocktail, then subjected to sonication. To prepare cell membrane fractions, the cell lysate was spun down to remove cellular debris (6000 × *g*, 30 min), followed by pelleting of the membranes (142,000 × *g*, 1 h). The membrane pellet was then resuspended in 0.25% of the original culture volume in buffer B (50 mM HEPES, pH 7.5/150 mM NaCl) with 10 mM imidazole and stored at −80 °C. To purify proteins from crude membrane fractions, the membranes were solubilized for 1 h in either 1% n-dodecyl-β-D-maltoside (DDM) for AglK and AglC or 1% Triton X-100 for AglB and centrifuged (142,000 × *g*), and the resulting supernatant was incubated with Ni-NTA resin for 3 h. The resin was then washed (buffer B + 25 mM imidazole/0.05% DDM), the proteins eluted (buffer B + 250 mM imidazole/0.05% DDM), and the samples dialyzed (buffer B + 0.05% DDM). All purification steps were performed at 4 °C.

Synthesis of (C55–60) (S)-dolichols

For the preparation of short chain (C55–60) (*S*)-dolichols, polyprenols were first extracted from the leaves of *Rhus typhina* (staghorn sumac) as described previously,[46] which affords a distribution of C50–65 isomers. The isolated mixture was further fractionated to obtain a 5:1 mixture of C55:60 linear polyprenols. (C55–60) (*S*)-Dolichols were prepared by

regioselective asymmetric hydrogenation according to a previously reported protocol.[47] A sample of linear polyprenols (C55–60 5:1, 0.08 g, 0.10 mmol) was dissolved in toluene and concentrated by rotary evaporation. (*S*)-tol-binap $RuCl_2$ (*p*-cymene) (0.01 g, 0.01 mmol) and KOH (0.001 g, 0.02 mmol) were added to the polyprenols and dried under vacuum. N_2 was then backfilled into the flask and the vacuum was reestablished for 1 h. In a separate flask, *n*-propanol (anhydrous, 10 mL) was added to activated 4 Å molecular sieves and sparged with N_2 for 30 min. The sparged *n*-propanol was transferred to the flask containing polyprenols, (*S*)-tol-binap $RuCl_2$ (*p*-cymene) and KOH, which was immediately purged of atmosphere and backfilled with N_2. This process was repeated three times, after which a dry stir bar was added and the reaction was allowed to proceed for 8 h under N_2. NMR analysis was used to monitor the reaction. Upon completion, silica gel was added to the mixture, and flash column chromatography (toluene:EtOAc, 49:1 v/v) afforded the desired (*S*)-dolichols (5:1 C55–60 mixture, 0.072 g, 90% yield) as a colorless oil. ^{1}H NMR (600 MHz, $CDCl_3$): δ 5.17–5.07 (m, 10 H), 3.75–3.62 (m, 2 H), 2.12–1.94 (m, 38 H), 1.68 (s, 20 H), 1.60 (s, 18 H), 1.40–1.10 (m, 10 H), 0.92–0.82 (m, 3 H). The long chain (C85–105) (*S*)-dolichols were a kind gift from Kuraray Co. Ltd. Japan, where they had been prepared through the C5 homologation of C80–100 polyprenols extracted from the leaves of *Gingko biloba*.[48]

Synthesis of (C55–60) and (C85–105) (S)-dolichyl-phosphates

Phosphorylation was carried out essentially as described previously.[49] The dolichol mixture (0.003 mmol) was dissolved in dry THF (1 mL) under argon, followed by addition of 1-*H* tetrazole (0.006 g, 0.008 mmol) and diisopropylamino-di-(2-cyanoethoxy)-phosphine (0.03 g, 0.009 mmol). The solution was stirred for 2 h, then cooled to −30 °C, after which 30% hydrogen peroxide (0.2 mL) was added dropwise and the mixture was warmed to room temperature. The mixture was then diluted with 10% Na_2SO_3 (5 mL) and extracted with EtOAc (3 × 10 mL). The combined organic extract was washed with 5% $NaHCO_3$ (2 mL), brine (5 mL), dried ($MgSO_4$), and concentrated by rotary evaporation. The resulting oil was dissolved in dry MeOH (5 mL) and a catalytic amount of NaOMe was added and allowed to incubate for 2 d. Flash column chromatography with a gradient of EtOAc to EtOAc:MeOH (5:2 v/v) afforded the desired dolichol phosphate as a colorless oil. The ^{1}H and ^{31}P NMR values for the long (C85–105) (*S*)-dolichyl-phosphate matched published values.[50] For the short (C55–60) dolichyl-phosphate: TLC ($CHCl_3$:MeOH:H_2O:NH_4OH (50% in H_2O), 65:25:4:0.5 v/v): R_f=0.43; ^{1}H NMR (400 MHz, $CDCl_3$:CD_3OD, 2:1 v/v): δ 5.05–4.95 (m, 10H), 3.85–3.80 (m, 2H), 1.98–1.84 (m, 39H), 1.56 (s, 23H), 1.50–1.44 (m, 12H), 0.79 (d, *J* = 6.4 Hz, 3H); ^{31}P NMR (130 MHz, $CDCl_3$:CD_3OD, 2:1 v/v): δ 8.14. Note the long and short (*S*)-dolichols employed in these studies are derived from plant sources and therefore include three E-isoprene units, in contrast to the non-plant derived linear polyprenols that include two E-isoprenes and an additional Z-isoprene unit. Also, the ω-isoprene in archaeal dolichols is saturated.[10,13]

Synthesis of (C55–60) Dolichyl-PP-GlcNAc (1) and (C85–105) Dolichyl-PP-GlcNAc (2)

The acetylated GlcNAc precursors were synthesized following previously published protocols;[51,52] see Supplementary Figure 9 for the overall synthetic route. To a solution of GlcNAc-1-phosphate (**3**, 12.0 mg, 0.02 mmol) in dry DMF (1 mL), 1,1'-carbonyldiimidazole (CDI, 0.2 mmol) was added. After 6 h, MeOH was introduced to the reaction in a dropwise

manner to quench excess CDI, and the mixture was incubated for 30 min. Dolichyl-phosphate (C55–60 or C85–105, 0.014 mmol) in CH_2Cl_2 (1 mL) was then added, and the reaction was allowed to stir for 8 d at room temperature and monitored by TLC. Upon completion, the crude mixture was concentrated under vacuum and purified by silica gel chromatography ($CHCl_3$:MeOH:2 M NH_4OH, 85:14:1 v/v) to give the pure O-acetylated dolichyl-diphosphate-linked monosaccharides (**4**, **5**). To remove the O-acetyl protecting groups, the Dol-PP-GlcNAc compounds **4** and **5** were dissolved in MeOH (1 mL) and a solution of 25% NaOMe/MeOH (40 μL) was added dropwise to a final concentration of 1% NaOMe. The solution was stirred at room temperature for 30 min, then neutralized with Dowex 50WX8-200 ion-exchange resin (pyridinium form) and concentrated under vacuum to yield the desired dolichyl-diphosphate monosaccharides **1** and **2**. For short (C55–60) Dol-PP-GlcNAc (**1**): ($CHCl_3$:MeOH:H_2O:NH_4OH (50% in H_2O), 65:25:4:0.5 v/v): R_f=0.20; ESI-MS (*m/z*): $[M]^-$ calcd. for $C_{63}H_{106}NO_{12}P_2^{2-}$, 1131.5; found, 1131.3. ^{31}P NMR values (162 MHz, $CDCl_3$:CD_3OD, 2:1 v/v): δ −5.8, −8.6. NMR characterization of the long (C85–105) Dol-PP-GlcNAc (**2**) matched previously reported values.[50,51]

Synthesis of UDP-[^{3}H]Glc-2,3-diNAcA

UDP-[^{3}H]Glc-2,3-diNAcA was prepared through chemoenzymatic synthesis starting from UDP-GlcNAcA and the enzymes WbpB, WbpE, and WbpD from *P. aeruginosa* PAO1 as previously described.[32] After purification, the WbpB/WbpE product UDP-GlcNAc(3NH_2)A (1 mM) was incubated with [^{3}H]AcCoA (20 Ci/mmol, American Radiolabeled Chemicals), 50 mM HEPES, pH 7, and WbpD (0.5 μg) in a reaction volume of 1 mL at 30 °C for 5 min, followed by a chase with unlabeled AcCoA (0.75 mM) for 4 h. The radiolabeled product was purified on a Phenomenex Synergi C18 reversed-phase HPLC column as described previously.[32]

Glycosyltransferase assays

For AglK assays, dried (C55–60) or (C85–105) (*S*)-Dol-P (5 nmol) was resuspended in DMSO (3 μL) and vortexed, followed by the addition of 0.715% DDM (7 μL), 0.5 M HEPES, pH 7.5 (10 μL), 1 M $MgCl_2$ (1 μL), 100 mM DTT (2 μL), and AglK (0.5 μM), along with H_2O for a final volume of 100 μL. The reaction was then initiated by the addition of UDP-[^{3}H]GlcNAc (5 pmol, 20 Ci/mmol) and incubated at 25 °C. Aliquots (15 μL) were removed at various time points and quenched by dilution into $CHCl_3$:MeOH (2:1 v/v, 1.2 mL), followed by the addition of pure solvent upper phase (PSUP, 300 μL, composed of 15 mL $CHCl_3$, 240 mL MeOH, 1.83 g KCl in 235 mL H_2O).[53] The organic layer was extracted with PSUP (3 × 300 μL) and analyzed by scintillation counting. All other radiolabeled nucleotide diphosphate sugars were tested as substrates under identical conditions. Metal specificity was examined under similar conditions, varying divalent metal ion (10 mM $MgCl_2$, $MnCl_2$, $CaCl_2$, or EDTA). Preparative AglK reactions were carried out with 200 nmol (C55–60) (*S*)-Dol-P using unlabeled UDP-GlcNAc in a total volume of 1 mL. For AglC assays, dried Dol-P-GlcNAc (5 nmol) was resuspended in DMSO (3 μL) and vortexed, followed by the addition of 0.715% DDM (7 μL) and the reaction components as described for the AglK reaction. AglC (0.5 μM) was introduced, and the reaction was initiated with UDP-[^{3}H]Glc-2,3-diNAcA (1.62 nmol, 17.3 mCi/mmol) and incubated at 25 °C. All other radiolabeled UDP-sugars were tested at 5 pmol and 20 Ci/mmol, and dolichyl substrates

were tested at 5 nmol. Aliquots were removed and processed as described above. For preparative AglC reactions, 50 nmol Dol-P-GlcNAc was utilized along with unlabeled UDP-Glc-2,3-diNAcA in a total volume of 500 μL. The radiolabeled nucleotide sugars UDP-[^{3}H]-Glc, UDP-[^{3}H]GlcNAc, UDP-[^{3}H]Gal, UDP-[^{3}H]GalNAc, and GDP-[^{3}H]Man were purchased from American Radiolabeled Chemicals. Data reported are representative traces and when compared represent data from the same enzyme preparation performed on the same day.

Oligosaccharyl transferase assay

To assay the AglB OTase, DMSO (5 μL) and buffer C (50 μL, 100 mM HEPES, pH 7.5/280 mM sucrose/2.4% Triton X-100/20 mM $MnCl_2$) were added to a tube containing dried Dol-P-GlcNAc-[^{3}H]Glc-2,3-diNAcA (0.78 nmol, 17.3 mCi/mmol). Freshly purified AglB (1 μM) was then added, along with H_2O for a final volume of 100 μL. The reaction was initiated by the addition of peptide, either Ac(YKYNESSYK*p*NF)NH_2 (*p*NF = *para*-nitrophenylalanine) or Ac(YKYQESSYK*p*NF)NH_2 (0.2 μmol), which were synthesized on PAL-PEG-PS resin (Applied Biosystems) using standard Fmoc-based solid phase peptide synthesis protocols and purified to > 95% purity by C18 RP-HPLC. ESI-MS (m/z): $[M+H]^+$ calcd. for NES peptide, 1414.6; found, 1414.5. $[M+H]^+$ calcd. for QES peptide, 1428.6; found, 1428.5. Aliquots (10 μL) of the reaction mixture were removed at various time points and quenched in $CHCl_3$:MeOH:4 mM $MgCl_2$ (3:2:1 v/v, 1.2 mL), and the resulting aqueous layer was removed. The organic layer was further extracted with theoretical upper phase plus salt (TUPS, 2 × 600 μL, composed of 2.75% $CHCl_3$, 44% MeOH, and 1.55 mM $MgCl_2$).[53] The resulting aqueous layers were combined and analyzed by scintillation counting. Metal dependency was examined in the presence and absence of 10 mM $MnCl_2$ or with 10 mM EDTA.

Purification of Dol-P-saccharides and glycopeptides

Dol-P mono- and disaccharides were purified using a normal phase Varian Microsorb HPLC column and separating over 21–28% buffer E, in which buffer D is $CHCl_3$:MeOH (4:1 v/v) and buffer E is $CHCl_3$:MeOH:2 M NH_4OAc (10:10:3 v/v).[54] Elution of the Dol-P-saccharides was monitored by scintillation counting and/or TLC on silica plates using a solvent system of $CHCl_3$:MeOH:H_2O:NH_4OH (50% in H_2O) (65:25:4:0.5 v/v) and stained with ceric ammonium molybdate (CAM). Glycopeptides were purified using a C18 reverse phase column (YMC-Pack ODS-AQ, 120 Å, 3 mm, 100 × 3 mm) and separating over 30–38% buffer G, where buffer F is H_2O/0.1% TFA and buffer G is CH_3CN/0.1% TFA. The glycopeptide product was further characterized by high resolution ESI-MS.

NMR and MS characterization of Dol-P-saccharides

NMR spectra were acquired using either a Bruker 600 or Varian Inova 500 MHz spectrometer equipped with a 5 mm inverse broadband gradient probe. Samples were diluted in $CDCl_3$ for Dol-P or $CDCl_3$:CD_3OD (2:1 v/v) for Dol-P-GlcNAc and analyzed. The anomeric configuration of the Dol-P-GlcNAc glycosidic linkage was determined using a ^{31}P-decoupled ^{1}H pulse sequence. ESI-TOF mass spectra were obtained by the Mass Spectrometry Laboratory at the University of Illinois, Urbana-Champaign.

Analysis of lipid extracts from M. maripaludis

Lipid extraction from *M. maripaludis* cells was carried out as previously described.[24] A 10 g pellet of *M. maripaludis* was mixed with MeOH and $CHCl_3$ (2:1 v/v per 0.8 g cell pellet, 12.5 mL) and stirred for 24 h. The mixture was then centrifuged (360 × *g*) for 30 min at 4 °C, and the supernatant was collected, filtered, and extracted with $CHCl_3$ (10 mL). The organic layer was then concentrated under vacuum, and the residue was further purified by normal phase liquid chromatography on a YMC-Pack PVA-Sil-NP analytical column using mobile phases A ($CHCl_3$:MeOH, 4:1 v/v) and B ($CHCl_3$:MeOH:2 M NH_4OAc, 10:10:3 v/v) with a gradient of 1–100% B over 60 min. Column fractions were monitored by TLC ($CHCl_3$:MeOH:H_2O:NH_4OH (50% in H_2O), 65:25:4:0.5 v/v), concentrated and dissolved in $CHCl_3$:MeOH (4:1, v/v) + 0.1% formic acid, then directly injected onto a Finnegan LCQ Deca ESI-MS. A significant compound was observed by TLC and ESI-MS that corresponds with α- and ω-saturated (C55) ManNAc3NAm(6Thr)A-β1,4-Glc-2,3-diNAcA-β1,3-GalNAc-P-Dol, a compound containing a trisaccharide that exhibits the same mass as the first three sugars in the *M. maripaludis* asparagine-linked glycan.[27] TLC ($CHCl_3$:MeOH:H_2O:NH_4OH (50% in H_2O), 65:25:4:0.5 v/v): R_f= 0.39; ESI-MS (*m/z*): $[M-2H+Na]^-$ calcd. for $NaC_{87}H_{143}N_7O_{22}P$, 1692.0; found, 1691.9; $[M-2H]^{2-}$ calcd. for $C_{87}H_{143}N_7O_{22}P^-$, 834.5; found 834.6.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGEMENTS

This work was supported by a grant from the National Institutes of Health (GM039334 to B.I.) The authors are grateful to Ms. Marcie Jaffee for help with initial expression studies, Dr. J. Simpson (DCIF, MIT) for assistance with NMR acquisition, Prof. W. Whitman (University of Georgia) for the gift of both *M. voltae* genomic DNA and *M. maripaludis* cell pellets, and Prof. E. Swiezewska (Polish Academy of Sciences) for providing a sample of racemic short dolichols for preliminary studies.

References

1. Nothaft H, Szymanski CM. Protein glycosylation in bacteria: sweeter than ever. Nat. Rev. Microbiol. 2010; 8:765–778. [PubMed: 20948550]
2. Eichler J. Extreme sweetness: protein glycosylation in archaea. Nat. Rev. Microbiol. 2013; 11:151–156. [PubMed: 23353769]
3. Larkin A, Imperiali B. The expanding horizons of asparagine-linked glycosylation. Biochemistry. 2011; 50:4411–4426. [PubMed: 21506607]
4. Szymanski CM, Yao R, Ewing CP, Trust TJ, Guerry P. Evidence for a system of general protein glycosylation in *Campylobacter jejuni*. Mol. Microbiol. 1999; 32:1022–1030. [PubMed: 10361304]
5. Magidovich H, Eichler J. Glycosyltransferases and oligosaccharyltransferases in archaea: putative components of the N-glycosylation pathway in the third domain of life. FEMS Microbiol. Lett. 2009; 300:2005–2014.
6. Chaban B, Voisin S, Kelly J, Logan SM, Jarrell KF. Identification of genes involved in the biosynthesis and attachment of *Methanococcus voltae* N-linked glycans: insight into N-linked glycosylation pathways in archaea. Mol. Microbiol. 2006; 61:259–268. [PubMed: 16824110]
7. Abu-Qarn M, Eichler J. Protein N-glycosylation in archaea: defining *Haloferax volcanii* genes involved in S-layer glycoprotein glycosylation. Mol. Microbiol. 2006; 61:511–525. [PubMed: 16762024]

8. Mescher MF, Strominger JL. Purification and characterization of a prokaryotic glucoprotein from the cell envelope of *Halobacterium salinarium*. J. Biol. Chem. 1976; 251:2005–2014. [PubMed: 1270419]
9. Mescher MF, Hansen U, Strominger JL. Formation of lipid-linked sugar compounds in *Halobacterium salinarium*. Presumed intermediates in glycoprotein synthesis. J. Biol. Chem. 1976; 251:7289–7294. [PubMed: 1002689]
10. Lechner J, Wieland F, Sumper M. Biosynthesis of sulfated saccharides N-glycosidically linked to the protein via glucose. Purification and identification of sulfated dolichyl monophosphoryl tetrasaccharides from halobacteria. J. Biol. Chem. 1985; 260:860–866. [PubMed: 2857171]
11. Burda P, Aebi M. The dolichol pathway of N-linked glycosylation. Biochim. Biophys. Acta. 1999; 1426:239–257. [PubMed: 9878760]
12. Calo D, Guan Z, Naparstek S, Eichler J. Different routes to the same ending: comparing the N-glycosylaton processes of *Haloferax volcanii* and *Haloarcula morismortui*, two halophilic archaea from the Dead Sea. Mol. Microbiol. 2011; 81:1166–1177. [PubMed: 21815949]
13. Kuntz C, Sonnenbichler J, Sonnenbichler I, Sumper M, Zeitler R. Isolation and characterization of dolichol-linked oligosaccharides from *Haloferax volcanii*. Glycobiology. 1997; 7:897–904. [PubMed: 9363431]
14. Guan Z, et al. Distinct glycan-charged phosphodolichol carriers are required for the assembly of the pentasaccharide N-linked to the *Haloferax volcanii* S-layer glycoprotein. Mol. Microbiol. 2010; 78:1294–1303. [PubMed: 21091511]
15. Guan Z, Meyer B, Albers S-V, Eichler J. The thermoacidophilic archaeaon *Sulfolobus acidocaldarius* contains an unusually short, highly reduced dolichol phosphate. Biochim. Biophys. Acta. 2011; 1811:607–616. [PubMed: 21745590]
16. Yurist-Doutsch S, Chaban B, VanDyke DJ, Jarrell KF, Eichler J. Sweet to the extreme: protein glycosylation in archaea. Mol. Microbiol. 2008; 68:1079–1084. [PubMed: 18476920]
17. VanDyke DJ, et al. Identification of genes involves in the assembly and attachment of a novel flagellin N-linked tetrasaccharide important for motility in the archaeon *Methanococcus maripaludis*. Mol. Microbiol. 2009; 72:633–644. [PubMed: 19400781]
18. Voisin S, et al. Identification and characterization of the unique N-linked glycan common to the flagellins and S-layer glycoprotein of *Methanococcus voltae*. J. Biol. Chem. 2005; 280:16586–16593. [PubMed: 15723834]
19. Chaban B, Logan SM, Kelly JF, Jarrell KF. AglC and AglK are involved in biosynthesis and attachment of diacetylated glucuronic acid to the N-glycan in *Methanococcus voltae*. J. Bacteriol. 2009; 191:187–195. [PubMed: 18978056]
20. Shams-Eldin H, Chaban B, Niehus S, Schwarz RT, Jarrell KF. Identification of the archaeal *alg7* gene homolog (encoding N-acetylglucosamine-1-phosphate transferase) of the N-linked glycosylation system by cross-domain complementation in *Saccharomyces cerevisiae*. J. Bacteriol. 2008; 190:2217–2220. [PubMed: 18178736]
21. Jones MB, Rosenburg JN, Betenbaugh MJ, Krag SS. Structure and synthesis of polyisoprenoids used in *N*-glycosylation across the three domains of life. Biochim. Biophys. Acta. 2009; 1790:485–494. [PubMed: 19348869]
22. Naparstek S, Guan Z, Eichler J. A predicted geranylgeranyl reductase reduces the ω-position isoprene of dolichol phosphate in the halophilic archaeon, *Haloferax volcanii*. Biochim. Biophys. Acta. 2012; 1821:923–933. [PubMed: 22469971]
23. Heesen, S.t.; Lehle, L.; Weissmann, A.; Aebi, M. Isolation of the *ALG5* locus encoding the UDP-glucose:dolichyl-phosphate glucosyltransferase from *Saccharomyces cerevisiae*. Eur. J. Biochem. 1994; 224:71–79. [PubMed: 8076653]
24. Kaminski L, et al. AglJ adds the first sugar of the N-linked pentasaccharide decorating the *Haloferax volcanii* S-layer glycoprotein. J. Bacteriol. 2010; 192:5572–5579. [PubMed: 20802039]
25. Krogh A, Larsson B, von Heijne G, Sonnhammer EL. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J. Mol. Biol. 2001; 305:567–580. [PubMed: 11152613]

26. Hartley MD, Imperiali B. At the membrane frontier: a prospectus on the remarkable evolutionary conservation of polyprenols and polyprenyl-phosphates. Arch. Biochem. Biophys. 2011; 517:83–97. [PubMed: 22093697]
27. Kelly J, Logan SM, Jarrell KF, VanDyke DJ, Vinogradov EV. A novel *N*-linked flagellar glycan from *Methanococcus maripaludis*. Carbohydr. Res. 2009; 344:648–653. [PubMed: 19203750]
28. Behrens NH, Leloir LF. Dolichyl monophosphate glucose: an intermediate in glucose transfer in liver. Proc. Natl. Acad. Sci. U.S.A. 1970; 66:153–159. [PubMed: 5273893]
29. Tkacz JS, Herscovics A, Warren CD, Jeanloz RW. Mannosyltransferase activity in calf pancreas microsomes. Formation from guanosine diphosphate-d-[^{14}C]mannose of a ^{14}C-labeled mannolipid with properties of dolichyl mannopyranosyl phosphate. J. Biol. Chem. 1974; 249:6372–6381. [PubMed: 4371440]
30. Herscovics A, Warren CD, Jeanloz RW. Anomeric configuration of the dolichyl-d-mannosyl phosphate formed in calf pancreas microsomes. J. Biol. Chem. 1975; 250:8079–8084. [PubMed: 1176461]
31. O'Connor JV, Nunez HA, Barker R. α- and β-Glycopyranosyl phosphates and 1,2-phosphates. Assignments of conformations in solution by ^{13}C and ^{1}H NMR. Biochemistry. 1979; 18:500–507. [PubMed: 420796]
32. Larkin A, Imperiali B. Biosynthesis of UDP-GlcNAc(3NAc)A by WbpB, WbpE, and WbpD: enzymes in the Wbp pathway responsible for O-antigen assembly in *Pseudomonas aeruginosa* PAO1. Biochemistry. 2009; 48:5446–5455. [PubMed: 19348502]
33. Sharma CB, Lehle L, Tanner W. N-Glycosylation of yeast proteins. Characterization of the solubilized oligosaccharyl transferase. Eur. J. Biochem. 1981; 116:101–108. [PubMed: 7018901]
34. Imperiali B, Hendrickson TL. Asparagine-linked glycosylation: specificity and function of oligosaccharyl transferase. Bioorg. Med. Chem. 1995; 3:1565–1578. [PubMed: 8770382]
35. Glover KJ, Weerapana E, Numao S, Imperiali B. Chemoenzymatic synthesis of glycopeptides with PglB, a bacterial oligosaccharyl transferase from *Campylobacter jejuni*. Chem. Biol. 2005; 12:1311–1315. [PubMed: 16356848]
36. Jaffee MB, Imperiali B. Exploiting topological constrainsts to reveal buried sequence motifs in the membrane-bound N-linked oligosaccharyl transferases. Biochemistry. 2011; 35:7557–7567. [PubMed: 21812456]
37. Lizak C, Gerber S, Numao S, Aebi M, Locher KP. X-ray structure of a bacterial oligosaccharyltransferase. Nature. 2011; 474:350–355. [PubMed: 21677752]
38. Ferrante G, Ekiel I, Sprott GD. Structural characterization of the lipids of *Methanococcus voltae*, including a novel N-acetylglucosamine 1-phosphate diether. J. Biol. Chem. 1986; 261:17062–17066. [PubMed: 3782154]
39. Charnock SJ, Davies GJ. Structural analysis of the nucleotide-diphospho-sugar transferase, SpsA from *Bacillus subtilis*, in native and nucleotide-complexed forms. Biochemistry. 1999; 38:6380–6385. [PubMed: 10350455]
40. Coutinho PM, Deleury E, Davies GJ, Henrissat B. An evolving hierarchical family classification for glycosyltransferases. J. Mol. Biol. 2003; 328:307–317. [PubMed: 12691742]
41. Lairson LL, Henrissat B, Davies GJ, Withers SG. Glycosyltransferases: structures, functions, and mechanisms. Annu. Rev. Biochem. 2008; 77:521–555. [PubMed: 18518825]
42. Flint J, et al. Structural dissection and high-throughput screening of mannosylglycerate synthase. Nat. Struct. Mol. Biol. 2005; 12:608–614. [PubMed: 15951819]
43. Ramirez F, Marecek JF. Coordination of magnesium with adenosine 5'-diphosphate and triphosphate. Biochim. Biophys. Acta. 1980; 589:21–29. [PubMed: 6444521]
44. Igura M, et al. Structure-guided identification of a new catalytic motif of oligosaccharyltransferase. EMBO J. 2008; 27:234–243. [PubMed: 18046457]
45. Matsumoto S, et al. Crystal structure of the C-terminal globular domain of oligosaccharyltransferase from *Archaeoglobus fulgidus* at 1.75 Å resolution. Biochemistry. 2012; 51:4157–4166. [PubMed: 22559858]
46. Świeżewska E, et al. The search for plant polyprenols. Acta Biochim. Polon. 1994; 41:221–259. [PubMed: 7856395]

47. Wu R, Beauchamps MG, Laquidara JM, Sowa JR Jr. Ruthenium-catalyzed asymmetric transfer hydrogenation of allylic alcohols by an enantioselective isomerization/transfer hydrogenation mechanism. Angew. Chem. Int. Ed. 2012; 51:2106–2110.

48. Suzuki S, et al. Synthesis of mammalian dolichols from plant polyprenols. Tetrahedron Lett. 1983; 24:5103–5106.

49. Branch CL, Burton G, Moss SF. An expedient synthesis of allylic polyprenol phosphates. Synth. Commun. 1999; 29:2639–2644.

50. Imperiali B, Zimmerman JW. Synthesis of dolichylpyrophosphate-linked oligosaccharides. Tetrahedron Lett. 1990; 31:6485–6488.

51. Tai VWF, Imperiali B. Substrate specificity of the glycosyl donor for oligosaccharyl transferase. J. Org. Chem. 2001; 66:6217–6228. [PubMed: 11559166]

52. Sim MM, Kondo H, Wong C-H. Synthesis and use of glycosyl phosphites: an effective route to glycosyl phosphates, sugar nucleotides, and glycosides. J. Am. Chem. Soc. 1993; 115:2260–2267.

53. Folch J, Lees M, Stanley GHS. A simple method for the isolation and purification of total lipids from animal tissues. J. Biol. Chem. 1957; 226:497–509. [PubMed: 13428781]

54. Troutman JM, Imperiali B. *Campylobacter jejuni* PglH is a single active site processive polymerase that utilizes product inhibition to limit sequential glycosyl transfer reactions. Biochemistry. 2009; 48:2807–2816. [PubMed: 19159314]

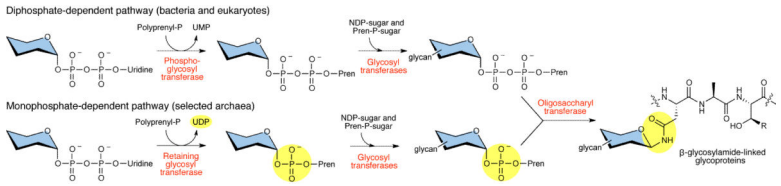

Figure 1. N-linked glycosylation across the three domains of life

The diphosphate-dependent pathway used by eukaryotes and bacteria is initiated by a phosphoglycosyl transferase, which generates a polyprenyl-PP-linked monosaccharide (Pren-PP-monosaccharide) (*top*). In this reaction, UMP is released as a byproduct. In contrast, the monophosphate-dependent pathway found in selected archaea begins with the action of a retaining glycosyltransferase to generate a polyprenyl-P-linked monosaccharide (Pren-P-monosaccharide) (*bottom*). Here, UDP is released as a byproduct. In both pathways, glycosyltransferases then complete assembly of the glycan, which is then transferred by an OTase to an acceptor protein. The sugar molecule is variable depending on the pathway and is depicted with blue shading to symbolize various carbohydrates.

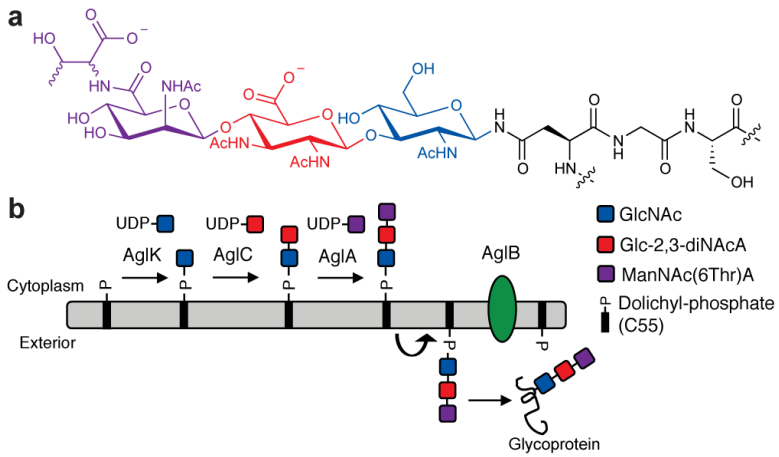

Figure 2. N-linked glycosylation in *M. voltae*

(**a**) The *M. voltae* N-linked glycan is ManNAc(6Thr)A-β1,4-Glc-2,3-diNAcA-β1,3-GlcNAc, where Glc-2,3-diNAcA is 2,3-diacetamido-2,3-dideoxy-D-glucuronic acid and the stereochemistry of the threonine residue is currently unknown. (**b**) N-linked glycan biosynthesis in *M. voltae* is initiated in the cytoplasm by the AglK-catalyzed transfer of GlcNAc to the membrane bound acceptor Dol-P. AglC transfers Glc-2,3-diNAcA to Dol-P-GlcNAc to form a β-1,3 linkage, presumably followed by the action of AglA. The resulting Dol-P linked trisaccharide is flipped from the cytoplasm to the exterior of the cell by a currently unidentified flippase and then transferred *en bloc* to a recipient protein by the OTase, AglB.

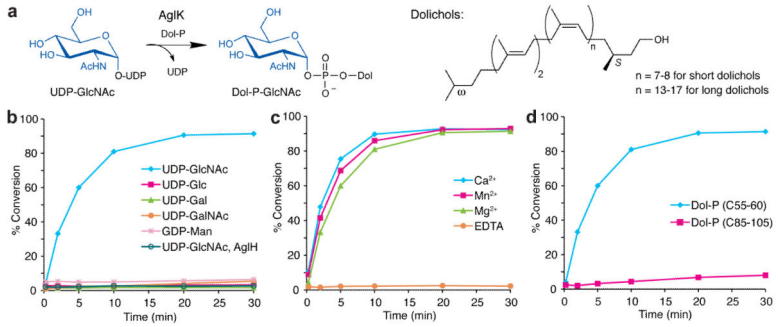

Figure 3. AglK is a Dol-P-GlcNAc synthase

(**a**) AglK catalyzes the transfer of GlcNAc from the donor substrate UDP-GlcNAc to Dol-P. (**b**) Glycosyl donor specificity of AglK and AglH using (C55-60) Dol-P and radiolabeled nucleotide sugars. Assay monitors the formation of radiolabeled Dol-P-monosaccharide over time. In this experiment, AglH activity was tested using only UDP-[^{3}H]GlcNAc and (C55-60) Dol-P. Traces represent AglK function unless otherwise indicated. (**c**) Metal dependency of AglK with UDP-[^{3}H]GlcNAc as the glycosyl donor and (C55-60) Dol-P as the polyprenyl-phosphate acceptor. (**d**) Polyprenyl-phosphate specificity of AglK with UDP-[^{3}H]GlcNAc. Assay data shown in (**b–d**) are representative traces from at least three independent experiments. Variability seen between runs was less than 5%.

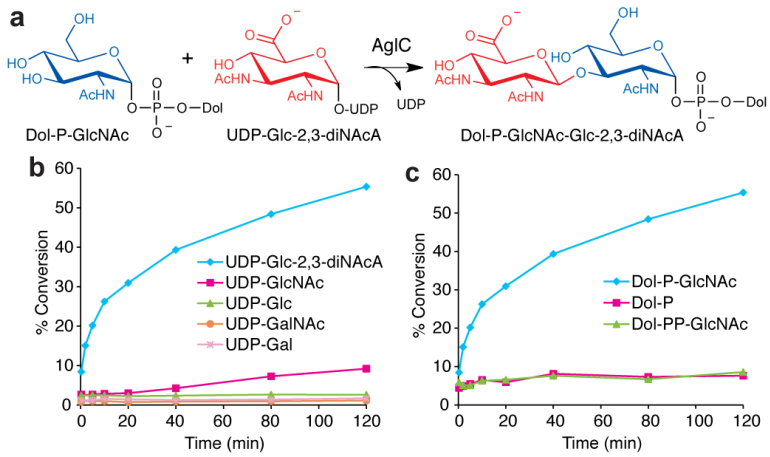

Figure 4. AglC is a UDP-Glc-2,3-diNAcA glycosyltransferase

(**a**) AglC assembles Dol-PGlcNAc-Glc-2,3-diNAcA from UDP-Glc-2,3-diNAcA and the AglK product Dol-P-GlcNAc. (**b**) Glycosyl donor specificity of AglC using Dol-P-GlcNAc and radiolabeled nucleotide sugars. Assay monitors the formation of radiolabeled Dol-P-disaccharide over time. (**c**) Polyprenyl-phosphate acceptor specificity of AglC using UDP-[^{3}H]Glc-2,3-diNAcA as the donor substrate. Assay data shown in (**b**,**c**) are representative traces from at least three independent experiments. Variability seen between runs was less than 5%.

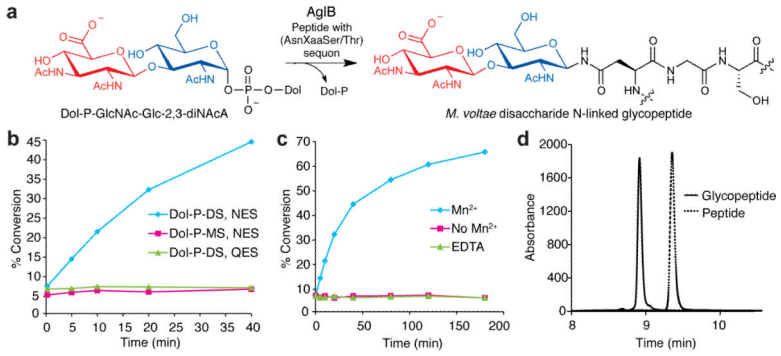

Figure 5. The OTase AglB utilizes Dol-P-disaccharide to generate an N-linked glycopeptide

(**a**) AglB catalyzes the transfer of the disaccharide GlcNAc-Glc-2,3-diNAcA from the AglC product Dol-P-GlcNAc-Glc-2,3-diNAcA to an acceptor peptide. (**b**) AglB activity assay using the peptide YKYNESSYK, based on the *M. voltae* flagellum protein FlaB2, as the acceptor substrate and Dol-P-GlcNAc (MS) or Dol-P-disaccharide (DS) as the donor substrates. A peptide lacking the key asparagine residue, YKYQESSYK, was also screened. (**c**) AglB assay performed with either the addition of $MnCl_2$, no exogenous $MnCl_2$, or EDTA. Assay data shown in (**b**,**c**) are representative traces from at least three independent experiments. Variability seen between runs was less than 5%. (**d**) RP-HPLC purification of the peptide and the glycopeptide (280 nm).