Production of homogeneous glycoprotein with multisite modifications by an engineered N-glycosyltransferase mutant

Received for publication, February 8, 2017, and in revised form, April 4, 2017 Published, Papers in Press, April 5, 2017, DOI 10.1074/jbc.M117.777383

Qitao Song †, Zhigang Wu ‡, Yueyuan Fan †, Woran Song †, Peiru Zhang †, Li Wang †, Faxing Wang †, Yangyang Xu †, Peng G. Wang ‡, and Jiansong Cheng ‡

From the ¶1 State Key Laboratory of Medicinal Chemical Biology and College of Pharmacy, Nankai University, Haihe Education Park, 38 Tongyan Road, Tianjin 300353, China and the ¶2 Department of Chemistry, Georgia State University, Atlanta, Georgia 30303

Edited by Gerald W. Hart

Naturally occurring N-glycoproteins exhibit glycoform heterogeneity with respect to N-glycan sequon occupancy (macroheterogeneity) and glycan structure (microheterogeneity). However, access to well-defined glycoproteins is always important for both basic research and therapeutic purposes. As a result, there has been a substantial effort to identify and understand the catalytic properties of N-glycosyltransferases, enzymes that install the first glycan on the protein chain. In this study we found that ApNGT, a newly discovered cytoplasmic N-glycosyltransferase from Actinobacillus pleuropneumoniae, has strict selectivity toward the residues around the Asn of N-glycosylation sequon by screening a small library of synthetic peptides. The inherent stringency was subsequently demonstrated to be closely associated with a critical residue (Gln-469) of ApNGT which we propose hinders the access of bulky residues surrounding the occupied Asn into the active site. Site-saturated mutagenesis revealed that the introduction of small hydrophobic residues at the site cannot only weaken the stringency of ApNGT but can also contribute to enormous improvement of glycosylation efficiency against both short peptides and proteins. We then employed the most efficient mutant (Q469A) other than the wild-type ApNGT to produce a homogeneous glycoprotein carrying multiple (up to 10) N-glycans, demonstrating that this construct is a promising biocatalyst for potentially addressing the issue of macroheterogeneity in glycoprotein preparation.

N-Glycosylation is one of the most prominent post-translational protein modifications and plays a role in various cellular functions including protein folding, sorting, and stability and cell-cell interactions. The modification is associated with all three domains of life (Eukarya, Bacteria, and Archaea) and has similar molecular mechanisms. In short, an oligosaccharide is assembled in a stepwise manner on a lipid anchor, giving rise to a lipid-linked oligosaccharide, and then transferred en bloc onto the acceptor asparagine residues by oligosaccharyltransferase (OST) (1–3).

Recently, an alternative N-glycosylation system has been discovered in a few species in the class of γ-proteobacteria (4–6). A striking difference from the canonical N-glycosylation is that this novel pathway begins with the attachment of the linking monosaccharide from nucleotide-activated donors to asparagine side chains, which is catalyzed by a cytoplasmic N-glycosyltransferase (NGT) (7). The resolved crystal structure of NGT from Actinobacillus pleuropneumoniae (ApNGT) revealed that the N-glycosyltransferase contains an N-terminal α-helical domain fold and a C-terminal GT-B fold with two Rossman-like domains. The extensive contacts between the α-helical domain and the GT-B domains generate a unique funnel-shaped groove adjacent to the UDP-hexose-binding site, which was proposed to accommodate the acceptor protein (8). Quite interestingly, ApNGT is neither phylogenetically nor structurally related to known OSTs. The GT-B domain of ApNGT is instead closely related to that of an O-GlcNAc transferase from Xanthomonas campestris (XcOGT) (8, 9), which involves in the regulation of the cellular protein O-GlcNAcylations (10).

So far, ApNGT is the most extensively investigated NGT-like N-glycosyltransferase and is capable of transferring both UDP- and GDP-activated glucose to an asparagine residue in the conserved Asn-X-Ser/Thr (where X ≠ Pro) sequon. UDP-Gal and UDP-Xyl can also be recognized by ApNGT, but substituted hexoses (GlcA, GlcNAc, GalNAc and Neu5Ac) are not tolerated (7). More recently, ApNGT has been employed for synthesis of polypeptides bearing defined complex oligosaccharides by coupling with an endoglicosidase (ENGase)-catalyzed transglycosylation. Because the Glc-Asn linkage is resistant to peptide-N-glycosidase (PNGase) cleavage, it was proposed to be able to improve the in vivo stability of glycopeptides/proteins (11).

As we know, bacterial OSTs recognize a more specific sequon that is extended by a negatively charged amino acid (Asp or Glu) at the –2 position relative to the Asn ((D/E)Y)NX(S/T); Y and X ≠ Pro) (12, 13). In contrast, ApNGT is

This work was supported by Natural Science Grants 21372130 and 81302682. The authors declare that they have no conflicts of interest with the contents of this article.

This article contains supplemental Table S1 and Figs. S1–S6.

1 Both authors contributed equally to this work.

2 To whom correspondence may be addressed: E-mail: pwang@nankai.edu.cn.

3 To whom correspondence may be addressed. E-mail: jiansongcheng@nankai.edu.cn.

4 The abbreviations used are: OST, oligosaccharyltransferase; NGT, N-glycosyltransferase; ApNGT, A. pleuropneumoniae NGT; XcOGT, X. campestris OGT; DMF, N,N-dimethylformamide; DIC, N,N′-diisopropylcarbodiimide.
capable of modifying polypeptides featuring sequences of eukaryotic N-glycoproteins despite no orthologous structure shared between ApNGT and OSTs, which highlights the hypothesis of convergent evolution of these two N-glycosylation systems (7). Heterogeneous proteins such as human erythropoietin (hEPO) and cholera toxin subunit B (CtxB) have been reported to be glycosylated when co-expressed with ApNGT, suggesting a novel route for glycoengineering in bacteria (14). In addition, the introduction of N-glycosylation sequons to increase carbohydrate content has also proved to be an effective strategy for improving properties of protein therapeutics (15, 16). Therefore, systematic investigation of position-specific residue preference within and surrounding the Asn- X-Ser/Thr sequon would facilitate the application of ApNGT in glycoprotein production. In this study we investigated the preference feature of ApNGT by screening three panels of synthetic peptide substrates and found a critical residue (Gln-469) being closely associated with the stringency of ApNGT. Site-saturated mutagenesis at this site gave rise to a series of valuable mutants with significantly enhanced glycosylation efficiency and expanded acceptor specificity. The most efficient mutant (Q469A) was successfully used for the preparation of homogeneous glycoprotein carrying multiple (up to 10) N-glycans, whereas the wild-type ApNGT produced heterogeneous glycoprotein.

Results

Investigation of position-specific residue preference of ApNGT

The residue preference within and surrounding the N-glycosylation sequon (position −1, +1, and +3) was investigated by screening three panels of synthetic peptide substrates in a systematic manner. For each position, at least five representative amino acids containing basic, acidic, hydroxyl, small hydrophobic, or aromatic side chains were evaluated. As shown in Fig. 1, A and B, both acidic residues and basic residues are disfavored for ApNGT at either the −1 or +1 position, whereas basic residues are particularly preferred at the +3 position. Also, residues with small side chains are relatively preferred at positions −1 and +1. Unexpectedly, the peptide with Gly at +1 position nearly abolished glycosylation. In light of the investigation, a shortest optimal sequence motif recognized by ApNGT would be given as “ANVT.K.” Further investigation showed that only the peptide optimized at all three positions (containing the optimal motif) could achieve the maximum glycosylation (supplemental Fig. S1).

Gln-469 contributes to acceptor specificity

Structural alignment of ApNGT and XcOGT, which share the closely related GT-B domains, showed that an extended loop-connecting sheet β10 and helix α19 (L19a19) sticks out toward UDP complexed with ApNGT (Fig. 2A). Residues Gln-469 within L19a19 and His-277 are structurally adjacent to each other and form the narrow “neck” of the acceptor binding funnel. Because the pair of residues locates above the UDP-binding site, the narrow entryway formed by them might block the access of amino acids with bulky side chains into the active site and thus plays a role in determining the residue preference surrounding the Asn of N-glycosylation sequons. His-277, which is conserved across all GT41 family members (classification according to CAZY database; Ref. 17), has been proposed to be involved in catalysis. The replacement of His-277 with Ala resulted in loss of ~95% catalytic activity (8). However, the mutation of Gln-469 to Ala led to an ~100-fold increase in the turnover rate constant ($k_{cat}$) and an ~162-fold increase in the catalytic efficiency ($k_{cat}/K_m$) against a peptide GGFLNTY-TIER (supplemental Table S1). We next explored the residue preference feature of Q469A in the same manner as done for the wild-type ApNGT previously. As shown in Fig. 1, the mutant Q469A and the wild-type ApNGT show a similar preference toward charged residues at all three positions, whereas the former exhibits better tolerance at both the −1 and +1 positions for residues harboring the bulky side chain, such as Leu, Phe, and the acylated Lys at the −1 position as well as Ile and Leu at the +1 position. In addition, the mutant Q469A exhibits better tolerance toward the Gln-X-Ser/Thr sequon (supplemental Fig. S2, A and B), which has been reported to be glycosylated by the wild-type ApNGT at an extremely low rate (>15,000-fold reduction in turnover rate as compared with Asn-X-Ser/Thr; Ref. 7).
Engineering a cytoplasmic N-glycosyltransferase

Site-saturated mutagenesis

To fully understand the role of Gln-469 in catalysis and determination of acceptor specificity, site-saturated mutagenesis was performed at this critical site. Finally, 17 mutants were successfully constructed, except two mutants (Q469M and Q469R) that were not expressed in *Escherichia coli*. The relative activity of the variants was assessed against a synthetic peptide containing the optimal sequence ANVTK. The residue preference around the N-glycosylation site for each mutant was also determined. As shown in Fig. 5, the apparent difference of preference at the +3 position (Table 1) revealed that the variant Q469A catalyzed much more productive glycosylation as expected. However, uniform glycoform could not be produced by both isoforms no matter by raising the ratio of enzyme to acceptor protein or elongating the reaction time.

Protein glycosylation

To compare the catalytic efficiency of the wild-type ApNGT and the mutant Q469A, which is the most efficient N-glycosyltransferase obtained in this study (Fig. 3), in protein glycosylation, a soluble fragment corresponding to amino acids 1205–1536 of a mature HMW1 adhesin (HMW1ct) was employed as a protein acceptor. HMW1 mediates adherence to respiratory epithelial cells and undergoes heavy N-glycosylation at 31 different sites (18). HMW1ct carries 12 Asn-Ser/Thr sequons (Fig. 4A), and 3 of them (AANVT_{P1}LNTT_{P2,3}GT and NINATSG_{P3}) have been identified as the modification sites of ApNGT (19). We first assessed whether the wild-type ApNGT and the mutant Q469A are capable of modifying all the potential N-glycosylation sites. As shown in Fig. 5, A–D, the mutant Q469A catalyzed much more productive glycosylation as expected. However, uniform glycoform could not be produced by both isoforms no matter by raising the ratio of enzyme to acceptor protein or elongating the reaction time.

To find out which sites are incompetent, site-directed mutagenesis was used to generate 12 mutants from a previously synthesized template that carries 12 inactive sequons with the

---

**Table 1**

The effect of substitutions at Gln-469 on the preference profile of ApNGT

The same amount of enzymes (1.4 \( \mu \)l) was used for all assays. Three sets of mixed peptides used were (GG\([G/A/S/V/L/K/F]\)NLTYTIER, GGS\([G/A/S/P/V/L/D/K/F]\)TYTIER and GGSNLT\([A/S/D/K/R/Y]\)TIER). --, products were not detected.

| Mutants    | −1 | +1 | +3 |
|------------|----|----|----|
| Wild type  | -- | Val| -- |
| Q469G      | Gly Ala Ser Val Leu Phe | Val Leu Phe | Ala Ser Lys Arg Tyr |
| Q469A      | Gly Ala Ser Val Leu Phe | Ala Ser Val Leu Phe | Ala Ser Lys Arg Tyr |
| Q469V      | Ala Ser Leu | Ala Ser Val Leu Phe | Ala Ser Lys Arg Tyr |
| Q469I      | -- | Ala Val Phe | Lys |
| Q469L      | -- | Gly Ala Ser Val Phe | -- |
| Q469S      | Ala Phe | Val | Lys |
| Q469N      | Phe | Val | Lys |
| Q469T      | -- | Val | -- |
| Q469H      | -- | Val | -- |
| Q469C      | -- | Val | -- |
| Q469E      | -- | Val | -- |
| Q469F      | -- | Val | -- |
| Q469Y      | -- | Val | -- |
| Q469D      | -- | -- | -- |
| Q469K      | -- | -- | -- |
| Q469W      | -- | -- | -- |
| Q469P      | -- | -- | -- |
Asn substituted by Ala. Each mutant bears a single Asn-X-Ser/Thr sequon. Finally, we detected five incompetent sequons (EANVTSAP1, QVNLSAQP4, NANGSGSP10, RVNITGDP11, and EPNNTITP12). As shown in Fig. 4B, mutant acceptors carrying P1, P4, P11, and P12 sequons were not glycosylated by the mutant Q469A at all, although that carrying P10 was partially glycosylated. Because charged residues are found in P1 (Glu at the 2 position) and P11 (Arg and Asp at the 2 and 4 positions, respectively) sites, the residue preference at these two distal positions (−2 and +4) was further explored in a multipeptide one-pot manner as described above. As shown in the supplemental Fig. S3, significant inhibition of charged, aromatic, and bulky uncharged residues was not observed at −2 and +4 positions for the Q469A mutant. According to this discovery, the inefficiency of P1 (EANVTS) and P11 (RVNITGD) might lead from some undesired secondary structures. By contrast, proline, and asparagine and glycine are disfavored at the −1 and +1 positions, respectively, by both the mutant Q469A and the wild-type enzyme (Fig. 1, A and B). Also, the threonine has been reported to be much preferred (>10-fold) by ApNGT, as opposed to serine, in position +2 (7). Thus, the potential N-glycosylation sites of P4 (QVNLSAQ), P10 (NANGSGS), and P12 (EPNNTIT) could be further optimized to enhance the glycosylation level.

After the five incompetent sites (P1, P4, P10, P11, and P12) were eliminated by converting the Asn of Asn-X-Ser/Thr sequons to Ala, the Q469A-catalyzed glycosylation yielded a uniform glycoprotein as shown in Fig. 5E. We next optimized the sequons of P4, P10, and P12 by replacing five disfavored residues with the preferred ones (QVNLTAQ, NANYTGS, and EANVITIT) in which and P1 and P11 remained inactive. Ultimately, a homogeneous glycoprotein carrying up to 10 N-glucoses was produced by using the mutant Q469A (Fig. 5F). In contrast, the wild-type ApNGT could not fulfill the uniform glycosylation for proteins bearing multiple N-glycosylation sequons. In comparison with Q469A, several
mutants, such as Q469G and Q469V, also exhibited analogous efficiency against HMW1ct beyond relatively flexible acceptor selectivity (supplemental Fig. S4 and Table 1).

**Donor specificity of the mutant Q469A**

Although Gln-469 structurally positions close to the UDP moiety and has been reported to make hydrogen-bond interactions with one phosphate group of UDP (via a water molecule) (8), the mutant Q469A possesses quite similar donor specificity to the wild-type ApNGT (active to both UDP-Glc and UDP-Gal but not their acetylated derivatives; data not shown).

**Residue preference feature of known N-glycoproteins in the PDB database**

Based on the amino acid composition analysis of a set of 2250 non-redundant N-glycan sites from 1433 glycoproteins in the Protein Data Bank crystallographic database (20), charged residues were found in quite low abundance at the positions adjacent to the occupied Asn (−1 and +1), although a strong preference for any single residue was not observed (Fig. 6). Because the mutant Q469A also dislikes sequons harboring charged residues at −1 and +1 positions, the most known N-glycan sites could be well-tolerated by Q469A.

**Discussion**

In this study the systematic investigation of position-specific residue preference within and around the N-glycan sequon indicated that ApNGT has strict specificity for residues especially at the −1 and +1 positions. The optimal N-glycosylation motif recognized by ApNGT (ANYTK) is unexpectedly similar to a previously reported one (PYNVTK) preferred by an archaeal OST from *Pyrococcus furiosus* particularly at positions +1 and +3 (21). Because the NGT involved N-glycosylation system has been proposed to evolved independently (22), it remains mysterious how ApNGT and the archaeal OST have such a similar residue preference feature. More interestingly, NGTs from either β- or γ-proteobacteria seem not to evolve independently of each other, suggesting that horizontal gene transfer might occur between the two classes (supplemental Fig. S5).

Gln-469, whose side chain sways into the UDP-binding pocket and forms the narrow neck of the acceptor-binding funnel with His-277, was herein established to be involved in determining the acceptor specificity of enzyme. The introduction of small hydrophobic residues at Gln-469, such as Gly, Ala, and Val, not only weakened the stringency of ApNGT but also contributed to an enormous improvement of glycosylation efficiency against both peptides and proteins. A plausible hypothesis is that the substitutions can facilitate the accommodation of peptides/proteins into the active site and thereby enhanced the turnover rate constant. The deduction can also be supported by the observation that a very short peptide (SNLTY) was well-glycosylated by the mutant Q469A instead of the wild-type ApNGT (supplemental Fig. S2, C and D). On the other hand, it is not difficult to understand why most ApNGT variants including the wild-type enzyme and even Q469F and Q469Y can well accept a peptide (GGANVTKTIER) containing the optimal motif ANVTK with small residues Ala and Val at the −1 and +1 positions, respectively (Fig. 3). In contrast to bacterial OSTs, which regulate acceptor specificity by requiring an additional acidic residue at −2 position, ApNGT evolved
Figure 6. Amino acid preferences in occupied N-glycan sites. 1433 glycoproteins containing N-linked GlcNAc were retrieved from the PDB database by using a web-based glycan searching tool (Glycan Modeling and Stimulation). Total 2250 non-redundant N-glycan sites derived from these glycoproteins were identified by using pdb2linucs (39) and used to generate the sequence logo with WebLogo (40). Neighboring residues located downstream (positions +1 to +5) and upstream (positions −1 to −3) from the occupied asparagine residue (position 0) are shown. The size of each letter represents the residue prevalence at the putative position.

(related strictly) selectivity by recruiting a critical residue (Gln-469) within the catalytic site. In addition, the introduction of variant residues at Gln-469 of ApNGT is structurally feasible, but exchange of the loop (L_{B1019}) with the corresponding fragment of XcOGT resulted in a completely inactive protein (data not shown), suggesting that the unique loop might be required for substrate binding and/or structure maintaining.

Further inspection of the potential acceptor-binding site of ApNGT revealed that residues of Phe-39 and His-272 introduce bulky bumps in the van der Waals surface of the funnel-shaped groove. Although the “bump-and-hole” strategies have been successfully applied for retuning the selectivity of proteins and protein-ligand engineering (23–25), replacement of them with a smaller alanine was previously demonstrated to decrease the enzyme activity to various extents (8).

Naturally occurring N-glycoproteins usually exhibit glycoform heterogeneity with respect to N-glycan sequon occupancy (macroheterogeneity) and glycan structure (microheterogeneity) (26–28). In recent years great efforts have been made for preparation of glycan-defined glycoproteins including in vivo glycoengineering and in vitro chemical or chemoenzymatic synthesis (28–34). Additionally, introducing extra N-linked glycosylation sites into therapeutic proteins, such as erythropoietin, has also been proven to improve their serum half-life and therapeutic efficiency (35).

In this study the engineered Q469A (the most efficient mutant) has been proven to efficiently modify proteins with multiple N-glycosylation sites and, therefore, presents a promising biocatalyst for potentially addressing the issue of glycoprotein macroheterogeneity. Meanwhile, sequon-optimization was also demonstrated to be a useful strategy for the generation of uniform glycoprotein. Moreover, the relatively extended acceptor selectivity of the mutant Q469A would allow fewer alterations toward acceptor proteins for reaching uniform glycosylation.

On the other hand, inaccessible Asn-X-Ser/Thr sequon or nearby exceptional secondary structures will obviously inhibit the glycosylation. For example, incompetent P1 and P11 sites probably resulted from structural elements because no disfavored residues are found at −1, +1, and +3 positions. Therefore, known the three-dimensional structure would greatly benefit the N-glycosylation site engineering either for modifying an existing Asn-X-Ser/Thr sequon or inserting a new one.

Taken together, Gln-469, which is conserved in NGTs of subgroup 2 (supplemental Fig. S6), was demonstrated to be a critical amino acid for ApNGT to maintain its stringency. Introduction of small hydrophobic residues at the site has created a series of efficient mutants with expanded selectivity on Asn-X-Ser/Thr sequon. Because signals that control the efficiency of N-linked glycosylation at individual Asn residues have not been fully defined, efficient N-glycosylation machines, such as the mutant Q469A, are always valuable for either in vivo or in vitro glycosylation methods. Additionally, this is the first report that homogeneous glycoprotein carrying multiple (up to 10) N-glycans were in vitro enzymatically prepared.

Experimental procedures

Materials

UDP-Glc and UDP-Gal used were purchased from Sigma. UDP-GlcNAc and UDP-GalNAc were enzymatically synthesized as previously described (36).

Cloning and mutant construction

apNGT was cloned from genomic DNA of A. pleuropneumoniae ATCC 27088D (purchased from ATCC, Manassas, VA) and then cloned into pET15b. hMW1ct and its isoform bearing 12 inactive Asn-X-Ser/Thr sequons with the Asn substituted by Ala were codon-optimized (for the E. coli expression system) and synthesized by Genevix (Suzhou, China). Mutagenesis was performed using a fast site-directed mutagenesis kit (TransGen, Beijing, China). All mutations were confirmed by DNA sequencing.

Expression and purification of recombinant proteins

The E. coli BL21 (DE3) strains, transformed with the corresponding expression plasmids, were cultured in LB medium supplemented with ampicillin (100 µg ml⁻¹). Overexpression of the target protein was achieved by inducing the E. coli culture...
Engineering a cytoplasmic N-glycosyltransferase

with 0.1 mM isopropyl-β-D-thiogalactopyranoside (IPTG) when the A_{600} reached 0.8 followed by incubating at 18 °C for 20 h. The cells were harvested by centrifugation at 6000 × g for 10 min, resuspended in 100 mM Tris-HCl buffer (pH 8.0), and broken by sonication. After centrifugation (13,523 × g, 20 min) of the cell lysate, the supernatant was applied onto a Ni^{2+}-nitrilotriacetic acid affinity column (Qiagen, Hilden, Germany). The column was pre-equilibrated with 8 column volumes of binding buffer (5 mM imidazole, 0.5 M NaCl, and 20 mM Tris-HCl (pH 8.0)) before the lysate was loaded. After washing with 8 column volumes of washing buffer (50 mM imidazole, 0.5 M NaCl, and 20 mM Tris-HCl (pH 8.0)), the target protein was eluted with elution buffer (200 mM imidazole, 0.5 M NaCl, and 20 mM Tris-HCl (pH 8.0)). Elution fractions were analyzed by 12% SDS-PAGE and Coomassie staining. The concentration of purified enzymes was obtained using a BCA Protein Assay kit (Pierce) with bovine serum albumin as a protein standard.

Peptide synthesis

Wang resin (GL Biochem, Shanghai, China) (1 g) was swelled with DCM (10 ml) for 30 min. l-Arg (l-type arginine) (1 mmol eq), DIC (1 mmol), and 4-dimethylaminopyridine (1 mmol) in DCM (10 ml) were added to the resin and bubbled with N₂ for 3 h. Then pyridine (1 ml) and acetic anhydride (1 ml) were added and removed after 30 min. The resin was washed with DCM and DMF. The protecting group Fmoc was removed subsequently by 20% piperidine in DMF (monitored by a ninhydrin test). The second amino acid L-Glu (3 mmol), 1-hydroxybenzo-triazole (3 mmol), and DIC (3 mmol) were added to the resin and bubbled with N₂ for 1 h. The resin was washed with DCM and DMF (monitored by the ninhydrin test; no blue appeared), whereupon the remaining amino acids were added sequentially repeating the above steps. Acp (6-aminocaproic acid) (3 mmol), 1-hydroxybenzotriazole (3 mmol), and DIC (3 mmol) in DMF were added and bubbled with N₂ for 1 h. Then FITC (3 mmol) and ethylidissopropylamine (3 mmol) in DMF were added and bubbled with N₂ for 4 h (monitored by the ninhydrin test; no blue appeared). At last, 95% TFA in water was added and oscillated for 2 h to give the crude peptide product. Purification by preparative HPLC afforded target compound (purity ≥98%) following by lyophilization to give a pale yellow solid powder ultimately.

Enzymatic assays for glycosylation of peptides and proteins

For systematic investigation of position-specific residue preference within and surrounding the N-glycan sequon, assays were performed in a total volume of 30 μl containing 1 mM concentrations of various peptides, 20 mM UDP-Glc, 9.2 μM ApNGT (wild type), or a 0.4 μM concentration of mutant Q469A and buffered with 50 mM Tris-HCl (pH 8.0) at 37 °C for 45 min. Reactions were then stopped by the addition of 20 μl of glycine stop solution (pH 10.6) and analyzed by reverse phase HPLC (RP-HPLC). Phenomenex Aegis PEPTIDE 3.6 μm XB-C18, 250 × 4.6-mm column, gradient elution from 14–42% acetonitrile over 30 min). Glycosylated peptides were further identified by mass spectrometry, which was performed on an Autoflex III TOF/TOF200 mass spectrometer (Bruker Daltonics, Leipzig, Germany). Apparent kinetic parameters were obtained by varying the concentration of FITC-labeled peptide from 5 to 80 μM and keeping the UDP-glucose concentration fixed (20 mM). The FITC-labeled peptide/glycopeptide was detected by excitation at 495 nm and emission at 520 nm. All assays were carried out in duplicate, and the data were fitted into the Michaelis-Menten equation using Grafit 6.0 (Erithacus Software Ltd). Multipeptide one-pot assays were carried out in 50 mM Tris-HCl (pH 8.0) containing a 1 mM mixture of various peptides, 20 mM UDP-Glc, and 1.4 μM enzyme at 37 °C for 45 min. The glycosylated product profile was analyzed by MALDI-TOF mass spectrometry. 10 mg ml⁻¹ α-cyano-4-hydroxy-cin-namic acid (Sigma) was used as the matrix and premixed with samples in a ratio of 1:1 (v/v).

The assays for glycosylation of proteins were performed in a total volume of 200 μl containing 12.8 μM various HMW1c derivatives, 5 mM UDP-Glc, appropriate amounts of ApNGT (WT), or the mutant Q469A and buffered with 50 mM Tris-HCl (pH 8.0). Glycosylated products were analyzed by mass spectrometry, which was performed on an AutoflexIII LRF200-CID mass spectrometer (Bruker Daltonics). Samples were concentrated using centrifugal ultrafiltration devices (Amicon Ultra 0.5 ml Centrifugal Filters, Merck-Millipore). 10 mg ml⁻¹ sian-pinic acid (Sigma) was used as the matrix and premixed with samples in a ratio of 1:1 (v/v).

Author contributions—Q. S. and Z. W. conducted most of the experiments and analyzed the results. Y. F., W. S., and P. Z. performed and analyzed the experiments shown in Fig. 4. L. W., F. W., and Y. X. conducted the experiments shown in Fig. 6. P. G. W. and J. C. conceived the idea for the project and wrote the paper.

Acknowledgment—We acknowledge Y. Liang for assistance in mass spectrometry analysis.

References

1. Dell, A., Galadari, A., Sastre, F., and Hitchen, P. (2010) Similarities and differences in the glycosylation mechanisms in prokaryotes and eu-karyotes. Int. J. Microbiol. 2010, 148178
2. Abu-Qarn, M., Eichler, J., and Sharon, N. (2008) Not just for Eukarya anymore: protein glycosylation in bacteria and archaea. Curr. Opin. Struct. Biol. 18, 544–550
3. Burda, P., and Aebi, M. (1999) The dolichol pathway of N-linked glycosylation. Biochem. Biophys. Acta 1426, 239–257
4. Grass, S., Lichti, C. F., Townsend, R. R., Gross, J., and St Geme, J. W., 3rd. (2010) The Haemophilus influenzae HMW1C protein is a glycosyltransferase that transfers hexose residues to asparagine sites in the HMW1 adhesin. PLoS Pathog. 6, e1000919
5. McCann, J. R., and St Geme, J. W., 3rd. (2014) The HMW1C-like glycosyltransferases: an enzyme family with a sweet tooth for simple sugars. PLoS Pathog. 10, e1003977
6. Fleckenstein, J. M., Roy, K., Fischer, J. F., and Burkitt, M. (2006) Identification of a two-partner secretion locus of enterotoxigenic Escherichia coli. Infect. Immun. 74, 2245–2258
7. Naegeli, A., Michaud, G., Schubert, M., Lin, C. W., Lizak, C., Darbre, T., Reymond, I. L., and Aebi, M. (2014) Substrate specificity of cytoplasmic N-glycosyltransferase. J. Biol. Chem. 289, 24521–24532
8. Kawai, F., Grass, S., Kim, Y., Choi, K. J., St Geme, J. W., 3rd, and Yeo, H. J. (2011) Structural insights into the glycosyltransferase activity of the Acti-nobacillus pleuropneumoniae HMW1C-like protein. J. Biol. Chem. 286, 38546–38557
9. Martinez-Fléites, C., Macauley, M. S., He, Y., Shen, D. L., Vocadlo, D. J., and Davies, G. J. (2008) Structure of an O-GlcNAc transferase homolog
Engineering a cytoplasmic N-glycosyltransferase

provides insight into intracellular glycosylation. Nat. Struct. Mol. Biol. 15, 764–765
10. Clarke, A. J., Hurtado-Guerrero, R., Pathak, S., Schüttelkopf, A. W., Borodkin, V., Shepherd, S. M., Ibrahim, A. F., and van Aalten, D. M. (2008) Structural insights into mechanism and specificity of O-GlcNAc transferase. EMBO J. 27, 2780–2788
11. Lomino, J. V., Naegeli, A., Orwenyo, J., Amin, M. N., Aebi, M., and Wang, L. X. (2013) A two-step enzymatic glycosylation of polypeptides with complex N-glycans. Bioorg. Med. Chem. 21, 2262–2270
12. Kowarik, M., Young, N. M., Numao, S., Schulz, B. L., Hug, I., Callewaert, N., Mills, D. C., Watson, D. C., Hernandez, M., Kelly, J. F., Wacker, M., and Aebi, M. (2006) Definition of the bacterial N-glycosylation site consensus sequence. EMBO J. 25, 1957–1966
13. Gerber, S., Lizak, C., Michaud, G., Bucher, M., Darbre, T., Aebi, M., Raymond, J. L., and Locher, K. P. (2013) Mechanism of bacterial oligosaccharidase: in vitro quantification of sequon binding and catalysis. J. Biol. Chem. 288, 8849–8861
14. Naegeli, A., Neupert, C., Fan, Y. Y., Lin, C. W., Poljak, K., Papini, A. M., Schwarz, F., and Aebi, M. (2014) Molecular analysis of an alternative N-glycosylation machinery by functional transfer from Actinobacillus pleuropneumoniae to Escherichia coli. J. Biol. Chem. 289, 2170–2179
15. Sinclair, A. M., and Elliott, S. (2005) Glycoengineering: the effect of glycosylation on the properties of therapeutic proteins. J. Pharm. Sci. 94, 1626–1635
16. Mazola, Y., Chinea, G., and Musacchio, A. (2011) Integrating bioinformatics tools to handle glycosylation. PLoS Comput. Biol. 7, e1002285
17. Cantarel, B. L., Coutinho, P. M., Rancurel, C., Bernard, T., Lombard, V., and Henrissat, B. (2009) The carbohydrate-active EnZymes database (CAZy): an expert resource for glycogenomics. Nucleic Acids Res. 37, D233–D238
18. Gross, J., Grass, S., Davis, A. E., Gilmore-Erdmann, P., Townsend, R. R., and St Gene, J. W., 3rd. (2008) The Haemophilus influenzae HMW1 adhesin is a glycoprotein with an unusual N-linked carbohydrate modification. J. Biol. Chem. 283, 26010–26015
19. Choi, K. J., Grass, S., Paek, S., St Gene J. W., 3rd, Yeo, H. J. (2010) The Actinobacillus pleuropneumoniae HMW1C-like glycosyltransferase mediates N-linked glycosylation of the Haemophilus influenzae HMW1 adhesin. PLoS ONE 5, e15888
20. Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N., and Bourne, P. E. (2000) The Protein Data Bank. Nucleic Acids Res. 28, 235–242
21. Iguira, M., and Kohda, D. (2011) Quantitative assessment of the preferences for the amino acid residues flanking archaeal N-linked glycosylation sites. Glycobiology 21, 575–583
22. Schwarz, F., and Aebi, M. (2011) Mechanisms of N-linked protein glycosylation. Curr. Opin. Struct. Biol. 21, 576–582
23. Baud, M. G., Lin-Shiao, E., Cardote, T., Tallant, C., Pschibul, A., Chan, K. H., Zengerle, M., Garcia, J. R., Kwan, T. T., Ferguson, F. M., and Cuili, A. (2014) Chemical biology: a bump-and-hole approach to engineer controlled selectivity of BET bromodomain chemical probes. Science 346, 638–641
24. Clackson, T., Yang, W., Rozamus, L. W., Hatada, M., Amara, J. F., Rollins, C. T., Stevenson, L. F., Magari, S. R., Wood, S. A., Courage, N. L., Lu, X., Cerasoli, F., Jr, Gilman, M., and Holt, D. A. (1998) Redesigning an FKBP-ligand interface to generate chemical dimersizers with novel specificity. Proc. Natl. Acad. Sci. U.S.A. 95, 10437–10442
25. Doerr, A. (2015) A bumpy, holey method to probe proteins. Nat. Methods 12, 14
26. Spahn, P. N., and Lewis, N. E. (2014) Systems glycobiology for glycoengineering. Curr. Opin. Biotechnol. 30, 218–224
27. Rudd, P. M., and Dwek, R. A. (1997) Glycosylation: heterogeneity and the 3D structure of proteins. Crit. Rev. Biochem. Mol. Biol. 32, 1–100
28. Rich, J. R., and Withers, S. G. (2009) Emerging methods for the production of homogeneous human glycoproteins. Nat. Chem. Biol. 5, 206–215
29. Grogan, M. J., Pratt, M. R., Marcaurelle, L. A., and Bertozzi, C. R. (2002) Homogeneous glycopeptides and glycoproteins for biological investigation. Annu. Rev. Biochem. 71, 593–634
30. Wang, P., Dong, S., Shieh, J. H., Peguero, E., Hendrickson, R., Moore, M. A., and Danishefsky, S. J. (2013) Erythropoietin derived by chemical synthesis. Science 342, 1357–1360
31. Yang, Z., Wang, S., Halim, A., Schulz, M. A., Frodin, M., Rahman, S. H., Vester-Christensen, M. B., Behrens, C., Kristensen, C., Vakhрушев, S. Y., Bennett, E. P., Wandall, H. H., and Clausen, H. (2015) Engineered CHO cells for production of diverse, homogeneous glycoproteins. Nat. Biotechnol. 33, 842–844
32. Wu, Z., Jiang, K., Zhu, H., Ma, C., Yu, Z., Li, L., Guan, W., Liu, Y., Zhu, H., Chen, Y., Li, S., Li, J., Cheng, J., Zhang, L., and Wang, P. G. (2016) Site-directed glycosylation of peptide/protein with homogeneous O-linked eukaryotic N-glycans. Bioconjug. Chem. 27, 1972–1975
33. Izumi, M., Makimura, Y., Dedola, S., Seko, A., Kanamori, A., Sakono, M., Ito, Y., and Kajihara, Y. (2012) Chemical synthesis of intentionally misfolded homogeneous glycoprotein: a unique approach for the study of glycoprotein quality control. J. Am. Chem. Soc. 134, 7238–7241
34. Schwarz, F., Huang, W., Li, C., Schulz, B. L., Lizak, C., Palumbo, A., Numao, S., Neri, D., Aebi, M., and Wang, L. X. (2010) A combined method for producing homogeneous glycoproteins with eukaryotic N-glycosylation. Nat. Chem. Biol. 6, 264–266
35. Elliott, S., Ergie, J., Browne, J., Lorenzini, T., Busse, L., Rogers, N., and Ponting, I. (2004) Control of hHuEPO biological activity: the role of carbohydrate. Exp. Hematol. 32, 1146–1155
36. Muthana, M. M., Qu, J., Li, Y., Zhang, L., Yu, H., Ding, L., Malekan, H., and Chen, X. (2012) Efficient one-pot multienzyme synthesis of UDP-sugars using a promiscuous UDP-sugar pyrophosphorylase from Bifidobacterium longum (BLUSP). Chem. Commun. (Camb) 48, 2728–2730
37. Beavis, R. C., and Chait, B. T. (1990) High-accuracy molecular mass determination of proteins using matrix-assisted laser desorption mass spectrometry. Anal. Chem. 62, 1836–1840
38. Beavis, R. C., and Chait, B. T. (1989) Cinnamic acid derivatives as matrices for ultraviolet laser desorption mass spectrometry of proteins. Rapid Commun. Mass Spectrom. 3, 432–435
39. Lütteke, T., Frank, M., and von der Lieth, C. W. (2004) Data mining the protein data bank: automatic detection and assignment of carbohydrate. Nat. Chem. Biol. 1357–1360
40. Crooks, G. E., Hon, G., Chandonia, J. M., and Brenner, S. E. (2004) WebLogo: a sequence logo generator. Genome Res. 14, 1188–1190