

RATIONAL DESIGN OF COMPONENTS OF THE OLIGO-SACCHARYLTRANSFERASE-CATALYSED ASPARAGINE-LINKED GLYCOSYLATION

Field of the invention

The present invention is directed to methods for identifying or designing (a) a potential oligosaccharide donor, (b) a potential oligosaccharyltransferase (OST), (c) a potential consensus sequence motif polypeptide, and/or (d) a potential glycosylation inhibitor for use in the oligosaccharyltransferase (OST)-catalysed asparagine-linked ("N-linked") glycosylation, comprising the steps of generating a three-dimensional model of the catalytic domain and/or the polypeptide binding site of the oligosaccharyltransferase (OST) of *Campylobacter lari*, and designing or selecting a potential component selected from (a) to (d) which optimizes the stereochemical complementarity of said three-dimensional model(s) and the potential component.

Relevant background of the invention

It is estimated that more than half of all eukaryotic proteins are glycoproteins, which implies that specific amino acid side chains are chemically modified with carbohydrates. The most abundant form of these modifications is asparagine-linked ("N-linked") glycosylation, which affects a multitude of cellular functions that range from protein folding, quality control, sorting and secretion to organism development and host-pathogen interactions. Asparagines facing the lumen of the endoplasmic reticulum (ER) are specifically glycosylated when located in the consensus "sequon" Asn-X-Ser/Thr, where X may be any amino acid except proline. The reaction takes place at the membrane surrounding the ER and is catalysed by the enzyme oligosaccharyltransferase (OST), a hetero-oligomeric protein complex embedded in the ER membrane of higher eukaryotes (see Fig. 1b). A hallmark of N-linked glycosylation is its broad specificity with respect to the polypeptide substrate, which is a direct consequence of the short recognition sequon. This characteristic distinguishes OST from glycosyltransferases that modify serine or threonine residues (O-linked glycosylation) and exhibit a higher specificity for their protein substrates.

The key step in OST-catalysed glycosylation is the formation of an N-glycosidic linkage between the amide nitrogen of an acceptor asparagine and the C1 carbon of the first saccharide moiety of a lipid-linked oligosaccharide (LLO) donor (see Fig. 1a). This results in the *en bloc* transfer of the oligosaccharide onto the acceptor asparagine. Details of the underlying reaction mechanism are poorly understood. This is due to the absence of structural insight into OST at high resolution, but also to the complex chemical nature of the LLO substrate, its low abundance in biological samples, and its insolubility in water. In contrast, crystal structures of various soluble O-glycosyltransferases have been published and their reaction mechanisms were investigated in great detail. For OST, the currently accepted model suggests that glycosylation sequons are recognized when located in unfolded protein segments, which can occur during protein translocation into the ER or after translocation is completed. The central catalytically active component within OST is the STT3 subunit, whereas the other subunits are thought to assist and refine the process by facilitating OST complex assembly or by interacting with a subset of acceptor proteins or the LLO substrate, leading to an increased number of accessible and modified glycosylation sites.

N-linked glycosylation is not restricted to eukaryotes. Homologous processes are found in archaea and in defined taxa of proteobacteria. However, prokaryotes and eukaryotic kinetoplastids contain a single-subunit OST enzyme that is homologous to the STT3 subunit of higher eukaryotes. The best-studied prokaryotic N-glycosylation process is mediated by the protein glycosylation locus *pgl* from the bacterium *Campylobacter jejuni* (Szymanski et al. (1999) Molecular Microbiology 32, 1022–1030). The locus contains an integral membrane protein termed PglB that shares significant sequence similarity with eukaryotic STT3, suggesting a common membrane topology and reaction mechanism (see Fig. 1b). This gene cluster is sufficient for catalyzing protein glycosylation when transferred into *Escherichia coli* cells. OST-catalysed prokaryotic protein glycosylation of sequon-containing protein substrates is an economic, effective and convenient way of glycosylating recombinantly produced proteins (Wacker et al. (2002), Science 298, 1790–1793). N-linked protein glycosylation can be engineered with diverse O antigen lipopolysaccharide structures of non-*C. jejuni* origin in *E. coli* (Feldman et al. (2005) PNAS 102(8), 3016–3021), thus allowing for the prokaryotic transfer of eukaryotic N-glycans to recombinant protein substrates. Glover et al. (2005, Chemistry & Biology 12, 1311–1315) demonstrated for the first time the *in vitro* protein glycosylation using *E. coli* cell membranes comprising overexpressed PglB and an undecaprenyl pyrophosphate bound oligosaccharide. In 2006 Kowarik et al. (2006, EMBO J. 25(9), 1957–1966) further

defined the bacterial N-glycosylation site consensus sequence by showing that the substrate specificity of bacterial OST is extended to a negatively charged amino acid in the -2 position of the acceptor asparagine, resulting in the consensus sequon Asp/Glu-X₁-Asn-X₂-Ser/Thr (wherein X₁ and X₂ are both not proline; SEQ ID NO: 3). By using a peptide substrate library, Chen et al. (2007, Biochemistry 46, 5579 -5585) confirmed the necessity for a negative charge in the -2 position of the acceptor asparagine and identified the sequence DQNAT (SEQ ID NO: 4) as the optimal substrate for *C. jejuni* PglB.

From the above it follows that the prokaryotic oligosaccharyltransferase (OST) has a broad, because small sequon-based specificity for protein substrates and can be used to transfer eukaryotic, prokaryotic as well as synthetic N-glycans. Essentially the prokaryotic OST-based N-glycosylation system requires three components, (a) an oligosaccharide donor, preferably a lipid- or undecaprenyl pyrophosphate-linked oligosaccharide donor, (b) a prokaryotic oligosaccharyltransferase (OST), (c) a potential consensus sequence motif polypeptide substrate and last but not least a suitable physiological microenvironment, e.g. cell membranes *in vitro* or *in vivo*.

The problem underlying the present information is that one cannot predict or design components essential for the very versatile prokaryotic OST-based N-glycosylation system beyond the information given for already known OST-components. In addition, there is no clue what structural requirements a potential OST glycosylation inhibitor must have. Such inhibitors would be expected to have pronounced biological effects and could be of great medical, diagnostic and scientific value. The other problem is that up to now it had not been possible to provide a three-dimensional model of the catalytic domain and the polypeptide binding site of an OST that could have provided the scientific community with insight regarding the possible variation of the components involved in OST-mediated glycosylation.

The above problems have been solved by the provision of the three-dimensional X-ray structure of a bacterial OST, the PglB protein from *Campylobacter lari* (sharing 56% sequence identity with PglB of the *C. jejuni*) in complex with the acceptor hexapeptide DQNATF (SEQ ID NO: 5). *C. lari* PglB is active when co-expressed with the *C. jejuni* *pgl* cluster in *E. coli* cells, as evidenced by glycosylation of an acceptor protein containing a consensus sequon (see Fig. 2). For its structural analysis, *C. lari* PglB was co-crystallized with the hexapeptide DQNATF (SEQ ID NO: 5) that contains the sequon

glycosylated in the *in vivo* assay, which had been identified as the optimal acceptor sequence for *C. jejuni* PglB (see Chen et al. above). The structure of *C. lari* PglB (712 amino acid residues, SEQ ID NO: 1) was determined using a combination of experimental phasing and molecular replacement, making use of the previously determined structure of the periplasmic domain of *C. jejuni* PglB. Co-crystals of PglB were small, fragile, and diffracted X-rays anisotropically, with the best native data extending to 3.4 Å resolution. The structure was refined to R/R_{free} values of 23.8 and 27.1 %, respectively (Table 2). Further details of the structure of *C. lari* PglB are provided in the experimental section below.

This new three-dimensional structure provides insight into the molecular basis for sequon recognition and reveals a catalytic site that is formed by the transmembrane domain of the protein and features conserved, acidic side chain residues and a bound divalent cation. These results suggest for the first time a mechanism for amide nitrogen activation and glycosylation and provide a reasoned approach for identifying and designing new oligosaccharide donors, new oligosaccharyltransferase variants (OST), new consensus sequence motif polypeptides, as well as OST glycosylation inhibitors, all of which have utility in recombinant glycoprotein production, diagnostics, medicine and as scientific tools.

In view of the above, a first aspect of the invention relates to a method for identifying a potential component for the oligosaccharyltransferase (OST)-catalysed asparagine-linked ("N-linked") glycosylation selected from the group consisting of

- (a) a potential oligosaccharide donor, preferably a lipid-linked oligosaccharide (LLO) or an undecaprenyl pyrophosphate bound oligosaccharide donor,
- (b) a potential oligosaccharyltransferase (OST),
- (c) a potential consensus sequence motif polypeptide, and
- (d) a potential glycosylation inhibitor,

comprising the steps of

- (i) using the atomic coordinates of Table 1, preferably ± 2 , more preferably ± 1.5 , most preferably ± 1.0 Å root mean square deviation (rmsd) from the backbone atoms, for generating a three-dimensional model of the catalytic domain of the oligosaccharyltransferase (OST) of *Campylobacter lari*, comprising at least one, two, three, four, five, six, seven, most preferably all amino acids D56, R147, D154, D156, E319, R375, Y468, and H485, and/or, preferably and

- (ii) using the atomic coordinates of Table 1, preferably ± 2 , more preferably ± 1.5 , most preferably ± 1.0 Å root mean square deviation (rmsd) from the backbone atoms, for generating a three-dimensional model of the polypeptide binding site of the oligosaccharyltransferase (OST) of *Campylobacter lari*, comprising at least one, two, three, four, five, most preferably all of amino acids M318, R331, W463, W464, D465, and I572,
- (iii) preferably performing whole body translations and/or rotations on the coordinates of the amino acids of the three-dimensional models of (i) and/or (ii),
- (iv) using said three-dimensional model(s) of (i), (ii) and/or (iii) for designing or selecting at least one of potential components (a) to (d),
- (v) providing at least one of said potential components (a) to (d), and
- (vi) contacting at least one of said potential components (a) to (d) with the further functional components necessary for an oligosaccharyltransferase (OST)-catalysed asparagine-linked ("N-linked") glycosylation,
- (vii) identifying a functional component selected from the group consisting of
 - (A) a functional oligosaccharide donor, preferably a functional lipid-linked oligosaccharide (LLO) donor or an undecaprenyl pyrophosphate bound oligosaccharide donor,
 - (B) a functional oligosaccharyltransferase (OST),
 - (C) a functional consensus sequence motif polypeptide, and
 - (D) a functional glycosylation inhibitor.

In a preferred embodiment, in step (ii) the three-dimensional model of the polypeptide binding site of the oligosaccharyltransferase (OST) of *Campylobacter lari* comprises at least two, preferably at least three, more preferably at least four, most preferably all of amino acids M318, A331, W463, W464, D465 and I572.

The atomic coordinates of Table 1 for use in the methods of the present invention are shown in Fig. 7. The X-ray coordinates of the OST of *C. lari*, in particular of the catalytic site and/or the polypeptide binding site complexed with the optimized polypeptide substrate DQNAT (SEQ ID NO: 4) provide the skilled person with the three-dimensional information necessary for identifying a potential component for OST catalysis and catalytic inhibition. The spatial restrictions in combination with the chemical functional nature of the individual atoms of the amino acids involved in catalytic action and polypeptide binding, e.g. electron densities, position of van der Waals forces, ionic interactions, hydrophobic interactions, etc., inform the person skilled in computer-

assisted molecular modelling of the structural and spatial prerequisites of (a) a potential oligosaccharide donor, preferably a lipid-linked oligosaccharide (LLO) or an undecaprenyl pyrophosphate bound oligosaccharide donor; (b) a potential oligosaccharyltransferase (OST), (c) a potential consensus sequence motif polypeptide, and/or (d) a potential glycosylation inhibitor.

As described before bacterial OSTs have a broad specificity for oligosaccharide donor molecules. With the coordinate information and the method of the invention, the repertoire of useful oligosaccharide donors can be rationally designed and extended without having to revert to trial and error synthetic strategies. Also, the OST itself can be rationally varied without rendering the catalytic site and the polypeptide binding pocket non-functional. This OST variation is useful, for example for modifying the catalytic potential, polypeptide substrate specificity and/or oligosaccharide donor specificity of OSTs. Moreover, the consensus motif of the polypeptide oligosaccharide acceptor can be rationally varied and designed, thus leading to the broadening of the OST utility such as e.g. glycosylation of eukaryotic sites. Last but not least, the inventive three-dimensional X-ray model provides an excellent basis for designing potential glycosylation inhibitors, which can be expected to be physiologically active by interrupting, modifying or slowing OST activity. These inhibitors have a great potential for providing scientific, diagnostic and therapeutic tools.

The term "root mean square deviation" or "rms deviation" or "rmsd" means the square root of the arithmetic mean of the square of the deviations from the mean. In the context of atomic objects the numbers are given in angstroms (Å). It is a way of expressing the deviation or variation from a trend or object.

The method of the invention comprises the step of using the atomic coordinates of Table 1, preferably ± 2 , more preferably ± 1.5 , most preferably ± 1.0 Å root mean square deviation (rmsd) from the backbone atoms, for generating a three-dimensional model of the catalytic domain of the oligosaccharyltransferase (OST) of *Campylobacter lari*, comprising at least one, two or three, preferably at least four, more preferably at least five or six, most preferably seven or all of amino acids D56, R147, D154, D156, E319, R375, Y468, and H485, and/or, preferably and using the atomic coordinates of Table 1, preferably ± 2 , more preferably ± 1.5 , most preferably ± 1.0 Å root mean square deviation (rmsd) from the backbone atoms, for generating a three-dimensional model of the polypeptide binding site of the oligosac-

charyltransferase (OST) of *Campylobacter lari*, comprising at least one or two, preferably at least three, more preferably at least four or five, most preferably all of amino acids M318, R331 (or A331), W463, W464, D465, and I572.

At the polypeptide binding site W463, W464 and D465 form strong hydrogen bonds to the β -hydroxyl group of T in the bound acceptor sequon. These residues strongly contribute to acceptor sequon binding and are the reason why an S or T is located at the +2 position of the acceptor asparagine (N-X-**S/T**). R331 forms a salt bridge to the negatively charged D of the acceptor polypeptide and therefore contributes to acceptor sequon binding. R331 is responsible for the requirement of a negatively charged amino acid in the -2 position of the acceptor asparagine and is responsible for the extension of the consensus sequon for bacterial N-linked protein glycosylation (**D/E**-X₁-N-X₂-**S/T**; SEQ ID NO: 3) (Fig. 4a). R331 is only conserved in bacterial OSTs and can serve as a target for changing the substrate specificity of PglB towards the recognition of eukaryotic glycosylation sites. In fact, when R331 is mutated to A, the resulting glycosylation site AQNAT (SEQ ID NO: 8) in an acceptor protein, e.g. modified scFv fragment 3D5 originally containing the sequon DQNAT (SEQ ID NO: 4), can be glycosylated, whereas this site does not serve as a substrate for the wild type enzyme (Fig. 4b).

At the catalytic site, D56, D154 and E319 seem responsible for the coordination of the bound divalent metal ion, which is essential for catalysis. More importantly, D56 and E319 form hydrogen bonds to the amido group of the asparagine side chain of the bound acceptor sequon. This interaction causes a rotation of the C-N bond in the amido group, which is important for the nucleophilic activation of the nitrogen. D156 and R147 stabilize the hydrogen bonding network and R375 complexes the negative charge of one of the phosphates of a bound lipid-linked oligosaccharide (LLO).

For example, in the polypeptide binding site W463, W464 and D456 could be replaced by two H and an E (WWD $\rightarrow$ HHE). To keep the requirement of the negative charge in the -2 position of the acceptor asparagine, R331 could be replaced by K. To create the requirement of a positive charge in this position, R331 could be replaced by D or E. To overcome the requirement of the negative charge in the -2 position of the acceptor asparagine, R331 can be replaced by A (Fig. 4). At the catalytic site D154 and D156 can be replaced by E each. E319 could be replaced by D and D56 could be replaced by E, respectively (see below). To modify PglB activity, for example, D56, D154 and E319 could be replaced by alanines or the corresponding amino function (N or Q). H485 could be replaced by W as W appears in eukaryotic OSTs at this position.

The structure coordinates of PglB from *C. lari* as listed in Table 1 in Fig. 7 are accessible for download to the skilled person from the pdb database. (The research collaboratory for structural bioinformatics (RCSB) Protein Database (PDB)). With the help of a computer and freely accessible structure programs such as PyMOL or commercially available structure programs the skilled person can easily generate three-dimensional models useful for the claimed method.

Those of skill in the art will understand that a set of structure coordinates for a protein, protein/substrate or protein/inhibitor complex or a portion thereof is a relative set of points that defines a shape in three dimensions. Thus, it is possible that an entirely different set of coordinates could define a similar or identical shape. For this reason it is preferred to perform whole body translations and/or rotations on the coordinates of the amino acids of the three-dimensional models (i) and/or (ii) obtained from the atomic coordinates of Table 1. These variations in coordinates may be generated by mathematical manipulations of the structure coordinates, for example manipulation by crystallographic permutations of the structure coordinates, fractionalization or matrix operations to sets of structure coordinates or any combination of the above.

Next, the above generated three-dimensional models are used for designing or selecting at least one of the potential components of the OST glycosylation or a potential inhibitor. Various computational analyses are necessary to determine whether a molecule such as a specific oligosaccharide donor, a modified OST, an oligosaccharide acceptor consensus sequence motif polypeptide or a glycosylation inhibitor is sufficiently designed to rationally predict its functionality in the OST-catalysed reaction. Spatial, functional and chemical considerations such as the nature, position of atoms, rotational degree of freedom, electron density, steric hindrance, van der Waals, ionic and hydrophobic interactions, etc. will have to be made. Such analysis may be carried out conveniently on computers by standard software applications such as CCP4 (COLLABORATIVE COMPUTATIONAL PROJECT, NUMBER 4. 1994. "The CCP4 Suite: Programs for Protein Crystallography". Acta Cryst. D50, 760-763.)

Current computational molecular similarity applications permit comparisons between different structures, different conformations of the same structure and different parts of the same structure. The comparison procedure is typically divided into four steps: (1) loading the structural information, (2) defining atom equivalence on these structures, (3) performing a fitting (superimposition) operation and (4) analysis of the results. Each

structure is identified by a name. One structure is then identified as the potential OST component or OST inhibitor (i.e. the target or fixed structure), all remaining structures are working structures (i.e. moving structures). When a rigid fitting method is used the working structure is translated and rotated to obtain an optimum fit (spatial and functional complementarity) with the target structure, e.g. an OST inhibitor is the fixed structure and the OST amino acids are translated and rotated to obtain the optimum fit. The fitting operation uses an algorithm that computes the optimum translation and rotation to be applied to the moving structure, such that the rmsd of the fit over the specified pairs of equivalent atom is an absolute minimum. After superimposition of the two structures an rmsd value can be calculated for specific sets of equivalent atoms.

The potential functional component or inhibitor of the OST reaction selected or designed according to the inventive method as described above will provide the skilled person with a reasonable expectation of success when verifying its functionality in a routine OST activity assay. For said purpose the potential component must be provided by purchase, modification of purchased materials, chemical and/or recombinant synthesis, etc. This potential component will then have to be contacted with the further functional components necessary for an OST-catalysed asparagine-linked ("N-linked") glycosylation, of course under conditions that allow for OST activity. Preferred OST activity assays are described in (1) 2005 Chemistry & Biology 12, 1311-1315, (2) 2006 Science 314, 1148-1150, (3) 2007 Biochemistry 46, 5579-5585, (4) 2007 Glycobiology 11, 1175-1182 and (5) 2011 Glycobiology 5, 575-583. Whether or not the potential components or inhibitor is OST active is preferably verified in comparison to positive or negative standards. For example, functionality of the OST assay is established with a known oligosaccharide acceptor polypeptide, e.g. the hexapeptide DQNATF (SEQ ID NO: 5), and then the potential functional consensus sequence motif polypeptide is substituted for the hexapeptide and glycosylation of the substitute polypeptide is determined. This simple OST assay system can be adapted for identifying any functional OST components, preferably one selected from the group consisting of (A) a functional oligosaccharide donor, preferably a functional lipid-linked oligosaccharide (LLO) or an undecaprenyl pyrophosphate bound oligosaccharide donor, (B) a functional oligosaccharyltransferase (OST), (C) a functional consensus sequence motif polypeptide, and (D) a functional glycosylation inhibitor.

In a second aspect, the present invention relates to a method for designing a potential component for the oligosaccharyltransferase (OST)-catalysed asparagine-linked ("N-

linked”) glycosylation selected from the group consisting of (a) a potential oligosaccharide donor, preferably a lipid-linked oligosaccharide (LLO) or an undecaprenyl pyrophosphate bound oligosaccharide donor, (b) a potential oligosaccharyltransferase (OST), (c) a potential consensus sequence motif polypeptide, and (d) a potential glycosylation inhibitor, comprising the steps of

- (i) using the atomic coordinates of Table 1, preferably ± 2 , more preferably ± 1.5 , most preferably ± 1.0 Å root mean square deviation (rmsd) from the backbone atoms, for generating a three-dimensional model of the catalytic domain of the oligosaccharyltransferase (OST) of *Campylobacter lari*, comprising at least one, two or three, preferably at least four or five, more preferably at least six or seven, most preferably all amino acids D56, R147, D154, D156, E319, R375, Y468, and H485, and/or, preferably and
- (ii) using the atomic coordinates of Table 1, preferably $+2$, more preferably $+1.5$, most preferably $+1.0$ Å root mean square deviation (rmsd) from the backbone atoms, for generating a three-dimensional model of the polypeptide binding site of the oligosaccharyltransferase (OST) of *Campylobacter lari*, comprising at least one or two, preferably at least three, more preferably at least four or five, most preferably all of amino acids M318, R331, W463, W464, D465, and I572,
- (iii) preferably performing whole body translations and/or rotations on the coordinates of the amino acids of the three-dimensional models of (i) and/or (ii),
- (iii.1) using said three-dimensional model of (i), (ii) and/or (iii) for assessing the stereochemical complementarity between said three-dimensional model(s) (i), (ii) and/or (iii) and a known or potential component for the oligosaccharyltransferase (OST)-catalysed asparagine-linked (“N-linked”) glycosylation selected from an oligosaccharide donor, preferably a lipid-linked oligosaccharide (LLO) donor or an undecaprenyl pyrophosphate bound oligosaccharide donor, a consensus sequence motif polypeptide, and a potential glycosylation inhibitor, or
- (iii.2) varying at least one amino acid in said three-dimensional model of (i), (ii) and/or (iii) and using said varied three-dimensional model of (i), (ii) and/or (iii) for assessing the stereochemical complementarity between said three-dimensional models (i), (ii) and/or (iii) and a known or potential component for the oligosaccharyltransferase (OST)-catalysed asparagine-linked (“N-linked”) glycosylation selected from an oligosaccharide (LLO) donor, preferably a lipid-linked oligosaccharide (LLO) donor or an undecaprenyl pyrophosphate bound oligosaccharide donor, a consensus sequence motif polypeptide, and a potential glycosylation inhibitor,

- (iv) optimizing said stereochemical complementarity in an iterative approach by observing changes in the three-dimensional model of (iii.1), (iii.2) or the component for the oligosaccharyltransferase (OST)-catalysed asparagine-linked ("N-linked") glycosylation,
- (v) designing a potential component selected from (a) to (d) which optimizes said stereochemical complementarity of said three-dimensional model(s) and potential component,
- (vi.1) optionally providing the optimized potential component, and
- (vi.2) contacting at least one of said potential components (a) to (d) with the further functional components necessary for an oligosaccharyltransferase (OST)-catalysed asparagine-linked ("N-linked") glycosylation,
- (vi.3) identifying a functional component selected from the group consisting of (A) a functional oligosaccharide donor, preferably a functional lipid-linked oligosaccharide (LLO) or an undecaprenyl pyrophosphate bound oligosaccharide donor, (B) a functional oligosaccharyltransferase (OST), (C) a functional consensus sequence motif polypeptide, and (D) a functional glycosylation inhibitor.

In a preferred embodiment, in step (ii) the three-dimensional model of the polypeptide binding site of the oligosaccharyltransferase (OST) of *Campylobacter lari* comprises at least two, preferably at least three, more preferably at least four, most preferably all of amino acids M318, A331, W463, W464, D465 and I572.

This method is basically very similar to the method of the first aspect except that in the method directly above the potential OST component is designed by optimizing its stereochemical complementarity to the three-dimensional models with or without whole body translations and rotations in an iterative approach by observing changes in the three-dimensional models or the component for the oligosaccharyltransferase (OST)-catalysed asparagine-linked ("N-linked") glycosylation, when varying at least one amino acid in at least one of said three-dimensional models.

Once the designed potential OST component is selected based on its optimized stereochemical complementarity to said three-dimensional model(s) it can optionally be verified in an OST assay, preferably by providing the optimized potential component (by chemical and/or recombinant synthesis, purchase, modification of known compounds, etc.), contacting said optimized potential component with the further functional components necessary for an oligosaccharyltransferase (OST)-catalysed asparagine-linked

("N-linked") glycosylation. In a last optional step the functional component of the OST reaction or an inhibitor thereof is identified by its impact on the OST reaction. Typically, positive and negative reference components are used to verify OST assay activity.

In a preferred embodiment of the methods of the present invention for identifying or designing potential OST components, the specific three-dimensional catalytic site model of step (i) further comprises one or more, preferably at least 5, more preferably at least 10, most preferably all of the amino acids selected from the group having residues located within Van der Waals distance to the bound peptide of SEQ ID NO: 2, preferably selected from those within a distance of 5 Å to said peptide, more preferably selected from the group consisting of T53, T54, N55, D56, N146, R147, Y152, E315, T316, I317, M318, E319, V320, N321, R331, L374, R375, Y433, S435, V438, W463, W464, D465, G482, H485, I572, V575.

In a further aspect the present invention relates to a machine-readable medium comprising, e.g. storing

- (i) the atomic coordinates of Table 1, preferably ± 2 , more preferably ± 1.5 , most preferably ± 1.0 Å root mean square deviation (rmsd) from the backbone atoms, preferably comprising at least one, two or three, preferably at least four or five, more preferably at least six or 7, most preferably all of amino acids D56, R147, D154, D156, E319, R375, Y468, and H485, and/or, preferably and
- (ii) the atomic coordinates of Table 1, preferably ± 2 , more preferably ± 1.5 , most preferably ± 1.0 Å root mean square deviation (rmsd) from the backbone atoms, for generating a three-dimensional model of the polypeptide binding site of the oligosaccharyltransferase (OST) of *Campylobacter lari*, comprising at least one or two, preferably at least three, more preferably at least four or five, most preferably all of amino acids M318, R331 (or A331), W463, W464, D465, and I572,
- (iii) preferably the atomic coordinates of (i) or (ii) modified by performing whole body translations and/or rotations on said coordinates.

The above medium is particularly useful for a variety of purposes such as computer-assisted drug design, drug discovery and the X-ray crystallographic analysis of OSTs from other bacteria.

In the following the present invention will be further illustrated with reference to specific embodiments and experiments which are not intended to be interpreted as limiting the scope of the invention as presented by the appended claims.

SEQ ID NO: 1 lists the 712 amino acids of the oligosaccharyltransferase (OST) of *C. lari* PglB.

MELQQNFTDNNISIKYTCILILIAFAFVLCRLYWVAWASEFYEFFFNNDQLMITTNDGYAFAEGAR
 DMIAGFHQPNDLSYFGSSSLSTLTWLYSILPFSFESIILYMSTFFASLIVVPIILIAREYKLTTY
 GFIAALLGSIANSYNNRTMSGYYDTDMLVLVLPMLILLTFIRLTINKDIFTLLLSPIFIMIYLWW
 YPSSYSLNFMIGLFLGLYTLVFHRKEKIFYLAIALMIIALSMLAWQYKLALIVLLFAIFAFKEEK
 INFYMIWALIFISISILHLSGGLDPVLYQLKFYVFKASDVQNLKDAAFMYFNVNETIMEVNTIDP
 EVFMQRISSSVLVFILSFIGFILLCKDHKSMLLALPMLALGFMALRAGLRFTIYAVPVMALGFGY
 FLYAFFNFLEKKQIKLSLRNKNILLILIAFFSISPALMHIYYYKSSTVFTSYEASILNDLKNKAQ
 REDYVVAWWDYGYPIRYYSVDKTLIDGGKHLGKDNFFSSFVLSKEQIPAANMARLSVEYTEKSEK
 ENYPDVLKAMVKDYNQTSKDFLESNDKNFKFDTNKTRDVYIYMPYRMLRIMPVVAQFANTNPD
 NGEQEKSLLFFSQANAI AQDKTTGSVMDNGVEIINDFRALKVEGASIPLKAFVDIESITNGKFYY
 NEIDSKAQIYLLFLREYKSFVILDESLYNSAYIQMFLNQQYDQDLFEQVTNDTRAKIYRLKR

SEQ ID NO: 2 lists the amino acids of the hexapeptide DQNATF{pNO₂} (where F{pNO₂} is paranitro-phenylalanine) representing the optimised oligosaccharide acceptor substrate for the OST. This hexapeptide was crystallized together with the OST of *C. lari* PglB to give the atomic structure coordinates of Table 1 below, the statistics of which are provided in Table 2.

Table 1 is shown in Fig. 7. It lists the atomic structure coordinates of the crystallized oligosaccharyltransferase (OST, chain A) of *Campylobacter lari* complexed with peptide sequence DQNATF{pNO₂} (SEQ ID NO: 2) (chain B), the optimal substrate for OST glycosylation and a bound divalent metal ion (chain C), useful for generating a three-dimensional model of the catalytic domain of the oligosaccharyltransferase (OST) of *Campylobacter lari*. The table contains following information:

COLUMNS	CONTENTS
1-6	Atom
7-11	Atom serial number
13-16	Atom name
17	Alternate location indicator
18-20	Residue name
22	Chain identifier
23-26	Residue sequence number
27	Code for insertion of residues
31-38	Orthogonal coordinates for X in Angstroms
39-46	Orthogonal coordinates for Y in Angstroms

47-54	Orthogonal coordinates for Z in Angstroms
55-60	Occupancy
61-66	Temperature factor (Default = 0.0)
73-76	Segment identifier, left-justified
77-78	Element symbol, right-justified
79-80	Charge on the atom

Table 2 lists the X-ray data collection and refinement statistics for Table 1 in Fig. 7.

A. Data collection statistics

Data set	Native	EMP1	EMP2	EMP3
Beamline / detector	MD2 at SLS S06SA/PX1 (Mar225)	HighRes at SLS S06SA/PX1 (Pilatus)	MD2 at SLS S06SA/PX1 (Mar225)	MD2 at SLS S06SA/PX1 (Mar225)
Software	XDS/HKL	XDS	HKL	HKL
Wavelength (Å)	1.0	1.0	1.0	1.0
Space group	P2 ₁ 2 ₁ 2 ₁	P2 ₁ 2 ₁ 2 ₁	P2 ₁ 2 ₁ 2 ₁	P2 ₁ 2 ₁ 2 ₁
Unit cell: a (Å)	85.06	85.5	86.1	87.8
b (Å)	116.1	116.4	117.0	119.4
c (Å)	175.04	175.2	174.8	169.9
Resolution	30-3.4	30-4.45	30-3.8	30 – 4.2
Crystal positions collected	12	3		4
Completeness (%)	99.3% (97.6%)	99.6% (100%)	92.3 (68.3)	96.9 (84.2)
Redundancy	9.6 (8.7)	11.1	9.2 (7.3)	9.2 (7.2)
<I/σ(I)>	13.2 (1.3)	10.8 (2.25)	11.1 (0.8)	13.2 (2.6)
R _{merged} -F (XDS) (%)	13.9 (132.8)	15.2 (86.3)		
R _{merged} (HKL) (%)			13.3 ^a	16 (50.7)

^a No R_{merged} factors indicated by HKL due to severe anisotropy

B. refinement statistics (native data)

Resolution (Å)	30 – 3.4
No. of reflections working set (test set)	21834 (2000)
R _{work} / R _{free} (%)	23.8 / 27.1
rmsd from ideality	
bond lengths (Å)	0.011
bond angles (°)	1.475
Average B factor (Å ²)	
PglB	129
Peptide	117
Ramachandran analysis (Molprobit)	
Ramachandran favored	82.6 %
Ramachandran outliers	1.5 %

Figures

Fig. 1 shows schematically the process of N-linked protein glycosylation. **a.** Proteins containing acceptor sequons (NxT/S in eukaryotes, D/Ex₁Nx₂ T/S (SEQ ID NO: 3) in bacteria) are glycosylated at the asparagine residue by the action of OST. The chemical reaction includes an activation of the amide nitrogen and the formation of a glycosidic bond. **b.** Similarities and distinct characteristics of N-linked glycosylation in bacteria (left panel) and eukaryotes (right panel). The central enzyme of OST is STT3 (PglB in bacteria), which is conserved in sequence and structure in all domains of life. The lipid moieties and the oligosaccharide are similar but non-identical in bacteria and eukaryotes, as indicated. Both contain an isoprenoid-pyrophosphate moiety that is hydrolysed after the glycosylation step.

Fig. 2 shows the activity of *C. lari* oligosaccharyltransferase PglB. *E. coli* cells were transformed with a combination of three separate plasmids: (i) The glycosylation machinery of *C. jejuni* that generates LLO, but contains inactivated PglB; (ii) An acceptor protein (modified scFv fragment 3D5) containing the sequon DQNAT (SEQ ID NO: 4) and (iii) Functional *C. lari* PglB. Note that the co-crystal structure presented here contains *C. lari* PglB bound to an acceptor peptide containing the sequence used in the assay. Glycosylation of the acceptor protein was analyzed from periplasmic extracts, whereas expression of PglB was analyzed from whole cell extracts. Proteins were analyzed by immunoblot using anti-c-Myc antibody detecting scFv substrate (top), glycan-specific antiserum hR6 (middle), or anti-HA antiserum detecting PglB (bottom). PglB constructs are indicated above the lanes: Vector control (ev), wild type (wt), or mutations in single letter code. Glycosylation of 3D5 results in a mobility shift from the unmodified form (g0) to the glycosylated form (g1). Functional PglB is partially auto-glycosylated at N535 and N556, resulting in two additional bands (g1 and g2). All experiments were performed at least in triplicate, and representative samples are shown.

Fig. 3 is a picture showing the structure and substrate binding cavities of *C. lari* PglB. **a.** Ribbon diagram of PglB in two orientations, with TM and periplasmic domains. The presumed position of the membrane is indicated by lines and the location of the bound substrate peptide is shown. **b.** Line representation of PglB with bound acceptor peptide. Two cavities are present on opposite sides of the protein (as marked), providing access for the substrates. The cavities are connected by a porthole that accommodates the side chain of the acceptor asparagine (not visible in this presentation mode).

Fig. 4 shows glycosylation sequon recognition by PglB. **a.** A picture showing the peptide binding site and sequon recognition in PglB. The acceptor peptide is shown in line format and individual amino acids are labeled in three letter code. PglB residues contributing to specific sequon binding are in ball and stick format and are labeled in single letter code including amino acid numbers. Hydrogen bonds between the WWD motif and the +2 Thr of the acceptor peptide are indicated by dashed lines. **b.** Activity of *C. lari* oligosaccharyltransferase PglB and PglB mutant R331A against different glycosylation sequons: *E. coli* cells were transformed with a combination of three separate plasmids: (i) The glycosylation machinery of *C. jejuni* that generates LLO, but contains inactivated PglB; (ii) An acceptor protein (modified scFv fragment 3D5) containing either the sequon DQNAT (SEQ ID NO: 4) or the sequon AQNAT (SEQ ID NO: 8) and (iii) Functional *C. lari* PglB or PglB mutant R331A. Glycosylation of the acceptor protein was analyzed from periplasmic extracts, whereas expression of PglB was analyzed from whole cell extracts. Proteins were analyzed by immunoblot using anti-c-Myc antibody detecting scFv substrate (top), glycan-specific antiserum hR6 (middle), or anti-HA antiserum detecting PglB (bottom). PglB constructs are indicated above the lanes and acceptor sequons are indicated above the lines. Glycosylation of 3D5 results in a mobility shift from the unmodified form (g0) to the glycosylated form (g1). Functional PglB is partially auto-glycosylated at N535 and N556, resulting in two additional bands (g1 and g2).

Fig. 5 shows the catalytic site and activation of amide nitrogen. **a.** Selected side chains of PglB are in ball and stick format and are labeled in single letter code including amino acid numbers. The acceptor peptide is shown in line format and the activated asparagine is labeled. The catalytic divalent cation is shown as a sphere. Dashed lines indicate hydrogen bonds or ligand-metal interactions as suggested by distance measurements. **b.** Chemical structure of the catalytic site, indicating interactions as in **a.** Thin dashed lines indicate protein and peptide backbones. **c.** Presumed mechanism of activation of the amido group nitrogen. Free asparagine side chain features delocalization/conjugation of the electron pair of the nitrogen, as indicated by resonance formulae. When bound to PglB, the amido group of the acceptor asparagine may form hydrogen bonds with the carboxyl groups of the catalytically essential D56 and E319, requiring rotation around the C-N bond (arrow). This would result in breaking of the conjugation and an increased nucleophilicity of the nitrogen.

Fig. 6: Proposed glycosylation mechanism. a. Surfaces of TM and periplasmic domains of PglB are shown. Bound acceptor peptide is in ball and stick format, and black

lines indicate the N- and C-termini. The chemical structure of bacterial LLO is shown schematically in white to highlight the presumed interactions of the pyrophosphate group with the divalent cation (M^{2+}) and with the conserved R375, while allowing for a collinear arrangement of the attacking and leaving groups of the nucleophilic substitution. An arrow indicates the attack of the activated amide nitrogen. A predominantly hydrophobic groove is indicated on the surface of PglB, where the isoprenoid moieties are expected to enter the lipid bilayer. **b.** Proposed three step mechanism of PglB-catalyzed glycosylation. The molecular events leading from one step to the next are indicated next to the arrows. The observed crystal structure reflects the top state, with the acceptor peptide bound to the protein and the C-terminal half of the external loop EL5 ordered. The bottom left state reflects the ground state, with no substrates bound and with the external loop EL5 disordered, indicated by dashed lines. In the bottom right state, LLO from *C. jejuni* (black line for isoprenoid moieties, P for phosphate, ellipsoids for saccharide moieties) is bound and the acceptor asparagine is glycosylated.

Fig. 7: see comment on Table 1 above.

Examples

Methods

In vivo complementation

To analyze the activity of PglB from *C. lari* *in vivo* the gene encoding pglB was amplified from the *pgl* gene cluster of *Campylobacter lari* isolate (sample kindly provided by H. Hächler, Luzern, Switzerland) genomic DNA by polymerase chain reaction (PCR) and was cloned into a pMLBAD plasmid (Lefebvre & Valvano, *Appl. Environ. Microb.* **68**, 5956-5964 (2002) with a C-terminal HA-tag fused to PglB, resulting in the plasmid pMIK71. For complementation studies pMIK71 or pMLBAD empty vector were transformed into *E. coli* SCM6 cells carrying the plasmids pCL21 (2011 Bioconjug Chem. **3**, 488-496) or pCL64 and pACYC*pgl*_{mut} (Wacker et al., *Science* **298**, 1790-1793 (2002). pCL21 encodes for the expression of the single-chain Fv fragment of 3D5 carrying a DQNAT (SEQ ID NO: 4) glycosylation site in the linker region and a C-terminal Myc-tag fused to 3D5. In pCL64 the DQNAT site of pCL21 was replaced by an AQNAT (SEQ ID NO: 8) glycosylation site. pACYC*pgl*_{mut} encodes for the biosynthesis of the *C. jejuni* lipid-linked oligosaccharide (LLO) with an inactivated *C. jejuni* pglB gene (W458A and D459A). A 5 mL pre-culture was inoculated from a single clone and grown over night at 37°C in LB medium. The

main culture was inoculated to an optical density (A_{600}) of 0.05 in 15 mL LB medium and grown at 37°C to A_{600} of 0.5. The culture was induced by addition of arabinose to 0.1 % (w/v) and grown for 4 h at 24°C. For extraction of periplasmic proteins an equivalent of 1 mL culture volume with an A_{600} of 3.0 was harvested by centrifugation, resuspended in 150 μ L extraction buffer, consisting of 30 mM Tris-HCl, pH 8.5; 20 % (w/v) sucrose; 1 mM EDTA and 1 mg/mL lysozyme (Sigma) and incubated for 1 h at 4 °C. A final centrifugation step yielded periplasmic proteins in the supernatant. Glycosylation of 3D5 and expression of PglB were analyzed by SDS–PAGE (performed according to Lämmli). Immunodetection was performed with anti-c-Myc monoclonal antibody (Calbiochem) and anti-glycan serum hR6 (Amber S. and Aebi M., personal communication) to observe glycosylated 3D5. Immunodetection of *C. lari* PglB was performed with anti-HA antiserum (Santa Cruz).

Mutagenesis study

Mutant PglB was generated by QuickChange method. Plasmid pCL64 was generated by ligation of phosphorylated, double stranded DNA of oligonucleotides CTAGCGGTGGTGGTGGTTCTGGTGGTGGTGGCCAGAACGCCA and CCGGTGGCGTTCTGGGCACCACCACCAGAACCCACCACCG into the plasmid pCL21 digested with *NheI* and *AgeI*. The resulting plasmids of all constructs were validated by DNA sequencing. The mutant PglB variants were cloned into pMLBAD as above and used in complementation assays.

PglB purification

The gene encoding pglB was cloned into a modified pBAD (Invitrogen) expression plasmid with a C-terminal decahistidine affinity tag fused to PglB, resulting in the plasmid pSF2. Due to the applied cloning strategy PglB carried the mutation K2E and the plasmid was confirmed by DNA sequencing (Microsynth). PglB from *C. lari* was overexpressed from pSF2 in *Escherichia coli* BL21-Gold (DE3) (Stratagene) cells in a 30 L fermentor (Infors). Cells were grown at 37 °C in Terrific Broth medium supplemented with 1 % glycerol (w/v) to an optical density (A_{600}) of 10.0 before the culture was induced by the addition of 0.1 % arabinose (w/v) for 2h. All following steps were performed at 4 °C unless specified differently. Cells were harvested by centrifugation, resuspended in 25 mM Tris-HCl, pH 8.0; 250 mM NaCl and disrupted in a M-110L microfluidizer (Microfluidics) at 15.000 psi external pressure. Membranes were pelleted by ultracentrifugation at 100.000 g for 0.5 h. PglB was solubilized in 25 mM Tris-HCl, pH 8.0; 250 mM NaCl; 10 % glycerol (v/v) and 1% N-dodecyl- β -D-maltopyranoside (w/v) (DDM, Anatrace) for 1 h.

All subsequent buffers contained DDM as detergent. The supernatant was supplemented with 25 mM imidazole and loaded onto a NiNTA superflow affinity column (Qiagen), washed with 60 mM imidazole before PglB was eluted with 200 mM imidazole. The protein was desalted into 10 mM MES-NaOH, pH6.5; 100 mM NaCl; 0.5 mM EDTA; 3% glycerol (v/v); 3% polyethylene glycol 400 (v/v) and concentrated to 7-10 mg/mL in an Amicon Ultra-15 concentrator (Millipore) with a molecular weight cut off of 100 kDa.

Native crystals

The peptide Ac-DQNATF{4NO₂}-NH₂ (SEQ ID NO: 6) was added to concentrated PglB to a final concentration of 0.75 mM, incubated for 0.5 h, and crystallized by vapor diffusion in sitting drops at 20 °C against a reservoir of 100 mM glycine, pH 9.4; 50 mM magnesium acetate; 6 % Dimethyl sulfoxide (DMSO) (v/v) and 23-34 % (v/v) polyethylene glycol 400. The protein to reservoir volume ratio in the sitting drop was 2:1. Crystals typically appeared after three to four weeks and matured to full size within six weeks. Crystals were directly flash-frozen by immersion in liquid nitrogen prior to data collection.

Heavy metal derivatives

Native crystals were soaked for 30-60 min in 1 mM ethyl mercury phosphate (EMP) prior to back-soaking and flash-freezing by immersion in liquid nitrogen.

Data collection

Crystals belonged to the space group P2₁2₁2₁, with one PglB-peptide complex in the asymmetric unit. Native data were collected at the micro-diffractometer beamline X06SA at the Swiss Light Source (SLS, Villigen) because not all sections of the crystals diffracted equally well. EMP2 and EMP3 derivative data sets (see Table 2 above) were collected at the same station, whereas EMP1 was collected at the high resolution station of the same beam line. Data were processed and merged with XDS (Kabsch, W. Xds. *Acta Crystallogr D* **66**, 125-132, (2010) or HKL2000 (HKL Research, Inc.).

Structure determination

The structure was determined using a combination of molecular replacement using the periplasmic domain of *C. jejuni* PglB (pdb code 3AAG) as a search model and Phaser (McCoy et al., Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658-674, 2007) on the one hand and multiple isomorphous replacement with anomalous scattering (MIRAS) using SHARP (Global Phasing Limited) on the other. The process of phase calculation and model building (using O; Jones et al., *Acta Crystallogr.* **A47**, 110-119,

1991) and refinement (using Phenix; Adams et al., *Acta Crystallogr D* **66**, 213-221, 2010) was iterated, starting with the periplasmic domain and extending into the best-ordered regions of the TM domain (TM1-4 and TM10-13) followed by TM5-9. The locations of three cysteines in the TM domain (indicated by Hg anomalous peaks) served as starting points for tracing until very good density allowed placement of bulky residues, confirming the sequence register. The final structure excludes two disordered loops of PglB (residues 283-306 and residues 605-607) as well as the C-terminal polyhistidine tag. Data collection and refinement statistics are given in Table 2 above.

Results and conclusions

Structure of C. lari PglB

In agreement with earlier predictions (Kelleher & Gilmore, *Glycobiology* **16**, 47-62, 2006) the X-ray structure revealed that PglB consists of two domains (Fig. 3a), a transmembrane (TM) domain comprising residues 1-432 and a periplasmic domain comprising residues 433-712. In addition to the covalent linkage, the two domains interact extensively through non-covalent interactions, primarily provided by the first external loop (EL1) of the TM domain that forms two helices positioned parallel to the membrane plane (EL1-h1 and EL1-h2,). The periplasmic domain reveals a mixed α/β structure that harbors the most conserved sequence motif of the protein family (see below). The structures of two related periplasmic domains, one from *C. jejuni* PglB and another from *Pyrococcus furiosus* AglB, have been reported previously (Maita et al., *J. Biol. Chem.* **285**, 4941-4950, 2010; Igura et al., *Embo J* **27**, 234-243, 2008). However, these isolated domains were catalytically inactive and unable to bind acceptor peptides. The present structure of full-length PglB provides a molecular basis for this observation by revealing that the TM domain is indispensable both for peptide binding and catalysis.

In contrast to the periplasmic domain, the TM domain features a novel fold, with 13 TM segments connected by relatively short cytoplasmic and external (extracytoplasmic) loops, with the notable exception of the long external loops EL1 and EL5. Based on sequence alignments and TM segment predictions, the observed topology appears to be conserved in the STT3 protein family. TM1-4 and TM10-13 form the sequon-binding and catalytic sites and provide the bulk of the interface with the periplasmic domain, whereas TM5-9 are short, parallel TM helices normal to the membrane plane. They may interact with the oligosaccharide or bactoprenol moieties or provide a spacer for the remote attachment of EL5 (see below). Whereas EL1 is well-ordered and firmly interacts with the periplasmic domain, EL5 is only partially ordered, with 25 residues poorly or completely

disordered in the electron density maps. The ordered section of EL5 interacts both with the periplasmic domain and with the bound acceptor peptide, suggesting a crucial role in the recruitment and binding of the sequon. It is conceivable that in the absence of peptide, EL5 is completely flexible and disordered, accounting for the finding that no crystals of PglB grew in the absence of acceptor peptide. In the peptide-bound state, PglB forms two large cavities above the membrane surface, both accessible from the periplasmic space, but located at opposite sides of the protein (Fig. 3). The left-side cavity provides access for acceptor proteins as suggested by the presence of bound peptide in the structure, whereas the right-side cavity harbors the catalytic residues (see below) and probably serves as the binding pocket for the LLO substrate. The two cavities are connected by a small porthole, through which the acceptor asparagine of the bound peptide reaches from the peptide-binding site into the presumed catalytic site.

Acceptor sequon binding and recognition

The structure of PglB was determined in complex with the hexapeptide DQNATF (SEQ ID NO: 5), and clear density for the peptide was observed in a location that placed the acceptor asparagine some 15 Å above the membrane surface (Fig. 3a). The peptide-binding and catalytic sites are the best-ordered regions of the structure, and the electron density allowed an unambiguous assignment of the acceptor peptide register. Almost 80 % of the contact surface (as calculated by areaimol; Bailey, S., *Acta Crystallogr D* **50**, 760-763, 1994) of the peptide is buried at the interface of the TM and periplasmic domains, suggesting a very tight binding along with a firmly imposed conformation. The hexapeptide forms a loop that almost completes a 180° turn; accordingly, polypeptide substrates have to present their glycosylation sequons in sufficiently large, flexible and surface-exposed loops, because the peptide-binding cavity of PglB does not appear to fit fully folded protein domains. A significant part of the interaction of the TM domain with the peptide is provided by the external loop EL5, which also features a methionine residue that resembles a saddle for the peptide. The observed conformation of the peptide would be incompatible with a proline residue at the +1 position, in agreement with the observation that +1 prolines are not allowed in glycosylation sequons. A hallmark of N-linked glycosylation is the requirement of a serine or threonine at the +2 position of the acceptor sequon. The PglB structure provides a molecular explanation by revealing that the β-hydroxyl group of the +2 Thr of the bound peptide forms three hydrogen bonds, one with each of the side chains of the "WWD motif", which is strictly conserved in STT3 proteins (Figs. 4). The WWD motif is located in the periplasmic domain, and the interaction of the two tryptophan and the aspartate side chains saturate

the hydrogen-bonding capacity of the β -hydroxyl group, a functional group that is only present in serines and threonines. The arrangement physically separates the +2 Thr from the acceptor asparagine, and it is assumed that the WWD motif defines the polypeptide substrate specificity, but is not directly involved in the catalytic activity of the enzyme. Notably, the structure can also explain preferences and deviations at the +2 position of glycosylation sequons. The γ -methyl group of the +2 Thr is in Van der Waals contact with the conserved isoleucine I572 of PglB (3.6 Å distance to the γ -methyl group of I572, Fig. 4). This stabilizing interaction is absent if a serine is in the +2 position. This may explain that acceptor sequons containing a +2 Thr are glycosylated 40 times more efficiently than if they contain a +2 serine (Bause, E., *Biochem. Soc. T12*, 514-517, 1984). The structure suggests that the non-natural, S-configured threonine would cause a steric clash with I572. S-configured threonine is indeed not allowed at the +2 position, with a 15'000-fold reduction in glycosylation efficiency compared to R-configured threonine (Breuer et al., *Febs Lett.* **501**, 106-110, 2001). I572 is conserved in bacteria and has been suggested to be part of a MxxI motif (SEQ ID NO: 7; Maita et al., *J. Biol. Chem.* **285**, 4941-4950, 2010). However, the corresponding residue in the archaeal AglB protein was found to be a lysine (Igura et al., *Embo J27*, 234-243, 2008) and sequence alignments with eukaryotic STT3 homologues reveal no clear conservation of I572, suggesting that residues other than isoleucine can provide contacts to the +2 Thr in homologous proteins. The PglB structure can also explain allowed deviations from the consensus sequons: The acceptor sequence N-X-C, present in ~2.2% of experimentally determined glycosylation sites of the mouse glycoproteome (Zielinska et al., *Cell* **141**, 897-907, 2010), is probably allowed because the β -sulfhydryl group of cysteine can form similar hydrogen bonds as a β -hydroxyl group. Glycines, alanines and valines have also been reported at the +2 position of glycosylated sequons, albeit only at low abundance (Zielinska et al., *Cell* **141**, 897-907, 2010; ; Schwarz et al., *Glycobiology* **21**, 45-54, 2011; Valliere-Douglass et al. *J. Biol. Chem.* **284**, 32493-32506, 2009). These residues can in principle be accommodated in the binding pocket of PglB because they are equal in size or smaller than threonine. However, glycosylation of sequons such as N-G-X with X being larger than threonine, or of T/S-X-N ("reverse sequons") (Valliere-Douglass et al. *J. Biol. Chem.* **284**, 32493-32506, 2009) cannot be explained by the structure of PglB. Compared to eukaryotic enzymes bacterial OST have an additional requirement for the acceptor sequon: Glycosylation is only efficient if a negatively charged residue (Asp or Glu) is present at the -2 position, resulting in a consensus sequon D/E-x₁-N-x₂-S/T (SEQ ID NO: 3; Kowarik et al., *Embo. J.* **25**, 1957-1966, 2006). In PglB the arginine residue R331 provides a salt bridge to the -2 Asp of the acceptor peptide (Fig. 4a), thereby

strengthening the PglB-peptide interaction. R331 is conserved in bacteria, but not in eukaryotes, where no requirement for a negative charge at the -2 position is observed. The extended sequon recognition may reflect the need for tighter peptide binding in bacteria, where the local concentration of the acceptor polypeptide is likely lower than in eukaryotes. In fact, mutation of R331 to Ala results in a reduced glycosylation efficiency of the 3D5 acceptor protein containing a DQNAT glycosylation site (Fig. 4b). However, the R331A mutant enables glycosylation of 3D5 containing an AQNAT (SEQ ID NO: 8) site, which is not glycosylated by wild type PglB (Fig. 4b). Therefore, the R331A mutant can be used to selectively occupy glycosylation sites that are not used by wild type PglB. Consequently, a combination of wild type enzyme and the R331A mutant allows the site specific attachment of different glycans within the same glycoprotein.

Catalytic site

The catalytic pocket is located in the right-side cavity of PglB (Fig. 3b) and is marked by a bound cation, located ~8Å above the membrane boundary. Due to the high concentration of magnesium salt in the crystallization solution, it was modelled as Mg²⁺. Like all OSTs PglB is only functional with a bound divalent cation (Imperiali & Rickert, *P. Natl. Acad. Sci. USA* **92**, 97-101, 1995; Sharma et al., *Eur. J. Biochem.* **116**, 101-108, 1981). The physiological cation has been suggested to be Mn²⁺, but PglB is also active in Mg²⁺ (unpublished), a property that has previously been observed for other metal-dependent glycosyltransferases (Liu & Mushegian, *Protein Sci.* **12**, 1418-1431, 2003). The catalytic pocket of PglB features three acidic side chains (D56, D154, E319) that are provided by the TM domain and that coordinate the M²⁺ (Fig. 5a). At the current resolution, water molecules that might be additional ligands of M²⁺ cannot be modelled. The residues located in the catalytic pocket are conserved in STT3 proteins. The aspartates D154 and D156 belong to a previously reported D-X-D motif and mutation of either aspartate to alanine in the mannosyltransferase GPI-MT-1, a member of the same glycosyltransferase family as PglB (GT-C) (Lairson et al., *Annu. Rev. Biochem.* **77**, 521-555, 2008; Liu & Mushegian, *Protein Sci.* **12**, 1418-1431, 2003), abolished the activity of this enzyme (Maeda et al., *Embo J* **20**, 250-261, 2001). In contrast, D56 and E319 have not been previously identified as catalytic relevant, but their carboxyl groups are interacting both with the metal ion and the amido group of the acceptor asparagine. To confirm the catalytic involvement of the three acidic, M²⁺-binding residues, they were mutated individually to alanines and the activity of the resulting PglB mutants was tested in a complementation assay (Fig. 2). Even though OST is not limiting in our assay, the mutation D154A reduced the observed glycosylation yield by >50%, the mutants D56A

and E319A reduced it by >90%, and the double mutant D56A/E319A was completely inactive.

There is a controversial discussion on how the amido group of the acceptor asparagine might be activated to perform a nucleophilic attack on the C1 carbon of the LLO substrate, the key step in N-linked glycosylation. Amides are poor nucleophiles because the free electron pair of the nitrogen is conjugated to the double bond of the carbonyl group (Fig. 5c). As a consequence, the N-C bond has double bond character, and the nucleophilicity of the nitrogen is low. To explain the reactivity of the amido group, specific conformations of the acceptor peptide such as a " β -turn" or an "Asx-turn" have been proposed, invoking direct involvement of the β -hydroxyl group of the +2 Ser/Thr for increased nucleophilicity of the amide group (Bause & Legler, *Biochem. J.* **195**, 639-644, 1981; Imperiali et al., *J. Am. Chem. Soc.* **114**, 7942-7944, 1992). Given the firm binding of the +2 Thr to the WWD motif in our PglB structure, such a mechanism can be ruled out. Instead, the structure of PglB presents a distinct possibility for explaining amide nitrogen activation: The two catalytically essential, acidic residues D56 and E319 are optimally positioned to form hydrogen bonds with the amide protons of the acceptor asparagine. Forming such hydrogen bonds would require a rotation of the N-C bond of the amide group, thereby abolishing the delocalization of the free electrons of the nitrogen atom and breaking the conjugation with the carbonyl group (Fig. 5c). Not only would this increase the electronegative nature of the amide nitrogen (by polarizing the N-H bonds and increasing the electron density on the nitrogen), but it would also generate an sp^3 hybridized nitrogen with a reactive lone pair optimally positioned for the nucleophilic attack on the C1 carbon of the activated oligosaccharide substrate (LLO). The energy barrier for rotating the N-C bond in most amides is estimated to be 16-20 kcal/mol, and the 270° amide conformation shown in Fig 5c has been calculated to have an energy of ~ 18.6 kcal/mol relative to the planar conformation (Wiberg & Breneman, *J Am Chem Soc.* **114**, 7942-7944, 1992). Hence it would take 1-2 low barrier hydrogen bonds (Cleland & Kreevoy, *Science* **264**, 1887-1890, 1994) (each worth ~ 10 kcal/mol) to provide sufficient energy to permanently break the conjugation of the carboxamido group of the acceptor asparagine. The carboxylates of D56 or E319 might provide such interactions in the transition state of the glycosylation reaction, although it will require a higher resolution structure to reliably measure lengths of hydrogen bonds. Mutating D56 to asparagine (D56N) has an even more pronounced inhibitory effect than the truncation to alanine, and the E319Q mutant is completely inactive (Fig. 2). This demonstrates that the negative charges provided by the carboxyl groups of D56 and E319 are essential for catalysis and the acidic side chains cannot be replaced by the

corresponding iso-electronic amides. Steric effects might explain the increased inhibition of D56N relative to D56A and of E319Q compared with E319A.

Glycosylation mechanism

Given that PglB is active even when solubilized in detergent (used for purification and crystallization) the provided structure has likely captured a functionally competent state. The glycosylation reaction occurs with inversion of configuration at the substituted C1 carbon of the first sugar moiety. The LLO substrate was modelled into the PglB structure such that the di-N-acetyl-bacillosamine moiety is properly aligned for a nucleophilic attack by the activated amide nitrogen, while the leaving pyrophosphate group is in contact with the divalent metal ion and the conserved R375 (Fig 6a). This arrangement places the additional saccharide moieties in the right hand cavity of PglB, where they can interact with surface residues from both the TM and the periplasmic domains. The arrangement also places the C2 substituent of the first saccharide moiety, a N-acetyl group present in LLOs of bacteria and eukaryotes, in the vicinity of a conserved tyrosine residue (Y468), where density consistent with a bound water molecule is observed. When modelled as shown in Fig. 6a, the lipid tail of the LLO is located in a mostly hydrophobic groove on the surface of PglB, pointing its isoprenoid moieties into the lipid bilayer. The function of the bound divalent cation in PglB thus appears to be two-fold: On the one hand, it orients the acidic side chains that interact with the acceptor asparagine, and, on the other hand, it stabilizes the leaving group of the substitution (lipid-pyrophosphate), thus accelerating the reaction. This would be distinct from metal-dependent, configuration-inverting glycosyltransferases of the GT-A family, where the metal ion only serves the stabilization of the leaving group (see Lairson above). With the acceptor peptide present in the structure and the LLO molecule tentatively modelled, a basic, three-state catalytic cycle for PglB-catalysed glycosylation can be proposed (Fig. 6b). A critical element of the proposed mechanism is the engagement and disengagement of the external loop EL5, which is suspected to be flexible and disordered in the absence of bound acceptor peptide (ground state). Upon binding of peptide, this loop becomes partially ordered and pins the peptide against the periplasmic domain, thereby restricting its motion. Because the essential E319 is part of EL5, this simultaneously results in the formation of the catalytic site, where the acceptor Asn is correctly oriented and activated. This state can only be reached if a consensus sequon from a flexible, exposed protein loop is inserted in the binding pocket. In the following step, LLO binding is expected, whereupon the activated amide nitrogen performs a nucleophilic attack on the first saccharide moiety, resulting in glycosylation. Once the

glycosidic bond is formed, the newly attached sugars are tightly pressed against PglB (specifically against Ile317 and His485), causing steric tension that can be released by disengagement of EL5. This opens the acceptor Asn porthole and allows the glycopeptide to dissociate from the enzyme. Subsequent cleavage of the lipid-linked pyrophosphate anhydride and the folding of the glycosylated protein domain likely provide the main contributions to the driving force of the reaction. It is noted that PglB might also bind LLO before binding peptide, and there is no experimental evidence suggesting a strict sequence of events.

Claims

1. A method for identifying a potential component for the oligosaccharyltransferase (OST)-catalysed asparagine-linked ("N-linked") glycosylation selected from the group consisting of
 - (a) a potential oligosaccharide donor, preferably a lipid-linked oligosaccharide (LLO) or an undecaprenyl pyrophosphate bound oligosaccharide donor,
 - (b) a potential oligosaccharyltransferase (OST),
 - (c) a potential consensus sequence motif polypeptide, and
 - (d) a potential glycosylation inhibitor,comprising the steps of
 - (i) using the atomic coordinates of Table 1, preferably ± 2 , more preferably ± 1.5 , most preferably ± 1.0 Å root mean square deviation (rmsd) from the backbone atoms, for generating a three-dimensional model of the catalytic domain of the oligosaccharyltransferase (OST) of *Campylobacter lari*, comprising at least three, preferably at least four, more preferably at least six, most preferably all of amino acids D56, R147, D154, D156, E319, R375, Y468, and H485, and/or, preferably and
 - (ii) using the atomic coordinates of Table 1, preferably ± 2 , more preferably ± 1.5 , most preferably ± 1.0 Å root mean square deviation (rmsd) from the backbone atoms, for generating a three-dimensional model of the polypeptide binding site of the oligosaccharyltransferase (OST) of *Campylobacter lari*, comprising at least two, preferably at least three, more preferably at least four, most preferably all of amino acids M318, R331, W463, W464, D465, and I572,
 - (iii) preferably performing whole body translations and/or rotations on the coordinates of the amino acids of the three-dimensional models of (i) and/or (ii),
 - (iv) using said three-dimensional model(s) of (i), (ii) and/or (iii) for designing or selecting at least one of potential components (a) to (d),
 - (v) providing at least one of said potential components (a) to (d), and
 - (vi) contacting at least one of said potential components (a) to (d) with the further functional components necessary for an oligosaccharyltransferase (OST)-catalysed asparagine-linked ("N-linked") glycosylation,
 - (vii) identifying a functional component selected from the group consisting of

- (A) a functional oligosaccharide donor, preferably a functional lipid-linked oligosaccharide (LLO) or an undecaprenyl pyrophosphate bound oligosaccharide donor,
 - (B) a functional oligosaccharyltransferase (OST),
 - (C) a functional consensus sequence motif polypeptide, and
 - (D) a functional glycosylation inhibitor.
2. The method of claim 1, wherein in step (ii) the three-dimensional model of the polypeptide binding site of the oligosaccharyltransferase (OST) of *Campylobacter lari* comprises at least two, preferably at least three, more preferably at least four, most preferably all of amino acids M318, A331, W463, W464, D465 and I572.
3. A method for designing a potential component for the oligosyltransferase (OST)-catalysed asparagine-linked ("N-linked") glycosylation selected from the group consisting of
- (a) a potential oligosaccharide donor, preferably a lipid-linked oligosaccharide (LLO) or an undecaprenyl pyrophosphate bound oligosaccharide donor,
 - (b) a potential oligosaccharyltransferase (OST),
 - (c) a potential consensus sequence motif polypeptide, and
 - (d) a potential glycosylation inhibitor,
- comprising the steps of
- (i) using the atomic coordinates of Table 1, preferably ± 2 , more preferably ± 1.5 , most preferably ± 1.0 Å root mean square deviation (rmsd) from the backbone atoms, for generating a three-dimensional model of the catalytic domain of the oligosaccharyltransferase (OST) of *Campylobacter lari*, comprising at least three, preferably at least four, more preferably at least six, most preferably all of amino acids D56, R147, D154, D156, E319, R375, Y468, and H485, and/or, preferably and
 - (ii) using the atomic coordinates of Table 1, preferably ± 2 , more preferably ± 1.5 , most preferably ± 1.0 Å root mean square deviation (rmsd) from the backbone atoms, for generating a three-dimensional model of the polypeptide binding site of the oligosaccharyltransferase (OST) of *Campylobacter lari*, comprising at least two, preferably at least three, more preferably at least four, most preferably all of amino acids M318, R331, W463, W464, D465, and I572,

- (iii) preferably performing whole body translations and/or rotations on the coordinates of the amino acids of the three-dimensional models of (i) and/or (ii),
 - (iii.1) using said three-dimensional model of (i), (ii) and/or (iii) for assessing the stereochemical complementarity between said three-dimensional model(s) (i), (ii) and/or (iii) and a known or potential component for the oligosaccharyltransferase (OST)-catalysed asparagine-linked ("N-linked") glycosylation selected from an oligosaccharide donor, preferably a lipid-linked oligosaccharide (LLO) or an undecaprenyl pyrophosphate bound oligosaccharide donor, a consensus sequence motif polypeptide, and a potential glycosylation inhibitor, or
 - (iii.2) varying at least one amino acid in said three-dimensional model of (i), (ii) and/or (iii) and using said varied three-dimensional model of (i), (ii) and/or (iii) for assessing the stereochemical complementarity between said three-dimensional models (i), (ii) and/or (iii) and a known or potential component for the oligosaccharyltransferase (OST)-catalysed asparagine-linked ("N-linked") glycosylation selected from an oligosaccharide (LLO) donor, preferably a lipid-linked oligosaccharide (LLO) donor or an undecaprenyl pyrophosphate bound oligosaccharide donor, a consensus sequence motif polypeptide, and a potential glycosylation inhibitor,
- (iv) optimizing said stereochemical complementarity in an iterative approach by observing changes in the three-dimensional model of (iii.1), (iii.2) or the component for the oligosaccharyltransferase (OST)-catalysed asparagine-linked ("N-linked") glycosylation,
- (v) designing a potential component selected from (a) to (d) which optimizes said stereochemical complementarity of said three-dimensional model(s) and potential component,
 - (vi.1) optionally providing the optimized potential component, and
 - (vi.2) contacting at least one of said potential components (a) to (d) with the further functional components necessary for an oligosaccharyltransferase (OST)-catalysed asparagine-linked ("N-linked") glycosylation,
 - (vi.3) identifying a functional component selected from the group consisting of
 - (A) a functional oligosaccharide donor, preferably a functional lipid-linked oligosaccharide (LLO) or an undecaprenyl pyrophosphate bound oligosaccharide donor,
 - (B) a functional oligosaccharyltransferase (OST),

- (C) a functional consensus sequence motif polypeptide, and
 - (D) a functional glycosylation inhibitor.
4. The method of claim 3, wherein in step (ii) the three-dimensional model of the polypeptide binding site of the oligosaccharyltransferase (OST) of *Campylobacter lari* comprises at least two, preferably at least three, more preferably at least four, most preferably all of amino acids M318, A331, W463, W464, D465 and I572.
 5. The method according to any of claims 1 to 4, wherein the specific three-dimensional catalytic site model of step (i) further comprises one or more, preferably all of the amino acids selected from the group having residues located within Van der Waals distance to the bound peptide of SEQ ID NO: 2, preferably selected from the group consisting of Thr53, Thr54, Asn55, Asp56, Asn146, Arg147, Tyr152, Glu315, Thr316, Ile317, Met318, Glu319, Val320, Asn321, Arg331, Leu374, Arg375, Tyr433, Ser435, Val438, Trp463, Trp464, Asp465, Gly482, His485, Ile572, Val575.
 6. The method according to any of claims 1 to 4, wherein the specific three-dimensional catalytic site model of step (i) further comprises one or more, preferably all of the amino acids selected from the group having residues located within Van der Waals distance to the bound peptide of SEQ ID NO: 2, preferably selected from the group consisting of Thr53, Thr54, Asn55, Asp56, Asn146, Arg147, Tyr152, Glu315, Thr316, Ile317, Met318, Glu319, Val320, Asn321, Ala331, Leu374, Arg375, Tyr433, Ser435, Val438, Trp463, Trp464, Asp465, Gly482, His485, Ile572, Val575.
 7. The method according to any one of claims 1 to 6, wherein the potential component is an oligosaccharide donor (a), preferably a lipid-linked oligosaccharide (LLO) or an undecaprenyl pyrophosphate bound oligosaccharide donor.
 8. The method according to any one of claims 1 to 6, wherein the potential component is a oligosaccharyltransferase (OST) (b).
 9. The method according to any one of claims 1 to 6, wherein the potential component is a potential consensus sequence motif polypeptide (c).
 10. The method according to any one of claims 1 to 6, wherein the potential component is a glycosylation inhibitor (d).

11. Machine-readable medium comprising
 - (i) the atomic coordinates of Table 1, preferably ± 2 , more preferably ± 1.5 , most preferably ± 1.0 Å root mean square deviation (rmsd) from the backbone atoms, preferably comprising at least one, two or three, preferably at least four or five, more preferably at least six or 7, most preferably all of amino acids D56, R147, D154, D156, E319, R357, Y468, and H485, and/or, preferably and
 - (ii) the atomic coordinates of Table 1, preferably ± 2 , more preferably ± 1.5 , most preferably ± 1.0 Å root mean square deviation (rmd) from the backbone atoms, for generating a three-dimensional model of the polypeptide binding site of the oligosaccharyltransferase (OST) of *Campylobacter lari*, comprising at least one or two, preferably at least three, more preferably at least four or five, most preferably all of amino acids M318, R331, W463, W464, D465, and I572,
 - (iii) preferably the atomic coordinates of (i) or (ii) modified by performing whole body translations and/or rotations on said coordinates.
12. The machine-readable medium according to claim 11, comprising
 - (ii) the atomic coordinates of Table 1, preferably ± 2 , more preferably ± 1.5 , most preferably ± 1.0 Å root mean square deviation (rmd) from the backbone atoms, for generating a three-dimensional model of the polypeptide binding site of the oligosaccharyltransferase (OST) of *Campylobacter lari*, comprising at least one or two, preferably at least three, more preferably at least four or five, most preferably all of amino acids M318, A331, W463, W464, D465, and I572.