THE JOURNAL OF BIOLOGICAL CHEMISTRY VOL. 285, NO. 7, pp. 4941–4950, February 12, 2010

Comparative Structural Biology of Eubacterial and Archaeal Oligosaccharyltransferases*[S]

Received for publication, November 3, 2009, and in revised form, November 18, 2009 Published, JBC Papers in Press, December 9, 2009, DOI 10.1074/jbc.M109.081752

Nobuo Maita[1], **James Nyirenda**[2], **Mayumi Igura**[3], **Jun Kamishikiryo, and Daisuke Kohda**[3,4]

From the Division of Structural Biology, Medical Institute of Bioregulation, Kyushu University, Maidashi 3-1-1, Higashi-ku, Fukuoka 812-8582, Japan

Oligosaccharyltransferase (OST) catalyzes the transfer of an oligosaccharide from a lipid donor to an asparagine residue in nascent polypeptide chains. In the bacterium *Campylobacter jejuni*, a single-subunit membrane protein, PglB, catalyzes *N*-glycosylation. We report the 2.8 Å resolution crystal structure of the C-terminal globular domain of PglB and its comparison with the previously determined structure from the archaeon *Pyrococcus* AglB. The two distantly related oligosaccharyltransferases share unexpected structural similarity beyond that expected from the sequence comparison. The common architecture of the putative catalytic sites revealed a new catalytic motif in PglB. Site-directed mutagenesis analyses confirmed the contribution of this motif to the catalytic function. Bacterial PglB and archaeal AglB constitute a protein family of the catalytic subunit of OST along with STT3 from eukaryotes. A structure-aided multiple sequence alignment of the STT3/PglB/AglB protein family revealed three types of OST catalytic centers. This novel classification will provide a useful framework for understanding the enzymatic properties of the OST enzymes from Eukarya, Archaea, and Bacteria.

Protein *N*-glycosylation is an important posttranslational modification that occurs in all domains of life (1). The enzyme that creates the oligosaccharide-asparagine linkage is oligosaccharyltransferase (OST).[5] OST catalyzes the *en bloc* transfer of a preassembled oligosaccharide from a lipid carrier to asparagine residues in the glycosylation consensus (Asn-*X*-Thr/Ser, where *X* represents any amino acid except for Pro) of polypeptide chains (2–4). OST is a multisubunit membrane protein complex in higher eukaryotes. Yeast (*Saccharomyces cerevisiae*) OST consists of eight different subunits: Ost1p, Ost2p, Ost3p/Ost6p, Ost4p, Wbp1, Swp1, Stt3p, and Ost5p (5), where Ost3p and Ost6p are paralogs that are present in two distinct OST isoforms (6). The cryoelectron microscopy structure of the digitonin-solubilized OST complex from yeast provided the relative arrangement of Ost1p, Wbp1, and Stt3p on the lumenal side of the complex (7, 8).

Stt3p is the catalytic subunit of the yeast OST enzyme (9). The vertebrate, insect, and plant equivalents are the two paralog proteins, STT3A and STT3B, which define distinct OST isoforms (10, 11). In lower eukaryotes, such as trypanosomatids, OST is a single-polypeptide membrane protein (3), and these single-subunit OST proteins consist of STT3 (staurosporine and temperature sensitivity 3) alone. In fact, the STT3s from *Trypanosoma cruzi*, *Trypanosoma brucei*, and *Leishmania major* can each function as an OST enzyme when transferred into *stt3*-deficient yeast cells (12–15). The prokaryotic OST is also a single-polypeptide protein. The STT3 homologs, PglB (protein glycosylation B) and AglB (archaeal glycosylation B), comprise the bacterial OST and the archaeal OST, respectively (16–20). Multiple STT3/PglB/AglB paralogs also exist in some single-subunit OSTs. The trypanosomatids *L. major* and *T. brucei* contain four and three STT3 paralogs, respectively. In contrast, their related species, the trypanosomatid *T. cruzi*, contains one STT3 species. Similarly, the archaeon *Pyrococcus furiosus* contains two copies of AglB, but the bacterium *Campylobacter jejuni* contains one PglB species. Thus, the existence of multiple OST isoforms containing or consisting of different STT3/PglB/AglB paralogs raises the interesting question of the functional differences between these OST isoforms in an organism. The OST isoforms containing STT3A and STT3B in mammalian cells have different enzymatic properties and play complementary roles in the cotranslational and posttranslational *N*-glycosylation of proteins (21). The STT3 paralogs of *T. brucei* have distinct glycan donor specificities (13).

The roles of the subunits other than STT3/PglB/AglB in the multisubunit OST are not still clear. Ost1p/Ribophorin I was proposed to regulate the delivery of a set of proteins to the catalytic center of STT3 (22, 23). The Swp1-Wbp1-Ost2p subcomplex was suggested to bear the second regulatory binding site for the selection of glucosylated oligosaccharide donors (24). Structural information has greatly assisted in the elucidation of the molecular functions of these subunits. The crystal structure of the lumenal domain of yeast Ost6p (166 residues) revealed a thioredoxin-like fold (25). Indeed, yeast Ost6p has disulfide oxidoreductase activity and may prevent the nascent

* The experiments at SPring-8 were approved by the Japan Synchrotron Radiation Research Institute as Proposal 2007A1204. The experiments at the Photon Factory were approved by the High Energy Accelerator Research Organization (KEK) as Proposals 2007G208 and 2009G021.

[S] The on-line version of this article (available at http://www.jbc.org) contains supplemental Figs. S1–S6.

The atomic coordinates and structure factors (code 3AAG) have been deposited in the Protein Data Bank, Research Collaboratory for Structural Bioinformatics, Rutgers University, New Brunswick, NJ (http://www.rcsb.org/).

[1] Present address: Institute for Enzyme Research, Tokushima University, 2-24, Shinkura-cho, Tokushima 770-8501, Japan.

[2] Supported by a Japanese government scholarship from the Ministry of Education, Culture, Sports, Science and Technology (MEXT) of Japan.

[3] Supported by Grants-in-Aid for Scientific Research on Priority Areas and the Targeted Proteins Research Program from MEXT of Japan.

[4] To whom correspondence should be addressed. Tel.: 81-92-642-6968; Fax: 81-92-642-6764; E-mail: kohda@bioreg.kyushu-u.ac.jp.

[5] The abbreviations used are: OST, oligosaccharyltransferase; CC, central core; GST, glutathione *S*-transferase; IS, insertion; sPglB, soluble domain of PglB; TLS, translation/libration/screw; MES, 4-morpholineethanesulfonic acid.

polypeptide from forming disulfide bonds during the cotranslational *N*-glycosylation (25).

The primary sequences of the STT3/PglB/AglB proteins share a common architecture. A multispan transmembrane region exists in the N-terminal half. The membrane topologies of yeast Stt3p and mouse STT3A were experimentally deduced (26). These eukaryotic STT3s have 11 transmembrane helices and an overall $N_{cytoplasm}$-C_{lumen} orientation. The C-terminal half of the primary sequence forms a globular domain bearing a well conserved five-residue motif, WWDYG. The side chain carboxylate group of the central aspartate in the WWDYG motif probably functions as a catalytic base for the OST reaction (9, 16, 17). The entire primary sequences of eukaryotic STT3 are highly homologous across eukaryotes, including animals, plants, fungi, and protists, indicating the essential biological role of *N*-glycosylation in eukaryotic cells. By contrast, archaeal AglB shows remarkable sequence diversity, which may reflect a wider range of oligosaccharide structures than those found in eukaryotic glycoproteins (20, 27). *N*-Glycosylation of extracellular proteins, as well as *O*-glycosylation, facilitates the adaptation of archaeal organisms to the extreme environments where they thrive (28). Few bacteria have the protein *N*-glycosylation system. Until recently, *N*-glycosylation was reported exclusively in *Campylobacterales*, among bacteria, and the best characterized species is the human enteropathogenic bacterium, *Campylobacter jejuni* (29). *N*-Glycosylation is important for the virulence of this organism by increasing its adherence to and invasion of host cells. Recently, a comparative genomic analysis of two ϵ-proteobacterial species from a deep sea vent added new PglB sequences belonging to orders other than *Campylobacterales* (30). The multiple sequence alignment revealed a moderate level of sequence conservation (supplemental Fig. S1).

A meaningful multiple sequence alignment of STT3, PglB, and AglB across the three domains of life is virtually impossible, except for the vicinity of the WWDYG motif, due to the very low sequence homology among them. In such cases, reference to three-dimensional structures frequently helps to create a more reliable alignment. We previously reported the crystal structure of the C-terminal globular domain of *P. furiosus* AglB protein (PF0156, a longer paralog of the two AglBs) (31, 32). The crystal structure revealed the putative catalytic site by identifying a local structure consisting of the WWDYG motif and a long, kinked helix adjacent to the motif. We found a pair of Asp and Lys residues spaced three residues apart (D*XX*K, where *X* can be any residue) that is conserved in yeast Stt3p, the two *Pyrococcus* AglB paralogs, and *Campylobacter* PglB and named it the "DK" motif. The identification of the new motif enabled the extension of the alignment from the vicinity of the WWDYG motif to 100-residue segments, including the WWDYG motif and its flanking regions. This alignment seems valid between eukaryotic STT3 and archaeal AglB, because *in vivo* mutational studies indicated that the Asp and Lys residues in the motif were catalytically important in yeast Stt3p and in *L. major* STT3–1 (14, 31). The multiple sequence alignment using the recently published bacterial PglB sequences (supplemental Fig. S1), however, raised the possibility of an improper alignment of the D*XX*K sequence for *Campylobacter* PglB. To answer this question, we have determined the crystal structure of the C-terminal globular domain of *Campylobacter* PglB in the present study, and compared it with the previously determined structure of *Pyrococcus* AglB. Due to their high structural similarity, the structure-based sequence alignment yielded an accurate sequence alignment between *Campylobacter* PglB and *Pyrococcus* AglB. In fact, the counterpart of the D*XX*K sequence was found to be the M*XX*I sequence in the bacterial PglB. This finding provided the novel classification of the catalytic center of OST and unexpected insights into the evolutionary relationship between the OSTs from the three domains of life.

EXPERIMENTAL PROCEDURES

Crystallization of the Globular Domain of C. jejuni PglB—The cloning, expression, purification, and crystallization of the C-terminal soluble, globular domain (sPglB) of *C. jejuni* PglB protein (Q5HTX9_CAMJR) will be published elsewhere. Briefly, the expression plasmid encoding the C-terminal soluble domain (residues 428–713) was constructed by inserting a PCR product from genomic DNA (ATCC700819D) into the pGEX-6P-1 vector (GE Healthcare). The GST-fused sPglB protein was expressed by the addition of isopropyl-β-D-thiogalactopyranoside at 310 K in the *Escherichia coli* BL21(DE3)pLysS strain (Novagen) in selenomethionine core medium (Wako). The cells were disrupted by sonication. After centrifugation, the GST-sPglB protein in the supernatant was adsorbed to glutathione-Sepharose 4B resin (GE Healthcare) and was cleaved by GST-fused 3C protease on the resin. The sPglB was eluted, concentrated, and subjected to reductive methylation of the lysine residues, as described (33). The methylated sPglB protein was purified by gel filtration, using a Superdex75 column (GE Healthcare) and then by anion exchange chromatography, using a Resource Q column (GE Healthcare). The protease cleavage yielded N-terminally extended sPglB, containing five extra residues (Gly-Pro-Leu-Gly-Ser), and the reductive lysine methylation resulted in the dimethylation of the ϵ-amino group of all 24 lysine residues and the N-terminal α-amino group. Selenomethionyl derivative, methylated sPglB crystals grew from a hanging drop with a 1:1 volume ratio (total volume, 1 μl) of the protein stock solution (10 mg ml^{-1}, 10 mM Tris-HCl, pH 8.0) and the reservoir solution (0.1 M sodium cacodylate, pH 6.5, 18% polyethylene glycol 8000, 0.2 M calcium acetate) at 293 K.

Data Collection and Structure Determination—The sPglB crystals were cryoprotected by soaking in 0.1 M MES, pH 6.5, 18% polyethylene glycol 8000, and 0.2 M calcium acetate and then in 0.1 M MES, pH 6.5, 30% polyethylene glycol 8000, and 0.2 M calcium acetate. The diffraction data were recorded at BL-38B1 and BL-44XU, SPring-8 (Harima, Japan). The data were processed with *HKL*-2000 (34) up to a resolution of 2.8 Å and yielded the space group $P6_4$. The crystal contained two molecules in the asymmetric unit cell (52.4% solvent, $V_M = 2.58$ $Å^3$ Da^{-1}). A site search for the selenium atoms with *autoSHARP* (35) found 12 of 16 selenium sites in an asymmetric unit. The six selenium sites in each molecule were used for phasing. Phases were further improved by the density modification method, using *DM* and *Solomon* from the *CCP4* package (36), with a mean figure of merit value of 0.86. Refinement

TABLE 1
Data collection and refinement statistics

	Se-Peak	Se-Edge	Se-Remote1	Se-Remote2
Data collection				
X-ray source	BL-38B1	BL-38B1	BL-38B1	BL-44XU
Wavelength (Å)	0.97881	0.97909	0.94000	1.1000
Space group	$P6_4$	$P6_4$	$P6_4$	$P6_4$
Unit cell dimensions (Å)				
$a = b$	115.118	115.131	115.126	115.258
c	88.777	88.903	88.854	88.883
Resolution (Å)	30-3.0	30-3.2	30-3.2	30-2.8
Outer shell (Å)	3.11-3.0	3.31-3.2	3.31-3.2	2.9-2.8
Mosaicity (degrees)	0.193	0.234	0.209	0.206
No. of reflections				
Observed	200,768	83,950	85,597	63,435
Unique	13,487	11,117	11,112	16,601
Completeness (%)	99.7 (99.5)[a]	99.7 (99.5)	99.6 (99.7)	99.9 (100)
Redundancy	14.9 (14.9)	7.6 (7.6)	7.7 (7.8)	3.8 (3.8)
$\langle I \rangle / \langle \sigma I \rangle$	11.2 (4.58)	10.3 (3.38)	10.8 (4.09)	9.3 (3.37)
R-sym (%)[b]	13.3 (47.0)	12.3 (42.2)	12.6 (35.9)	8.5 (50.3)
Phasing (by *autoSHARP*)				
No. of selenium sites (found/all)	12/16			
Phasing power				
Anomalous	0.914	0.672	0.087	0.118
Dispersive (centric/acentric)		0.550/0.718	0.551/0.794	0.568/0.723
Overall figure of merit (initial/after *DM*)		0.416/0.860		
Refinement				
Resolution (Å)				20-2.8
No. of reflections (work/test)				15,279/1,276
R-work/R-free (%)[c]				24.51/26.00
$\langle B \rangle$ (Å^2)				33.28
Root mean square deviation				
Bonds (Å)				0.002
Angles (degrees)				0.335
Ramachandran plot analysis (%)				
Most favored				77.7
Additional allowed				19.6
Generously allowed				2.6
Disallowed				0.0

[a] The highest resolution shell is shown in parentheses.
[b] R-sym $= \Sigma|I_i - \langle I_i \rangle|/\Sigma I_i$, where I_i is the observed intensity, and $\langle I_i \rangle$ is the average intensity obtained from multiple observations of symmetry-related reflections.
[c] R-work $= \Sigma_{hkl}||F_o| - |F_c||/\Sigma_{hkl}|F_o|$. 7.7% of the reflections were excluded for R-free calculation.

was performed using the program *CNS* (37), and manual rebuilding was carried out with the program *COOT* (38). Translation/libration/screw (TLS) refinement was performed using the program *REFMAC5* in the *CCP4* package at the final round. The TLS groups were determined by the TLS Motion Determination server (available on the Washington University Web site). The final R factors obtained were R_{work}/R_{free} 24.5/26.0%. The data collection and refinement statistics are summarized in Table 1. In chain A, the N-terminal five extra residues as well as residues 428–435, 515–518, and 583–589 were disordered. Chain B displayed a higher degree of disorder than chain A. The two structures in the asymmetric unit are similar, with a Cα root mean square deviation value of 0.55 Å over 210 residues. Chain A was used for further analysis, and the figures were generated with the *PyMOL* version 1.1 (available on the World Wide Web). Superposition of the two structures was performed by the program *GASH* (available on the Protein Data Bank Japan Web site) (39).

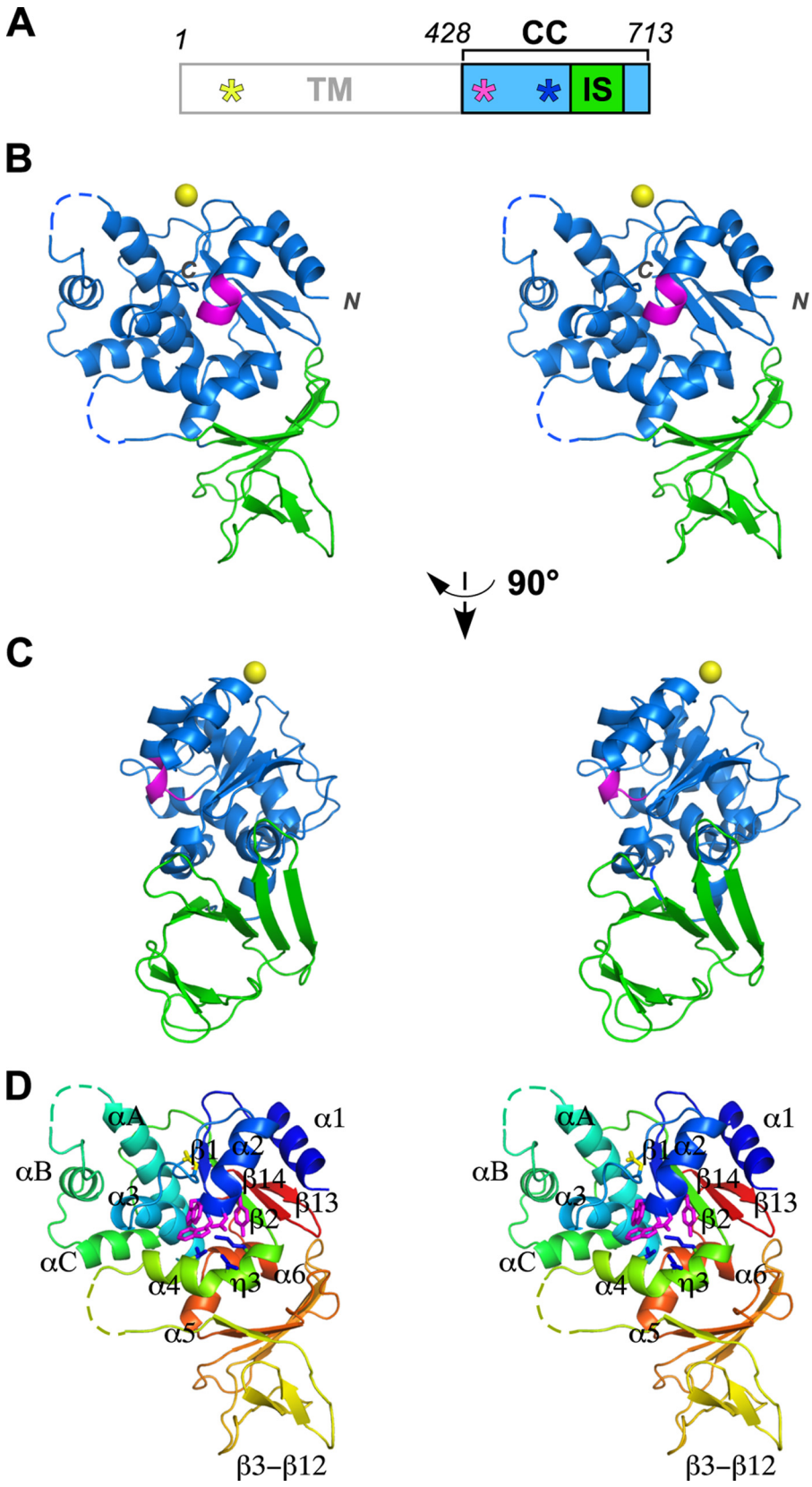
Multiple Sequence Alignment—The amino acid sequences of the oligosaccharyltransferases (STT3/AglB/PglB) were retrieved from the InterPro 17.0 data base (available on the World Wide Web). The family IPR003674 contains 278 sequences, of which 154 sequences belong to Eukarya, 84 to Archaea, and 40 to Bacteria. The multiple sequence alignment was performed with the program *MAFFT*, version 6 (available on the World Wide Web) (40).

Preparation of the E. coli Membrane Fractions Containing the Full-length PglB—A codon-optimized DNA sequence of *C. jejuni* PglB (residues 1–713, Q5HTX9_CAMJR) for expression in *E. coli* cells was obtained from GenScript in the pUC57 plasmid. The entire coding sequence was subcloned into pET-41b(+) (Novagen) between the NdeI and SalI sites, with a C-terminal His_8 tag. The PglB variants were generated using a KOD plus mutagenesis kit (TOYOBO). PglB and its variants were expressed in the *E. coli* BL21GOLD(DE3) strain (Novagen). The transformed *E. coli* cells were grown at 310 K in LB medium, supplemented with 30 mg/liter kanamycin. When the A_{600} reached 0.6, isopropyl-β-D-thiogalactopyranoside was added to a final concentration of 0.5 mM. After a 4-h induction at 310 K, the cells were harvested by centrifugation. The cell pellets from a 150-ml LB culture were resuspended in 20 ml of 50 mM Tris-HCl buffer, pH 7.4, containing 150 mM NaCl, 2 mM $MgCl_2$, and 1 mM phenylmethylsulfonyl fluoride, supplemented with complete protease inhibitor mixture (Roche Applied Science) according to the manufacturer's instructions. After cell disruption by sonication, the lysate was centrifuged at 5,700 $\times$ g for 10 min. The supernatant was transferred into eight polycarbonate tubes in 3-ml aliquots and was ultracentrifuged, using a Beckman Coulter Optima TLX ultracentrifuge equipped with a TLA 110 rotor, at 100,000 $\times$ g and at 4 °C for 1 h. The supernatant was discarded, and the pellets were kept at −80 °C until use. Each pellet was dissolved in 5 ml of

Triton buffer (20 mM Tris-HCl, pH 7.4, 150 mM NaCl, and 1% Triton X-100), incubated for 1 h on ice, and then divided into two polycarbonate tubes. After ultracentrifugation at 100,000 × *g*, the supernatant was collected as the membrane fraction. The amount of the PglB protein in the membrane fractions was quantified by SDS-PAGE and Western blotting with anti-His tag antibodies (mouse penta-His antibody, Qiagen) and goat anti-mouse IgG antibodies labeled with fluorescent IRDye 800CW (LI-COR). The fluorescence image was recorded with an Odyssey infrared imaging system (LI-COR) and was quantified using the Odyssey software, version 3.0.

Preparation of Lipid-linked Oligosaccharide from C. jejuni Cells—The glycerol stock (JCM catalog number 2013) of *C. jejuni* subsp. jejuni strain was obtained from the Microbe Division/Japan Collection of Microorganisms of the RIKEN Bioresource Center (Saitama, Japan). A microaerophilic atmosphere was set up in a 7-liter rectangular jar, with an AnaeroPack-MicroAero sachet (Mitsubishi Gas Chemical, Tokyo), which works as an oxygen absorber-CO_2 generator for microaerophilic cultivation. The frozen stock cells were streaked onto a CCDA plate (blood-free Campylobacter medium, Kanto Chemical, Tokyo) and incubated at 37 °C for 3 days in the rectangular jar. Some colonies were inoculated into LB medium and incubated with gentle shaking in the rectangular jar at 37 °C for 3 days. The cells were collected by centrifugation. Lipid-linked oligosaccharides were extracted twice from 0.5 g of the cell pellets with 20 ml of chloroform/methanol/water (10:20:3 (v/v/v), solvent A), using a Polytron homogenizer (PT1200E, Kinematica). The combined extracts were concentrated under a nitrogen gas stream. The dried residue was dissolved in 2 ml of solvent A and was stored at −20 °C.

Oligosaccharyltransferase Assay—The OST assay was performed by the PAGE method (41). A 5-μl aliquot of lipid-linked oligosaccharides in solvent A was transferred to a 1.5-ml plastic tube and dried in a SpeedVac concentrator (ThermoSavant). Eight microliters of buffer (50 mM Tris-HCl, pH 7.5, containing 1 mM dithiothreitol, 10 mM $MnCl_2$, and 0.02% Tween 20) were added, and the



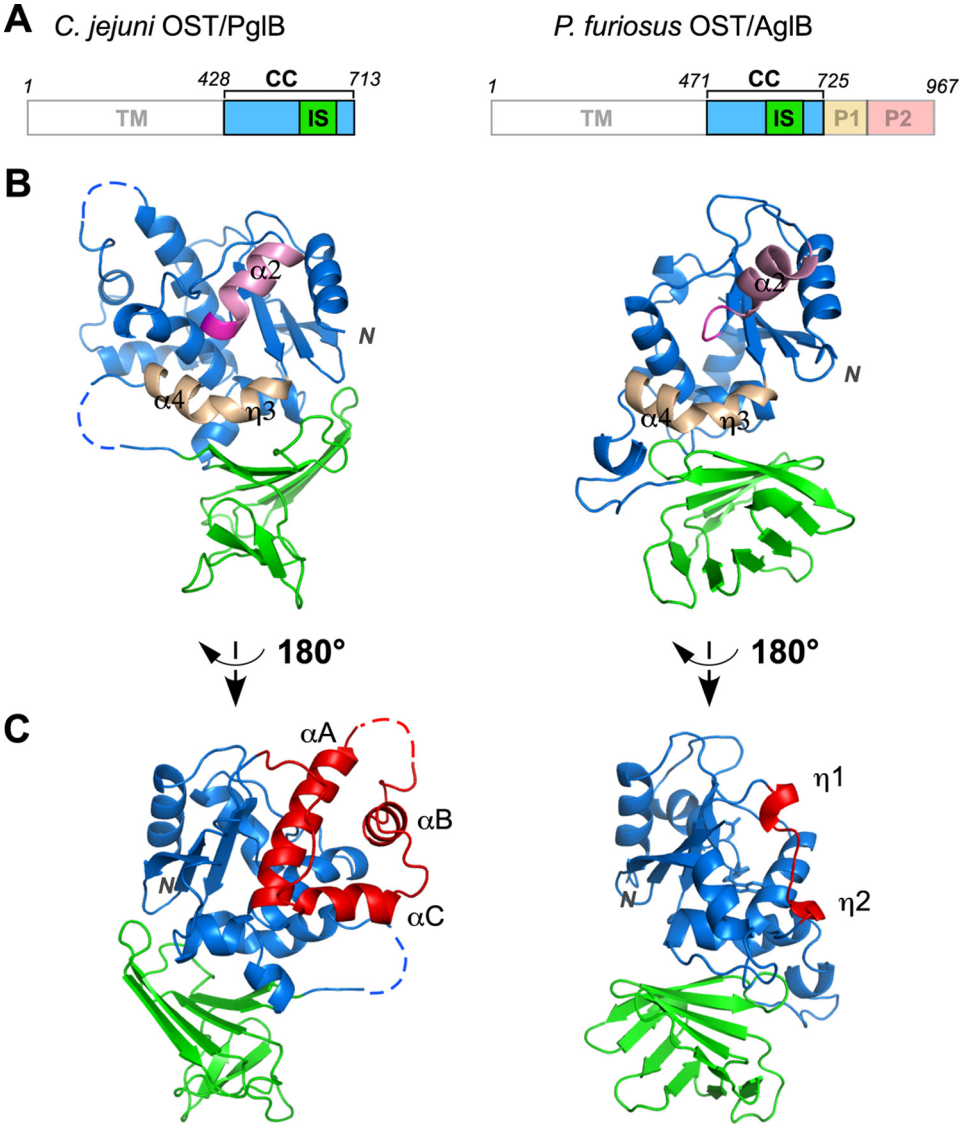


FIGURE 2. **Comparison of the structures of PglB and AglB.** *A*, domain structures of *C. jejuni* PglB and *P. furiosus* AglB. PglB and AglB are the catalytic subunits of the bacterial and archaeal oligosaccharyltransferases, respectively. P1 and P2 are additional domains of AglB. *B*, overall structures of the C-terminal domains of PglB (*left*) and AglB (*right*). The P1 and P2 domains of AglB were omitted for clarity. The WWDYG motif is *colored magenta*. The helix following the WWDYG motif is *colored pink*. The long, kinked helix that constitutes the active site of OST is *colored light brown*. *C*, different view from *B*. The portion of the CC domain of PglB that lacks a counterpart in the structure of AglB is *colored red* (*left*). The short stretch located at the corresponding position of AglB is also *colored red* (*right*). This segment is not solvent-exposed and is actually covered by the P1 and P2 domains in AglB.

assay tube was sonicated in a bath-type sonicator (120 watts) for 5 min. One microliter of 30 μM carboxytetramethylrhodamine-peptide solution and 3 μl of the *E. coli* membrane fraction containing the full-length PglB or its mutants were added. The acceptor peptide sequence (carboxytetramethylrhodamine-Ala-Asp-Gln-Asn-Ala-Thr-Tyr-Lys-COOH, 8 residues; the glycosylation consensus is underlined) was an optimized sequence for PglB (42). The reaction mixture (total 12 μl) was incubated for 2 h at 37 °C. The reaction was stopped by the addition of 2.4 μl of 5× SDS sample buffer. The fluorescence image of the SDS-polyacrylamide gel was recorded with an LAS-3000 multicolor image analyzer (Fuji Film) and was quantified using the Image-Gauge software (Fuji Film).

RESULTS AND DISCUSSION

Overall Structure of the C-terminal Globular Domain of PglB—The *C. jejuni* oligosaccharyltransferase is a single-subunit membrane protein called PglB. It consists of 11 deduced transmembrane helices in the N-terminal half (427 residues) and a globular domain in the C-terminal half (286 residues) of the primary sequence (Fig. 1*A*). The C-terminal globular domain was expressed as a fusion protein in *E. coli* and was purified after the removal of the GST tag. The crystal structure of the C-terminal domain of *C. jejuni* PglB was determined at 2.8 Å resolution (Fig. 1). A typical electron density is shown in supplemental Fig. S2. The overall structure can be described as one central core domain (CC domain; *blue*) with an inserted β-structure (IS domain; *green*). A cation (*yellow sphere*) binds to the CC domain.

A comparison with the previously determined structure of the archaeal oligosaccharyltransferase, *P. furiosus* AglB (31), revealed a striking level of folding similarity, despite their low sequence homology. The C-terminal globular domain (497 residues) of the AglB protein is much longer than that (286 residues) of the PglB protein (Fig. 2*A*). The AglB protein consists of four structural domains, CC, IS, peripheral domain 1 (P1), and peripheral domain 2 (P2). The P1 and P2 domains are both β-sheet-rich domains and encircle the CC domain. The counterparts of the P1 and P2 regions in AglB are missing in the PglB protein.

FIGURE 1. **Crystal structure of the C-terminal globular domain of PglB.** *A*, domain structure of *C. jejuni* PglB. *TM*, transmembrane domain; *CC*, central core domain, residues 428–592 plus 675–713; *IS*, insertion domain, residues 593–674. The positions of the *XX*D motif, WWDYG motif, and MI motif are indicated by *yellow*, *magenta*, and *blue asterisks*, respectively. *B*, stereo view of the overall structure of the C-terminal domain of PglB (residues 436–713). *Dashed lines* indicate the residues that were not included in the final model because of incomplete electron density. The WWDYG motif is *colored magenta*. The bound metal cation is shown as a *yellow sphere*. The side-chain carboxylate group of Glu^{451} and the backbone carbonyl group of Thr^{557} are involved in the coordination. This metal ion is not directly involved in the catalysis, because it is far away from the catalytic site. Note that the position of the bound metal ion in PglB is different from that in the previously determined structure of AglB. *C*, different view from *B*. *D*, same view as *B* but with the secondary structure elements defined in supplemental Fig. S4. The side chains of the WWDYG motif (*magenta*), MI motif (*blue*), and the conserved Asp (*yellow*; see supplemental Fig. S1) were drawn.

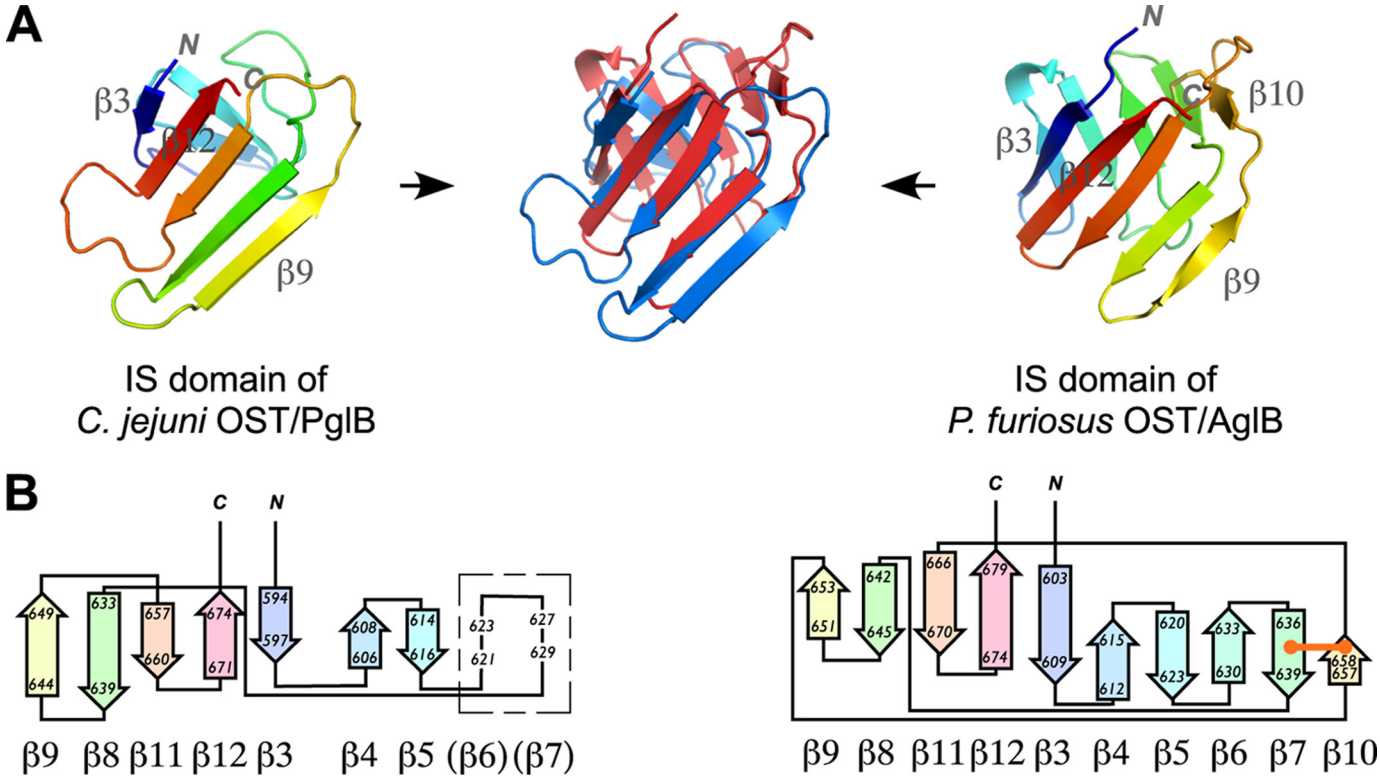


FIGURE 3. **Structural comparison of the IS domains of PglB and AglB.** *A*, the overlaid structures (blue, PglB; *red*, AglB) are placed in between the IS structures of PglB (*left*) and AglB (*right*). *B*, topology diagram of the IS domains, generated by the PDBsum server. The location of the two putative β-strands suggested by structural similarity between PglB and AglB in a region with high temperature factors is shown in a *box*. The disulfide bond is shown as an *orange line*.

Interestingly, the C terminus of the PglB protein exactly corresponds to the domain boundary between CC and P1 of AglB. The structural superposition of the CC + IS portion resulted in a Cα root mean square deviation value of 3.0 Å, despite sequence identity as low as 13%, over the 167 aligned residues (supplemental Fig. S3). The high overall similarity of the CC + IS structure facilitated the structure-aided sequence alignment (supplemental Fig. S4). The short stretch containing two successive 3_{10} helices, η1 and η2, of AglB is replaced by a long sequence containing three α-helices, αA, αB, and αC, in PglB (Fig. 2*C*, *red*). This long sequence forms a lining structure on one side of the CC domain in PglB, and its shorter counterpart is covered by the P1 and P2 domains in AglB.

Comparison of the IS Domains—The IS domain of PglB is clearly the counterpart of the IS domain of AglB (Fig. 2). The Cα root mean square deviation is 2.4 Å over 68 residues (Fig. 3*A*). In a previous report, we described the topology of the IS domain of *Pyrococcus* AglB as a 10-stranded antiparallel β-barrel structure (31). The reanalysis using the PDBsum server (available on the EMBL-European Bioinformatics Institute Web site), however, demonstrated that the IS domain is a "distorted barrel," because hydrogen bonds are not formed between the edge strands. Thus, we now describe the IS domain of *Pyrococcus* AglB as a barrel-like structure. The IS domain of PglB is regarded as a more distorted barrel, because there are two gaps in the β-β contacts. We describe the IS domain of *Campylobacter* PglB as a β-sandwich structure. Fig. 3*B* shows the topology diagram of the IS domains of the two structures. The IS domain of PglB contains seven β-strands, but we infer two additional strands in the region with high temperature factors, considering the overall structural similarity between PglB and AglB. We thus describe the IS domain of PglB as a nine-stranded structure consisting of almost antiparallel strands but with one parallel orientation. The AglB IS structure contains an extra, 10th strand, which is connected to the neighboring strand by a disulfide bond. The contact area between the CC and IS domains is about 2,000 $Å^2$ in the PglB structure and 1,500 $Å^2$ in the AglB structure. The wide contacts suggest that the relative orientation of the two domains is fixed in solution as well as in the crystal.

It is reasonable to assume that the IS domain has some important functions, considering the conservation of the IS domain across the two domains of life. A similarity search based on secondary structure matching (available on the EMBL-European Bioinformatics Institute Web site) for the PDB and SCOP data bases generated several β-barrel/sandwich-containing proteins. Some of them are sugar-binding proteins, but most of the sugar-binding sites are found in structures other than β-barrel/sandwich structures. One interesting example is the N-terminal three tandem β-barrel domains of the cation-independent mannose 6-phosphate receptor. The crystal structure revealed that the third barrel domain has a binding site for a mannose 6-phosphate molecule (43). In the three-dimensional structure, the IS domain is far away from the putative catalytic center in the CC domain (Fig. 2). We postulate that the IS domain recognizes the distal end sugar residues of the lipid-linked oligosaccharide substrates for the selection of *in vivo* oligosaccharide donors in Bacteria and Archaea. It would be interesting to determine whether a similar β-barrel/sandwich structure is present in the eukaryotic STT3 protein. Secondary

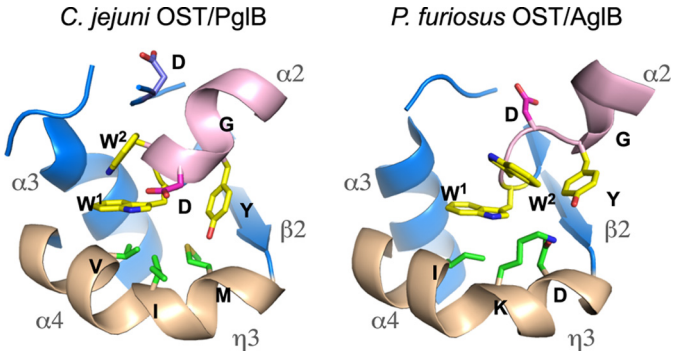


FIGURE 4. **Close-up views of the putative catalytic sites of PglB and AglB.** The side chains of the two Trp residues and one Tyr residue of the WWDYG motif are *colored yellow*, and that of the Asp residue is shown in *red*. The helix α2 following the WWDYG motif is *colored pink*. The long, kinked helix, η3 + α4, containing the DK or MI motif is *colored light brown*. The side chains of the DK and MI motifs are *green*. The side chain of the Asp residue conserved in the PglB protein family (Asp^{475}) is also shown in the *left panel*.

structure predictions, however, did not predict any repetitive short strands, but rather helices, in the corresponding region in the eukaryotic STT3 proteins, such as those from yeast and *Trypanosoma*. We suggest that a different helical structure of the STT3 protein or other subunit(s) performs the same function in the eukaryotic OST. The structure determination of a eukaryotic STT3 will clarify this point.

Comparison of the Putative Catalytic Sites—We focused our attention on the catalytic center of PglB and compared it with that of AglB. The conserved WWDYG motif (*magenta*) is located at the N terminus of the helix α2 (*pink*) and is close to the kinked helix η3 + α4 (*light brown*), as shown in Figs. 1 and 2. The putative catalytic site is formed between the WWDYG motif and the kinked helix (Fig. 4). The side chains of the Trp^1 residue and the Tyr residue of the WWDYG motif occupy the same positions with similar side chain orientations in the two structures, but the side chains of the Trp^2 and Asp residues protrude in opposite directions. The N-terminal part of the WWDYG motif in AglB adopts a rare left-handed helical conformation, which is apparently stabilized by the neighboring molecule in the crystals. By contrast, the corresponding part forms a typical right-handed helix in PglB. This mirror image relationship of the two conformations results in the opposite directions of the side chains of the Trp^2 and Asp residues (Fig. 4) and consequently the different orientations of the helix α2 (Fig. 2*B*). Without the crystal packing effect, the directions of the side chains of the Trp^2 and Asp residues in AglB would be the same as those found in PglB. This consideration implies that the correct configuration of the catalytic residues is that found in the PglB crystal. Alternatively, the two conformations of the WWDYG motif in the two structures reflect some plasticity of the active site structure, which enables large conformational changes during the OST reaction cycle.

A pair of Asp and Lys residues is present on the kinked helix in the AglB structure (Fig. 4). Instead of these residues, Met and Ile residues occupy the same positions in PglB. This finding prompted us to reexamine the multiple sequence alignment of the STT3/PglB/AglB protein family in more detail. We confirmed that the D*XX*K sequence is widely conserved in Eukarya and a portion of Archaea (supplemental Fig. S5). This motif was proved to be catalytically important by *in vivo* mutational studies; the substitution of the Asp and Lys residues in the motif each led to a lethal phenotype in yeast Stt3p and *L. major* STT3-1 (14, 31). The archaeal domain can be divided into two major phyla, Crenarchaeota and Euryarchaeota. All of the AglB proteins from Crenarchaeota and the class Thermococci (including *Pyrococcus*) of Euryarchaeota contain the typical DK motif. The D*XX*K sequence may be extended to D*XXKXXX*(M/I). The AglB proteins of other classes of Euryarchaeota possess a variant of the original DK motif, D*XXMXXX*(K/I). We refer to this as the "DM" motif. The three residues that define the motifs face the same side of the kinked helix, and thus the side chains come into close proximity (Fig. 4). This situation makes the second and third positions in the patterns interchangeable, and thus the DK and DM motifs are regarded as being functionally identical in OST catalysis. Some AglBs seem to contain a relaxed version, D*XXMX*(5,13)K, where *X*(*m*,*n*) denotes that at least *m* and at most *n* residues of any type may occur at this position, but the significance is not clear. There is the third set of classes in Euryarchaeota, in which the AglBs do not contain either the DK motif or the DM motif. We found that this archaeal group AglB possesses a pattern similar to that found in the bacterial PglB, M*XX*(K/I)*XXX*W. The archaeal and bacterial patterns can be combined into M*XXIXXX*(I/V/W), which we call the "MI" motif. Since the chemical properties of the side chains of the MI motif are very different from those of the DK/DM motif, the functional commonality is currently unclear. In summary, a careful sequence alignment of the STT3/PglB/AglB protein family revealed three types of conserved motifs on the kinked helix in the putative active site by reference to the two crystal structures of PglB and AglB.

Apart from the WWDYG motif, the sequence alignment of the PglB protein family indicates two strictly conserved residues, a conserved Asp and a conserved Gln (supplemental Fig. S1). The Asp residue is located near the catalytic site (Fig. 4), and the Gln residue is situated on the opposite face of the molecule. It is conceivable that the conserved Asp residue is involved in the catalysis, but the direction of the side chain is opposite to the catalytic site. Future studies are needed to uncover possible roles of these conserved residues in the structure and function of PglB.

Another Catalytic Motif in the Transmembrane Region—STT3/PglB/AglB is a new member of the glycosyltransferase GT-C superfamily (44). The members of the GT-C clan (CL0111 in the Pfam data base, available on the World Wide Web) are diverse glycosyltransferases that possess 8–13 predicted transmembrane helices and a D*X*D signature in the first lumenal/extracellular loop. This motif is also found in a mannosyltransferase, which catalyzes the *O*-mannosylation of proteins using D-mannose-P-dolichol as a sugar donor. Therefore, the role of this motif is estimated to be the binding to dolichyl-(pyro)phosphate via a bound metal ion, and thus it is indirectly involved in the catalysis. An E*X*D sequence is found in STT3 and a subset of AglB, belonging to Crenarchaeota (supplemental Fig. S6). A D*X*D sequence exists in another subset of AglB, belonging to the class Thermococci. The conservation of the

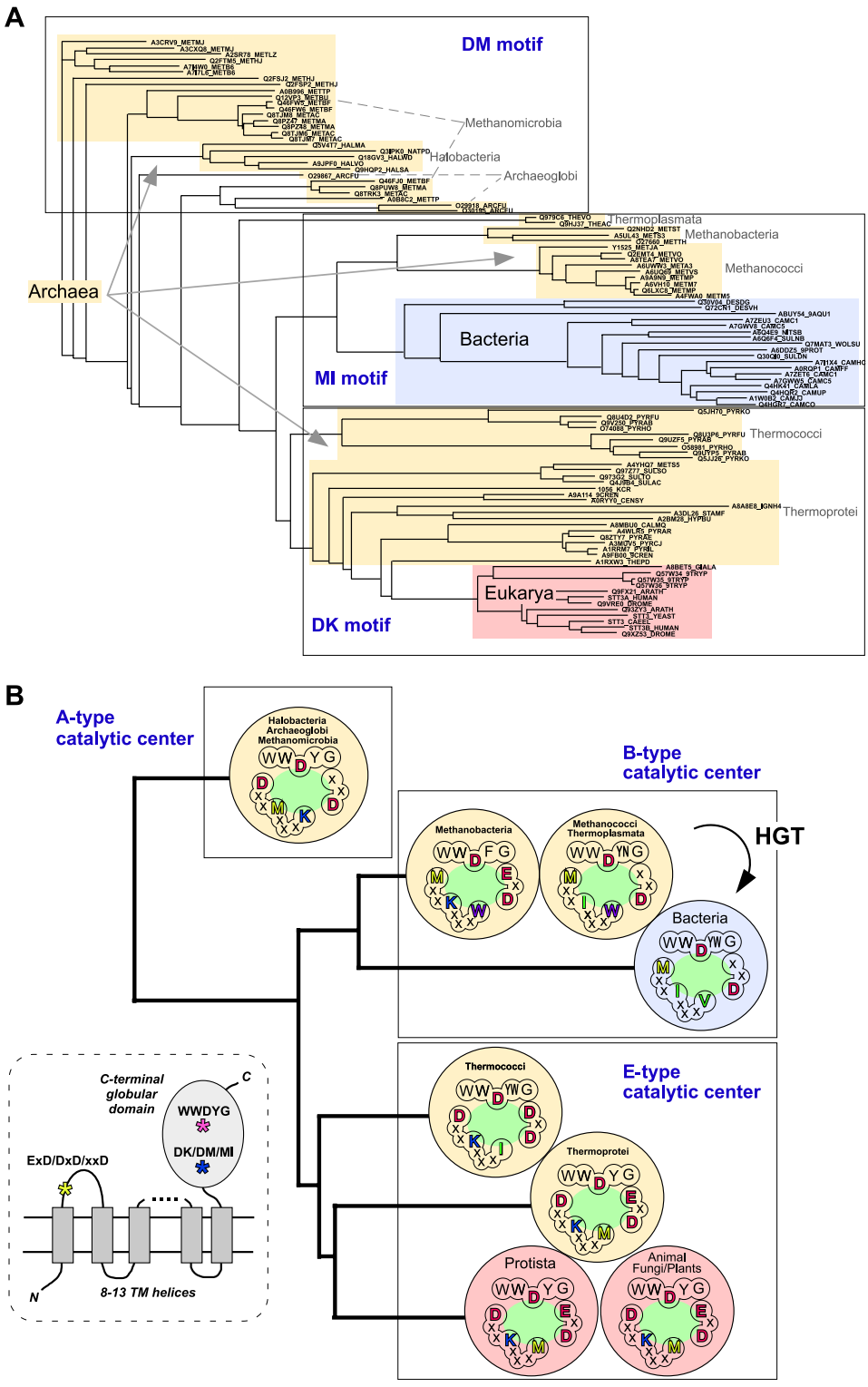


FIGURE 5. **Three types of catalytic centers in the oligosaccharyltransferases.** *A*, phylogenetic tree of the STT3/PglB/AglB proteins, based on the full-length sequences. The multiple-sequence alignment was performed by the program *MAFFT*, and the results were rendered by the program *ATV* (available on the Phylosoft Web site). The STT3/PglB/AglB entries in the phylogenetic tree can be classified into the three groups in the *boxes*, according to their type of DK/DM/MI motif. *B*, the catalytic centers (*green oval*) can be grouped into three types, assuming that the catalytic center of the OST enzyme is composed of the amino acid residues that define the three conserved motifs, WWDYG, DK/DM/MI, and EXD/DXD/XXD. The distribution of the three types is superimposed on the phylogenetic tree of STT3/PglB/AglB. Note that multiple STT3/PglB/AglB paralogs in an organism contain the same type of catalytic center, in principle. Eukaryotic STT3 and bacterial PglB exclusively have the E-type and B-type catalytic centers, respectively. Archaeal AglB is divided into three groups, which have either the A-, B-, or E-type catalytic center. Bacteria probably acquired the *pglB* gene by horizontal gene transfer (*HGT*) from the archaeal B-type. The *inset* shows the overall topology of the STT3/PglB/AglB proteins and the locations of the three catalytic motifs.

first Asp is not strict in the remainder of the AglB proteins and all of the PglB proteins, and thus the signature pattern must be relaxed to *XXD*, but the Asp at the third position is absolutely conserved in all STT3/PglB/AglB proteins. The importance of the diacidic motifs was shown by *in vivo* mutational studies of the EXD motif in yeast Stt3p and *L. major* STT3–1; the substitution of the Asp residue in the motif led to a lethal phenotype in yeast Stt3p, and that of the Glu and Asp residues each had the same lethal effect in *L. major* STT3–1 (14).

Three Types of OST Catalytic Centers—Fig. 5*A* shows a phylogenetic tree constructed from the full-length sequences of the STT3/PglB/AglB proteins. The multiple sequence alignment is difficult to interpret on a residue-by-residue basis due to the very low sequence homology, but the construction of a phylogenetic tree using the overall similarities is significant. Unexpectedly, the entries from the archaeal domain reside on different branches of the phylogenetic tree, as shown in the three boxes in Fig. 5*A*. This seemingly complicated situation can be simplified by identifying all of the entries in each box with the same type of DK/DM/MI motif. The clear grouping on the phylogenetic tree, based on the full-length sequences, confirms the significance of the conservation of the DK/DM/MI motif. We assumed that the catalytic center of the OST enzyme consists of the three short motifs. First, the catalytic motif, WWDYG, is highly conserved, but a variation is seen at the fourth position; Tyr is replaced by Trp, Asn, or Phe in some organisms (supplemental Fig. S5). Second, the DK, DM, or MI motif resides on the kinked helix in the active site. Third, the EXD, DXD, or *XXD* motif exists in the loop region that connects the first and second transmembrane helices. We classified the catalytic centers of the OST enzymes into three types (Fig. 5*B*). The E-type catalytic center consists of the

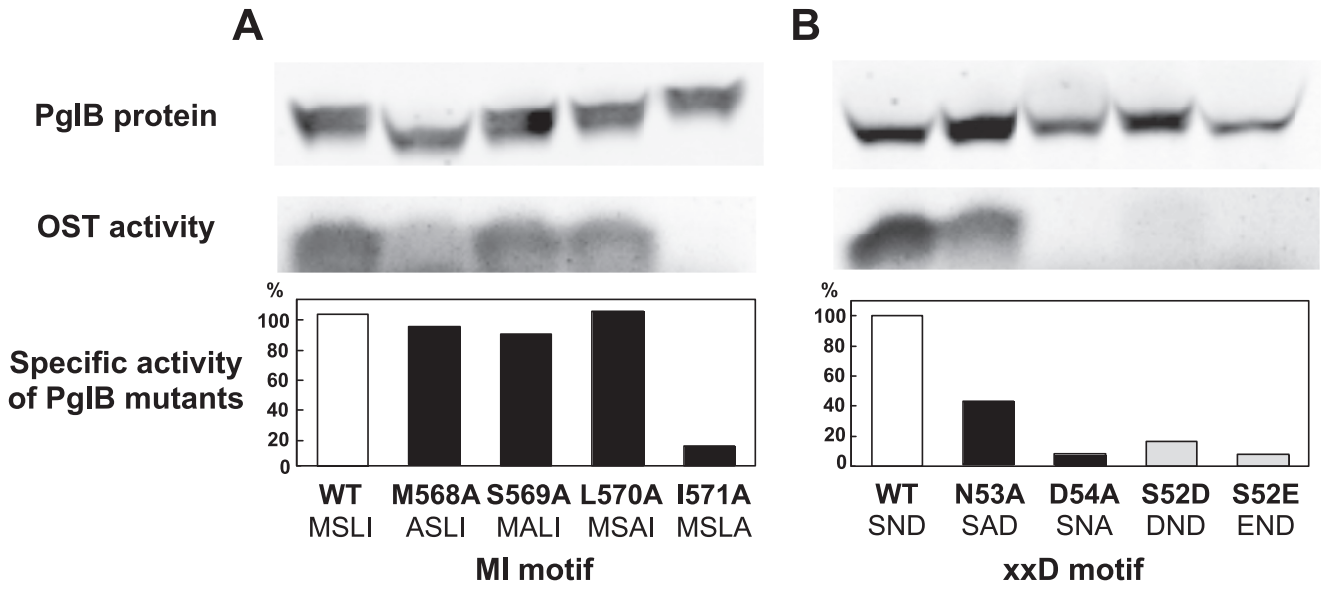


FIGURE 6. **Effect of amino acid substitutions on the oligosaccharide transfer activity of *C. jejuni* PglB.** *A*, mutations of the MI motif (residues Met^{568}-Ser^{569}-Leu^{570}-Ile^{571}); *B*, mutations of the *XX*D motif (Ser^{52}-Asn^{53}-Asp^{54}). The full-length PglB, with the wild-type sequence and its mutated sequences, was expressed in *E. coli*, and the Triton X-100-solubilized membrane fractions were prepared. The amount of PglB in the membrane fractions was quantified by a Western blot analysis, using an anti-His tag antibody. It was essential to use a secondary antibody labeled with an infrared fluorescent dye for the accurate quantification and a wide linear dynamic range. The oligosaccharide transfer assay was performed by the PAGE method (41). The assay solution was the mixture of the membrane fraction containing PglB, crude lipid-linked oligosaccharide donors (*cLLO*) extracted from *C. jejuni* cells, and an acceptor peptide containing the Asp-Gln-Asn-Ala-Thr sequence and a fluorescent dye, TAMRA. The fluorescence image of the SDS-polyacrylamide gel was recorded and quantified. The specific activity was calculated as the percentage of the wild-type PglB.

WWD(Y/W)G, DK, and (E/D)*X*D motifs. The B-type catalytic center consists of the WWD(Y/W/N/F)G, MI, and *XX*D motifs. Finally, the A-type catalytic center consists of the WWDYG, DM, and *XX*D motifs. Eukaryotic STT3 exclusively has the E-type catalytic center, whereas the bacterial PglB contains the B-type only. Archaeal AglB is divided into three groups, bearing either the A-, B-, or E-type catalytic center. The E-type catalytic center must have a very ancient origin, before the divergence of Archaea and Eukarya, because the E-type exists in both the archaeal and eukaryal domains. In contrast, the B-type catalytic center in Bacteria probably originated from horizontal gene transfer from the B-type Archaea, considering the limited distribution of the *N*-glycosylation system in the bacterial domain.

Site-directed Mutagenesis to Analyze Amino Acid Importance in the Catalytic Motifs—We used site-directed mutagenesis to address the role of the newly identified MI motif. Because the C-terminal globular domain alone has no catalytic activity, the full-length PglB protein was expressed in recombinant *E. coli* membrane fractions and used for an *in vitro* oligosaccharide transfer assay. We mutated four consecutive residues in the MI motif (Fig. 6*A*). The three alanine mutations, including the Met residue, had nearly the same specific activities as the wild type PglB, but the alanine mutation of the Ile residue led to substantially reduced activity.

We also tested the alanine mutations of the *XX*D motif (Fig. 6*B*), which resides in the first loop in the transmembrane region (*inset* of Fig. 5). The substitution of the second residue in the *XX*D motif, Asn^{53}, retained moderate activity, but the mutation of the Asp residue resulted in nearly complete loss of the activity. Together with the inactive mutations in the WWDYG motif reported previously (16, 17), the three catalytic motifs all contribute to the constitution of the B-type catalytic center of PglB. Finally, we converted the *XX*D motif to the D*X*D and E*X*D motifs, unique to the E-type catalytic center. These changes led to the substantial loss of the oligosaccharide transfer activity (Fig. 6*B*). This implies that intimate coordination between the three catalytic motifs is necessary to constitute the functional catalytic centers of OST.

Comparative Structure Biology of OST in the Three Domains of Life—The comparative biology viewpoint (Fig. 5) provides some interesting ideas for future experiments to elucidate the catalytic mechanism of OST. (i) Two crystal structures of the C-terminal globular domain of OST are now available: one from *Campylobacter jejuni* PglB and the other from *Pyrococcus furiosus* AglB. As a eukaryotic model, the *Pyrococcus* structure is more suitable than the *Campylobacter* structure, because the former has the E-type catalytic center. (ii) Some archaea, such as Halobacteria, contain the A-type catalytic center. The structure determination is desired because no crystal structure of this type is available. (iii) Some genomes encode more than two STT3/PglB/AglB paralogs. With few exceptions, the paralogs in an organism have the same type of catalytic center. It would be interesting to investigate why a certain organism with multiple OSTs does not have those with different types of catalytic centers.

Conclusions—The structural comparison of two distantly related catalytic subunits of the oligosaccharyltransferases, bacterial PglB and archaeal AglB, highlighted the common architecture of the catalytic center, an inserted β-structure domain, and new, conserved catalytic motifs, beyond sequence comparison. The discovery of the new motifs was impossible without the identification of the equivalent positions in the catalytic center by reference to the structures, because the amino acid residues found at the equivalent posi-

tions have vastly different chemical properties. The site-directed mutagenesis study demonstrated the involvement of the newly identified MI motif in the oligosaccharide transfer reaction catalyzed by PglB as well as the DK motif in the eukaryotic STT3. We propose that the new DK/DM/MI motifs, in addition to the known WWDYG and E*X*D/D*X*D/*XX*D motifs, compose the three distinct types of catalytic centers of the OST enzyme (Fig. 5*B*). This novel classification will provide a useful framework for understanding the common and distinct enzymatic properties of the OST enzymes from Eukarya, Archaea, and Bacteria.

Acknowledgments—We thank Masumi Otsu (Laboratory for Technical Support, Medical Institute of Bioregulation, Kyushu University) for the DNA sequencing and Drs. Kaori Sasaki and Katsumi Maenaka for assistance in synchrotron data collection. We thank the staff at beamlines BL-38B1 and BL-44XU of SPring-8 and at beamlines BL-5A and AR-NW12A of the Photon Factory.

REFERENCES

1. Spiro, R. G. (2002) *Glycobiology* **12,** 43R–56R
2. Yan, A., and Lennarz, W. J. (2005) *J. Biol. Chem.* **280,** 3121–3124
3. Kelleher, D. J., and Gilmore, R. (2006) *Glycobiology* **16,** 47R–62R
4. Gavel, Y., and von Heijne, G. (1990) *Protein Eng.* **3,** 433–442
5. Knauer, R., and Lehle, L. (1999) *Biochim. Biophys. Acta* **1426,** 259–273
6. Schwarz, M., Knauer, R., and Lehle, L. (2005) *FEBS Lett.* **579,** 6564–6568
7. Li, H., Chavan, M., Schindelin, H., Lennarz, W. J., and Li, H. (2008) *Structure* **16,** 432–440
8. Chavan, M., Chen, Z., Li, G., Schindelin, H., Lennarz, W. J., and Li, H. (2006) *Proc. Natl. Acad. Sci. U.S.A.* **103,** 8947–8952
9. Yan, Q., and Lennarz, W. J. (2002) *J. Biol. Chem.* **277,** 47692–47700
10. Koiwa, H., Li, F., McCully, M. G., Mendoza, I., Koizumi, N., Manabe, Y., Nakagawa, Y., Zhu, J., Rus, A., Pardo, J. M., Bressan, R. A., and Hasegawa, P. M. (2003) *Plant Cell* **15,** 2273–2284
11. Kelleher, D. J., Karaoglu, D., Mandon, E. C., and Gilmore, R. (2003) *Mol. Cell* **12,** 101–111
12. Nasab, F. P., Schulz, B. L., Gamarro, F., Parodi, A. J., and Aebi, M. (2008) *Mol. Biol. Cell* **19,** 3758–3768
13. Izquierdo, L., Schulz, B. L., Rodrigues, J. A., Güther, M. L., Procter, J. B., Barton, G. J., Aebi, M., and Ferguson, M. A. (2009) *EMBO J.* **28,** 2650–2661
14. Hese, K., Otto, C., Routier, F. H., and Lehle, L. (2009) *Glycobiology* **19,** 160–171
15. Castro, O., Movsichoff, F., and Parodi, A. J. (2006) *Proc. Natl. Acad. Sci. U.S.A.* **103,** 14756–14760
16. Wacker, M., Linton, D., Hitchen, P. G., Nita-Lazar, M., Haslam, S. M., North, S. J., Panico, M., Morris, H. R., Dell, A., Wren, B. W., and Aebi, M. (2002) *Science* **298,** 1790–1793
17. Glover, K. J., Weerapana, E., Numao, S., and Imperiali, B. (2005) *Chem. Biol.* **12,** 1311–1315
18. Feldman, M. F., Wacker, M., Hernandez, M., Hitchen, P. G., Marolda, C. L., Kowarik, M., Morris, H. R., Dell, A., Valvano, M. A., and Aebi, M. (2005) *Proc. Natl. Acad. Sci. U.S.A.* **102,** 3016–3021
19. Chaban, B., Voisin, S., Kelly, J., Logan, S. M., and Jarrell, K. F. (2006) *Mol. Microbiol.* **61,** 259–268
20. Abu-Qarn, M., and Eichler, J. (2007) *Archaea* **2,** 73–81
21. Ruiz-Canada, C., Kelleher, D. J., and Gilmore, R. (2009) *Cell* **136,** 272–283
22. Wilson, C. M., Roebuck, Q., and High, S. (2008) *Proc. Natl. Acad. Sci. U.S.A.* **105,** 9534–9539
23. Wilson, C. M., and High, S. (2007) *J. Cell Sci.* **120,** 648–657
24. Kelleher, D. J., Banerjee, S., Cura, A. J., Samuelson, J., and Gilmore, R. (2007) *J. Cell Biol.* **177,** 29–37
25. Schulz, B. L., Stirnimann, C. U., Grimshaw, J. P., Brozzo, M. S., Fritsch, F., Mohorko, E., Capitani, G., Glockshuber, R., Grütter, M. G., and Aebi, M. (2009) *Proc. Natl. Acad. Sci. U.S.A.* **106,** 11061–11066
26. Kim, H., von Heijne, G., and Nilsson, I. (2005) *J. Biol. Chem.* **280,** 20261–20267
27. Voisin, S., Houliston, R. S., Kelly, J., Brisson, J. R., Watson, D., Bardy, S. L., Jarrell, K. F., and Logan, S. M. (2005) *J. Biol. Chem.* **280,** 16586–16593
28. Eichler, J. (2003) *Microbiology* **149,** 3347–3351
29. Szymanski, C. M., and Wren, B. W. (2005) *Nat. Rev. Microbiol.* **3,** 225–237
30. Nakagawa, S., Takaki, Y., Shimamura, S., Reysenbach, A. L., Takai, K., and Horikoshi, K. (2007) *Proc. Natl. Acad. Sci. U.S.A.* **104,** 12146–12150
31. Igura, M., Maita, N., Kamishikiryo, J., Yamada, M., Obita, T., Maenaka, K., and Kohda, D. (2008) *EMBO J.* **27,** 234–243
32. Igura, M., Maita, N., Obita, T., Kamishikiryo, J., Maenaka, K., and Kohda, D. (2007) *Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun.* **63,** 798–801
33. Walter, T. S., Meier, C., Assenberg, R., Au, K. F., Ren, J., Verma, A., Nettleship, J. E., Owens, R. J., Stuart, D. I., and Grimes, J. M. (2006) *Structure* **14,** 1617–1622
34. Otwinowski, Z., and Minor, W. (1997) *Methods Enzymol.* **276,** 307–326
35. Vonrhein, C., Blanc, E., Roversi, P., and Bricogne, G. (2007) *Methods Mol. Biol.* **364,** 215–230
36. Collaborative Computational Project 4 (1994) *Acta Crystallogr. D Biol. Crystallogr.* **50,** 760–763
37. Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T., and Warren, G. L. (1998) *Acta Crystallogr. D Biol. Crystallogr.* **54,** 905–921
38. Emsley, P., and Cowtan, K. (2004) *Acta Crystallogr. D Biol. Crystallogr.* **60,** 2126–2132
39. Standley, D. M., Toh, H., and Nakamura, H. (2005) *BMC Bioinformatics* **6,** 221
40. Katoh, K., Misawa, K., Kuma, K., and Miyata, T. (2002) *Nucleic Acids Res.* **30,** 3059–3066
41. Kohda, D., Yamada, M., Igura, M., Kamishikiryo, J., and Maenaka, K. (2007) *Glycobiology* **17,** 1175–1182
42. Chen, M. M., Glover, K. J., and Imperiali, B. (2007) *Biochemistry* **46,** 5579–5585
43. Olson, L. J., Dahms, N. M., and Kim, J. J. (2004) *J. Biol. Chem.* **279,** 34000–34009
44. Liu, J., and Mushegian, A. (2003) *Protein Sci.* **12,** 1418–1431