Structural elucidation of asparagine-linked oligosaccharides from three hyperthermophilic archaeal species

藤浪, 大輔

出版情報：九州大学, 2016, 博士（システム生命科学）, 課程博士
バージョン：
権利関係：

Structural elucidation of asparagine-linked oligosaccharides from three hyperthermophilic archaeal species

Daisuke Fujinami

2016

Division of Structural Biology,
Graduate school of Systems Life Sciences,
Kyushu University, Fukuoka, Japan

Acknowledgements

I would like to express this sincere to Professor Daisuke Kohda (Division of Structural Biology, Graduate school of Systems Life Science, Kyushu university,) for his cordial guidance, valuable discussion, and kind encouragement throughout the study.

I also wished to express Professor Kenji Sonomoto and Dr. Takuya Noguchi (Division of Applied Molecular Microbiology and Biomass Chemistry, Department of Bioscience and Biotechnology, Faculty of Agriculture, Graduate School, Kyushu University) for giving kindly access to Gas chromatography, Associate Professor Masaki Matsumoto (Division of Cell Biology, Department of Molecular and Cellular Biology, Medical Institute of Bioregulation, Kyushu University) for giving access to MASS spectrometry. My thanks also go to Dr. Shunsuke Matsumoto (Faculty of Life Sciences, Kyoto Sangyo University), Dr. James Nyirenda (Department of Chemistry, School of Natural Sciences, The University of Zambia), for the valuable time they rendered in teaching biological experiment kindly, Associate Professor Takashi Saitoh (Division of Application Pharmacy, School of Pharmacy, Hokkaido Pharmaceutical University), and Mr. Yuya Taguchi for assisting me in the NMR measurement and MS measurement, Ms. Mizuho Oda for the preparation of MASCOT sample, Professor Shinsuke Sando and Dr. Hiroshi Nonaka (Department of Chemistry and Biotechnology, Graduate School of Engineering, The University of Tokyo) for the preparation of hydrochloric acid in butanol.

I am grateful to our present and previous research group members Associate Professor Atsushi Shimada, Assistant Professor Kouta Mayanagi, Mr. Takahiro Yamasaki, Ms. Marie Ishikawa, Ms. Yuki Kawasaki. Ms. Siqin Bala, Mr. Shigekazu Koya, Ms. Xiling Han, Mr. Hajime Motomura, Ms. Miki Otsu, Dr. Sam-Geun Kong, Dr. Tetsuya Satoh, Dr. Akira Takano, Dr. Mayumi Igura, Dr. Rei matsuoka, Dr. Hiromi Ogino, Mr. Hiroyuki Oda, Mr. Keisei Izumi, Mr. Yuki Nakamura and Ms. Atsuko Yamaguchi.

Finally, I am great thankful to all my family and friends for their kindly support.

Prior Publication

Much of this thesis work has been published in two papers (Fujinami et al., 2014 and Fujinami et al., 2015), and represents the joint work and various contributions of the coauthors of the paper.

Publication lists

１． Structural elucidation of an asparagine-linked oligosaccharide from the hyperthermophilic archaeon, *Pyrococcus furiosus*
Carbohydrate Research, 38, 30-36 (2014)
Daisuke Fujinami, Masaki Matsumoto, Takuya Noguchi, Kenji Sonomoto, and Daisuke Kohda.

２． Structural elucidation of an asparagine-linked oligosaccharide from the hyperthermophilic archaeon, *Archaeoglobus fulgidus*
Carbohydrate Research, 413, 55-62 (2015)
Daisuke Fujinami, James Nyirenda, Shunsuke Matsumoto, and Daisuke Kohda

３． Comparative Analysis of Archaeal Lipid-linked Oligosaccharides That Serve as Oligosaccharide Donors for Asn Glycosylation.
The Journal of Biological Chemistry, 291, 11042-11054 (2016)
Yuya Taguchi, Daisuke Fujinami, and Daisuke Kohda

Abbreviations

ABEE	*p*-aminobenzoic acid ethyl ester
AglB	Archaeal glycolsylation B
COSY	Correlation Spectroscopy
CBB	Coomassie Brilliant Blue
Dol-P	Dolichol-monophosphate
Dol-PP	Dolichol-diphosphate
ESI-MS	Electrospray ionization tandem mass spectrometry
Gal	Galactose
GalNAc	N-acetylgalactosamine
Glc	Glucose
GlcNAc	N-acetylglucosamine
HSQC	Heteronuclear single quantum coherence
HMBC	Heteronuclear multiple bond coherence
LLO	Lipid-linked oligosaccharide
Man	Mannose
ManNAc	N-acetylmannosamine
NAc	N-acetyl
NMR	Nuclear Magnetic Resonace
NOE	Nuclear overhauser effect
NOESY	NOE spectroscopy
OST	Oligosaccharyltransferase
PAS	Periodic acid-Schiff
ROESY	Rotating frame NOESY
STT	Staurosporine and temperature sensitivety
TAMRA	5/6-carboxytetramethylrhodamin
TFA	Trifluoroacetic acid
TIC	Total ion chromatogram
TOCSY	Totally correlated spectroscopy
Tris	tris(hydroxylmethyl)aminomethane
Triton X-100	t-octylphenoxypolyethoxyethanol
XIC	Extracted ion chromatogram
Xyl	Xylose

Contents

Abstract

N-glycosylation is one of the most common post-translational protein modifications, occurring in all three domains of life, eukaryote, archaea and eubacteria. Oligosaccharyltransferase (OST) is a transmembrane enzyme, which catalyzes the transfer of an oligosaccharide chain preassembled on lipid carrier as a lipid-linked oligosaccharide (LLO) to the side chain amide nitrogen of acceptor asparagine residue on the consensus sequon, Asn-X-Thr/Ser (where X is any amino acid except for proline). Reflecting the essential roles of N-glycosylation in eukaryotes, the canonical core glycan structure, $Glc_3Man_9GlcNAc_2$-Asn, is well conserved in eukaryotic species. On the other hand, the glycan structures in archaea and eubacteria show a far greater variety of monosaccharide composition and their linkages.

In previous research, the crystal structures of archaeal OST from *Archaeoglobus fulgidus*, *Pyrococcus furiosus* have been determined and provide insight into the acceptor glycosylation sequon recognition. However, the donor LLO recognition mechanism is unclear due to lacks of donor LLO chemical structure information. Here, we elucidated three N-linked oligosaccharide structures from three hyperthermophilic archaeons, *A. fulgidus*, *P. furiosus,* and *Pyrobaculum calidifontis* by NMR to reveal the donor LLO recognition mechanism of OST.

Glycopeptide samples for NMR analyses were produced by *in vitro* oligosaccharide-transfer reaction carried out by addition of a structurally defined acceptor peptide into detergent solubilized archaeal membrane fractions, which contains both of oligosaccharyltransferase and LLO. To facilitate the NMR analysis, the archaeal cells were grown in a rich medium supplemented with ^{13}C-glucose for LLOs metabolically labeling.

The N-linked oligosaccharide of *A. fulgidus* and *P. furiosus* showed unique structures: in *A. fulgidus* N-glycan, the reducing end monosaccharide residue that directly linked to the Asn residue is glucose, which is not very common since the reducing end monosaccharide residue usually possesses an N-acetyl group at the C-2 position. On the other hand, the *P. furiosus* N-linked oligosaccharide has two blanching xylose residues at third mannose and fourth glucose residues. Regarding *P. calidifontis* N-linked oligosaccharide, which contains nine mannoses, one di-N-acetyl glucosamine and one di N-acetyl glucuronic acid residues, shares the typical signature of eukaryotic N-glycan: blanching mannose substructure and chitobiose core ($GlcNAc_2$-Asn) at reducing end.

The comparison of the archaeal N-linked oligosaccharide structures with eukaryotic ones

provides insight into the evolution of the N-glycosylation system, such as the N-glycan biosynthesis and the N-glycan processing. The obtained chemical structure information makes the donor LLO chemical synthesis and the investigation of OST-LLO interaction possible.

Chapter 1-Introduction

1.1

N-glycosylation in eukaryote

N-glycosylation is one of the common post-translational modifications of membrane and secretory proteins, existing in all three domains of life, eukaryote, archaea, and eubacteria (*1*,*2*). In the eukaryote, this essential and highly conserved modification takes place in 50% of the total proteins on the inner face of the endoplasmic reticulum (ER) membrane, coupling with the ribosomal protein translocation via translocation channel (translocon) formed from Sec proteins (*3*) (Figure 1-1). Eukaryotic N-glycosylation has been well studied in yeast and human. Yeast oligosaccharyltransferase (OST) is membrane-bound enzyme, which composed of eight different membrane-spanning proteins, and catalyzes the transfer of the tetradecasaccharide ($Glc_3Man_9GlcNAc_2$-) preassembled on a lipid carrier as a lipid-linked oligosaccharide (LLO) to the amide side nitrogen of the asparagine residue in the consensus sequon, Asn-X-Thr/Ser (where X is any amino acid residues except for proline). The LLO is synthesized by a series of membrane-bound glycosyltransferases in the *Alg* family (asparagine-linked glycosylation), which catalyze the transfer of the each monosaccharide residue on a dolicyldiphosphate carrier (Figure 1-1, upper panel). The first seven-monosaccharide residues (two N-acetyl glucosamines and five mannoses) were assembled on the cytoplasmic face of the ER membrane, and then, flipped from cytoplasmic to the lumenal face of the ER by flippase (Rft1). After flipping, heptasaccharide is further elongated by four mannose and three glucose residues. The tetradecasaccharide attached on the protein is processed by a series of glycosidases and glycosyltransferases in the both of ER and Goldi, resulting in the vast diversity of the N-linked glycan structures (*4*) (Figure 1-1, middle panel). The N-linked glycans have important roles; assisting the proper protein folding in ER, intracellular targeting, cellar trafficking, cell signaling, and the immune response (*5*). In human, reflecting these essential roles of N-linked glycan, mutation of OST or other enzymes involved in N-glycosylation system causes congenital disorders of glycosylation syndrome (CDG disease), resulting in malfunction of several different organ systems (*6*).

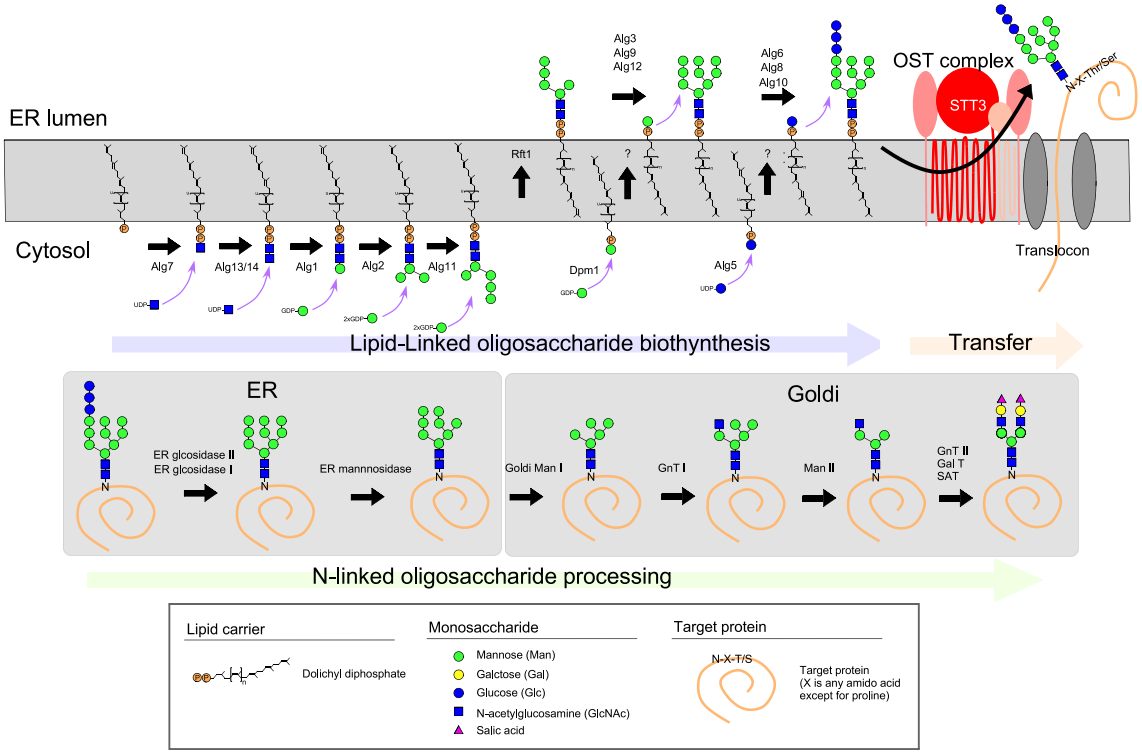


Figure 1-1. The N-glycosylation system in *Saccharomyces cerevisiae*.

1.2

N-glycosylation in prokaryote

In contrast to the eukaryote, archaea and eubacteria exhibit the great diversity in their N-linked oligosaccharide structures for monosaccharide compositions and their linkage positions (Figure 1-2). Furthermore, the lipid carrier structure is different among three domains of life: a dolichol-diphosphate in eukaryote, and a polyprenol-diphosphate in eubacteria. Archaea use two different types of lipid carriers: the Crenarchaeota uses the dolichol-diphosphate, and the Euryarchaeota uses the dolichol-monophosphate (*7*). This difference implies that N-glycosylation systems of the Crenarchaeota are closer to eukaryotic systems. Prokaryoitc N-glycan biosynthesis pathway has been studied in three euryarchaeal spices, *Haroferax volcanii*, *Methanococcus voltae* and *Methanococcus maripaludis* and one eubacterial spice, *Campylobacter jejuni* (*2*) (Figure 1-3). The prokaryotic N-glycan are synthesized by their unique sets of membrane-bound glycosyltransferases in the *Agl* (Archaeal glycolsylation) family (in archaea) and the *Pgl* (Protein glycosylation) family (in eubacteria) (Figure 1-3), which are clustered in their genome. Most of the glycosyltransferase gene identifications have been performed based on the glycosyltransferase gene deletion approach, giving arises to the defect of N-glycan structure or the lack of monosaccharide reside. And the rest glycosyltransferases (AlgK, C, A involved in *M. voltae* N-glycan biosynthesis) have been characterized by *in vitro* analysis using recombinant glycosyltransferases and a suite of synthetic and semisynthetic LLO substrate (*13*). In terms of the oligosaccharide processing, the addition of a mannose residue to the hexuronic acid residue in a protein-bound tetrasaccharide chain has reported in haloarchaeaon *Haloferax volcanii* (*8*) (Figure 1-3). Thus, the oligosaccharide processing steps are probably widespread in prokaryotes. Regarding OST, archaeal and bacterial OST are composed by single transmembrane protein aliased as AglB in archaea (Archaeal glycolsylation B) and PglB in bacteria (Protein glycosylation B), which are homologues of catalytically active protein called STT3 (Staurosporine and temperature sensitivity B, Figure 1-1) in the eukaryotic OST complex. All STT3/AglB/PglB proteins share the typical signature of a hydrophobic, N-terminal, membrane-embedded part with 11 or 13 predicted transmembrane helices and a globular C-terminal domain that is located in the lumen of the ER in eukaryotes or in the periplasm of eubacteria bacteria or in the exterior of archaea. Similar to eukaryote, AglB transfers the oligosaccharide (LLO) to the asparagine residue in the consensus sequon, Asn-X-Thr/Ser (where X is any amino acid residues except for proline). In contrast, PglB uses an extended and thus more restrictive version of the sequon,

Asp/Glu-X-Asn-X-Thr/Ser (where X is any amino acid residues except for proline).

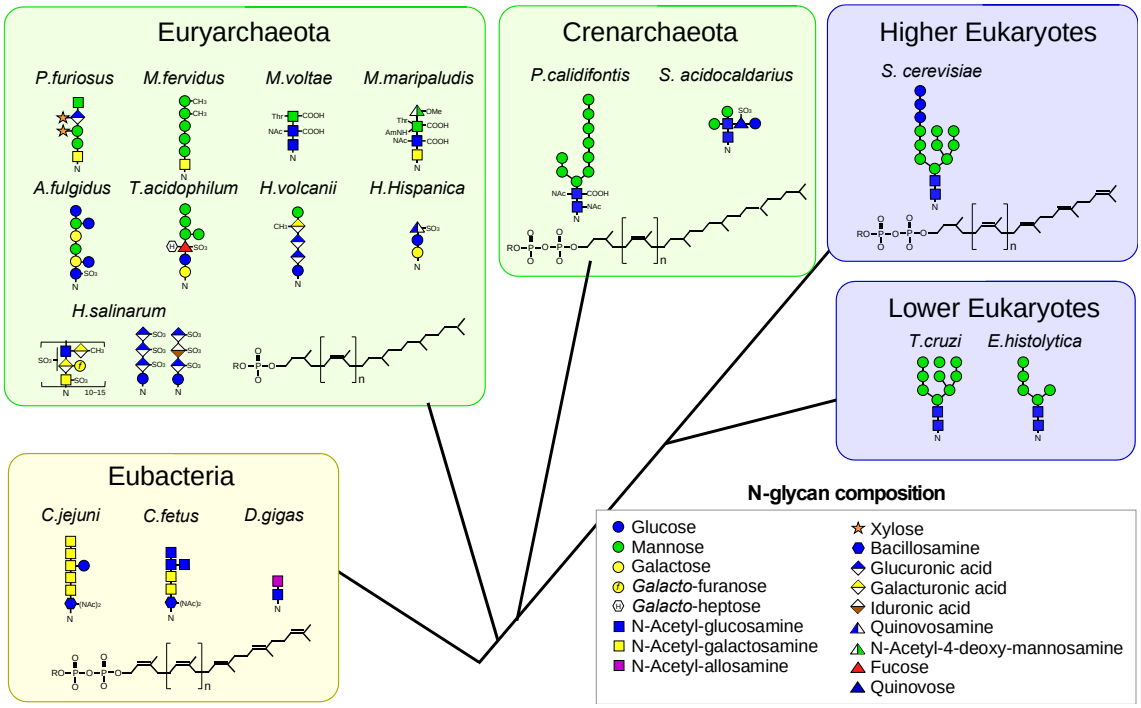


Figure 1-2. The N-linked oligosaccharide and lipid carrier structures in archaea and eubacteria species.

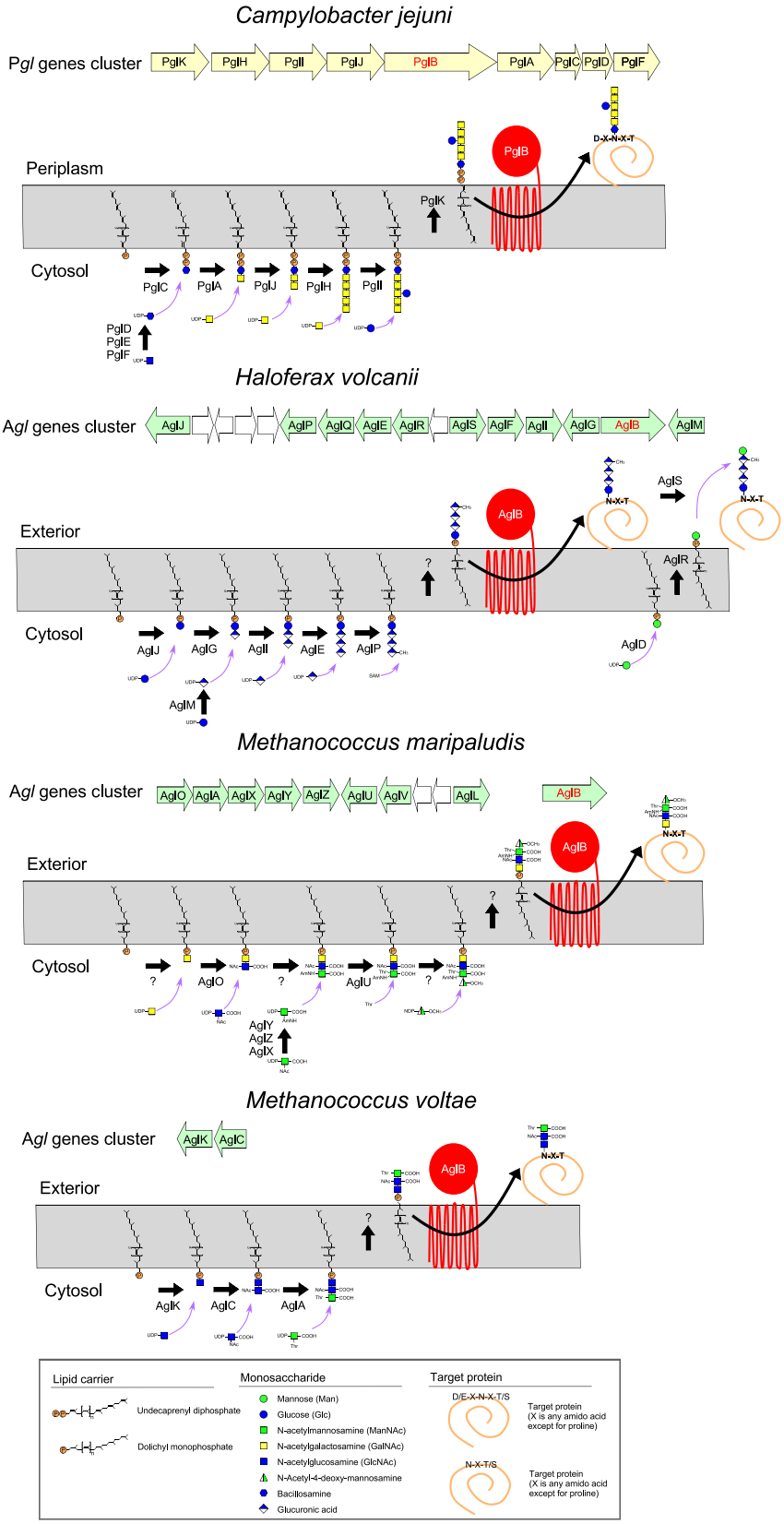
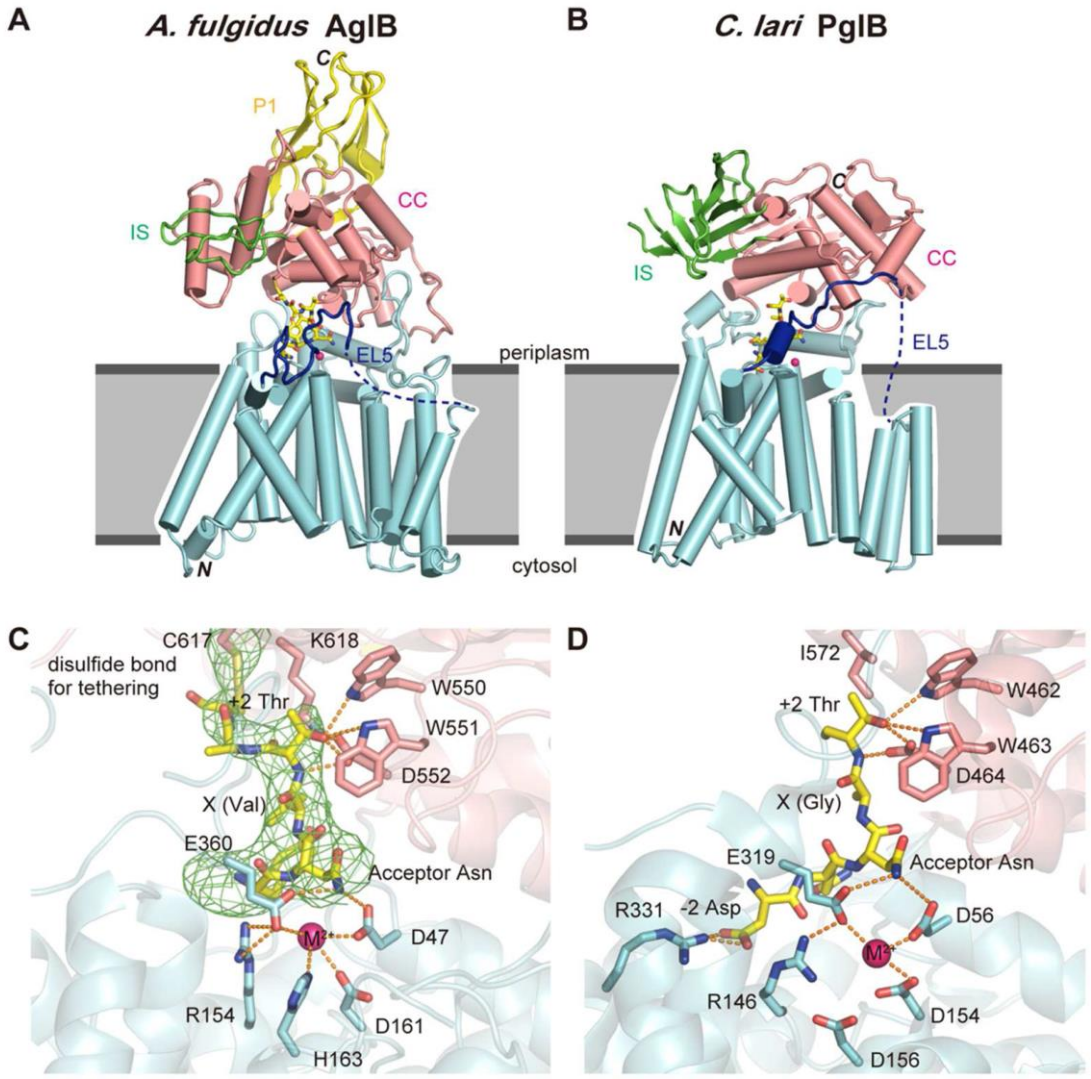


Figure 1-3. The N-glycosylation system in three archaea (*H. volcanii*, *M. voltae* and *M. maripaludis*) and one eubacteria species (*C. jejuni*).

1.3

Structural biology of oligosaccharyltransferase

The first OST crystal structure was reported for C-terminal globular domain of AglB from hyperthermophilic archaeon, *Pyrococcus friosus* in 2008 (*9*). Following *Pf*AglB structure, the full length structure of the bacterial PglB of *Campylobacter lari* in complex with acceptor peptide (*10*) and the archaeal AglB of *Archaeoglobus fulgidus* in apo state (*11*) and in complex with acceptor peptide (*12*) were reported. These structures revealed glycosylation sequon (N-X-T for AglB and D-X-N-X-T for PglB) recognition mechanism (Figure 1-4). The carboxylate groups of the two acidic residues (D47 and E360 in *Af*AglB, D56 and E316 in *Cl*PglB) contact the asparagine side chain amide in the sequon, and simultaneously coordinate a divalent metal ion. The WWD motif in C-terminal globular domain contacts threonine residue at the +2 position via hydrogen bonds. In *Cl*PglB, aspartic acid residue at the -2 position is recognized by R331 via electrostatic interaction. In contrast to well-studied acceptor substrate recognition, the donor LLO recognition mechanism is still unclear due to lack of LLO chemical structure and OST crystal structure in complex with LLO.



Fugure 1-4. Crystal structure of the *Af*AglB-sequon in the complex with their acceptor peptide, and comparison with the *Cl*PglB structure in the complex with a sequon peptide. Overall structures of *Af*AglB (A) and *Cl*PglB (PDB 3RCE) (B), both in the acceptor peptide (shown by *yellow* stick) bound states. The N-terminal transmembrane region consists of 13 transmembrane (TM) helices (*cyan*), and the C-terminal globular domain comprises structural units referred to as CC (*salmon*), IS (*green*), and P1 (*yellow*). Panel (C) and (D) show the closed view of acceptor peptide recognition pockets. This figure is cited from S. matsumoto *etal*, biochemistry, (2016) *in press* (*12*).

1.4

Motivation and outline of my research

1. Structural elucidation of N-linked oligosaccharides from three hyperthermophilic archaea species

Our aim is to determine structures of LLOs from two hyperthermophilic archaeons, *P. friosus* (chapeter 2) and *A. fulgidus* (chapter 3). The obtained structure makes the donor LLO (water-soluble LLO analog) chemical synthesis possible, which could be used for AglB co-crystallization and the investigation of OST-LLO interaction.

The conventional method for the chemical structure determination of N-linked oligosaccharide as follows: cell culturing, glycoprotein isolating by lectin purification from cell lysate, glycoproteins digesting by protease treatment, glycopeptide separating by HPLC, analyzing chemical structure by NMR. However, as a result of using conventional method, we end up generating mixture of heterogeneous N-linked oligosaccharides whose structures are changed from primary structure by post-attachment oligosaccharide processing, such as a monosaccharide trimming and a monosaccharide addition. This method seems to not suit our propose that is the structure determination of LLO serving as a donor substrate for AglB. In this study, we established new sample preparation method to produce intact N-linked oligosaccharides (chapter 2 to 4, material and methods) and applied to three hyperthermophilic archaeal species, *A. fulgidus* (chapter 2), *P. furiosus* (chapter 3) and *Pyrobaculum calidifontis* (chapter 4). In new established methods, glycopeptide samples were produced by *in vitro* oligosaccharide-transfer reaction carried out by addition of a structurally defined acceptor peptide into detergent solubilized archaeal membrane fractions, which contains both of oligosaccharyltransferase and LLO. The N-linked oligosaccharide of *A. fulgidus* and *P. furiosus* showed unique structures: in *A. fulgidus* N-glycan, the reducing end monosaccharide residue that directly linked to the Asn residue is glucose, which is not very common since the reducing end monosaccharide residue usually possesses an N-acetyl group at the C-2 position and the *P. furiosus* N-linked oligosaccharide has two blanching xylose residues at third mannose and fourth glucose residues. On the other hand, the *P. calidifontis* N-linked oligosaccharide shares the typical signature of eukaryotic N-glycan: blanching mannose substructure and chitobiose core ($GlcNAc_2$-Asn) at reducing end. The structural similarity makes the gene identification involved in N-glycan biosynthesis possible by homology search with counterpart eukaryotic genes. We also prepared *P. calidifontis* N-linked oligosaccharide from glycoprotein by using conventional method. The N-linked oligosaccharide

structure recovered from *P. calidifontis* glycoproteins was 1-residue shorter than *in vitro* enzymatically synthesized one, suggesting the first case of the N-glycan trimming on proteins in archaea.

2. Development of NMR structure analysis of carbohydrate

Nuclear magnetic resonance (NMR) is powerful tool for structural analysis of the oligosaccharide. NMR is superior to other methods, such as mass-spectrometry and X-ray analysis, in terms of carbohydrate structural investigation and provides stereo-chemical information: *gluco*/*galacto*/*manno* configuration, anomeric configuration and glycosidic linkage position. The classical carbohydrate NMR measurement is performed using ^{1}H nuclei spin systems, which have higher sensitivity and abundance. However, low chemical shift dispersion due to the dynamic behavior of oligosaccharide and multiplicity of constituent monosaccharides make ^{1}H resonance overlap severe. On the other hand, ^{13}C nuclei spin systems show the large chemical shift dispersion, whereas their lower sensitivity and abundance require high sample concentration. Here, we applied ^{13}C isotope labeling to N-linked oligosaccharide and performed NMR analysis based on ^{1}H and ^{13}C nuclei spin systems. In spite of low sample concentration, ^{13}C labeled carbohydrate showed high quality spectrum for low sensitivity measurements, such as ^{1}H-^{13}C HMBC and ^{13}C-^{13}C correlation spectroscopy. These oligosaccharide ^{13}C labeling approach broadens carbohydrate NMR analysis and makes it reliable.

1.5

Reference

(1) A. Larkin, and B. Imperiali, Biochemistry 21 (2011), pp. 4411-4426.

(2) K.F. Jarrell, Y. Ding, B.H. Meyer, S.V. Albers, L. Kaminski, and J. Eichler, Microbiol. Mol. Biol. Rev. 2 (2014), pp. 304-341.

(3) M. Aebi, Biochimica et Biophysica Acta (BBA) - Molecular Cell Research 11 (2013), pp. 2430-2437.

(4) E.S. Trombetta, Glycobiology *9* (2003), pp. 77R-91R.

(5) K.W. Moremen, M. Tiemeyer, and A.V. Nairn, Nat. Rev. Mol. Cell Biol. 7 (2012), pp. 448-462.

(6) T. Marquardt, and J. Denecke, Eur. J. Pediatr. 6 (2013), pp 359-79.

(7) Y. Taguchi, D. Fujinami, and D. Kohda, J. Biol. Chem. 21 (2016), pp 11042-11054.

(8) C. Cohen-Rosenzweig, S. Yurist-Doutsch, and J. Eichler, J. Bacteriol. 24 (2012), pp 6909-6916.

(9) M. Igura, N. Maita, J. Kamishikiryo, M. Yamada, T. Obita, K. Maenaka, K., and D. Kohda, EMBO J. *1* (2008), pp 234-243.

(10) C. Lizak, S. Gerber, S. Numao, M. Aebi, and K.P. Locher, Nature 7351 (2011), pp 350-355.

(11) S. Matsumoto, A. Shimada, J. Nyirenda, M. Igura, Y. Kawano, and D. Kohda, Proc. Natl. Acad. Sci. U. S. A. 44 (2013), pp 17868-17873.

(12) S. Matsumoto, Y. Taguchi, A. Shimada, M. Igura, M., and D. Kohda, Biochemistry (2016), *in press*

(13) A. Larkin, M. M. Chang, G. E. Whitworth, B. Imperiali, Nat. Chem. Biol., 9 (2013), pp. 367-373

Chapter 2- Structural elucidation of an asparagine-linked oligosaccharide from the hyperthermophilic archaeon, *Pyrococcus furiosus.*

2.1 Material and Methods

2.1.1 Preparation of membrane fractions and OST reaction

Pyrococcus furiosus cells were grown anaerobically in rich medium at 98 °C, as described (*1*). The cells were disrupted by an osmotic shock treatment, and Triton X-100 solubilized membrane fractions were prepared. Typically, 2–3 mL membrane fractions were obtained from 1-g wet cells. The acceptor peptide was custom-synthesized with an N-terminal TAMRA (5/6-carboxytetramethylrhodamine) fluorophore, and a C-terminal biotin group attached to the side-chain amino group of the C-terminal Lys residue (Hayashi Kasei, Osaka, Japan). The amino acid sequence is Ala-Ala-Tyr-Asn-Val-Thr-Lys-Arg-Lys, where the underlined sequence is the N-glycosylation sequon. The peptides resolved into two peaks, corresponding to the 5- and 6-isomers of the TAMRA fluorophore, in the reverse phase HPLC chromatogram. We used the isomer that eluted more slowly. The (glyco)peptide concentrations were determined by the absorbance of the TAMRA dye at 558 nm, with an extinction coefficient of 90,000 M^{-1} cm^{-1}. To 23-mL membrane fractions, 3 mL of 1 M Tris–HCl, pH 7.5, (final concentration 40 mM), 0.15 mL of 1 mM TAMRA-AAYNVTKRK-biotin (final 2 μM), protease inhibitor cocktail, EDTA-free (Roche, 7 tablets, according to the manufacturer's recommendations), and 50 mL water were added, and the mixture was incubated for 4 h at 65 °C. The reaction was stopped by cooling and stored at −20 °C until use. For a larger-scale preparation, 300 mL membrane fraction was used, and $ZnSO_4$ was added to the reaction mixture to a final concentration of 10 mM.

2.1.2 Purification of glycopeptides

Isolation of the biotin-tagged glycopeptide was performed using Streptavidin Mutein agarose (Roche). One milliliter of the Mutein resin was equilibrated in buffer A (0.1 M Tris–HCl, pH 7.5, 0.15 M NaCl, and 0.4 M $(NH_4)_2SO_4$). The reaction mixture (total volume 76 mL) was mixed with

38 mL of 1.2 M $(NH_4)_2SO_4$, and then with the equilibrated resin slurry. After an incubation for 1 h at 4 °C, the reaction mixture and resin were poured into an empty column, and the supernatant was removed as the flow-through. The resin was washed twice with buffer A, and the absorbed materials were eluted with buffer B (0.1 M Tris–HCl, pH 7.5, 0.15 M NaCl, and 5 mM biotin). The adsorption and elution was repeated using new Mutein resin. The eluted materials were combined and loaded on a TSKgel Amide-80 normal phase column (Tosoh, Tokyo, Japan). A linear gradient of solvent B was applied from 40% B to 60% B. Solvent A was 10 mM CH_3COONH_4, CH_3CN/H_2O = 9:1 (v:v), and Solvent B was 10 mM CH_3COONH_4, CH_3CN/H_2O = 1:9 (v:v). The peaks were collected separately (Fig. 1B) and loaded on a COSMOSIL 5C18-AR-II reverse phase column (Nacalai Tesque, Kyoto, Japan) in 0.1% trifluoroacetic acid (TFA) and acetonitrile. The peak from the reverse phase column was collected and dried in a SpeedVac concentrator (Thermo Savant).

2.1.3

Monosaccharide analyses

The monosaccharide composition was analyzed using an ABEE (*p*-aminobenzoic acid ethyl ester) labeling kit (J-Oil Mills, Tokyo, Japan). Acid hydrolysis of a glycopeptide sample (100–250 pmol) with 4 N TFA, N-acetylation, and conversion with ABEE were performed according to the manufacturer's recommendations. The resultant ABEE-converted monosaccharides in the aqueous layer were analyzed by reverse phase HPLC on a Honenpak C18 column (J-Oil Mills) at a flow rate of 1 mL/min, using isocratic elution with 0.2 M potassium borate buffer, pH 8.9, 7% acetonitrile, at 30 °C in a column oven with fluorescent detection (Ex. 305 nm/Em. 360 nm). The standard mixture contained 1 nmol of each monosaccharide.

The absolute configurations of the monosaccharides were analyzed as described (*2*). The (*S*)-2-butyl glycosides were generated from 15 μg of the glycopeptide sample, by acidic solvolysis with 0.2 mL of 2 N HCl in (*S*)-(+)-2-butanol at 80 °C for 24 h. After re-N-acetylation with acetic anhydride and trimethylsilylation with N,O-bis(trimethylsilyl)trifluoroacetamide (BSTFA, Tokyo Chemical Industry, Tokyo, Japan), the resultant trimethylsilyl derivatives were analyzed with an Agilent/HP 6890 GC system equipped with a flame-ion detector, using an SE-30 column (0.32 mm × 30 m, Agilent Technologies, CA), with a temperature gradient of 1 °C/min from 135 to 200 °C. The standard mixture contained 1 mg of each monosaccharide. Acidic solvolysis was performed with 0.2 mL of 2 N HCl in (±)-2-butanol at 80 °C for 8 h.

2.1.4

Electrospray ionization mass spectrometry

The peaks P1 and P3 were desalted by reverse phase HPLC chromatography in 0.1% TFA and acetonitrile. The eluted peaks were collected, dried in a SpeedVac, and dissolved in 0.1% formic acid and 50% methanol. For the direct-infusion electrospray ionization (ESI) analysis in the positive ion mode, TripleTOF 5600 and QSTAR Elite mass spectrometers (ABSciex) were used at flow rates of 1 μL/min and 5 μL/min, respectively. The triply charged precursor ions were selected and subjected to MS/MS analyses with a collision energy of 35.

2.1.5

NMR spectroscopy

NMR spectra were recorded on a Bruker Avance600 spectrometer equipped with a TXI cryoprobe in 99.96% $^{2}H_2O$ solutions at 25 °C, with an internal 100 μM DSS (sodium 2,2-dimethyl-2-silapentane-5-sulfonate) standard (δH 0). 13C chemical shift values relative to TMS (tetramethylsilane) were calculated using the chemical shift referencing ratio, Ξ = 0.25145004. COSY, TOCSY (mixing time 70 ms), NOESY (mixing time 600 ms), and ROESY (mixing time 225 ms) spectra were recorded using a 20 μM sample. Typically, 896 FIDs of 4096 points were acquired with 128–192 scans per FID. $^{1}H–^{13}C$ HSQC spectra were recorded with and without 13C composite pulse decoupling during acquisition, using a 110 μM sample. In total, 512 FIDs of 2048 points were acquired, with 256 scans per FID. NMR data were processed and displayed using the nmrPipe/nmrDraw program (*3*).

2.2 Results

Glycopeptides profiling by NP-HPLC and ESI-MS/MS

The Triton X-100-solubilized membrane fractions were prepared from cultured *P. furiosus* cells, and mixed with a peptide containing the N-glycosylation sequon, Asn-Val-Thr (Fig. 2-1A). The substrate peptide also possessed a fluorescent dye, TAMRA, attached to the N-terminus for detection, and biotin at the C-terminus as an affinity tag. The generated glycopeptides were isolated using streptavidin-based affinity chromatography, and then purified by normal phase HPLC chromatography (Fig. 2-1B). Three fractions (P1, P2, and P3) were recovered. Note that no unmodified peptide was detected in the chromatogram, even though an excess amount of the substrate peptide was present in the reaction mixture. The unreacted peptide was probably digested by endogenous proteases during the incubation, since it lacked the protective effects of the oligosaccharide moiety attached to the peptide. Finally, each fraction was desalted and concentrated by reverse phase HPLC chromatography.

In our previous study, the MALDI-MS/MS analysis suggested that the oligosaccharide structure of peak P3 consisted of seven monosaccharide residues (2 × Hex, 2 × HexNAc, HexA, 2 × Pent) (*4*). In the present study, the MALDI-MS analysis revealed that the oligosaccharide structures of peaks P1 and P2 consisted of five and six monosaccharide residues, lacking two or one pentose residues, respectively. We further analyzed peaks P1 and P3 by ESI-MS/MS. The MS/MS spectra of the precursor ions with m/z 865.71 for P1 and m/z 953.75 for P3 are shown (Fig. 2-2). There are two distinct mass ladders corresponding to the $[M+2H]^{2}+$ and $[M+3H]^{3}+$ fragment ions in each spectrum. The mass ladders of the glycopeptide P1 provided a fragmentation series composed of sequential losses of N-acetyl hexosamine, hexuronic acid, hexose, hexose, and N-acetyl hexosamine. The fragmentation pattern suggested that these residues are located linearly from the distal end. The mass ladders of the glycopeptide P3 showed more complex pattern, suggesting two branching pentose residues attached to the 5-residue core structure of P1.

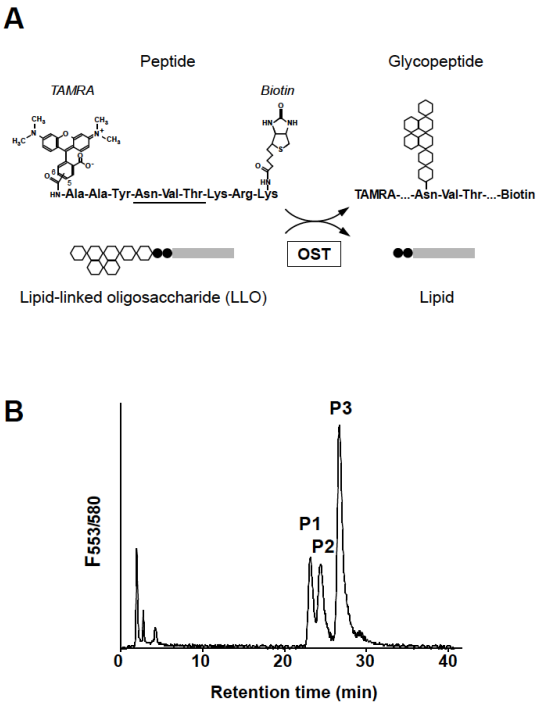


Figure 2-1. Production of the glycopeptides by enzymatic reaction, and purification by normal phase HPLC chromatography. (A) Schematic representation of the reaction catalyzed by the oligosaccharyltransferase (OST). A 9-residue peptide containing the N-glycosylation sequon (underlined) is the acceptor substrate, with a fluorescent dye, TAMRA, at the N-terminus, and a biotin tag attached to the side chain of the C-terminal Lys residue. The lipid-linked oligosaccharide (LLO) is the donor substrate. The products are a glycopeptide and a lipid-pyrophosphate. Hexagon, monosaccharide; circle, phosphate group; rectangle, dolichol. (B) Elution profile of the normal phase Amide-80 chromatography of the reaction mixture. The three eluted peaks are referred to as P1, P2, and P3. Elution was monitored with fluorescent detection (Ex. 553 nm/Em. 580 nm). Peaks P1, P2, and P3 correspond to glycopeptides consisting of five, six, and seven monosaccharide residues, respectively.

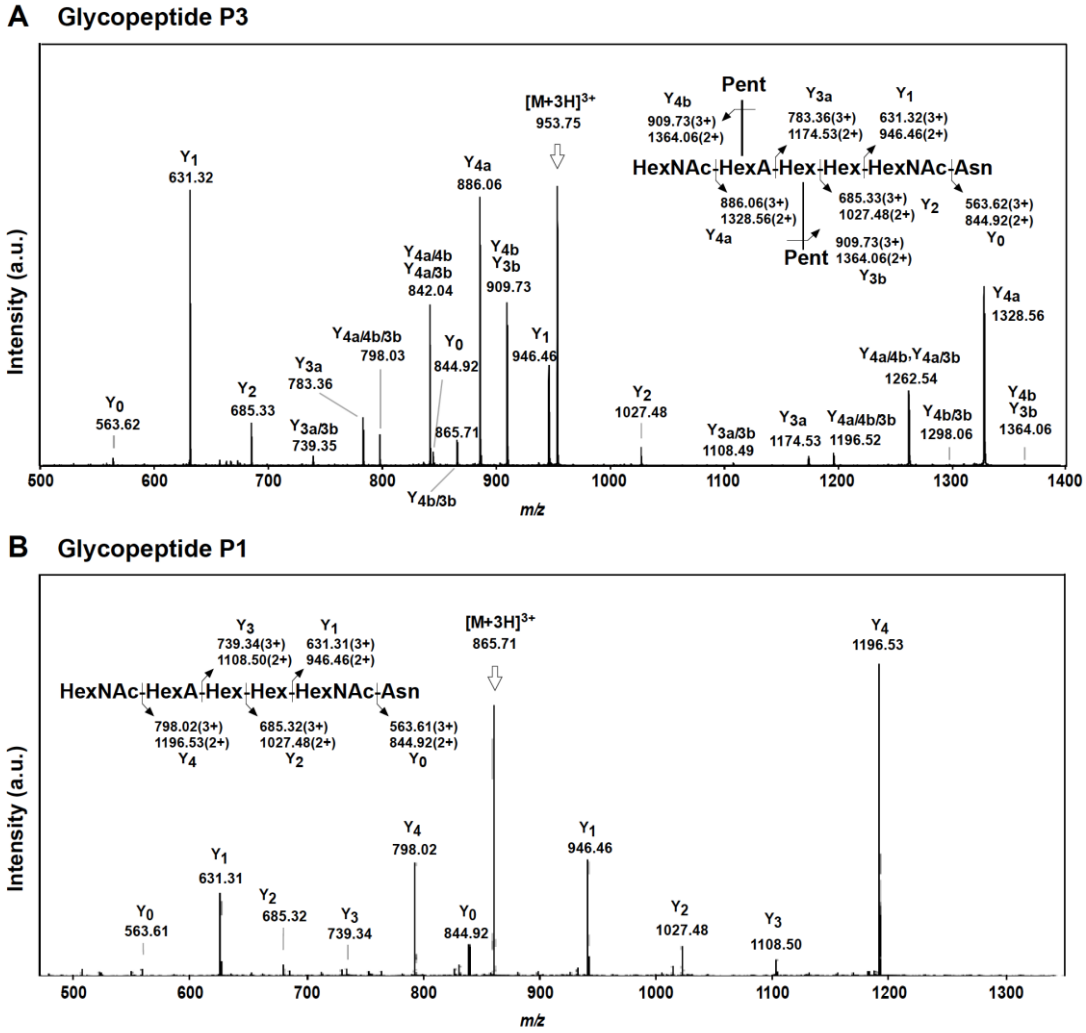


Figure 2-2. ESI-MS/MS analyses of glycopeptides P1 and P3. The precursor ions are marked by the vertical arrows. Assignments of the MS/MS fragment ions originating from single cleavages are illustrated as insets. The $[M+2H]^{2+}$ and $[M+3H]^{3+}$ ions are marked with 2+ and 3+, respectively. The fragment ions derived through double cleavages are shown in the spectrum, in addition to the single cleavage products. The expected *m/z* values were observed within 0.01.

The structure determination by NMR

The monosaccharide composition analysis of P3 identified the two N-acetyl hexosamine residues as GalNAc and ManNAc, and the two pentose residues as xylose. The two hexose residues were probably mannose, but the identification was not conclusive due to a small contamination peak corresponding to glucose. No data on the hexuronic acid residue were obtained, due to its decomposition during acid hydrolysis.

The NMR sample for ^{1}H NMR contained 20 μM glycopeptide derived from peak P3 (Fig. 2-1B). The sample was dissolved in 260 μL of 99.96% $^{2}H_2O$ in a Shigemi micro NMR tube. Five anomeric 1H resonances were readily identified in the downfield region of the water signal (ca. 4.7 ppm) in the one-dimensional (1D) ^{1}H spectrum, and labeled **a** to **e** (Fig. 2-3A). The 1D slices of the two-dimensional (2D) TOCSY spectrum indicated the existence of two additional anomeric resonances upfield of the water signal, which were labeled **f** and **g** (Fig. 2-3B). The NMR sample for natural-abundance ^{13}C NMR contained 110 μM glycopeptide. We recorded a ^{1}H–^{13}C HSQC spectrum (Fig. 2-4A). The assignments of the 1H resonances were transferred to the ^{13}C resonances.

The identities of the monosaccharides were deduced from the homonuclear ^{3}J coupling constants and ^{1}H chemical shifts in the 1D slices of the 2D TOCSY spectra (Fig. 2-3B). We propose the following assignments of the monosaccharide identities and anomeric protons: **a**, N-acetyl mannose; **b**, glucuronic acid; **c**, mannose; **d**, mannose; **e**, N-acetyl galactose; **f**, xylose; **g**, xylose. All seven residues are in the pyranose ring form. The substitution with the N-acetylamino group at the C-2 position was confirmed by the upfield shifts of the C-2 carbons to 52.6 ppm (**a**2) and 50.1 ppm (**e**2) in the HSQC spectrum (Fig. 2-4A). Note that the ^{13}C chemical shifts of the cross peaks of these C-2 carbons are aliased to 110–113 ppm, due to the spectral aliasing along the ^{13}C axis. The positions of the glycosidic linkages were then deduced using the inter-residue ^{1}H–^{1}H nuclear Overhauser effects (NOE) from the anomeric protons in the 2D NOESY spectrum (Fig. 3C). The NOEs between proton pairs **(a**1–**b**4, **b**1–**d**3, **c**1–**e**3, **d**1–**c**2, **f**1–**b**3, and **g**1–**d**2) established the inter-glycosidic linkages. The downfield shift of the 13C chemical shift of the linkage carbons (**b**4, **d**3, **e**3, **c**2, **b**3, and **d**2) to around 80 ppm in the ^{1}H–^{13}C HSQC spectrum confirmed the linkage positions of the monosaccharide residues (Fig. 2-4A). The ^{13}C chemical shift for the **e**1 anomeric carbon was considerably shifted upfield at 80 ppm, as compared to those of the other anomeric carbons at 95–105 ppm. Hence, monosaccharide residue **e** (GalNAc) is N-glycosidically linked to Asn. The NMR data of the heptasaccharide N-glycan are summarized

in Table 2-1. We also performed the spectral analysis of the glycopeptide derived from peak P1 (data not shown). The chemical shift data of the pentasaccharide N-glycan are summarized in Table 2-2.

The anomericity of each monosaccharide residue was determined from the homonuclear $^{3}J_{H1–H2}$ values. The relatively large $^{3}J_{H1–H2}$ values revealed that monosaccharide residues **e** (GalNAc), **f** (Xyl), and **g** (Xyl) are in the β-configuration. The strong $^{1}H–^{1}H$ NOEs observed between the H-1, H-3, and H-5 resonances confirmed that these monosaccharide residues are β-linked pyranosides. On the other hand, the relatively small $^{3}J_{H1–H2}$ showed that the monosaccharide residue **b** (GlcA) is in the α-configuration. As for the manno-configurated monosaccharide residues (**a**, **c**, and **d**), the $^{3}J_{H1–H2}$ values do not offer clues to discriminate the anomeric configuration. Instead, we used the heteronuclear $^{1}J_{C1–H1}$ values, obtained by a comparison of two $^{1}H–^{13}C$ HSQC spectra recorded with and without ^{13}C composite pulse decoupling during acquisition (Fig. 2-4B, Table 2-1). The typical $^{1}J_{C1–H1}$ values for the α- and β-configurations are 170 Hz and 160 Hz, respectively (*5*, *6*). Hence, monosaccharide residues **a** (ManNAc, 177 Hz), **c** (Man, 171 Hz), and **d** (Man, 172 Hz) are all in the α-configuration. The $^{1}J_{C1–H1}$ values of the other monosaccharide residues are fully consistent with the anomeric configurations deduced from the homonuclear $^{3}J_{H1–H2}$ values.

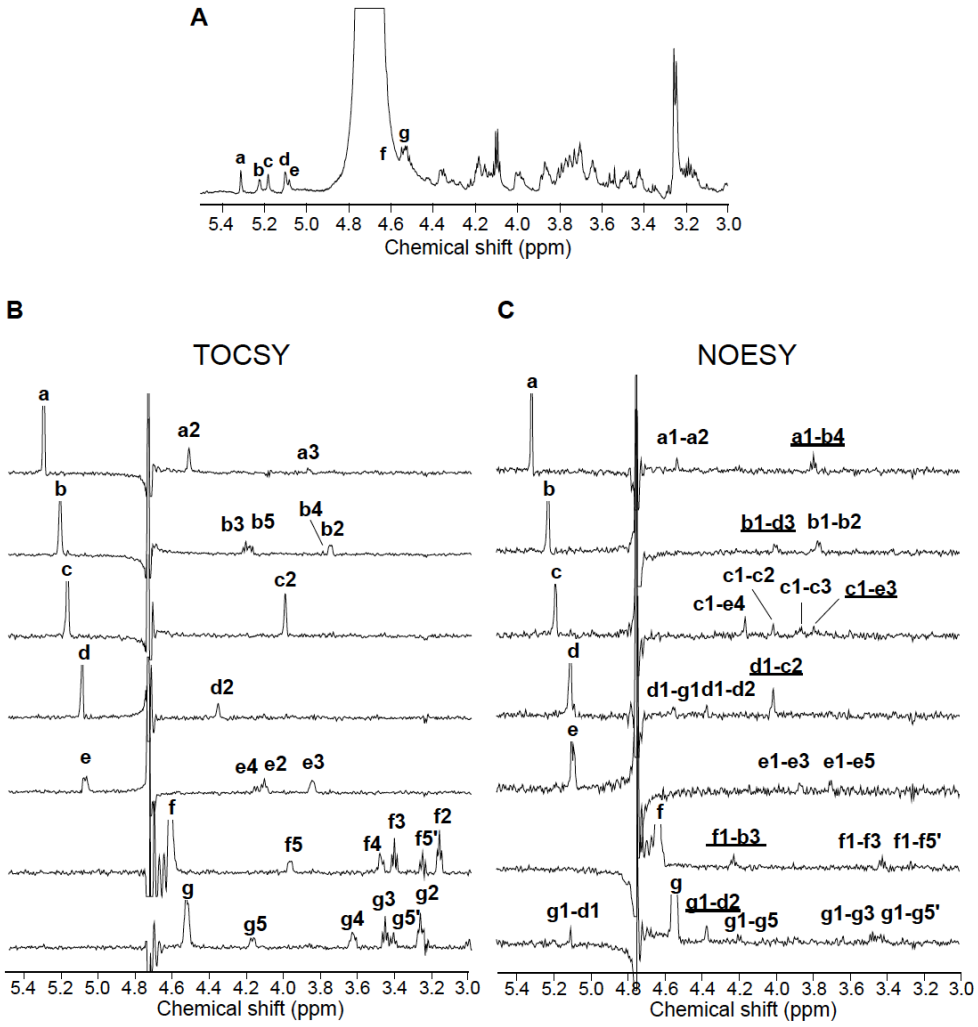


Figure 2-3. ^{1}H NMR spectra of the glycopeptide P3. (A) 1D NMR spectrum with water suppression using presaturation. Seven anomeric proton resonances are labeled, **a**–**g**. (B) 1D slices of the 2D TOCSY spectrum with a mixing time of 70 ms. (C) 1D slices of the 2D NOESY spectrum with a mixing time of 600 ms. The NOEs used to establish the inter-glycosidic linkages are underlined.

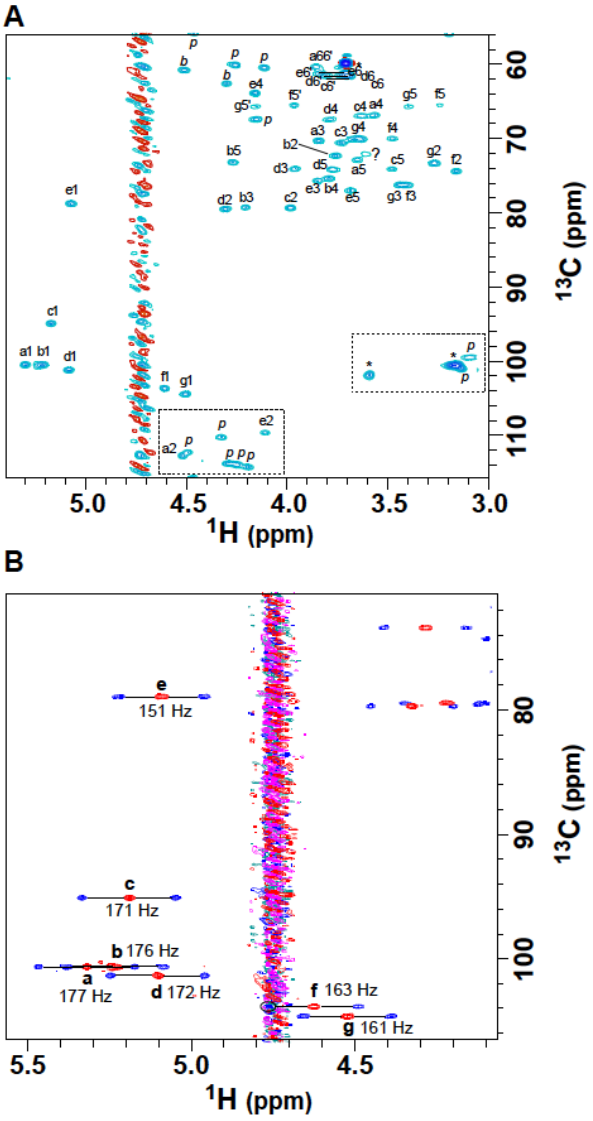


Figure 2-4. ^{1}H–^{13}C HSQC spectra of the glycopeptide P3. (A) HSQC spectrum. The two regions enclosed by dashed lines contain cross peaks aliased by 60 ppm along the ^{13}C axis. The cross peaks derived from the peptide moiety of the glycopeptide are marked with *p*, those from the biotin moiety are marked with *b*, and those from impurities are marked with asterisks. (B) Superposition of two HSQC spectra with (*red*) and without (*blue*) ^{13}C composite pulse decoupling during acquisition, to determine the $^{1}J_{C1–H1}$ values.

Table 2-1. ^{1}H and ^{13}C NMR chemical shifts and $^{1}J_{C1\text{-}H1}$ values of the N-linked heptasaccharide glycan on the glycopeptide P3 from *Pyrococcus furiosus*.

	H-1	H-2	H-3	H-4	H-5, H-5’	H-6, H-6’
Residue	C-1	C-2	C-3	C-4	C-5	C-6
	$^{1}J_{C1\text{-}H1}$					
a	5.31	4.57	3.88	3.57	3.74	3.88
α-ManNAc	102.2	54.3	72.0	68.5	74.5	62
	177					
b	5.22	3.76	4.22	3.79	4.20	-
-4)-α-GlcA	102.2	74.0	81.0	77.1	74.9	-
	176					
c	5.18	4.00	3.74	3.65	3.5	3.71,3.83
-2)-α-Man	96.6	81.0	72.3	68.6	75.7	63.2
	171					
d	5.10	4.36	3.99	3.81	3.78	3.63, 3.74
-3)-α-Man	102.9	81.2	75.8	71.8	75.9	63.2
	172					
e	5.09	4.12	3.86	4.16	3.69	3.75, 3.80
-4)-β-GalNAc	80.5	51.8	77.4	65.6	78.7	62.0
	151					
f	4.61	3.18	3.41	3.5	3.98, 3.27	-
β-Xyl	105.3	76.1	78.0	71.8	67.0	-
	163					
g	4.54	3.29	3.47	3.64	4.19, 3.42	-
β-Xyl	106.3	75.0	78.0	71.8	67.2	-
	161					

Chemical shifts in ppm relative to TMS and ^{1}J coupling constants in Hz measured in $^{2}H_2O$ at 25 °C. Average errors of ±0.02 ppm for ^{1}H, ±0.1 ppm for ^{13}C, and ±0.5 Hz for $^{1}J_{C1–H1}$. Additional chemical shifts for the N-acetyl groups are δH/δC = 1.92/22.9 and 1.97/22.4 (Me).

Table 2-2. ^{1}H and ^{13}C NMR chemical shifts of the N-linked pentasaccharide glycan on the glycopeptide P1 from *Pyrococcus furiosus*.

	H-1	H-2	H-3	H-4	H-5	H-6, H-6’
Residue	C-1	C-2	C-3	C-4	C-5	C-6
a	5.27	4.47	4.0	3.61	3.76	3.76, 3.84
α-ManNAc	101.8	55.2	71.2	68.5	74.7	62.2
b	5.25	3.61	3.95	3.72	4.05	-
-4)-α-GlcA	102.4	73.9	75.9	78.7	74.9	-
c	5.24	3.95	3.73	3.75	3.47	3.66, 3.87
-2)-α-Man	96.6	81.4	72.3	68.8	75.6	63.5
d	5.0	4.12	3.89	3.7	3.78	3.71, 3.82
-3)-α-Man	104.6	72.0	79.9	68.8	75.8	63.3
e	5.1	4.11	3.86	4.17	3.69	3.72
-4)-β-GalNAc	80.5	51.8	77.4	65.6	78.7	62.5

Chemical shifts in ppm relative to TMS measured in $^{2}H_2O$ at 25 °C. Average errors of ±0.02 ppm for ^{1}H and ±0.1 ppm for ^{13}C. Additional chemical shifts for the N-acetyl groups are δH/δC = 1.92/22.9 and 1.97/22.4 (Me).

The absolute configuration determination

The absolute configurations of the monosaccharide residues other than **b** (GlcA) were determined to be the d configuration, from the gas–liquid chromatography (GLC) retention times of the trimethylsilylated (*S*)-2-butyl glycoside derivatives. The absolute configuration of **b** (GlcA) was estimated by the dependence of the ^{13}C NMR chemical shifts on the absolute configurations of the neighboring monosaccharides (*7*) (Table 2-3).

Table 2-3. Estimation of the absolute configulation of the glucuronic acid residue using the dependence of ^{13}C chemical shifts on the absolute configulation of neighboring monosaccharides.

Disaccharide unit	Carbon position	^{13}C chemical shift (ppm)		Chemical shift change induced by the glyocosidic bond formation (ppm)		
		Pentasaccharide[a]	Monosaccharide[b]	Experimental[c]	Theoretical[d]	
					D-Ald*p*-D-Man*p*	L-Ald*p*-D-Man*p*
α-GlcA*p*-(1→3)-α-D-Man*p*	GlcA C1	102.4	93.0	+9.4	**+8.2 ± 0.5**	+3.5 ± 1.1
	Man C2	72.0	71.4	+0.6	**+0.1 ± 0.6**	-3.3 ± 0.8
	Man C3	79.9	72.1	+7.8	+8.6 ± 1.5	+5.8 ± 2.1
	Man C4	68.8	68.3	+0.5	**-0.6 ± 0.6**	-1.7 ± 0.2

[a] Data are from Table 2-2.

[b] Data are from Ref. 6.

[c] Difference between the chemical shifts of the carbons in the pentasaccharide and the monosaccharide.

[d] Data are from Ref. 7. The D-Ald*p*-D-Man*p* configulation is more consistent with the experimental data than the L-Ald*p*-D-Man*p* configulation for the three carbon positions, GlcA C1, Man C2, and Man C3 (in bold). Ald*p* is a residue of aldopyranose.

2.3
Disucussion

On the basis of our data, we concluded that the N-linked heptasaccharide of *Pyrococcus furiosus* has the structure shown in Figure 2-5. The N-glycans of archaea and eubacteria exhibit high degrees of structural variation (*8*, *9*, *10*, *11*). We have now added another novel structure to the known prokaryotic N-glycans. The reaction mixture contained penta- and hexasaccharides (Fig. 2-1B), which lacked two or one branching xylose residues. It is not clear whether these shorter oligosaccharide structures represent natural products in vivo or artificial products due to the contamination of biosynthetic or degraded intermediates of LLO during the membrane fraction preparation. In any case, the xylose residues are not essential for serving as the substrate of the *P. furiosus* OST enzyme.

For many years, there was no experimental evidence for the processing of N-glycans after their transfer onto proteins in archaea or eubacteria. In those cases, the oligosaccharide structures in glycopeptides derived from protease-digested glycoprotein(s) were those recognized by OST. Recently, the addition of a mannose residue to the hexuronic acid residue in a protein-bound tetrasaccharide chain was detected in *Haloferax volcanii* (*12*). Thus, the modifications of the N-glycan structures after the transfer onto proteins are probably more widespread than previously expected in prokaryotes. Considering the possibility of the post-transfer modifications, we produced glycopeptides by an in vitro enzymatic OST reaction for the chemical structure determination. The N-glycan structure revealed in the present study will be useful for future studies of the interactions of OST with LLO or glycopeptides.

The use of enzymatic reaction products has several advantages. First, the structure of the peptide portion of the glycopeptides is known, which permits straightforward NMR and MS analyses. Second, the peptide portion can be tagged or labeled, which also greatly facilitates the purification of the glycopeptides. Third, the enzymatic reactions can be optimized. For instance, we found that the addition of 10 mM Zn^{2+} increased the glycopeptide yield three-fold. In contrast, a disadvantage of this approach is the limited availability of the glycopeptides (for example, 75 μg from 100 L cell culture in this study). In fact, we could not measure the insensitive ^{1}H–^{13}C HMBC spectrum, which is useful to connect an anomeric proton and the carbon atom on the opposite side of the glycosidic linkage. We tried to grow *P. furiosus* cells using ^{13}C-glucose as the sole carbon source, but the cultivation was unsuccessful, as described previously (*13*).

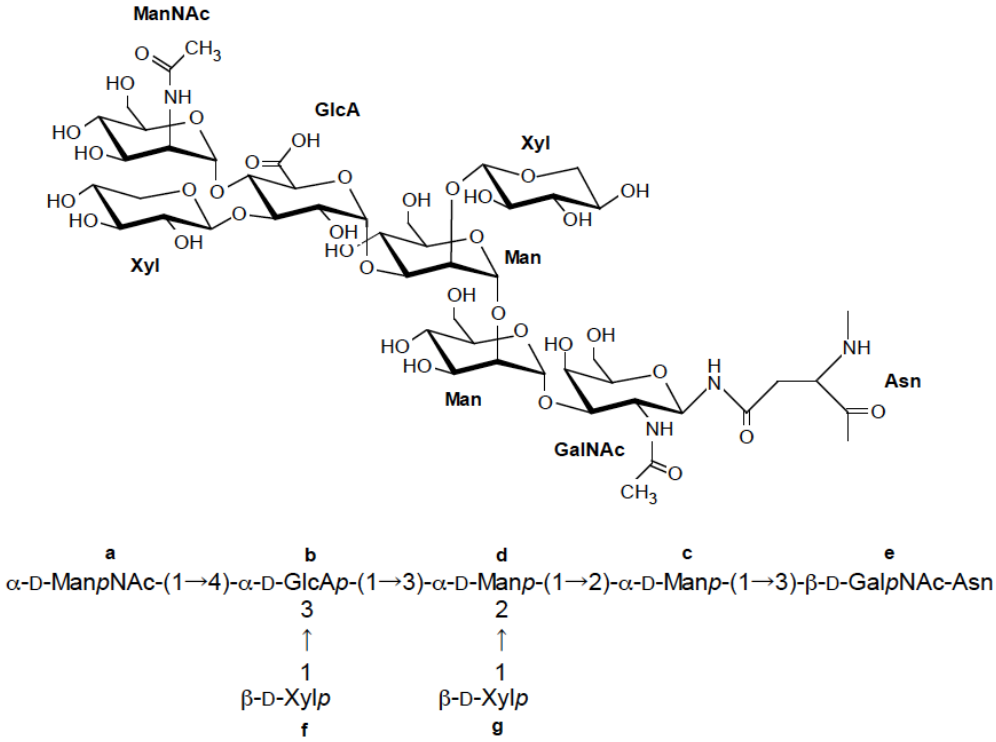


Figure 2-5. Structure of the N-linked heptasaccharide glycan on the glycopeptide P3, produced by the oligosaccharyltransferase of the euryarchaeon, *Pyrococcus furiosus*.

2.4

Reference

(1) D. Kohda, M. Yamada, M. Igura, J. Kamishikiryo, K. Maenaka, Glycobiology, 17 (2007), pp. 1175–1182

(2) G.J. Gerwig, J.P. Kamerling, J.F. Vliegenthart, Carbohydr. Res., 77 (1979), pp. 10–17

(3) F. Delaglio, S. Grzesiek, G.W. Vuister, G. Zhu, J. Pfeifer, A. Bax, J. Biomol. NMR, 6 (1995), pp. 277–293

(4) M. Igura, N. Maita, J. Kamishikiryo, M. Yamada, T. Obita, K. Maenaka, D. Kohda, EMBO J., 27 (2008), pp. 234–243

(5) K. Bock, C. Pedersen, J. Chem. Soc., Perkin Trans. 2 (1974), pp. 293–297

(6) P.K. Agrawal, Phytochemistry, 31 (1992), pp. 3307–3330

(7) A.S. Shashkov, G.M. Lipkind, Y.A. Knirel, N.K. Kochetkov, Magn. Reson. Chem., 26 (1988), pp. 735–747

(8) D. Calo, L. Kaminski, J. Eichler, Glycobiology, 20 (2010), pp. 1065–1076

(9) A. Larkin, B. Imperiali, Biochemistry, 50 (2011), pp. 4411–4426

(10) F. Schwarz, M. Aebi, Curr. Opin. Struct. Biol., 21 (2011), pp. 576–582

(11) G. Palmieri, M. Balestrieri, J. Peter-Katalinic, G. Pohlentz, M. Rossi, I. Fiume, G. Pocsfalvi, J. Proteome Res., 12 (2013), pp. 2779–2790

(12) L. Kaminski, Z. Guan, M. Abu-Qarn, Z. Konrad, J. Eichler, Biochim. Biophys. Acta, 2012 (1820), pp. 1664–1670

(13) G. Fiala, K.O. Stetter, Arch. Microbiol., 145 (1986), pp. 56–61

Chapter 3- Structural elucidation of an asparagine-linked oligosaccharide from the hyperthermophilic archaeon, *Archaeoglobus fulgidus.*

3.1 Material and Methods

3.1.1 Preparation of the membrane fractions and OST reaction

A. fulgidus strain DSM 4304 (NBRC 100126) was obtained from the National Institute of Technology and Evaluation (NITE) Biological Resource Center, Tokyo, Japan. The *A. fulgidus* cells were grown anaerobically without shaking, at 80 °C for 3 days. The culture medium was a simplified version of the predefined medium (NBRC Medium No. 1019). See Table 3-1 for details. For ^{13}C isotope labeling, 0.85 g of ^{13}C-glucose (99 atom%, ISOTEC) was added to the medium, and the amounts of yeast extract and Na_2SO_4 were reduced (1.0 → 0.15 g and 2.4 → 1.5 g per liter of the culture medium, respectively). The cells were disrupted by sonication, and Triton X-100-solubilized membrane fractions were prepared according to the procedure used for Pyrococcus furiosus cells (*1*). The amino acid sequence of the acceptor peptide was Ala-Ala-Tyr-*Asn*-*Val*-*Thr*-Lys-Arg-Lys, where the N-glycosylation sequon is indicated in italics. The acceptor peptide was synthesized with an N-terminal TAMRA (5/6-carboxytetramethylrhodamine) fluorophore, and a C-terminal biotin group attached to the side-chain amino group of the C-terminal Lys residue (Hayashi Kasei, Osaka, Japan). The synthesized peptide eluted as two peaks in the reverse-phase HPLC chromatography, due to the 5- and 6-isomers of the TAMRA fluorophore. We used the peak that eluted more slowly. The (glyco)peptide concentrations were determined by the absorbance of the TAMRA fluorophore at 558 nm, with an extinction coefficient of 90,000 M^{-1} cm^{-1}. To prepare the glycopeptide product for the NMR analysis, 16.5 mL of the membrane fractions prepared from 9 L cell culture, 60 mL of buffer (50 mM Tris–HCl, pH 7.5, containing 0.02% (v/v) Tween 20 and 10 mM $MnCl_2$), 2 mL of 0.1 mM acceptor peptide (final concentration 2 μM), and 21 mL water were mixed and incubated for 2 h at 65 °C. The reaction was stopped by cooling on ice, and stored at −20 °C until use.

Table 3-1

Composition of the culture medium for *Archaeoglobus fulgidus*

Composition	This study (non-label) [e]	This study (^{13}C label) [e]	NBRC Medium 1019
KP buffer[a]	5 ml	5 ml	5 ml
$MgCl_2 \cdot 6H_2O$	0.75 g	0.75 g	0.75 g
$CaCl_2 \cdot 2H_2O$	0.15 g	0.15 g	0.15 g
NH_4Cl	0.54 g	0.54 g	0.54 g
NaCl	30 g	30 g	30 g
Na_2SO_4	2.4 g	1.5 g	2.4 g
sodium lactate	2.2 g	2.2 g	2.2 g
yeast extract	Wako, 1g	Wako, 0.15 g	Difco, 0.5 g
^{13}C glucose (99 atom %)	-	ISOTEC, 0.85 g	-
Fe (II)	$FeCl_2 \cdot 4H_2O$, 2mg	$FeCl_2 \cdot 4H_2O$, 2mg	$Fe(NH_4)_2(SO_4)_2 \cdot 7H_2O$, 2 mg
Trace element solution[b]	2 ml	2 ml	2 ml
Vitamin solution[c]	-	-	2 ml
Resazurin	-	-	1 mg
Na_2CO_3	-	-	1 g
$Na_2S \cdot 9H_2O$[d]	0.25 g	0.25 g	0.36 g
water	Tap water, 1L	Tap water, 1L	Distilled water, 1L

[a] KP buffer: KH_2PO_4, 11.9 g and K_2HPO_4, 2.1 g in 100 ml of distilled water.

[b] Trace element solution in this study: $CoCl_2 \cdot 6H_2O$, 0.024 g; $ZnCl_2$, 0.1 g; $CuCl_2 \cdot 2H_2O$, 0.025 g; H_3BO_3, 0.01g; $NiCl_2 \cdot 6H_2O$ 0.12 g in 1 L of distilled water. The trace element solution of the NBRC Medium 1019: see http://www.nbrc.nite.go.jp/NBRC2/NBRCMediumDetailServlet?NO=1019

[c] Vitamin solution of the NBRC Medium 1019: see http://www.nbrc.nite.go.jp/NBRC2/NBRCMediumDetailServlet?NO=1019

[d] Na_2S solution contains $Ns_2S \cdot 9H_2O$, 5 g in 100 ml of distilled water.

[e] KP and Na_2S solutions are put into 100-ml glass bottles capped with an aluminum crimp seal and a

rubber septum, then autoclaved and cooled to room temperature. While the culture medium (1 L) is still hot after autoclaving, the KP (5 ml) and Na_2S solutions (5 ml) are anaerobically added with plastic syringes with needles, prior to inoculation.

3.1.2

Immunoblotting analysis

The C-terminal soluble globular domains of the three *A. fulgidus* AglB proteins produced in *E. coli* cells (*2*, *3*, *4*) were used to raise rabbit polyclonal antibodies. The proteins were fractionated on an SDS-PAGE gel and electroblotted onto a PVDF membrane (Millipore), using a Trans-Blot Semi-Dry Transfer System (Bio-Rad), and reacted with the anti-AfAglB-L, anti-AfAglB-S1, and anti-AfAglB-S2 antisera. The proteins were visualized with an ECL Advance Western Blotting Detection Kit (GE Healthcare).

3.1.3

Purification of the glycopeptides

The glycopeptide for the NMR analysis was purified according to the method used for *Pyrococcus furiosus* (*1*). Briefly, the biotin-tagged glycopeptide was adsorbed to Streptavidin Mutein Agarose (Roche) in the presence of 0.4 M $(NH4)_2SO_4$, and eluted in the presence of 5 mM biotin. The eluted glycopeptide mixture was separated on a TSKgel Amide-80 normal-phase column (4.6 mm×25 cm, Tosoh, Tokyo, Japan). Solvent A was 10 mM CH_3COONH_4, CH_3CN:H_2O=9:1 (v:v), and Solvent B was 10 mM CH_3COONH_4, CH_3CN:H_2O=1:9 (v:v). A linear gradient of solvent B was applied from 40% B to 60% B. The main peak was collected and loaded on a COSMOSIL 5C18-AR-II reverse-phase column (4.6 mm×25 cm, Nacalai Tesque, Kyoto, Japan), run in 0.1% trifluoroacetic acid (TFA) and 10% acetonitrile for desalting. The materials eluted by a linear gradient of 10–70 % acetonitrile were collected and dried in a SpeedVac concentrator (Thermo Savant).

The glycopeptide sample for the chemical modification analysis was produced by the in vitro oligosaccharyl transfer reaction, using the His-tag purified recombinant *Af*AglB-L from *E. coli* cells and the crude LLO extracted from cultured *A. fulgidus* cells, as described (*5*). The reaction mixture was directly loaded on a COSMOSIL Sugar-D normal-phase column (4.6 mm×25 cm, Nacalai Tesque, Kyoto, Japan). A linear gradient of solvent B was applied from 15% B to 80% B. The eluted materials were collected and dried in a SpeedVac.

3.1.4

Monosaccharide analyses

The monosaccharide composition was analyzed with an ABEE (p-aminobenzoic acid ethyl

ester) labeling kit (J-Oil Mills, Tokyo, Japan) using 950 pmol of the glycopeptide sample (*1*), according to the manufacturer's instructions. The absolute configurations of the monosaccharides were analyzed by acid solvolysis with (*S*)-2-butanol (*6*). The (*S*)-2-butyl glycosides were generated from 1 μg of the glycopeptide sample.

3.1.5

NMR spectroscopy

The NMR sample contained 110 μM ^{13}C-labeled glycopeptide and 0.1 mM DSS (sodium 2,2-dimethyl-2-silapentane-5-sulfonate) standard (δH 0 ppm), dissolved in 260 μL of 99.96% $^{2}H_2O$ in a micro NMR tube (Shigemi, Japan). The ^{1}H chemical shifts relative to tetramethylsilane (TMS) are equal to those relative to DSS. The ^{13}C chemical shift values relative to TMS were calculated using the chemical shift referencing ratio, Ξ=0.25145004. NMR spectra were recorded on a Bruker Avance600 spectrometer equipped with a TXI cryoprobe at 25 °C. All NMR spectra were recorded with 13C WURST decoupling (*7*) during acquisition, unless otherwise indicated. Two-dimensional ^{1}H–^{1}H COSY, 2D ^{1}H–^{1}H TOCSY (a $^{3}J_{HH}$ mixing time of 70 ms), 2D ^{1}H–^{1}H NOESY (NOE mixing time of 600 ms), and 2D ^{1}H–^{1}H ROESY (NOE mixing time of 225 ms) spectra were recorded. Typically, 512 FIDs of 4096 points were acquired, with 64 scans per FID. Two-dimensional ^{1}H–^{13}C CT (constant time)-HSQC (CT=44 ms) and 2D ^{1}H–^{13}C HMBC spectra (*8*) (Δ=50 ms for H–C long-range coupling evolution, without ^{13}C decoupling) were recorded. In total, 512 and 256 FIDs of 4096 and 2048 points were acquired with 16 and 512 scans per FID, respectively. The three-dimensional ^{1}H–^{13}C CCH-TOCSY spectrum (a $^{1}J_{CC}$ mixing time of 10 ms) was recorded (*9*). In total, 48×96 FIDs of 2048 points were acquired with 8 scans per FID. NMR data were processed and displayed using the nmrPipe/nmrDraw program (*10*).

3.1.6

Mass spectrometry

The dried glycopeptide sample was dissolved in 0.1% formic acid and 50% methanol. The direct-infusion electrospray ionization (ESI) mass spectrometry was performed on a QSTAR Elite mass spectrometer (ABSciex) in the positive ion mode. The ionspray voltage (IS) was set to 5.5 kV. Cesium iodine (Cs^{+}=132.9054) and the octapeptide iPD1 (H-Ala-Leu-Ile-Leu-Thr-Leu-Val-Ser-OH, $[M+H]^{+}$=829.5398, Bachem, Switzerland) were used as calibration standards. Typically, the mass accuracy and the FWHM (full width at half

maximum) were 1 ppm and 12,000 in the measured mass range of *m/z* 100–2000. The glycopeptide samples were introduced into the ES source at a flow injection rate of 5 μL min^{-1} with a syringe pump. The scan type of MS/MS analysis was set to the product ion mode. The triply-charged precursor ions were selected in the mass filter quadrupole (Q1) with a mass window of ± 0.7 amu, and subjected to MS/MS analyses with a collision energy of 20 V.

3.2 Results

AglB homologues expression in A. fulgidus cells

The use of the recombinant AglB proteins provides the ability to prepare a specific glycopeptide for each of the AglB paralogs. However, it is difficult to produce large amounts of glycopeptides for NMR analyses, due to the limited availability of crude LLO extracts from cultured cells. Instead, for large-scale production, we used the detergent-solubilized membrane fractions of cultured *A. fulgidus* cells. The Triton X-100-solubilized membrane fractions contain both AglBs and LLOs, and the addition of an acceptor peptide starts the oligosaccharyl transfer reaction. The chemically synthesized acceptor peptide had a fluorescent dye at the N-terminus and a biotin tag attached to the C-terminus, for sensitive detection and affinity purification, respectively. To facilitate the NMR analysis, we added ^{13}C-glucose to the culture medium for metabolic ^{13}C labeling of LLOs. Under our culture conditions, we found that the membrane fractions of A. fulgidus cells contained only AfAglB-L, and not AfAglB-S1 or AfAglB-S2 (Fig. 3-1). Thus, even though the *A. fulgidus* membrane fractions contained the LLOs specific for *Af*AglB-S1 (and probably *Af*AglB-S2), the glycopeptide analyzed in this study was the product of *Af*AglB-L.

Glycopeptides profiling by NP-HPLC

Preparative-scale normal-phase HPLC chromatography was performed, to separate the glycopeptide mixture after the biotin-tag affinity purification. The elution of the glycopeptides was monitored by the fluorescence of the peptide portion (Fig. 3-2A). The main peak (peak 3) was collected and desalted by successive reverse-phase HPLC chromatography, and was used for the following chemical and NMR analyses.

The monosaccharide composition analysis revealed that the glycopeptide contained glucose, mannose, and galactose in a molar ratio of about 2:1:1. The absolute configurations of all of the monosaccharide residues were determined to be the D configuration, based on the gas-liquid chromatography retention times of the trimethylsilylated (*S*)-2-butyl glycoside derivatives.

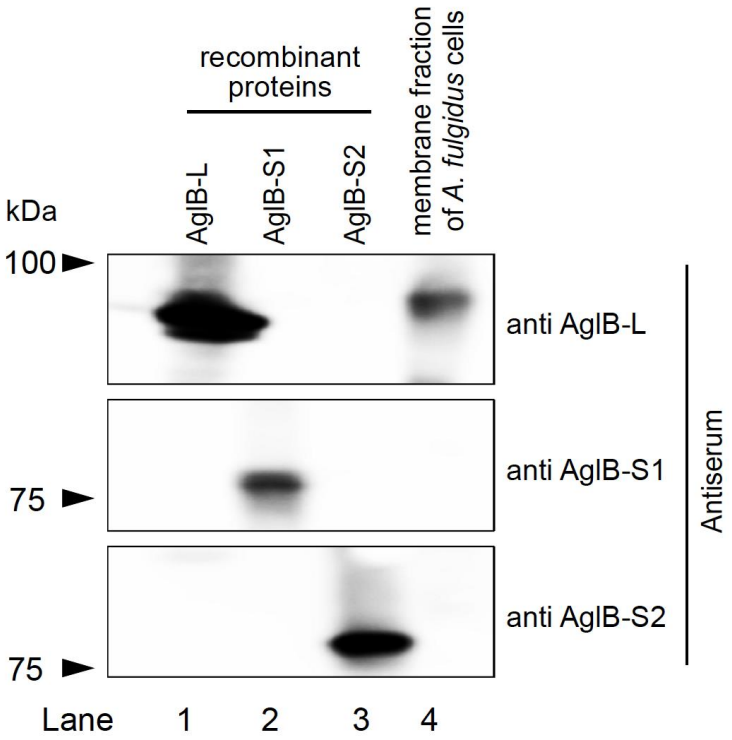


Fig. 3-1. Immunodetection of the three paralogous AglB proteins expressed in the membrane fractions of cultured *A. fulgidus* cells. As positive controls, the E. coli membrane fractions that contained each recombinant AglB protein were included (lanes 1, 2, and 3). The membrane fraction of cultured *A. fulgidus* cells was found to contain only AglB-L (lane 4).

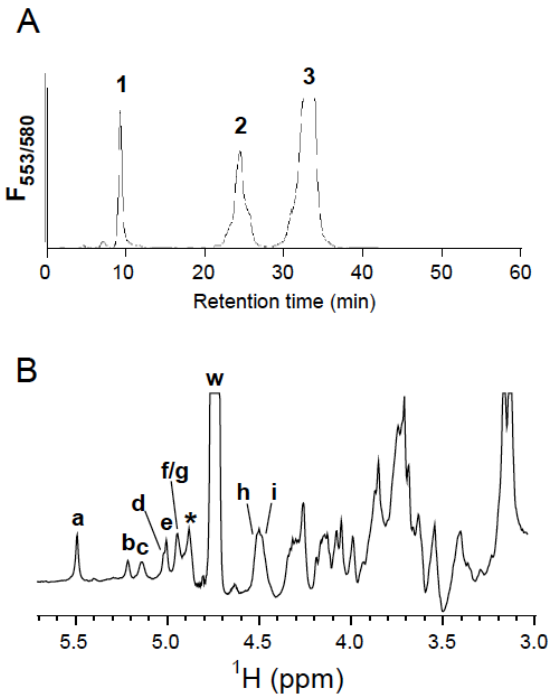


Fig. 3-2. Purification of the glycopeptides and 1D ^{1}H NMR spectrum. (A) Elution profile of the preparative-scale normal-phase HPLC, monitored by the fluorescence with Ex. 553 nm/Em. 580 nm. Peak 3 was collected for sugar and NMR analyses. Peak 1 was the unreacted peptide, and peak 2 was a mixture of glycopeptides modified with a sulfate group (see Fig. 3-6 and Fig. 3-7). (B) 1D ^{1}H NMR spectrum with water presaturation. The oligosaccharide portion of the glycopeptide was metabolically labeled with ^{13}C, but $^{1}J_{CH}$ coupling was suppressed by the ^{13}C decoupling during acquisition. Nine anomeric proton resonances are labeled, **a**–**i**. The **w** denotes the residual water signal, and the asterisk indicates the impurity peak of non-sugar origin.

The structure determination by NMR

The NMR sample contained 110 μM glycopeptide bearing a ^{13}C-labeled oligosaccharide, dissolved in 260 μL $^{2}H_2O$. Nine anomeric ^{1}H resonances were identified in the one-dimensional (1D) ^{1}H spectrum with ^{13}C decoupling during acquisition, and were labeled **a** to **i** (Fig. 3-2B). The resonance assignment was performed by analyzing the two-dimensional (2D) ^{1}H–^{1}H COSY, 2D ^{1}H–^{1}H TOCSY (Fig. 3-3), 2D ^{1}H–^{1}H NOESY, 2D ^{1}H–^{13}C HSQC (Fig. 3-4A), three-dimensional (3D) ^{1}H–^{13}C CCH-TOCSY (Fig. 3-4B), and 2D ^{1}H–^{13}C HMBC (Fig. 3-5) spectra. The ^{1}H and ^{13}C chemical shifts are summarized in Table 3-2.

The homonuclear $^{3}J_{HH}$ coupling constants and the ^{1}H–^{1}H TOCSY transfer patterns from the anomeric protons were used to identify each type of monosaccharide residue (Fig. 3-3). We propose the assignment of the anomeric protons and monosaccharide types as follows: **a**, mannose; **b**, glucose; **c**, galactose; **d**, mannose; **e**, mannose; **f**, glucose; **g**, glucose; **h**, galactose; **i**, glucose. All residues are in the pyranose ring form. The spin networks within residues were confirmed by the ^{13}C–^{13}C correlations in the 2D projection of 3D ^{1}H–^{13}C CCH-TOCSY (Fig. 3-4B).

The linkage positions of the monosaccharide residues were deduced from the downfield shift of the ^{13}C chemical shifts. The downfield-shift deviations from free monosaccharides indicate the positions of the linkage carbons (**a**2, **c**3, **d**6, **e**3/**e**6, **b**4, and **h**3/**h**4) of the non-terminal residues (orange bars, Fig. 3-5A). The anomeric carbons were also down-field shifted, except for the glucose **g** (cyan bars, Fig. 3-5A). In fact, the anomeric carbon of the glucose **g** was considerably shifted upfield at 79.8 ppm, as compared to those of the other anomeric carbons at 95–104 ppm, indicating that glucose **g** is linked to the amide nitrogen of the Asn residue in the peptide (**g**1 vs **a**1-**f**1, **h**1, and **i**1 in Fig. 3-4A). The linkages between the monosaccharide residues were confirmed by the inter-residual long-range $^{3}J_{CH}$ cross peaks in the HMBC spectrum (Fig. 3-5B). We noticed during the NMR analysis that the ^{13}C metabolic labeling was not uniform. The variations in the cross peak intensities in the ^{1}H–^{13}C HSQC (Fig. 3-4A) and the ^{1}H–^{13}C CCH-TOCSY (Fig. 3-4B) spectra suggested that the glucose and galactose residues were almost fully ^{13}C labeled, but the mannose residues were only partially labeled. The MS analysis of the NMR sample also suggested that the mannose labeling was fractional with respect to each carbon atom, with about 20% efficiency. This complex labeling pattern made the HMBC linkage analysis apparently complicated. Besides the intra-residual connections due to $^{1}J_{CH}$, (cyan cross peaks), $^{2}J_{CH}$ (green cross peaks), and $^{3}J_{CH}$ (orange cross peaks), 10 connections originating from

inter-residual magnetic transfers based on $^{3}J_{CH}$ (magenta cross peaks) were identified in the HMBC spectrum (Fig. 3-5B). These inter-residual $^{3}J_{CH}$ cross peaks were split into doublets along the ^{1}H axis by the one-bond $^{1}J_{CH}$ (corresponding to $\Delta\delta$H=0.27 ppm), in the cases where the observed ^{1}H belonged to the glucose/galactose residues, and were single peaks in the other cases where the observed ^{1}H belonged to the mannose residues. The down-field ^{13}C shifts and the inter-residual HMBC correlations established the inter-glycosidic linkages, as shown in Fig. 3-6. It should be noted that the two mannose residues, **d** and **e**, were assigned to the same mannose residue, without and with glucose **b**, respectively.

The anomeric configuration of each monosaccharide residue were determined from the heteronuclear $^{1}J_{C1–H1}$ values obtained from the HSQC spectra, recorded without ^{13}C decoupling during acquisition (data not shown). The large $^{1}J_{C1–H1}$ values (>168 Hz) indicated that residues **a** (Man), **b** (Glc), **c** (Gal), **d/e** (Man), and **f** (Glc) were in the α-configuration, whereas the small $^{1}J_{C1–H1}$ values (<160 Hz) indicated that residues **g** (Glc), **h** (Gal), and **i** (Glc) were in the β-configuration (*11*). The homonuclear $^{3}J_{H1–H2}$ values can be used to confirm the anomericity in the case of glucose and galactose residues: small $^{3}J_{H1–H2}$ values indicate the α-configuration (**b**, **c**, and **f**), and large $^{3}J_{H1–H2}$ values indicate the β-configuration (**g**, **h**, and **i**).

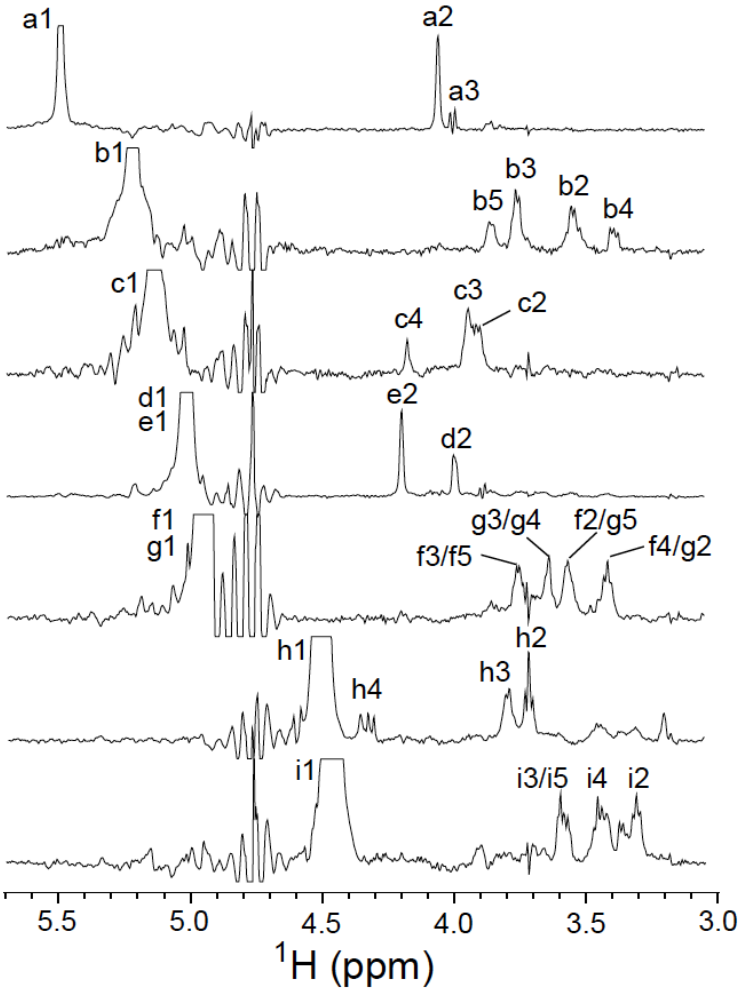


Fig. 3-3. 1D slices of the 2D ^{1}H–^{1}H TOCSY spectrum of the glycopeptide. ^{13}C decoupling was performed during acquisition.

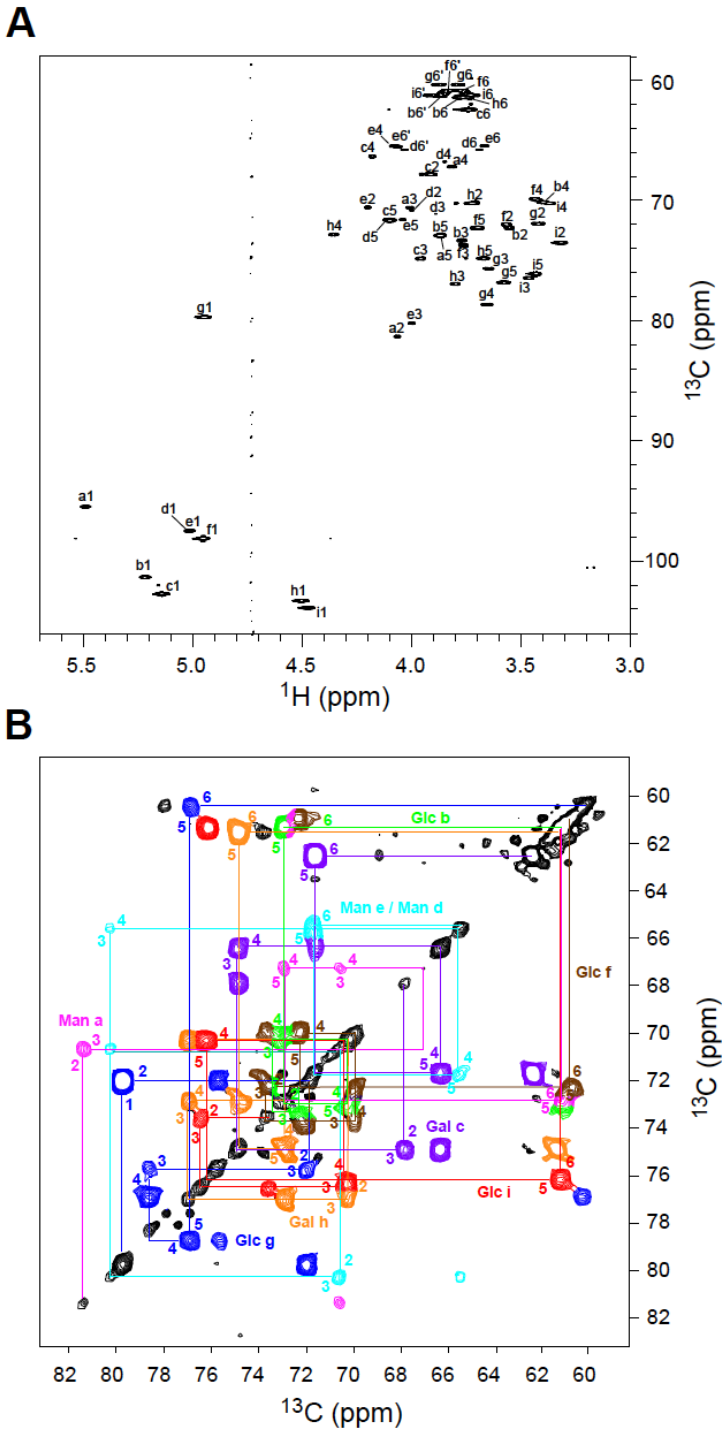


Fig. 3-4. ^{1}H–^{13}C and ^{13}C–^{13}C correlations of the glycopeptide. Only the oligosaccharide portion was metabolically labeled with ^{13}C, and thus was selectively observed. (A) The 2D ^{1}H–^{13}C CT-HSQC spectrum, and (B) a portion of the 2D projection of the 3D ^{1}H–^{13}C CCH-TOCSY spectrum. In (B), the ^{13}C–^{13}C correlations involving anomeric carbons are located outside of the displayed region.

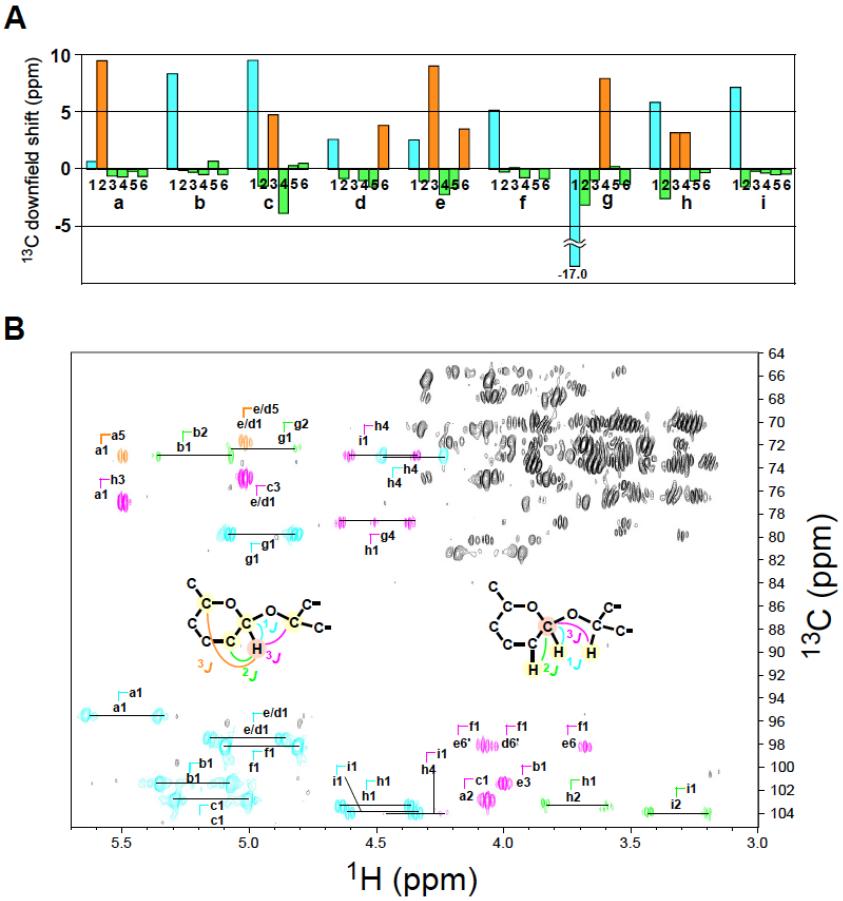


Fig. 3-5. Determination of the linkages between the monosaccharide residues. (A) The ^{13}C chemical shift deviations from free monosaccharides were plotted as a function of the carbon positions. The ^{13}C chemical shifts of the free monosaccharides in the pyranose conformations were obtained from the previous report (*11*). Large downfield shifts (>+ 3 ppm, orange bars) suggest the non-aromatic carbons involving in the glycosidic linkage. The anomeric carbons also showed large downfield shifts (>+ 2.5 ppm, cyan bars), except for mannose **a** and the Asn-linked glucose **g**. (B) A portion of the 2D ^{1}H–^{13}C HMBC spectrum. The cross peaks originating from the intra-residual spin couplings, $^{1}J_{CH}$, $^{2}J_{CH}$, and $^{3}J_{CH}$, are colored cyan, green, and orange, respectively. The cross peaks originating from the inter-residual spin coupling, $^{3}J_{CH}$, are colored magenta, and indicate the formation of the glycosidic bonds. Insets show the magnetic transfer routes schematically.

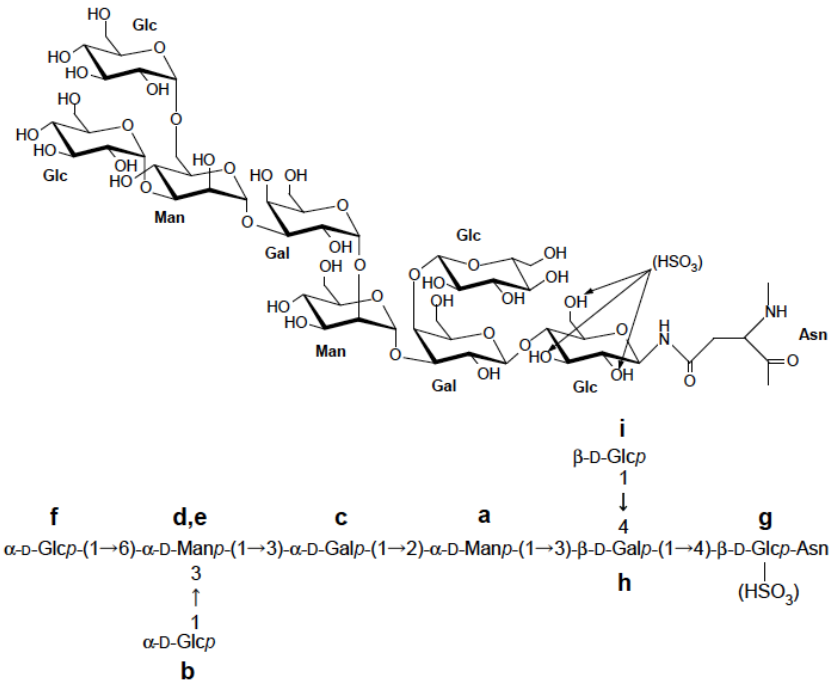


Fig. 3-6. Structure of the N-linked glycan produced by *Archaeoglobus fulgidus* AglB-L. The α1,3-linked glucose **b** is attached to the mannose **e** at the branching site. The mannose **d** is the unbranched mannose residue without the optional glucose **b**. The glucose **g** in the NMR sample lacked a sulfate group, but the glycopeptides in peak 2 (Fig. 3-2A) possessed a sulfate group attached to the glucose residue directly linked to the Asn residue.

Table 3-1

^{1}H and ^{13}C NMR chemical shifts and $^{1}J_{C1\text{-}H1}$ values of the N-linked glycan that is a product of *Archaeoglobus fulgidus* AglB-L

Residue	H-1 C-1 $^{1}J_{C1\text{-}H1}$	H-2 C-2	H-3 C-3	H-4 C-4	H-5 C-5	H-6, H-6' C-6
a	5.49	4.06	4.01	3.82	3.87	(3.76, 3.86)
-2)-α-Man	95.5	81.3	70.6	67.2	73	(61.3)
	176					
b	5.22	3.55	3.77	3.40	3.86	3.74, 3.86
α-Glc	101.4	72.3	73.4	70.2	73	61.3
	171					
c	5.15	3.92	3.95	4.18	4.10	3.73
-3)-α-Gal	102.8	67.8	74.9	66.4	71.6	62.5
	168					
d	5.03	4.00	3.89	3.85	4.08	3.68, 4.03
-6)-α-Man	97.5	70.8	71.2	66.9	71.7	65.8
	172					
e	5.02	4.20	4.00	4.07	4.04	3.66, 4.06
-3,6)-α-Man	97.5	70.6	80.3	65.6	71.6	65.5
	172					
f	4.95	3.57	3.76	3.43	3.70	3.77, 3.83
α-Glc	98.2	72.1	73.8	69.9	72.3	60.9
	171					
g	4.94	3.42	3.64	3.65	3.58	3.78, 3.88
-4)-β-Glc	79.8	72.0	75.7	78.7	76.9	60.4
	156					
h	4.50	3.71	3.80	4.35	3.67	3.75
-3,4)-β-Gal	103.3	70.3	77.0	72.9	74.8	61.4
	160					
i	4.47	3.31	3.46	3.37	3.44	3.71, 3.91
β-Glc	104.0	73.6	76.5	70.3	76.2	61.3

Chemical shifts in ppm relative to TMS, and ^{1}J coupling constants in Hz measured in $^{2}H_2O$ at 25 °C. Average errors of ±0.02 ppm for ^{1}H, ±0.1 ppm for ^{13}C, and ±0.5 Hz for $^{1}J_{C1\text{-}H1}$.

Sulfate modification detected by MS/MS analysis

A minor peak present in the chromatogram of the normal-phase HPLC was analyzed by MS spectrometry (peak 2 in Fig. 3-2A). The interpretation of the MS/MS spectra was not straightforward due to the ^{13}C labeling, but suggested the presence of a chemical modification of the oligosaccharide moiety. Thus, a non-labeled glycopeptide sample was prepared by the in vitro oligosaccharyl transfer reaction using the recombinant AfAglB-L and crude LLO, and was subjected to the MS analysis. Fig. 3-7 shows the MS/MS spectrum of the $[M+3H]^{3+}$ ion of a modified heptaoligosaccharide peptide. There are seven pairs of peaks with an *m/z* difference of 26.65, which suggests a modification with a sulfate group. Considering that the fragment ion Y_1 has its paired ion, Y_1–SO3, but Y_0 has no counterpart, the sulfate group is attached to the innermost glucose residue linked to the Asn residue in the peptide (inset in Fig. 3-7). Although it would be possible to determine the linkage position of the sulfate group by ^{13}C NMR, we did not perform this analysis, due to the limited availability of the ^{13}C-labeled sulfated glycopeptide sample.

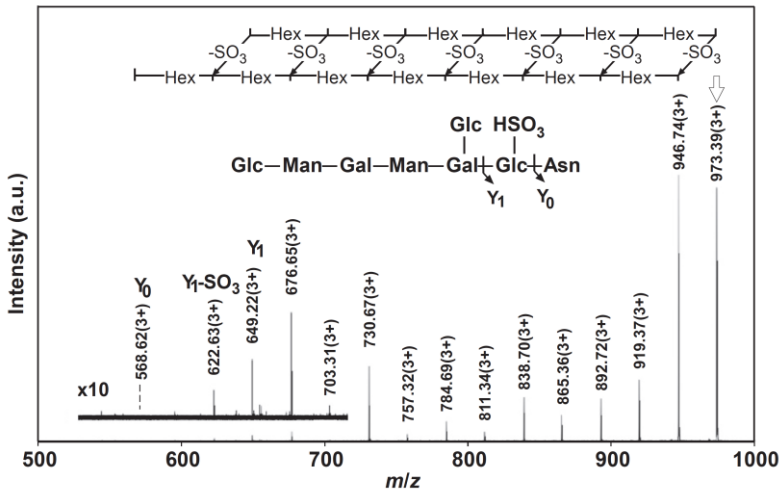


Fig. 3-7. ESI-MS/MS analysis of the glycopeptide bearing a sulfated heptaoligosaccharide. The monoisotopic precursor ion is marked by the vertical arrow. There are two series of $[M+3H]^{3+}$ fragment ions, originating from the sequential loss (Δ 54.023^{+}) of hexose residues. The constant *m/z* differences (Δ 26.653^{+}) between the two mass ladders suggest the modification of a sulfate group. The *inset* shows that the sulfate group is attached to the glucose residue directly linked to the Asn residue in the peptide. The expected *m/z* values were observed within 0.01 of the theoretical mass values.

3.3
Discussion

On the basis of our data, we concluded that the oligosaccharide chain that *A. fulgidus* AglB-L transfers to the Asn residues in proteins had the structure in Fig. 3-6. The archaeal N-glycans exhibit a high degree of structural variation (*12*). Here, we added another unique structure to the known archaeal N-glycan repertoire. In particular, the monosaccharide residue that is directly linked to the Asn residue is hexose, which is not very common, since the linking monosaccharide residue usually possesses an N-acetyl group at the C-2 position. This is interesting, because the involvement of the N-acetyl group in the oligosaccharyl transfer reaction was suggested for the eubacterial OST enzyme, PglB (*13*). The other known exceptions include the archaeal N-glycans from *Halobacterium salinarum* (glucose) (*14*), *Haloferax volcanii* (hexose) (*15*), and *Thermoplasma acidophilum* (galactose) (*16*). This feature is also seen in the second N-glycan structure that is specific for *A. fulgidus* AglB-S1 (*2*). Finally, the N-glycan structures expressed in vivo may be different due to post-transfer processing of the oligosaccharide chains on proteins. Such post-transfer modification was not known in Archaea for a long time, but the addition of the terminal mannose residue to the N-linked tetrasaccharide on the S-layer glycoprotein was recently discovered in *H. volcanii* (*17*).

We showed that *A. fulgidus* cells expressed only *Af*AglB-L, and not *Af*AglB-S1 or *Af*AglB-S2, under our culture conditions (Fig. 3-1). The alteration in the N-glycan structure of *H. volcanii* was reported, as an adaptation to different salinity concentrations of the culture medium (*18*). In a similar way, *A. fulgidus* may change its N-glycan structure to adapt to harsh external environments, by the induced expression of *Af*AglB-S1 and *Af*AglB-S2.

3.4

Reference

(1) D. Fujinami, M. Matsumoto, T. Noguchi, K. Sonomoto, D. Kohda, Carbohydr Res, 387 (2014), pp. 30–36

(2) S. Matsumoto, M. Igura, J. Nyirenda, M. Matsumoto, S. Yuzawa, N. Noda, et al, Biochemistry, 51 (2012), pp. 4157–4166

(3) S. Matsumoto, A. Shimada, D. Kohda, BMC Struct Biol, 13 (2013), p. 11

(4) J. Nyirenda, S. Matsumoto, T. Saitoh, N. Maita, N.N. Noda, F. Inagaki, et al, Structure, 21 (2013), pp. 32–41

(5) S. Matsumoto, A. Shimada, J. Nyirenda, M. Igura, Y. Kawano, D. Kohda, Proc Natl Acad Sci USA, 110 (2013), pp. 17868–17873

(6) G.J. Gerwig, J.P. Kamerling, J.F. Vliegenthart, Carbohydr Res, 77 (1979), pp. 10–17

(7) E. Kupce, R. Freeman, J Magn Reson A, 115 (1995), pp. 273–276

(8) D.O. Cicero, G. Barbato, R. Bazzo, J Magn Reson, 148 (2001), pp. 209–213

(9) K. Gehring, I. Ekiel, J Magn Reson, 135 (1998), pp. 185–193

(10) F. Delaglio, S. Grzesiek, G.W. Vuister, G. Zhu, J. Pfeifer, A. Bax, J Biomol NMR, 6 (1995), pp. 277–293

(11) P.K. Agrawal, Phytochemistry, 31 (1992), pp. 3307–3330

(12) K.F. Jarrell, Y. Ding, B.H. Meyer, S.V. Albers, L. Kaminski, J. Eichler, Microbiol Mol Biol Rev, 78 (2014), pp. 304–341

(13) M. Wacker, M.F. Feldman, N. Callewaert, M. Kowarik, B.R. Clarke, NL. Pohl, et al, Proc Natl Acad Sci USA, 103 (2006), pp. 7088–7093

(14) J. Lechner, F. Wieland, M. Sumper, J Biol Chem, 260 (1985), pp. 860–866

(15) M. Abu-Qarn, S. Yurist-Doutsch, A. Giordano, A. Trauner, H.R. Morris, P. Hitchen, et al, J Mol Biol, 374 (2007), pp. 1224–1236

(16) E. Vinogradov, L. Deschatelets, M. Lamoureux, G.B. Patel, T.L. Tremblay, A. Robotham, et al, Glycobiology, 22 (2012), pp. 1256–1267

(17) C. Cohen-Rosenzweig, S. Yurist-Doutsch, J. Eichler, J Bacteriol, 194 (2012), pp. 6909–6916

(18) Z. Guan, S. Naparstek, D. Calo, J. Eichler, Environ Microbiol, 14 (2012), pp. 743–753

Chapter 4- Structural elucidation of an asparagine-linked oligosaccharide from the hyperthermophilic archaeon, *Pyrobaculum calidifontis.*

4.1
Material and Methods

4.1.1
Preparation of the membrane fractions from cultured *P. calidifontis* cells

P. calidifontis strain JCM 11548 was obtained from Japan Collection of Microorganisms, the Microbe Division of RIKEN Bio resource Center (Tsukuba, Japan). *P. calidifontis* cells were grown aerobically at 90 ^{o}C in the medium consisted of 1.0 g yeast extract, 10 g of tryptone, and 3.0 g of $Na_2S_2O_3$ $5H_2O$ per liter of tap water, as described (*1*). For ^{13}C isotope labeling, 200 mg 1-^{13}C or U-^{13}C-glucose (99 atom%, ISOTEC) was add to the medium per liter. The cells were disrupted by sonication, centrifuged at 8,500 g for 10 min. The supernatant was collected and ultra centrifuged at 100,000 g for 1 h to collect the membrane fraction (the pellet) and cytosol fraction (the supernatant). The membrane fractions were solubilized by triton X-100 detergent, according to the procedure used for *Pyrococcus furious* cells (*31*). Finally, we prepared 20 mL of Triton X-100-solubilized membrane fractions from 30-liter cultured *P. calidifontis* cells.

4.1.2
Glycoprotein isolation from membrane fractions

Glycoproteins were isolated from Triton X-100 solubilized membrane fractions using a ConA-agarose (Wako, Japan), according to manufacture's instruction. Briefly, the 5 mL of the membrane fractions were loaded on a ConA-agarose (0.5-ml bed volume) column in the presence of 1% Triton X-100, washed five times with buffer (20 mM Tris-HCl, pH 7.5, containing 500 mM NaCl and 0.1 % Triton X-100), and eluted by 1 mL of the buffer A with 1 M methyl α-D-mannopyranoside. The glycoproteins were detected by PAS glycoproteins staining kit (Polyscience Inc).

4.1.3
LC-ESI-MS/MS analysis of the ConA enriched glycoproteins

The ConA enriched proteins were dissolved in SDS sample buffer, fractionated by SDS–polyacrylamide gel electrophoresis (PAGE), and stained with CBB. The lane of the stained gel was

sliced into 10 pieces, and proteins within these pieces were subjected to in-gel digestion with trypsin as described previously (*2*). The obtained peptides were dried and then dissolved in a solution containing 0.1% trifluoroacetic acid and 2% acetonitrile before nanoscale liquid chromatography (nanoLC)–MS/MS analysis with a system consisting of an LTQ Orbitrap Velos mass spectrometer (Thermo Fisher Scientific, Waltham, MA) coupled with a Advance UHPLC (Bruker) and HTC-PAL autosampler (CTC Analytics, Zwingen, Switzerland). Peptide separation was performed with an in-house pulled fused silica capillary (internal diameter, 0.1 mm; length, 15 cm; tip internal diameter, 0.05 mm) packed with 3-μm C18 L-column (Chemicals Evaluation and Research Institute, Japan). The mobile phases consisted of 0.1% formic acid (Solvent A) and 100% acetonitrile (Solvent B). Peptides were eluted with a gradient of 5–40 % Solvent B for 20 min at a flow rate of 300 nL/min. Full MS spectra were obtained with Orbitrap in the *m/z* range of 300–2000 with a resolution of 60,000. Auto gain control (AGC) was set to $1e^6$. Lock mass function was activated to minimize mass error during analysis. Following each survey scan the top nine most intense ions with multiple charged ions were selected for fragmentation at normalized collision energy of 35%. Fragment ion spectra produced via collision-induced dissociation (CID) were acquired in the ion trap. AGC was set to $1e^4$. All data were acquired with Xcalibur software v2.2. For glycoprotein identification, the generated peak lists were searched against *P. calidifontis* genome extracted from NCBInr database with the use of the MASCOT algorithm. Trypsin was selected as the enzyme used, the allowed number of missed cleavages was set at 2, and carbamidomethylation on cysteine was selected as the fixed modification. Oxidized methionine was searched as variable modifications. Precursor mass tolerance was 10 ppm, and tolerance of MS/MS ions was 0.8 Da. The threshold used for glycoprotein was a MASCOT score ≥ 100.

4.1.4

Generation of asparaginyl-oligosaccharides from glycoproteins

The asparaginyl-oligosaccharides were generated by Pronase E digestion of ConA enriched glycoproteins, according to (*3*). Briefly, 30 μg of Pronase E (Sawai Pharmaceutical, JAPAN) were added to 1 mL of the glycoprotein solution and then heated at 40 ^{o}C. Another 30 μg of Pronase E was added to the solution after 24 h. After further 36-h incubation time, the reaction mixture was centrifuged at 1200 g for 10 min. The brown supernatant was collected and dried in a SpeedVac concentrator (Thermo Savant). The dried pellet was redissolved in 0.1 mL of 0.1 M acetic acid and centrifuged at 1200 g for 10 min. The clean supernatant was collected and dialyzed to water by using Spectra/Por membrane tubing (MWCO 100-500 Da) for 24 h to remove free amino acids. The dialyzed solution was separated by gel-permeation chromatography utilizing Superdex 75 (GE healthcare). The eluted fractions were collected and used for following LC-ESI-MS/MS

analysis.

4.1.5

LC-ESI-MS/MS analysis of the asparaginyl-oligosaccharides

The asparaginyl-oligosaccharides were further separated on a COSMOSIL Sugar-D NP-HPLC column (4.6 mm × 25 cm, Nacalai Tesque, Kyoto, Japan). Solvent A was 10 mM CH_3COONH_4, CH_3CN:H_2O = 9:1 (v:v), and Solvent B was 10 mM CH_3COONH_4, CH_3CN:H_2O = 1:9 (v:v). A linear gradient of solvent B was applied from 30 % B to 100 % B. A post-column splitter was used to deliver 20% of the LC flow to the ESI source of a QSTAR Elite mass spectrometer (ABSciex). The parameter sets of mass spectrometer were set according to (*1*). The doubly charged negative precursor ion was selected and subjected to MS/MS analyses with collision energy of 55.

4.1.6

Peptide substrate synthesis

The amino acid sequence of the acceptor peptide is Ala-Ala-Tyr-Asn-Val-Thr-Lys-Arg-Lys, where the underlined sequence is the N-glycosylation sequon. The peptide was synthesized with an N-terminal TAMRA (5/6-carboxytetramethylrhodamine) fluorophore, and a C-terminal biotin group attached to the side-chain amino group of the C-terminal Lys residue (Toray Research Center, Japan). The synthesized peptide eluted as two peaks in the reverse phase HPLC chromatography due to the 5- and 6-isomers of the TAMRA fluorophore. We used the peak that eluted more slowly. The peptide concentrations were determined by the absorbance of the TAMRA fluorophore at 558 nm, with an extinction coefficient of 90,000 M^{-1} cm^{-1}.

4.1.7

Glycopeptides preparation by in vitro OST enzymatic reaction

By the OST enzymatic activity, lipid-linked oligosaccharide chain was transferred to acceptor peptide. In this experiment, glycopeptides were produced enzymatically by the addition of acceptor peptide substrate into membrane fractions, which contain both OST enzyme and donor lipid-linked oligosaccharide substrate. The reaction mixture was consisted of the 0.5 mL of 0.5 mM the acceptor peptide, the 20 mL of the Triton X-100-solubilized membrane fractions, and 79.5 mL of buffer (50 mM Tris-HCl, pH 7.5, containing 0.02 % (v/v) Tween 20, and 10 mM $MgCl_2$). OST enzyme reaction was performed at 65 ^oC for 1 h. OST reaction was stopped by cooling reaction mixture on ice. The product glycopeptide was adsorbed to Streptavidin Mutein Agarose (Roche) in the presence of 0.4 M $(NH_4)_2SO_4$, and eluted in the presence of 5 mM biotin.

The eluted glycopeptide mixture was further separated on a COSMOSIL Sugar-D NP-HPLC column (4.6 mm × 25 cm). Solvent A was 10 mM CH_3COONH_4, $CH_3CN:H_2O$ = 9:1 (v:v), and Solvent B was 10 mM CH_3COONH_4, $CH_3CN:H_2O$ = 1:9 (v:v). A linear gradient of solvent B was applied from 15% B to 80% B. The each three peaks were collected and loaded on a COSMOSIL 5C18-AR-II reverse-phase column respectively (4.6 mm × 25 cm, Nacalai Tesque, Kyoto, Japan) in 0.1% trifluoroacetic acid (TFA) and 10 % acetonitrile for desalting. The materials eluted by a linear gradient of 10-90 % acetonitrile were collected and dried in a SpeedVac concentrator (Thermo Savant).

4.1.8

Sugar analyses

The monosaccharide composition of the glycopeptide sample (0.1 nmol) was analyzed with an ABEE (*p*-aminobenzoic acid ethyl ester) labeling kit (J-Oil Mills, Tokyo, Japan), according to the manufacturer's instruction. The absolute configurations of the monosaccharides were analyzed by acid solvolysis with (*S*)-2-butanol (Tokyo Chemical Industry Co, JPN). The (*S*)-2-butyl glycosides were generated from 1 ug of the glycopeptide sample as described (*4*).

4.1.9

NMR spectroscopy

Dried glycopeptide were dissolved in 260 μl of 99.96 % ^2H_2O with 0.1 mM DSS (sodium 2,2-dimethyl-2-silapentane-5-sulfonate) standard (δH 0 ppm) and transferred into micro NMR tube (Shigemi, Japan). The concentration of glycopeptides S10 were 72 μM (U-^{13}C labeled glycan) and 50 μM (1-^{13}C labeled glycan) and glycopeptide S11 were 30 μM (U-^{13}C labeled glycan). ^{1}H chemical shifts relative to tetramethylsilane (TMS) are equal to those relative to DSS. ^{13}C chemical shift values relative to TMS were calculated using the chemical shift referencing ratio, Ξ = 0.25145004. NMR spectra were recorded on a Bruker Avance600 spectrometer equipped with a TXI cryoprobe at 25 °C. All NMR spectra were recorded with the ^{13}C WURST decoupling (*5*) during acquisition unless otherwise indicated. Two-dimensional ^{1}H-$_1$H COSY, 2D ^{1}H-^{1}H TOCSY (a $^3J_{HH}$ mixing time of 70 ms), 2D ^{1}H-^{1}H relayed TOCSY (a $^3J_{HH}$ mixing time of 90 ms and a delay time of 100 ms), 2D ^{1}H-^{1}H NOESY (an NOE mixing time of 600 ms) spectra were recorded. Typically, 512 FIDs of 4096 points were acquired with 40 scans per FID. Two-dimensional ^{1}H-^{13}C HSQC, 2D ^{1}H-^{13}C HSQC-TOCSY (a $^3J_{HH}$ mixing time of 70 ms), 2D ^{1}H-^{13}C HSQC-NOESY (an NOE mixing time of 600 ms) were recorded with a constant

time of CT=44 ms. Two dimensional ^{1}H-^{13}C HMBC spectra (delay time = 50 ms for H-C long-range coupling evolution) were recorded without ^{13}C decoupling (*6*). 512 FIDs of 4096 points were acquired with 16 scans per FID. Three-dimensional ^{1}H-^{13}C CCH-TOCSY (a $^{1}J_{CC}$ mixing time of 10 ms) was recorded (*7*). 48 × 96 FIDs of 2048 points were acquired with 8 scans per FID. NMR data were processed and displayed using the nmrPipe/nmrDraw program (*8*).

4.2

Results

Identification of P. calidifontis glycoproteins

P. calidifontis glycoproteins were enriched by ConA (Concanavalin A) affinity chromatography (Figures 4-1A), and detected by periodic acid-Schiff (PAS) staining on SDS-PAGE gels (Figure 4-1B). Many PAS-positive bands were detected in the membrane fractions (lane 4). The SDS-PAGE gel was sliced into 10 pieces (lane 9), in-gel digested with trypsin, and analyzed by nano LC-MS/MS (Figures 4-1C and 4-1D). The MASCOT search identified 130 glycoproteins with at least one potential N-glycosylation site (appendix Table I). One interesting example is the *P. calidifontis* oligosaccharyltransferase (Pcal_0997). An oligopeptide binding protein (Pcal_0975) is a homologue of the well-characterized *Sulfolobus solfataricus* P2 glycoprotein (SSO1273), with 20 N-glycosylation sites (*9*). The Pcal_0975 possesses three potential N-glycosylation sites (N287, N515 and N528), but only N528 is the conserved in SSO1273. An ABC transporter substrate-binding protein (Pcal_2112) and a maltooligosaccharide-binding protein (Pcal_1617) are homologous to the previously reported S. solfataricus P2 glycoproteins, SSO3066 and SSO1171, respectively (*9*).

Identification of glycopeptides generated by trypsin digestion

The large size of the *P. calidifontis* N-linked oligosaccharide hinders the detection of the most of the tryptic glycopeptides due to the out of the sensitive mass range. Thus, we focused the analysis of the $[M+2H]^{2+}$ ions of twelve small trypic glycopeptides (< *m/z* 1550) by nano LC-MS orifiles (appendix Table II). As an example, the averaged MS spectrum during 11.2-11.6 min in the reverse phase-HPLC chromatogram showed two $[M+2H]^{2+}$ peaks with *m/z* 1320.56 and 1401.58, which corresponded to the amino acid sequence, $V^{54}AN^{56}ISASVDK^{63}$, generated from a signal peptidase (Pcal_0153) modified with nonasaccharide (S9) and decasaccharide (S10), respectively (Figure 4-1C). These assignments were accomplished by the low-energy collision-induced dissociation (CID) MS/MS analysis (Figure 4-1D). The fragment ions, $Y4^+$ to $Y6^+$, in the MSMS spectrum of the parent ion with *m/z* 1401.58 indicated the cleavages of the peptide backbone (Figure 4-1D). The fragment ions, Y_A^+ and Y_B^+, showed the existence of the $HexA(NAc)_2$-$Hex(NAc)_2$ core structure attached to the peptide, and the mass ladder showed the eight hexose residues as the extension of the core disaccharide. We created a sequence logo

representation of the amino acid occurrence within and around the experimentally confirmed N-glycan occupied sites (Figure 4-1E). The statistical analysis confirmed the requirement of Asn at the 0 position and the requirement of Thr or Ser at the +2 position in the sequon. Even though the small number of instances was available, the exclusion of Pro at the +1 position was also confirmed.

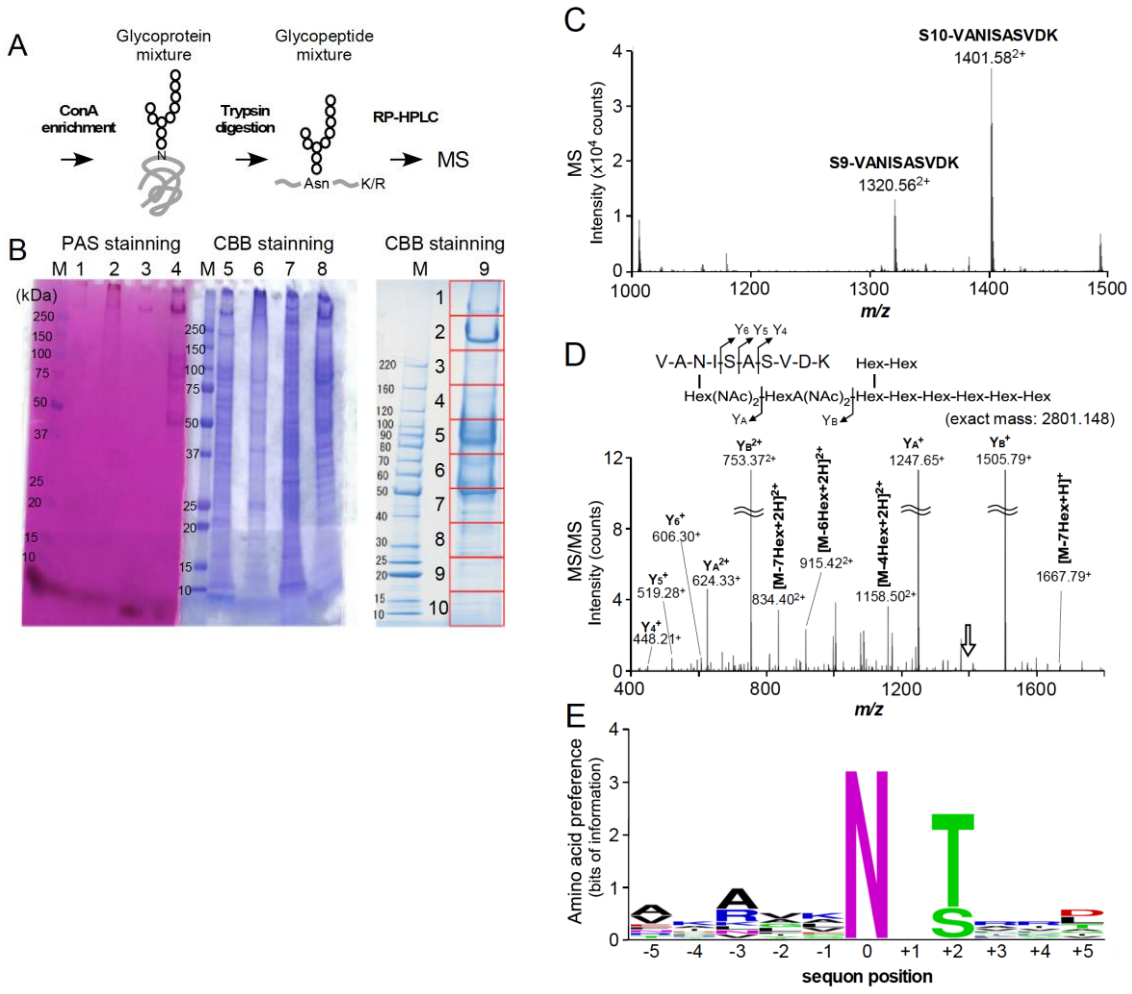


Figure. 4-1. NanoLC-MS/MS analyses of the ConA enriched glycoprotein (A) Schematic representations of nanoLC-MS/MS analyses. The glycoprotein mixture was prepared by ConA chromatography of *P. calidifontis* Triton X-100 solubilized membrane fraction and *in-gel* digested by trypsin. The tryptic glycopeptides were analyzed by reverse-phase nanoLC-MS/MS. (B) (*left panel*) SDS-PAGE (10–20 % gradient gel) image of the different fractions obtained from *P. calidifontis* cells. Lane 1 and 5, broad-range markers; lane 2 and 7, flow-thorough fraction of ConA chromatography of cytosolic fraction; lane 3 and 8, elution fraction of ConA chromatography of cytosolic fraction; lane 4 and 9, flow-thorough fraction of ConA chromatography of Triton X-100 solubilized membrane fractions; lane 5 and 10, elution fraction of ConA chromatography of Triton X-100 solubilized membrane fraction. Lane 1-5 and were stained by PAS staining and Lane 6-10 were stained by CBB staining. (*right panel*) SDS-PAGE image of elution fraction of ConA chromatography of membrane fraction for *in-gel* digestion. SDS-PAGE gel was sliced into 10 pieces as shown by the red rectangular. (C) The full MS spectrum from the mass scans acquired during 11.2-11.6 min. The two mass peaks at *m/z* 1320.56 and 1401.58 were assigned as the glycopeptides ($V^{54}AN^{56}ISASVDK^{63}$) with nona-saccharide S9

and deca-saccharide S10. (D) The ESI-MS/MS spectrum of precursor ion at *m/z* 1408.58. The sequential loss of nine monosaccharide residues and fragment ions derived from cleavage of the peptide backbone Y4 to Y6 were observed. (E) Statistical analysis of active N-glycosylation sites found in *P. calidifontis* glycoproteins. Fractional occurrence of amino acids in the region from −5 to +5 of the glycosylated Asn residue (16 sequons). The figure was prepared from the data of appendix Table II, by means of WebLogo (http://weblogo.berkeley.edu/logo.cgi).

N-linked oligosaccharide profile by NPLC-MS/MS analysis

First, the ConA enriched glycoprotein mixture was completely digested by Pronase E to produce asparaginyl-oligosaccharide mixture (Figure 4-2A). The generated asparaginyl-oligosaccharides were purified by gel filtration chromatography, and then subjected to normal phase (NP)-HPLC that directly connected to a mass spectrometer operated in negative ion mode, to obtain the monosaccharide composition. The total ion chromatogram (TIC, ion chromatogram created by taking the summed intensities of all ions) of the online ESI-MS analysis was shown in Figure 4-2B. The MS spectrum, averaged mass scans during 35-40 min of TIC, showed four $[M-2H]^{2-}$ ion peaks with *m/z* 802.27, 883.30, 964.33, and 1045.35, which were assigned to the asparaginyl-oligosaccharides containing 8, 9, 10, and 11 monosaccharide residues (theoretical *m/z* 802.2735, 883.299, 964.3263, and 1045.3527, respectively) (Figure 4-2C). For confirmation, the MS/MS spectrum of the precursor ion with *m/z* 964.3 (S10-Asn) showed a mass ladder corresponding to the sequential losses of four hexose residues (Figure 4-2D). Based on the MS peak intensities in Fig. 4-2C, we estimated the composition of the number of the monosaccharide residues (Figure 4-2E).

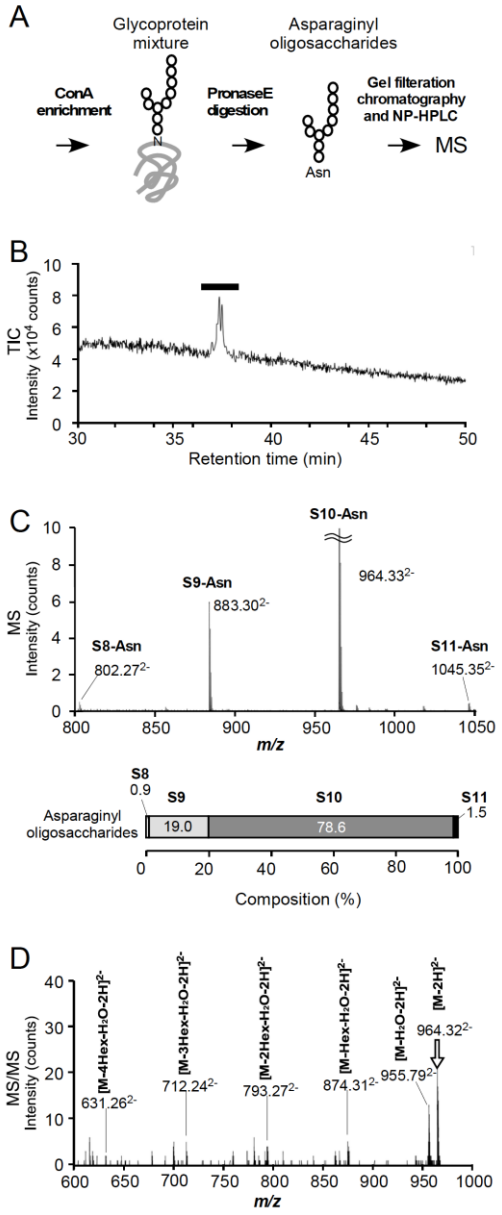


Figure 4-2. Normal-phase LC-MS/MS analyses of the asparaginyl-oligosaccharides (A) Schematic representations of asparaginyl-oligosaccharides preparation. The asparaginyl-oligosaccharides were liberated from ConA enriched glycoproteins by pronase E digestion. The asparaginyl-oligosaccharides were separated by gel filtration chromatography and analyzed by normal-phase LC-MS/MS. (B) Total ion chromatography of the asparaginyl-oligosaccharides. (C) The TOF-MS spectrum averaged from the mass scans acquired during 35- 40 min of the TIC, indicated by *horizontal red bars*, are shown. The TOF-MS spectrum was acquired in the negative ion mode. The 4 mass peaks at *m/z* 802.27, 883.33, 964.33, and 1097.91 were assigned as asparagineyl-oligosaccharides consist of eight, nine, ten, and eleven

monosaccharides respectively. The oligosaccharide composition in the asparaginyl-oligosaccharides in 100 % stacked charts was in below. (D) ESI-MS/MS spectrum. The precursor ion at *m/z* 964.32 was selected and marked by the *vertical arrow*. The sequential loss of four hexoses and H_2O loss were observed. The expected *m/z* values were observed within 0.01.

Preparation of glycopeptides by in vitro oligosaccharyl transfer reaction

For the NMR analysis of the N-glycan structure, the oligosaccharide chain was transferred from the lipid carrier to a structurally defined peptide by the action of the OST enzyme (Figure 4-3A). The Triton X-100-solubilized membrane fractions prepared from cultured *P. calidifontis* cells contains the oligosaccharide donor, LLO, as well as the OST enzyme, Pcal_0997. The substrate peptide, AAYNVTKRK, where underlined sequence is the sequon, was synthesized to possess a fluorescent dye, TAMRA (carboxytetramethylrhodamine), at the N-terminus for detection, and biotin at the C-terminus for purification. Simple mixing and incubation of the membrane fractions with the synthesized acceptor peptide generated the glycopeptides products *in vitro*. The glycopeptides were isolated using biotin/streptavidin-based affinity chromatography, and further purified by NP-HPLC chromatography (Figure 4-3B). The three peaks were collected and subjected to in-fusion ESI-MS/MS analysis. The three peaks were indentified as S9, S10, and S11-linked peptides. The result of the MSMS analysis of the S10-pep was reported previously (Taguchi et al, JBC 2016). Based on the peak fluorescent intensities, we estimated the composition of the number of the monosaccharide residues in the enzymatically prepared glycopeptides (upper bar graph, Figure 4-3C). In the previous study, we determined the composition of the number of monosaccharide residues in the lipid-linked oligosaccharide molecule (lower bar graph, Figure 4-3C). The two profiles are very similar to each other. This comparison suggested that the two quantification methods, fluorescence intensity *versus* mass intensity, provided the reliable estimate of the monosaccharide compositions, and the *P. calidifontis* OST had no biased preference for the oligosaccharide length in the donor substrate.

Monosaccharide analysis

The S10- and S11-pep peaks in Figure 4-3B were collected and desalted by reverse-phase HPLC chromatography, and subjected to monosaccharide composition analysis, absolute configuration analysis, and NMR analysis. The monosaccharide composition analysis of the S10-pep revealed that the hexose residues in both S10 and S11 were all mannose residues. The two modified residues, $HexA(NAc)_2$ and $Hex(NAc)_2$, were not identified in the monosaccharide composition analysis, because of the lack of the standard monosaccharides. The absolute configuration of the mannose residue was determined to be the D configuration, based on the gas-liquid chromatography retention times of the trimethylsilylated (*S*)-2-butyl glycolside derivatives (data not shown).

NMR analysis

The S10-pep was dissolved in 260 □L of $^{2}H_2O$ in a micro NMR tube at the concentration of 72 □M. To facilitate the NMR analysis, the *P. calidifontis* cell culture medium was supplemented with ^{13}C glucose, and consequently the S10 portion of the glycopeptide was uniformly labeled with ^{13}C. The labeling efficiency was about 90 % judged by MS. In the one-dimensional (1D) ^{1}H spectrum with ^{13}C decoupling during acquisition, 9 anomeric peaks were readily identified in the downfield region of the $^{1}H^{2}HO$ peak (ca. 4.7 ppm), and labeled **a**-**i** (Figure 4-4A). The remaining anomeric signal was found in the upfield region of the $^{1}H^{2}HO$ peak in the slice of the two-dimensional (2D) TOCSY spectrum (Figure 4-4B), and labeled **j**.

The ^{1}H-^{1}H TOCSY magnetization transfer patterns were used to identify the type of monosaccharide residues (Figure 4-4B). The efficient magnetic transfer from the anomeric resonances **e** and **j** to the H5 resonance suggested the all *trans* conformation of vicinal hydrogen atoms in the *gluco*-configurated sugar. The magnetic transfer in the remaining 8 residues stopped at the H2 resonances suggested the *gauche* conformation of the vicinal H2-H3 hydrogen atoms in the *manno*-configurated sugar. To complete the resonance assignment beyond H2 resonances in these *manno*-configurated residues, we used relayed-type TOCSY spectrum (Figure 4-4C), which is a variant of ^{1}H-^{1}H TOCSY that has a relay unit after the TOCSY mixing (*10*). Together with the results of MS and monosaccharide composition analyses, we propose the assignment of the anomeric protons and monosaccharide types as follows: the residues **a**-**d** and **f**-**i** are mannose residue, and the residues **e** and **j** are $Glc(NAc)_2$ or $GlcA(NAc)_2$.

The ^{13}C assignment was performed by reference to ^{1}H-^{13}C HSQC spectrum (Figure 4-5A). The considerable upfield shift at 78.9 ppm of the anomeric ^{13}C resonance **e**, as compared to those of the other anomeric ^{13}C resonances at 98.5 - 102.7 ppm, indicated that the residue **e** was linked to the amide nitrogen of the Asn residue in the peptide, and thus $Glc(NAc)_2$. By elimination, the residue **j** is $GlcA(NAc)_2$. The ^{13}C upfield shifts of **e**2, **e**3, **j**2, and **j**3 suggested the substitutions with the N-acetylamino group at the C2 and C3 positions of $Glc(NAc)_2$ and $GlcA(NAc)_2$ in the residues **e** and **j**, respectively.

The linkages between the monosaccharide residues were identified by the inter-residual long-range $^{3}J_{CH}$ cross peaks in the ^{1}H-^{13}C HMBC spectra (Figure 4-6). The analysis of the HMBC spectrum of the S10-pep containing uniformly ^{13}C-labeled sugars suffered from low signal intensities and heavy overlapping of cross peaks due to $^{1}J_{CC}$ splittings along the ^{13}C axis (Figure

4-6A). Thus, we prepared another S10-pep containing 1-^{13}C labeled sugars, and used it for the HMBC measurement, in which the $^{1}J_{CC}$ splittings were suppressed (Figure 4-6B). The linkage between the monosaccharide residues (**a**1-**c**2, **b**1-**i**3, **c**1-**f**2, **d**1-**h**3, **f**1-**i**6, **g**1-**b**2, **h**1-**a**2, **i**1-**j**4, and **j**1-**e**4) were readily identified. The downfield shifts of the ^{13}C chemical shift of the linkage carbons (**a**2, **b**2, **c**2, **e**4, **f**2, **h**3, **i**3, **i**6, and **j**4) in the ^{1}H-^{13}C HSQC spectrum confirmed the linkage position in the monosaccharide residues (Figure 4-5).

In summary, ^{1}H and ^{13}C resonance assignment was performed by analyzing ^{1}H-^{1}H TOCSY (Figure 4B), ^{1}H-^{1}H relayed TOCSY (Figure 4-4C), ^{1}H-^{13}C HSQC (Figure 4-5), and ^{1}H-^{13}C HMBC (Figure 4-6) spectra. ^{1}H-^{1}H COSY, ^{1}H-^{1}H NOESY, ^{1}H-^{13}C HSQC-TOCSY, ^{1}H-^{13}C HSQC-NOESY, and 3D ^{1}H-^{13}C CCH-TOCSY spectra were also recorded, and used to help and confirm the resonance assignment (data not shown). The ^{1}H and ^{13}C chemical shifts of the *P. calidifontis* N-linked decasaccharide are summarized in Table 4-1.

The anomeric configuration of each monosaccharide residues was determined from the heteronuclear $^{1}J_{C1\text{-}H1}$ values obtained from the ^{1}H-^{13}C HSQC spectrum recorded without ^{13}C decoupling during acquisition (data not shown). The large $^{1}J_{C1\text{-}H1}$ values (> 170 Hz) indicated that all mannose residues were in the □ configuration, whereas the small $^{1}J_{C1\text{-}H1}$ values (< 160 Hz) indicated that residues **e** and **j** were in the β configuration (Table 4-1). By comparing the ^{13}C chemical shift of **d**2, the ^{13}C downfield shift from 70.5 ppm in the S10-pep to 79.1 ppm in the S11-pep indicated that the additional mannose was linked to the C2 position of residue **d** (Figure 4-5, *inset*). Taken all together, we propose that N-linked glycan of *P. calidifontis* has the structure shown schematically in Figure 4-8 and in the chemical formula in Figure 4-7.

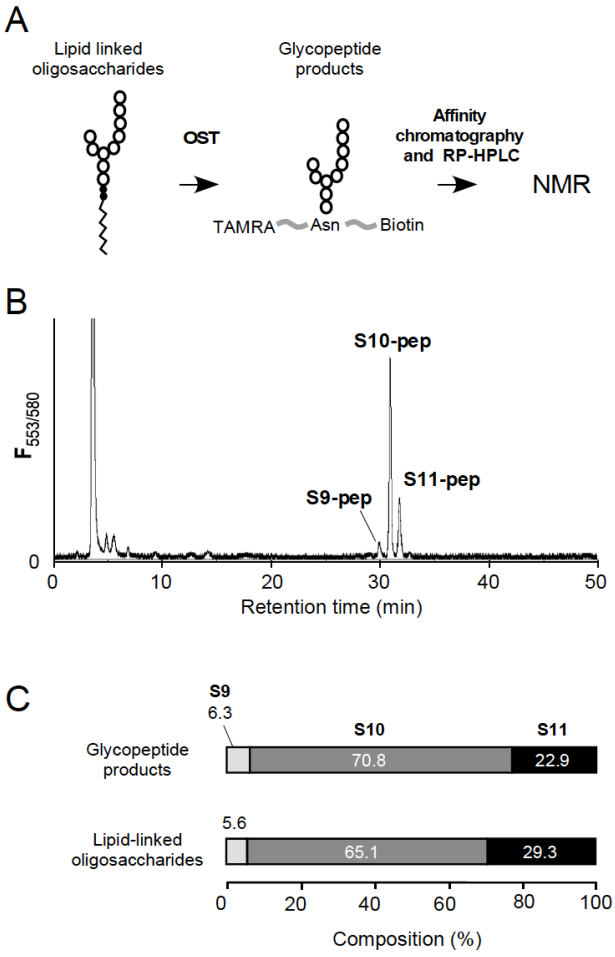


Figure 4-3. Preparation of large scale of oligosaccharide transferred to chemical synthesized peptide by OST reaction. (A) Schematic representations of enzymatically glycopeptide preparation. By the *in vitro* oligosaccharyltransferase reaction, the donor lipid-linked oligosaccharides from Triton X-100 solubilized membrane fraction were transferred to asparagine residue in the chemical synthesized acceptor peptide, which contains a fluorescent dye, TAMRA, at the N-terminus and biotin tag at C-terminus for sensitive detection and affinity purification, respectively. The glycopeptides were separated by biotin affinity chromatography and following normal-phase HPLC. (B) Elution profile of the normal-phase sugar-D chromatography, monitored by the fluorescence with Ex. 553 nm/Em. 580 nm. Peaks S10 and S11 were collected for NMR analyses. Peaks S9, S10, and S11 correspond to glycopeptides consisting of nine, ten, and eleven monosaccharide residues, respectively. (C) The oligosaccharide composition in the peptide transferred oligosaccharides derived from peak intensity of normal-phase HPLC, and the lipid-linked oligosaccharides (data from ref *1*) are shown in 100 % stacked charts.

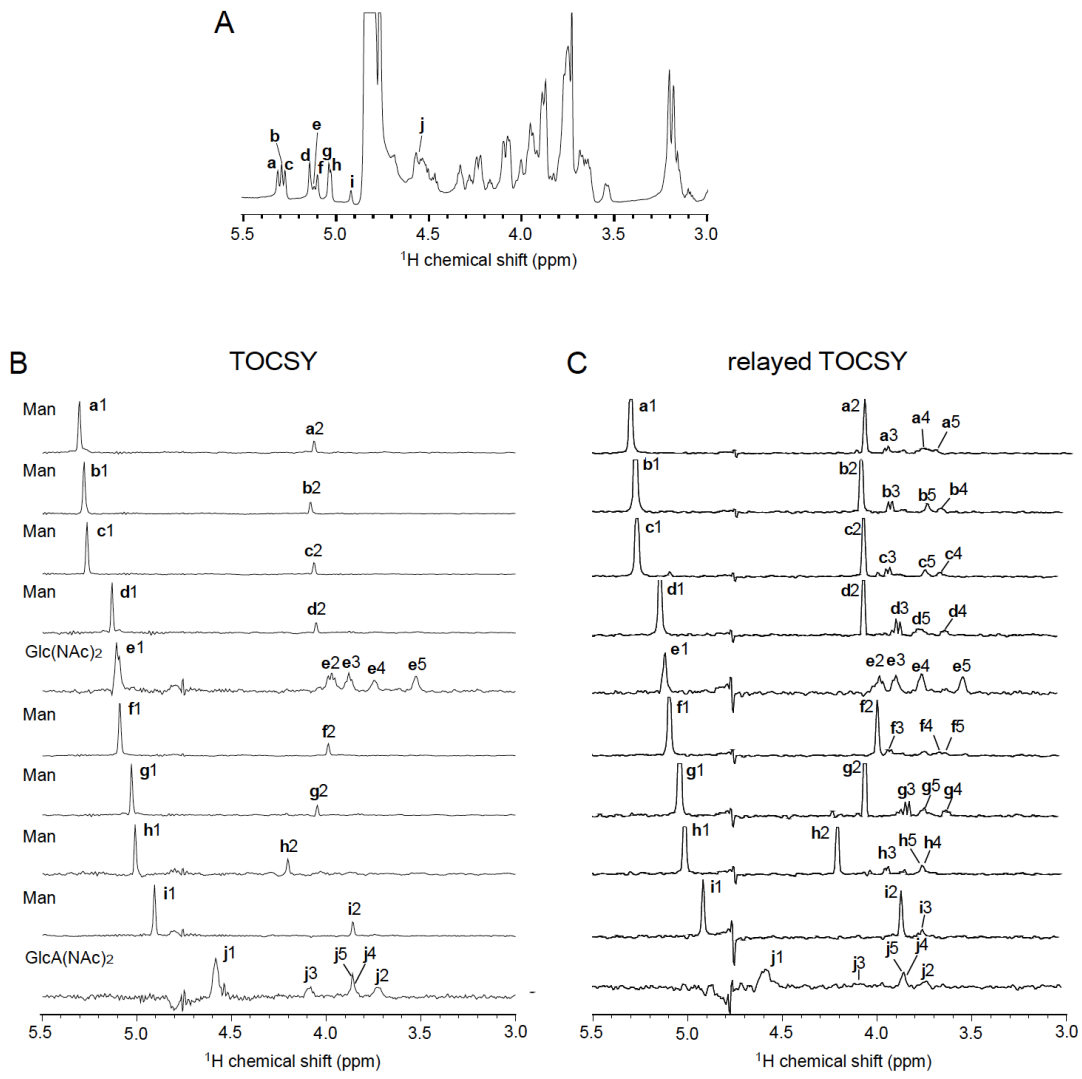


Figure 4-4. ^{1}H NMR spectrum of oligosaccharide S10 linked synthesized peptide. (A) 1D ^{1}H NMR spectrum with water presaturation and ^{13}C decoupling during acquisition. Ten anomeric proton resonance are labeled, **a-j**. (B) 1D slices of the 2D ^{1}H-^{1}H TOCSY spectra with ^{13}C decoupling during acquisition. TOCSY with a mixing time was set to 70 ms. Identified monosaccharide residues by TOCSY transfer patterns were labeled by *bold type*. (C) 1D slices of the 2D relayed TOCSY with a mixing time of 90 ms and a delay time of 100 ms.

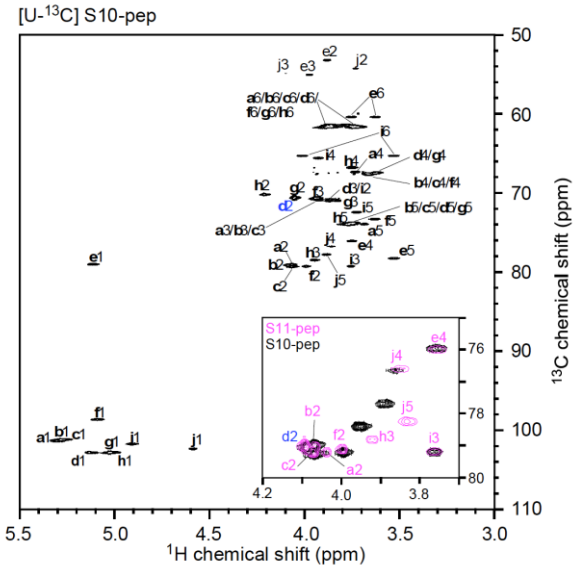


Figure 4-5. ^{1}H-^{13}C correlations of the glycopeptide. 2D ^{1}H-^{13}C HSQC spectrum of U-^{13}C labeled oligosaccharide S10 transferred peptide. A portion of HSQC spectrum of U-^{13}C labeled oligosaccharide S11 (magenta) and S10 (black) transferred peptide is shown as *inset*. The asterisk indicates the downfield shifted **d**2 cross peak, which indicated the linkage position of the additional mannose residue.

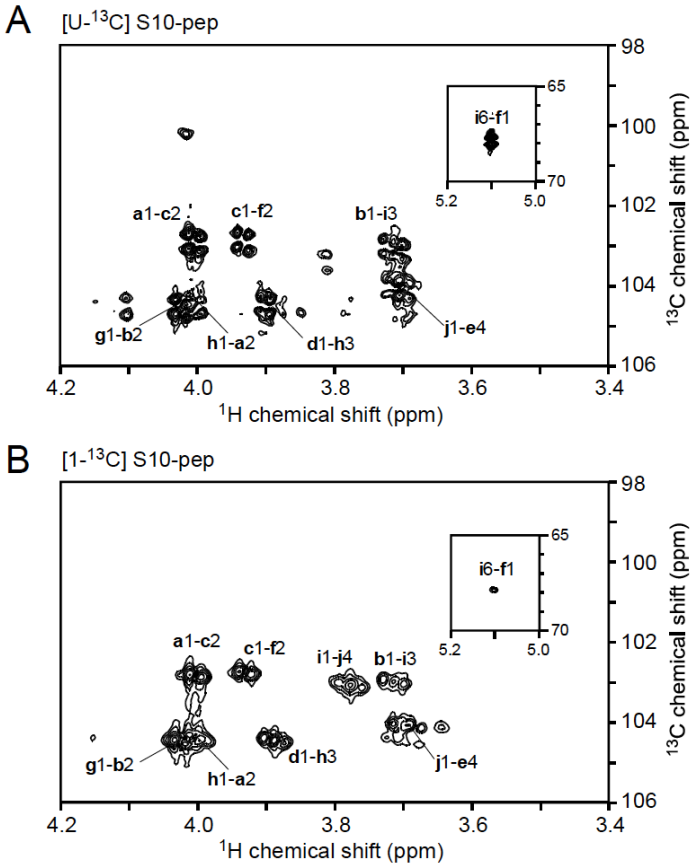
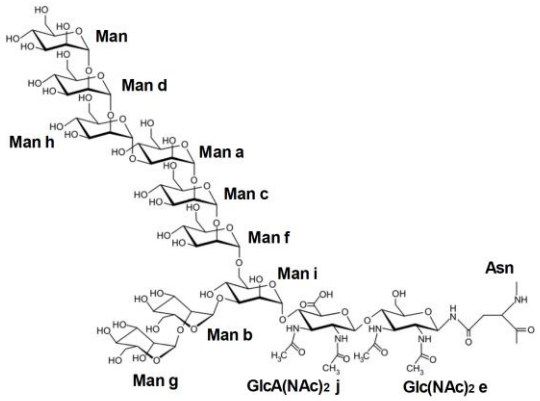


Figure4-6 (A) A portion of the 2D ^{1}H-^{13}C HMBC spectrum of uniform ^{13}C labeled S10 transferred peptide. Each cross peaks was split for ^{13}C dimension by $^{1}J_{CC}$ coupling and for ^{1}H dimension by $^{3}J_{CH}$ coupling. (B) A portion of the 2D ^{1}H-^{13}C HMBC spectrum of 1-^{13}C labeled S10 transferred peptide. Suppressing $^{1}J_{CC}$ splitting, specific ^{13}C labeling glycopeptides showed few peak overlapping and high signal intensity.



α-D-Man*p*-(1-2)-α-D-Man*p*-(1-2)-α-D-Man*p*-(1-3)-α-D-Man*p*-(1-2)-α-D-Man*p*-(1-2)-α-D-Man*p*
1
|
6
α-D-Man*p*-(1-2)-α-D-Man*p*-(1-3)-α-D-Man*p*-(1-4)-β-GlcANAc3NAc*p*-(1-4)-β-GlcNAc3NAc*p*-(1-*N*)-Asn

Figure 4-7. Structure of the N-linked undecasaccharide on the glycopeptide S11, transferred by the *P. calidifontis* oligosaccharyltransferase.

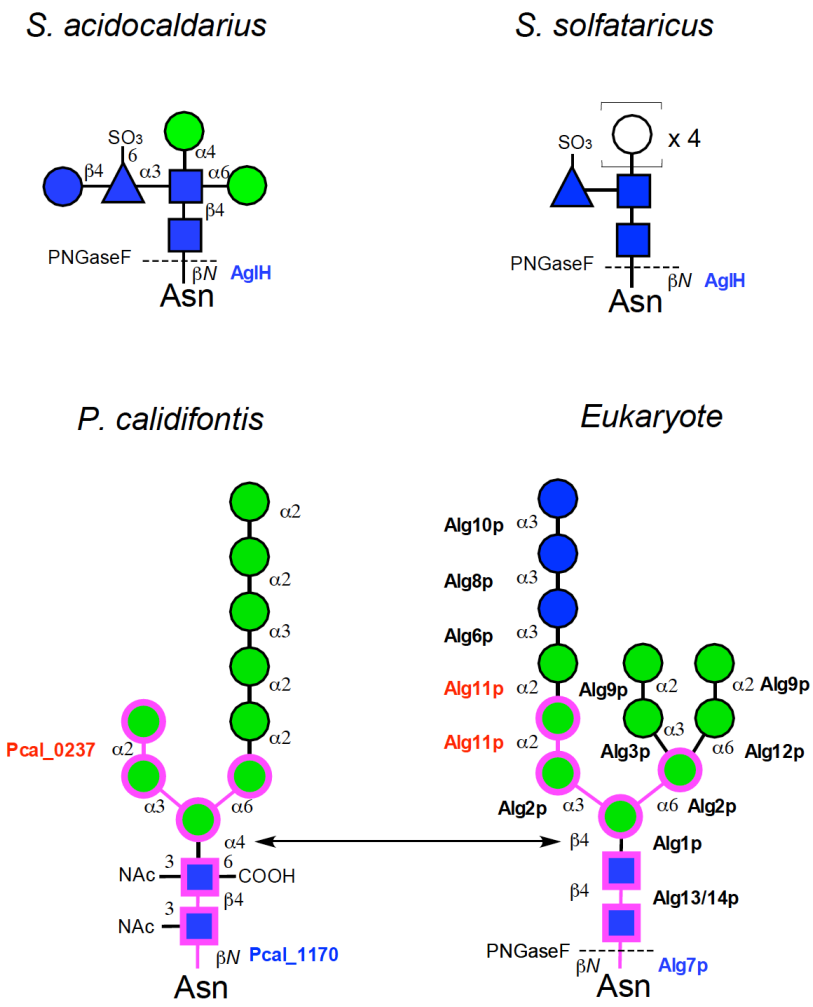


Figure 4-8. N-glycan structure of the Crenarchaeota, *P .calidifontis*, *S. acidocaldarius*, and *S. solfataricus* compared with eukaryotic one. The common substructures between Crenarchaea and Eukaryote were indicated by magenta line. The eukaryotic glycosyltransferase genes or predicted Crenarchaeal glycosyltransferase genes to be required for the assembly of each N-glycan are shown in *italic* at each structural linkage. The alg7 homologues and alg11 homologues were colored by blue and red, respectively. The number next to crenarchaeal gene name represents BLAST score between eukaryotic homologues.

Table 4-1. ^{1}H and ^{13}C NMR chemical shifts and $^{1}J_{C1-H1}$ values of the N-linked glycan from *Pyrobaculum calidifontis*.

	H-1	H-2	H-3	H-4	H-5	H-6, H-6’
Residue	C-1	C-2	C-3	C-4	C-5	C-6
	$^{1}J_{C1-H1}$					
a	5.31	4.07	3.96	3.74	3.68	3.89,3.78
-2)-α-Man	101.3	78.9	70.6	67.3	73.9	61.3
	172					
b	5.29	4.09	3.94	3.69	3.75	3.85,3.71
-2)-α-Man	101.1	79.0	70.8	67.5	73.8	61.4
	172					
c	5.27	4.07	3.93	3.67	3.75	3.85,3.71
-2)-α-Man	101.1	79.3	70.9	67.5	73.8	61.5
	172					
d	5.14	4.05	3.88	3.64	3.76	3.90,3.76
α-Man	102.7	70.5	70.8	67.4	73.8	61.6
	171					
e	5.13	3.89	3.97	3.75	3.54	3.76,3.63
-4)-β-GlcNAcNAc	78.9	53.2	55.0	76.0	78.2	60.3
	159					
f	5.10	4.00	3.94	3.67	3.64	3.88,3.74
-2)-α-Man	98.5	79.2	70.5	67.5	73.3	61.4
	172					
g	5.04	4.06	3.83	3.62	3.75	3.90,3.76
α-Man	102.7	70.5	70.8	67.3	73.8	61.6
	171					
h	5.02	4.22	3.95	3.75	3.77	3.86,3.75
-3)-α-Man	102.6	70.2	78.4	66.6	74.0	61.5
	171					
i	4.92	3.87	3.76	3.93	3.73	4.03,3.54
-3,6)-α-Man	101.6	70.8	79.2	65.5	72.4	65.3

	170				
j	4.59	3.74	4.10	3.86	3.88
-4)-β-GlcANAcNAc	102.3	54.2	54.8	76.6	77.6
	159				

Chemical shifts in ppm from an internal standard, DSS, and ^{1}J coupling constants in Hz measured in $^{2}H_2O$ at 25 °C. Average errors of ±0.02 ppm for ^{1}H, ±0.1 ppm for ^{13}C, and ±0.5 Hz for $^{1}J_{C1-H1}$.

4.3 Discussion

In the previous study, ten N-linked oligosaccharide structures from the Euryarhcaeota showed exhibit a high degree of the structural variation (*11-14*). On the other hand, in the crenarchaeota, only two N-linked oligosaccharide structures from *Sulfolobus acidocaldarius* and *Sulfolobus solfataricus* of the genus *Sulfolobus* have been reported. Similar to eukaryotes and unlike other euryarchaea, both of Sulfolobus N-linked oligosaccharides share chitobiose core (GlcNAc-GlcNAc-Asn) with eukaryotes and are released from the glycoprotein by peptide:N-glycosidase (PNGase) F (*9, 15*).

In this study, besides two Sulfolobus N-linked oligosaccharides, we elucidated novel N-linked oligosaccharide structure from hyper thermophilic archaeon *Pyrobaculum calidifontis* which also posses chitobiose core, as same as *Sulfolobus* and eukaryotic N-glycans, with three modifications; N-acetylamination at the C3 positions of the two GlcNAcs and oxidation at the C6 position of second GlcNAc to a carboxylic acid, resulting in protection from PNGaseF cleavage (data not shown). Furthermore, to our surprise, *P. calidifontis* N-glycan showed high-mannose glycan, which is typical of the eukaryotic N-glycan, with alpha -(1-3)- and alpha -(1-6)- mannose antennas branched from the third mannose (Figure 4-8). These structural similarities lead to hypothesis that N-glycosylation systems in eukaryotes and crenarchaea share common evolutionally origin. This hypothesis supported by comparative analysis of archaeal lipid-linked oligosaccharides, indicating that the crenarchaea use a eukaryotic type lipid carrier, dolicyl-diphosphate, for the oligosaccharide biosynthesis, in contrast to euryarchaeal dolicyl-monophosphate (*1*) and bacterial undecaprenyl-diphosphate (*16*). The N-linked oligosaccharide similarity between *P. calidifontis* and eukaryote is consistent with phylogenomic analysis of eukaryotic N-glycan biosynthesis glycolsyltransferases, implying that Alg7, STT3 and maybe also Alg2/11 likely have origins to TACK (Thaum-, Aig-, Cren-, and Korarchaeota) superphylum AglH, AglB and glycosyltransferase 1 (GT1) superfamily (*17*).

Pcal_1170 coded in *P. calidifontis* JCM 11548 genome belongs to the polyprenol phosphate:N-acetylhexosamine-1-phosphate transferase (HPT) family and possesses putative Mg2+ cofactor coordinating Asp-Asp motif, shown in other HPT family (*18*). Pcal_1170 shows 31 % sequence identity with *S. cerevisiae* Alg7 (GenePept accession number CAA68324) and 29 % sequence identity with *S. acidocaldarius* AglH (GenePept accession number AAY79520) (*19*) (Figure 4-9). Pcal_2037 belongs to the glycosyltransferase 1 (GT1) superfamily, possessing conserved EX_7E motif (30), and shows 29 % sequence identity with *S. cerevisiae* Alg11 (GenePept accession number CAA95916) and 19 % sequence identity with *S. cerevisiae* Alg2 (GenePept accession number CAA61199) (Figure 4-10). Thus, we assumed that Pcal_1170 and Pcal_2037 is involved in the first di-N acetyl glucosamice addition and blanching mannose extension (Figure

4-8).

We identified twelve N-glycosylation sequons from seven glycoproteins prepared by ConA affinity chromatography, as listed in appendix Table II. Fractional occurrence of amino acids in the region from −5 to +5 of the glycosylated Asn residue showed that proline is not allowed at position +1 and there is more containing Thr than Ser at position +2. Similar to other archaeal N-glycosylation sequons (*9, 20*), the presence of Asp or Glu residues at the −2 position was not required. Almost of previously characterized glycoproteins are cell surface appendages related proteins, such as S-layer, flagella and pili proteins (http://www.proglycprot.org, *21 and 9*). In this study, we added other member of glycoproteins, including transporter and membrane-bound enzyme (Appendix table I). Among these glycoproteins, our interest are oligosaccharyltransferase, AglB (Pcal_0097), which have five potential N-glycosylation sites, N527, N534, N538, N661, and N680 (Figure 4-11). Li, G *et al* showed that *S. cerevisiae* STT3 possessed potential N-glycosylation sequon N(535)NTWN(539)NT, lied downstream of conserved acceptor peptide binding WWD motif and N-glycosylation at N539 is essential for their catalytic activity (*22*). *P. calidifontis* AglB also have potential N-glycosylation sequon N(534)STIN(538)AT at a position similar to *S. cerevisiae* STT3, thus there is considerable interest in whether the N-glycosylation occurs at this sequon and has essential role for their activity or not.

The comparison of oligosaccharide profiles between N-linked oligosaccharide derived from ConA enriched glycoproteins and LLO (Figure 4-2E and 4-3C) suggested that terminal mannose was released from N-linked oligosaccharide attached to acceptor proteins. Although it was long thought that no post-transfer N-glycan processing occurs in Archaea, the two instance was discovered: a terminal mannose transfer to the four-residue N-glycan on a protein on the exterior side of the cell membrane in the halophilic archaeon, *Haloferax volcanii* (*23*), and a hexose transfer to the seven-residue N-glycan in *S. solfataricus* (*1*). Here, archaeal N-linked oligosaccharide processing is expanded to sugar trimming as well as sugar extending. But biological role of archaeal N-glycan processing is still unclear.

NMR analysis using ^{13}C spin system is effective for resonance assignment and glycoside linkage position determination of complex oligosaccharides, which shows severe signal overlapping of ^{1}H resoncance, such as high mannose type N-glycan (*24*). The ^{13}C labeling of carbohydrate have been reported in *S. cerevisiae* (*25*), *E.coli* (*26*), insect Sf9 cells (*27*), and mammalian cell lines (*28*), by using minimal medium containing ^{13}C labeled compounds as carbon sources. Kjellberg A *et al* also reported the ^{13}C labeling of *E.coli* O-antigenic polysaccharide by adding ^{13}C glucose [0.1% (w/v) final concentration] in the rich TY-medium, resulting in randomly ^{13}C enrichment of *ca.* 40% (*29*). We previously performed the ^{13}C labeling of N-linked oligosaccharide from hyperthermophilec archaeon, Archaeoglobus fulgidus by additing ^{13}C glucose [0.085% (w/v) final

concentration] in defined medium, showing nearly fully glucose/galaclose and fractionally 20% mannose 13C labeling efficiency (*13*). In this study, *P. calidifontis* cell culture in defined medium with only 0.02% (w/v) ^{13}C glucose produced ^{13}C labeling of N-glycan with 60% efficiency for all monosaccharide residues, which showed high quality ^{13}C-^{13}C correlation spectrum and ^{1}H-^{13}C HMBC spectrum in spite of a low sample concentration. These carbohydrate ^{13}C labeling approach would be use for any other organism species, whose minimal medium have not been defined yet.

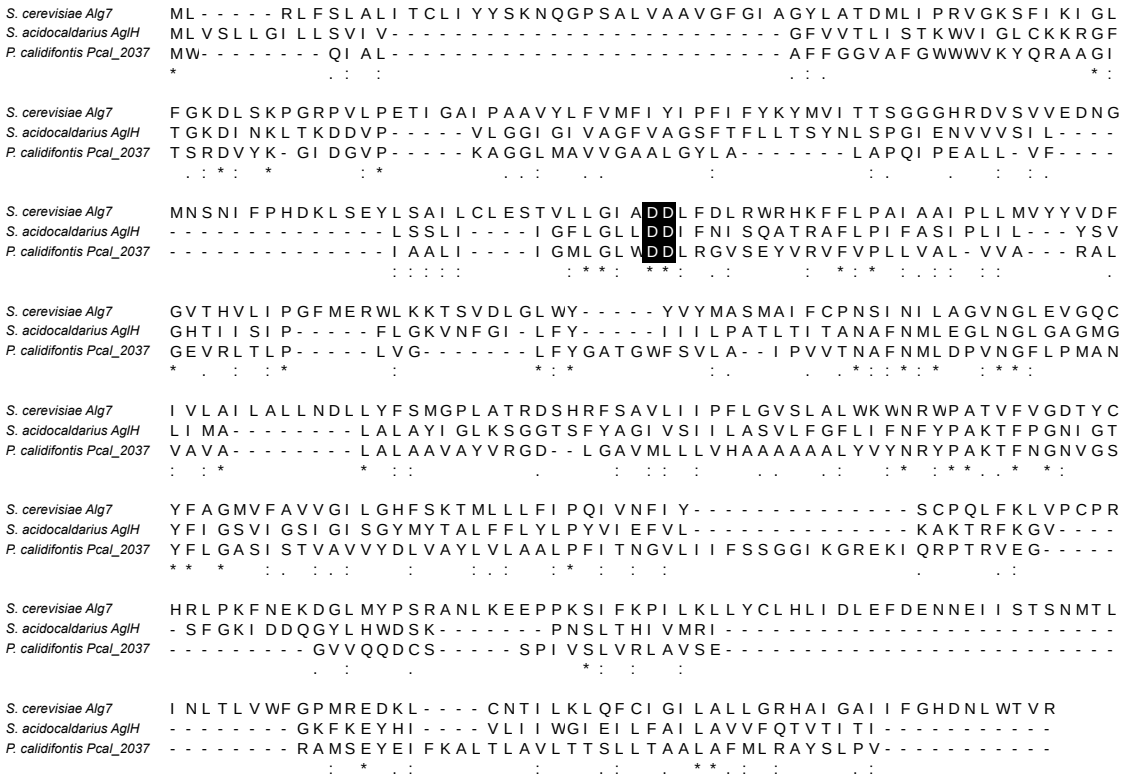


Figure 4-9. Sequence alignment of *S. cerevisiae* Alg7, *S. acidocaldarius* AglH, and *P. calidifontis* Pcal_1170. The amino acid sequences of *S. cerevisiae* Alg7 (GenePept accession number CAA68324), *S. acidocaldarius* AglH (GenePept accession number AAY79520), and *P. calidifontis* Pcal_1170 (GenePept accession number ABO08595) were aligned using the MAFFT program. *Asterisks* indicate identical amino acid residues among the three proteins, while *colons* and *periods* represent residues with conserved and semi-conserved substitutions, respectively. The dashes indicate gaps. Conserved Asp-Asp motif was shown by outlined character on a black background.

Figure 4-10. Sequence alignment of *S. cerevisiae* Alg11, and *P. calidifontis* Pcal_2037. The amino acid sequences of *S. cerevisiae* Alg11 (GenePept accession number CAA95916) and *P. calidifontis* Pcal_1170 (GenePept accession number ABO09452) were aligned using the MAFFT program. *Asterisks* indicate identical amino acid residues among the two proteins, while *colons* and *periods* represent residues with conserved and semi-conserved substitutions, respectively. The dashes indicate gaps. Conserved EX_7E motif was shown by outlined character on a black background.

Figure 4-11. Sequence alignment of *S. cerevisiae* STT3, and *P. calidifontis* AglB. The amino acid sequences of *S. cerevisiae* STT3 (GenePept accession number BAA06079) and *P. calidifontis* AglB (GenePept accession number ABO08422) were aligned using the MAFFT program. *Asterisks* indicate identical amino acid residues among the two proteins, while *colons* and *periods* represent residues with conserved and semi-conserved substitutions, respectively. The dashes indicate gaps. Conserved WWDYG motif was shown by outlined character on a black background. The position of the conserved N-glycolsylation sequon is shown by red box. The predicted N-glycosylation asparagine residues are shown by red character.

4.4

Reference

(1) Y. Taguchi, D. Fujinami, D. Kohda, J. Biol. Chem., 291 (2016), pp. 11042-11054

(2) M. Matsumoto, S. Hatakeyama, K. Oyamada, Y. Oda, T. Nishimura, K. I. Nakayama, Proteomics, 5 (2005), 5, pp. 4145-4151

(3) T. Kurihara, Biomed. Chromatogr., 23 (2009), pp. 516-523

(4) G. J. Gerwig, J. P. Kamerling, J. F. Vliegenthart, Carbohydr. Res., 77 (1979), pp. 10-17

(5) E. Kupce, J. Boyd, I. D. Campbell, J. Magn. Reson. B, 106 (1995), pp. 300-303

(6) D. O. Cicero, G. Barbato, R. Bazzo, J. Magn. Reson., 148 (2001), pp. 209-213

(7) K. Gehring, I. Ekiel, J. Magn. Reson., 135 (1998), pp. 185-193

(8) F. Delaglio, S. Grzesiek, G. W. Vuister, G. Zhu, J. Pfeifer, A. Bax, J. Biomol. NMR, 6 (1995), pp. 277-293

(9) G. Palmieri, M. Balestrieri, J. Peter-Katalinic, G. Pohlentz, M. Rossi, I. Fiume, G. Pocsfalvi, J. Proteome Res., 12 (2013), pp. 2779-2790

(10) F. Inagaki, I. Shimada, D. Kohda, A. Suzuki, A. Bax, J. Magn. Reson., 81 (1989), pp. 186-190

(11) K. F. Jarrell, Y. Ding, B. H. Meyer, S. V. Albers, L. Kaminski, J. Eichler, Microbiol. Mol. Biol. Rev., 78 (2014), pp. 304-341

(12) D. Fujinami, M. Matsumoto, T. Noguchi, K. Sonomoto, D. Kohda, Carbohydr. Res., 387 (2014), pp. 30-36

(13) D. Fujinami, J. Nyirenda, S. Matsumoto, D. Kohda, Carbohydr. Res., 413 (2015), pp. 55-62

(14) L. Kandiba, C. W. Lin, M. Aebi, J. Eichler, Y. Guerardel, Glycobiology, 26 (2016), pp.745-756

(15) U. Zahringer, H. Moll, T. Hettmann, Y. A. Knirel, G. Schafer, Eur. J. Biochem., 267 (2000), pp. 4144-4149

(16) A. Larkin, B. Imperiali, Biochemistry, 50 (2011), pp. 4411-4426

(17) J. Lombard, Biol. Direct, 11 (2016), 36-016-0137-2

(18) H. Shams-Eldin, B. Chaban, S. Niehus, R. T. Schwarz, K. F. Jarrell, J. Bacteriol., 190 (2008), pp. 2217-2220

(19) B. H. Meyer, H. Shams-Eldin, S. V. Albers, Extremophiles, 2016, *in press*

(20) E. Vinogradov, L. Deschatelets, M. Lamoureux, G. B. Patel, T. L. Tremblay, A. Robotham, M. F. Goneau, C. Cummings-Lorbetskie, D. C. Watson, J. R. Brisson, J. F. Kelly, M. Gilbert,

Glycobiology, 22 (2012), pp. 1256-1267

(21) A. H. Bhat, H. Mondal, J. S. Chauhan, G. P. Raghava, A. Methi, A. Rao, Nucleic Acids Res., 40 (2012), pp. D388-393

(22) G. Li, Q. Yan, A. Nita-Lazar, R. S. Haltiwanger, W. J. Lennarz, J. Biol. Chem., 280 (2005), pp. 1864-1871

(23) C. Cohen-Rosenzweig, S. Yurist-Doutsch, J. Eichler, J. Bacteriol., 194 (2012), pp. 6909-6916

(24) C. Fontana, H. Kovacs, G. Widmalm, J. Biomol. NMR, 59 (2014), pp. 95-110

(25) Y. Kamiya, S. Yamamoto, Y. Chiba, Y. Jigami, K. Kato, J. Biomol. NMR, 50 (2011), pp. 397-401

(26) H. F. Azurmendi, J. Vionnet, L. Wrightson, L. B. Trinh, J. Shiloach, D. I. Freedberg, Proc. Natl. Acad. Sci. U. S. A., 104 (2007), pp. 11557-11561

(27) W. J. Walton, A. J. Kasprzak, J. T. Hare, T. M. Logan, J. Biomol. NMR, 36 (2006), pp. 225-233

(28) J. W. Lustbader, S. Birken, S. Pollak, A. Pound, B. T. Chait, U. A. Mirza, S. Ramnarain, R. E. Canfield, J. M. Brown, J. Biomol. NMR, 7 (1996), pp. 295-304

(29) K. Gehring, I. Ekiel, J. Magn. Reson., 135 (1998), pp. 185-193

(30) M. K. O'Reilly, G. Zhang, B. Imperiali, Biochemistry, 45 (2006), pp. 9593-9603

(31) D. Kohda, M. Yamada, M. Igura, J. Kamishikiryo, K. Maenaka, Glycobiology, 17 (2007), pp. 1175–1182

Chapter 5-Overall discussion

5.1

Conclusion and discussion for future works

In this study, we elucidated three N-linked oligosaccharide structure from hyperthermophilic archaeal spices, *P. furiosus*, *A. fulgidus* and *P. calidifontis*. Our research goal is to reveal enzymatic mechanism of oligosaccharyltransferase. For investigating LLO recognition mechanism, the AglB crystal structure in complex with LLO must be required. Concerning LLO preparation for AglB-LLO co-crystallizing, LLO chemical synthesis is superior to *in vitro* enzymatic preparation using sets of recombinant glycosyltransferases (*1*) and preparation from natural source (*2*), in terms of high yield and high purity. Structural information of *P. furiosus*, and *A. fulgidus* N-glycan makes chemical synthesis of LLO water-soluble analogues possible. Now, we are making *A. fulgidus* LLO water-soluble analogues (Gal-[β1,4]-Glc-[α 1,P]-monophosphate-citronellol), in collaboration with Dr. Akihiro Ishiwata (RIKEN, wako).

P. calidifontis N-glycan showed eukaryotic chitobiose core and di-blanching mannose antenna. Next our interest is "how *P. calidifontis* N-glycan are synthesized *in vivo* and which glycosyltransferases are involved in N-glycan biosynthesis". Pcal_1170 shows 31 % sequence identity with *S. cerevisiae* Alg7 and Pcal_2037 shows 29 % sequence identity with *S. cerevisiae* Alg11. Thus, we assumed that Pcal_1170 and Pcal_2037 is involved in the first di-N acetyl glucosamine addition and blanching mannose extension (Figure 4-8 and 5-1). Archaeal N-glycan biosynthesis pathway has been studied in three euryarchaeal spices, *Haroferax volcanii*, *Methanococcus voltae* and *Methanococcus maripaludis* (*3*) (Figure 1-3). The identification of euryarchaeal glycosyltransferase genes have been performed based on the gene deletion, giving arises to the defect of N-glycan structure or the lack of monosaccharide residue. In contrast to euryarchaea, crenarchaeal N-glycan biosynthesis related gene deletion resulted in lethal phenotype (*5*, *6*) (except for S. acidocaldarius agl16, which transfers terminal glucose residue to blanched quinovose (*7*), Figure 5-1). It makes glycosyltransfere gene identification difficult, hence only one of the glycosyltransferases, *S. acidocaldarius* AglH, have been characterized by cross-domain complementation experiment, revealing that AglH is able to restore N-glycosylation in a conditional lethal *S. cerevisiae alg7* mutant (*5*, Figrue 5-1). To characterize Pcal_1170 and Pcal_2037, such cross-domain complementation experiment using a conditional lethal *S. cerevisiae alg7* or *alg11* mutant will be required. On the other hand, several glycosyltransferases

involved in *M. voltae* N-glycan biosynthesis have been characterized by *in vitro* analysis using a suite of synthetic and semisynthetic LLO substrate and recombinant glycosyltransferases (*1*, Figure 1-3). The N-linked oligosaccharide structure information that we obtained allows to synthesize donor LLO substrate for each glycosyltransferase and to establish *in vitro* glycosyltransferase reaction assay.

5.2 Industrial application

One of the fastest growing fields in the pharmaceutical industry is the market for therapeutic glycoproteins i.e., clotting factors, hormones, cytokines, antisera, enzymes, enzyme inhibitors, Ig-Fc-Fusion proteins, and monoclonal antibodies (*8*). There are two strategies for glycoprotein production: glycoprotein synthesis/semisynthesis (*9*) or recombinant/natural glycoprotein preparation from *S. cerevisiae*, insect Sf9 cells, mammalian cell lines, and other hosts. The production of recombinant protein in *E. coli* has many advantages, such as high-level production, fast growth, and inexpensive (*10*), whereas no post-translation modifications. However, several engineering of *E. coli* expression system were reported, aiming toward glycoprotein production in *E. coli*: bacterial N-glycosylated protein production in *E. coli* host by *pgl* gene cluster transfer (*4*), eukaryotic N-glycosylated protein production by similar approach to reference (*4*) with additional *in vitro* enzymatic reaction (*11*) and by the same manner as a reference (*11*) with substitution of eubacterial glycosyltransferase genes by eukaryotic counterparts (*12*) (Figure 5-1). However, there are problems that only a small fraction (<1%) of target protein was glycolsylated. Furthermore, low yield (50 μg/L) due to low expression level and low activity of each glycolsyltransferase are points of improvement. The finding new *P. calidifontis* glycolsyltransferase catalyzing eukaryotic N-glycan assembly will contribute toward developing pre-existing *E. coli* expression systems (Figure 5-1, Future application).

On the other hand, the application of glycosyltransferases for the synthesis of the complex and highly pure oligosaccharides is currently recognized as being the most effective (*13*). In oligosaccharide chemical synthesis, the controlling anomeric stereochemistry and proper position of glycosidic linkage require complicated processes, such as the extensive use of protecting groups and the preparation of specialized precursor compounds. In contrast, enzymatic synthesis of oligosaccharide do not require protecting groups or elaborate precursors, and they construct glycosidic linkages with perfect regio- and stereochemical control. The finding new

glycolsyltransferase will contribute toward the expansion of glycosyltransferase applications as a tool of enzymatic synthesis of oligosaccharide (*14*).

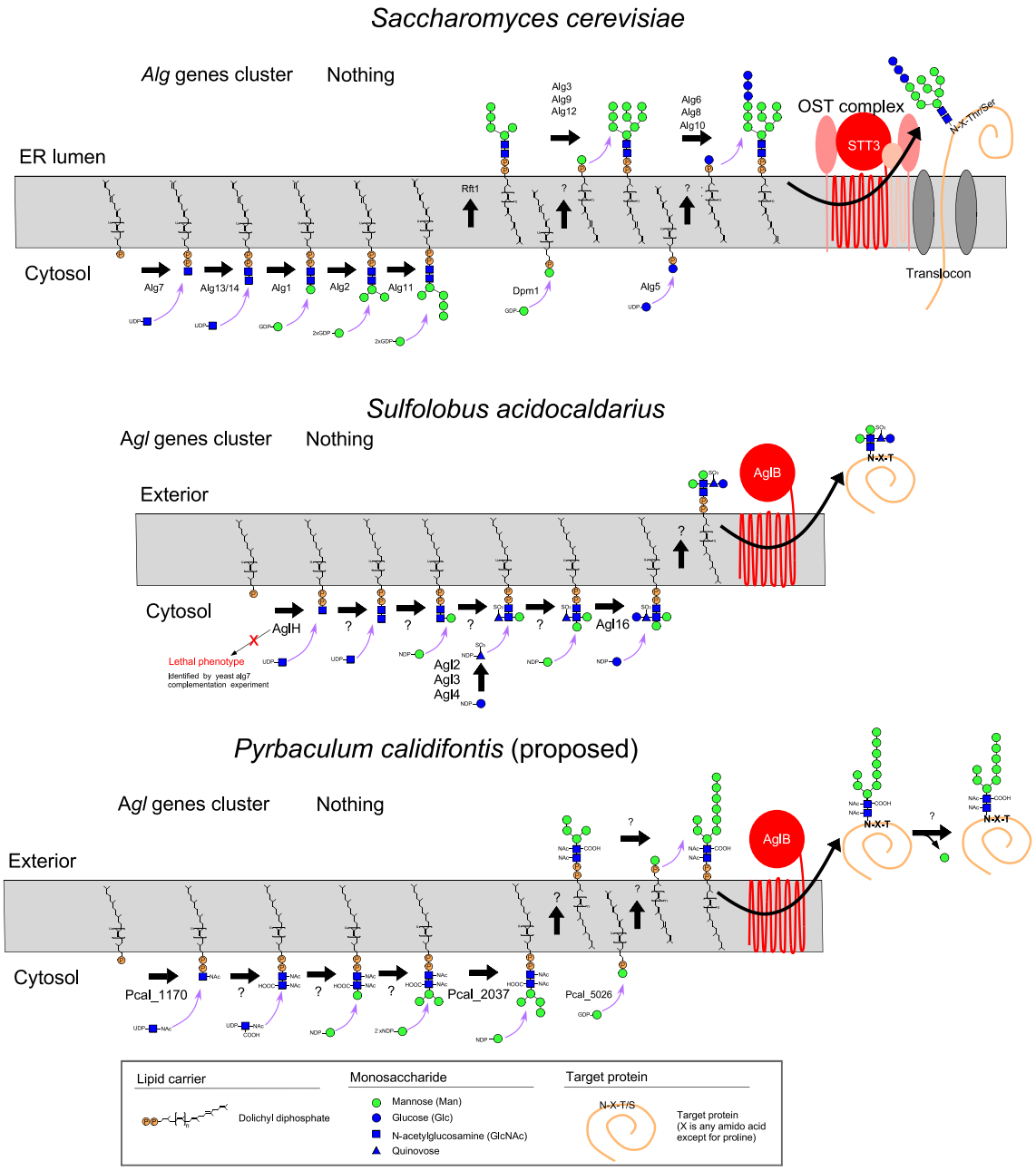


Figure 5-1. N-glycan biosynthesis pathway in eukaryote and crenarchaea.

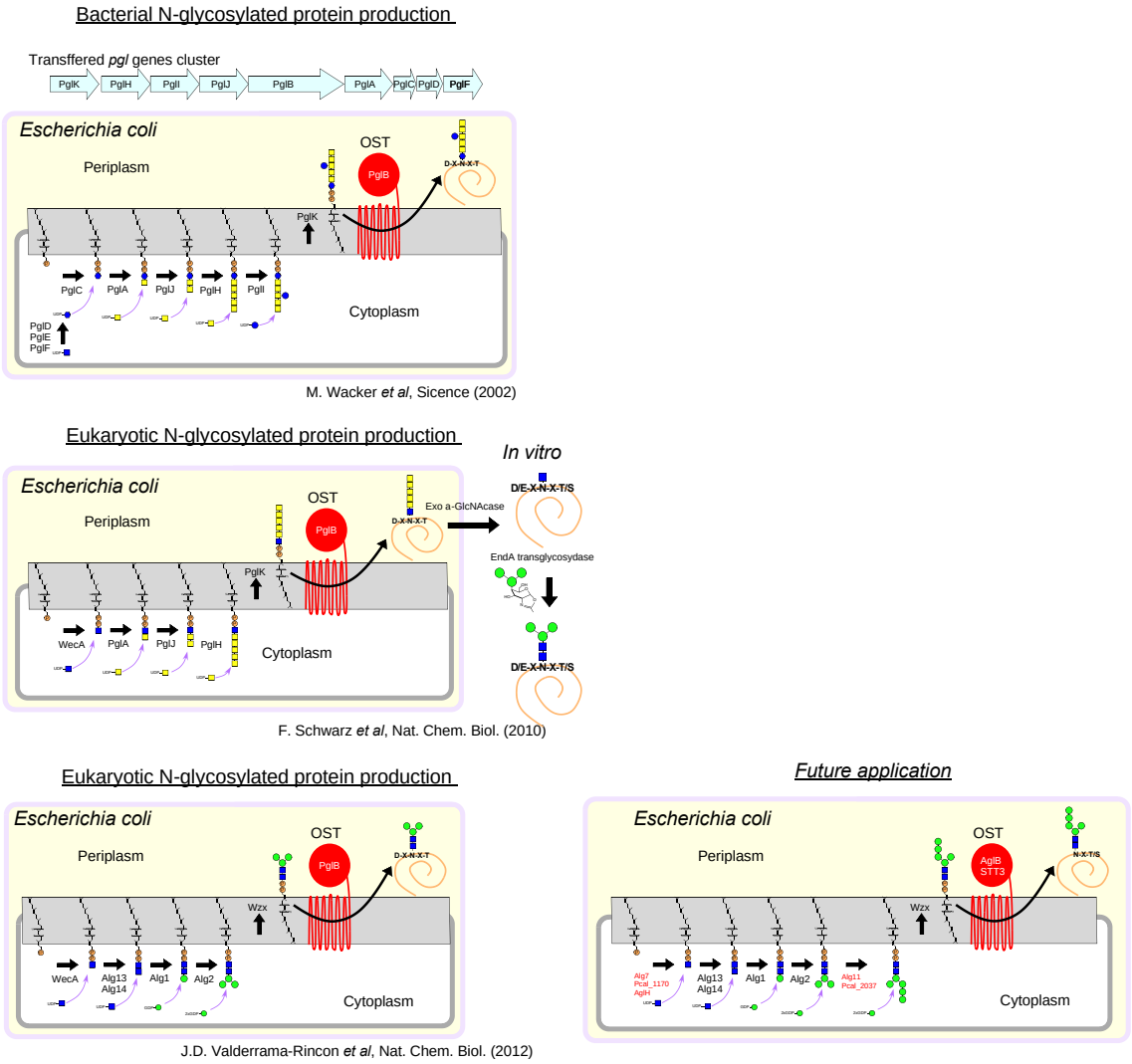


Figure 5-2. Schematic representation of the *E. coli* glycoprotein expression systems. The *pgl* gene cassette (*4*) or modified (*11*, *12*) gene cassette were transferred into *E. coli*. The acceptor protein is secreted to the periplasm space via translocation channel. A tailored LLO is assembled on lipid carrier by each glycosyltransferase, and then flipped to periplasmic side of the inner membrane. Finally, OST catalyze the transfer of oligosaccharide to the acceptor proteins.

5.3

Reference

(1) A. Larkin, M. M. Chang, G. E. Whitworth, B. Imperiali, Nat. Chem. Biol., 9 (2013), pp. 367-373

(2) Y. Taguchi, D. Fujinami, D. Kohda, J. Biol. Chem., 291 (2016), pp. 11042-11054

(3) K. F. Jarrell, Y. Ding, B. H. Meyer, S. V. Albers, L. Kaminski, J. Eichler, Microbiol. Mol. Biol. Rev., 78 (2014), pp. 304-341

(4) M. Wacker, D. Linton, P. G. Hitchen, M. Nita-Lazar, S. M. Haslam, S. J. North, M. Panico, H. R. Morris, A. Dell, B. W. Wren, M. Aebi, Science, 298 (2002), pp. 1790-1793

(5) B. H. Meyer, H. Shams-Eldin, S. V. Albers, Extremophiles, 2016, *in press*

(6) B. H. Meyer, S. V. Albers, Microbiologyopen, 3 (2014), pp. 531-543

(7) . H. Meyer, E. Peyfoon, C. Dietrich, P. Hitchen, M. Panico, H. R. Morris, A. Dell, S. V. Albers, J. Bacteriol., 195 (2013), pp. 2177-2186

(8) D. Ghaderi, M. Zhang, N. Hurtado-Ziola, A. Varki, Biotechnol. Genet. Eng. Rev., 28 (2012), pp. 147-175

(9) D. P. Gamblin, E. M. Scanlan, B. G. Davis, Chem. Rev., 109 (2009), pp. 131-163

(10) G. L. Rosano, E. A. Ceccarelli, Front. Microbiol., 5 (2014), pp. 172

(11) F. Schwarz, W. Huang, C. Li, B. L. Schulz, C. Lizak, A. Palumbo, S. Numao, D. Neri, M. Aebi, L. X. Wang, Nat. Chem. Biol., 6 (2010), pp. 264-266

(12) J. D. Valderrama-Rincon, A. C. Fisher, J. H. Merritt, Y. Y. Fan, C. A. Reading, K. Chhiba, C. Heiss, P. Azadi, M. Aebi, M. P. DeLisa, Nat. Chem. Biol., 8 (2012), pp. 434-436

(13) C. A. Weijers, M. C. Franssen, G. M. Visser, Biotechnol. Adv., 26 (2008), pp. 436-456

(14) H. Narimatsu, Glycoconj. J., 21 (2004), pp. 17-24

Appendix Table I. Glycoproteins of *Pyrobaculum calidifontis* Enriched by ConA Affinity Chromatography

protein name in NCBInr	Description (KEGG)	ORF	Accession	Num. of N glycosylation site	MASCOT Score	Mass (Da)	Peptide matches	Sequence coverage (%)
hypothetical protein	conserved hypothetical protein	Pcal_0034	gi|500174306	40	5782	290584	165	32
nitrate reductase subunit alpha	respiratory nitrate reductase alpha subunit apoprotein	Pcal_1907	gi|500176156	3	5382	144408	172	74
peptidase S8	peptidase S8 and S53	Pcal_2124	gi|500176372	5	240	135037	9	7
formate dehydrogenase	formate dehydrogenase alpha subunit	Pcal_0396	gi|500174663	2	196	130876	3	4
hypothetical protein	hypothetical protein	Pcal_0081	gi|500174353	9	738	127129	21	19
hypothetical protein	conserved hypothetical protein	Pcal_0032	gi|500174304	27	1360	126353	42	28
hypothetical protein	hypothetical protein	Pcal_1340	gi|500175595	14	338	122997	11	10
pullulanase	pullulanase / alpha-amylase	Pcal_1616	gi|500175866	9	2837	111898	72	46
hypothetical protein	hypothetical protein	Pcal_0429	gi|500174695	11	1536	106066	53	38
hypothetical protein	hypothetical protein	Pcal_0194	gi|500174463	7	2417	96841	59	52
peptide-binding protein	oligopeptide binding protein, putative	Pcal_0975	gi|500175233	3	8208	95805	211	54
multidrug RND transporter	membrane protein, putative	Pcal_1643	gi|500175894	13	701	94920	23	20
molybdopterin oxidoreductase	molybdopterin oxidoreductase	Pcal_1500	gi|500175752	4	664	94722	24	27
cytochrome c biogenesis protein	cytochrome c assembly protein	Pcal_1488	gi|500175740	4	892	92081	37	14
hypothetical protein	conserved hypothetical protein	Pcal_0090	gi|500174362	9	251	91884	9	8

peptide transporter	Oligosaccharyl transferase, STT3 subunit	Pcal_0997	gi|500175255	5	860	85494	25	20
hypothetical protein	conserved hypothetical protein	Pcal_1005	gi|500175263	11	954	83758	22	34
hypothetical protein	hypothetical protein	Pcal_0485	gi|500174749	10	282	83037	7	7
hypothetical protein	conserved hypothetical protein	Pcal_1154	gi|500175412	26	1121	81710	29	35
peptide transporter	extracellular solute-binding protein, family 5	Pcal_1231	gi|500175489	5	2563	80348	61	59
hypothetical protein	hypothetical protein	Pcal_2023	gi|500176271	4	681	79012	19	28
penicillin amidase	Penicillin amidase	Pcal_1846	gi|500176096	6	1114	78745	34	45
nitric-oxide reductase large subunit	nitric oxide reductase, NorZ apoprotein	Pcal_1908	gi|500176157	1	402	78701	13	14
cytochrome c biogenesis protein	cytochrome c assembly protein	Pcal_1915	gi|500176164	5	260	74721	6	9
hypothetical protein	conserved hypothetical protein	Pcal_0654	gi|500174913	10	656	73314	10	16
histidine kinase	putative signal-transduction protein with CBS domains	Pcal_1808	gi|500176059	1	455	71980	14	22
cytochrome c assembly protein	cytochrome c assembly protein	Pcal_1953	gi|500176202	6	143	71290	4	5
hypothetical protein	conserved hypothetical protein	Pcal_0033	gi|500174305	16	3765	69454	102	48
metallophosphoesterase	metallophosphoesterase	Pcal_0250	gi|500174519	6	1711	67734	50	51
arabinase	peptidase S49	Pcal_0153	gi|500174422	13	2450	67070	86	37
ATP synthase subunit A	Sodium-transporting two-sector ATPase	Pcal_2089	gi|500176337	2	110	66242	5	10
hypothetical protein	conserved hypothetical protein	Pcal_0670	gi|500174929	4	246	66188	9	17

hypothetical protein	Tetratricopeptide TPR_2 repeat protein	Pcal_1491	gi|500175743	11	562	66027	17	28
hypothetical protein	conserved hypothetical protein	Pcal_1330	gi|500175585	4	175	65916	3	7
chloride channel protein	Cl-channel, voltage-gated family protein	Pcal_1543	gi|500175794	1	157	61618	2	4
hypothetical protein	hypothetical protein	Pcal_1493	gi|500175745	14	638	60934	17	28
thermosome subunit	thermosome subunit	Pcal_1052	gi|500175310	1	185	60131	7	19
hypothetical protein	hypothetical protein	Pcal_0409	gi|500174675	3	755	59968	18	25
peptide ABC transporter substrate-binding protein	extracellular solute-binding protein, family 5	Pcal_0422	gi|500174688	2	3942	58860	109	78
peptidase	peptidase S49	Pcal_1261	gi|500175518	9	340	55771	13	17
hypothetical protein	conserved hypothetical protein	Pcal_2152	gi|500176400	1	189	55655	6	9
ABC transporter substrate-binding protein	extracellular solute-binding protein, family 1	Pcal_0248	gi|500174517	1	1930	55352	49	63
glycosyl transferase	glycosyl transferase, family 2	Pcal_0500	gi|500174761	5	513	55010	15	29
nitrate reductase subunit beta	respiratory nitrate reductase beta subunit	Pcal_1906	gi|500176155	2	2485	54773	65	75
hypothetical protein	hypothetical protein	Pcal_0193	gi|500174462	4	354	54740	12	21
peptidase M50	peptidase M50	Pcal_0618	gi|500174878	4	632	54108	27	25
hypothetical protein	hypothetical protein	Pcal_0138	gi|500174408	5	288	53808	11	29
hypothetical protein	hypothetical protein	Pcal_1093	gi|500175351	10	175	53654	6	17
transglutaminase	transglutaminase domain protein	Pcal_0413	gi|500174679	2	143	52890	3	7
FAD-dependent oxidoreductase	FAD dependent oxidoreductase	Pcal_0127	gi|500174397	1	933	52655	33	43

peptide ABC transporter permease	binding-protein-dependent transport systems inner membrane component	Pcal_0973	gi|500175231	3	1170	52569	43	42
ABC transporter substrate-binding protein	carbohydrate ABC transporter substrate-binding protein, CUT1 family	Pcal_2112	gi|500176360	1	2330	51528	55	58
iron ABC transporter substrate-binding protein	conserved hypothetical protein	Pcal_0650	gi|500174909	5	1187	51298	34	43
hypothetical protein	hypothetical protein	Pcal_1299	gi|500175556	2	140	51216	5	10
branched-chain amino acid ABC transporter substrate-binding protein	amino acid/amide ABC transporter substrate-binding protein, HAAT family	Pcal_0848	gi|500175107	1	2001	51062	51	71
geranylgeranyl reductase	2,3-di-O-geranylgeranylglyceryl phosphate reductase	Pcal_1073	gi|500175331	1	540	50514	27	44
sugar ABC transporter substrate-binding protein	maltooligosaccharide-binding protein	Pcal_1617	gi|500175867	1	1779	50328	51	48
hypothetical protein	conserved hypothetical protein	Pcal_1477	gi|500175729	1	455	49261	14	31
hypothetical protein	hypothetical protein	Pcal_0079	gi|500174351	5	559	48099	11	38
phosphate starvation protein PhoH	phosphate ABC transporter substrate-binding protein, PhoT family	Pcal_1882	gi|500176131	2	204	47562	5	13
ABC transporter substrate-binding protein	periplasmic binding protein	Pcal_1519	gi|500175771	6	1309	47477	34	49
ABC transporter	hypothetical protein	Pcal_0249	gi|500174518	3	3297	47465	70	34

nitrous oxidase accessory protein	Nitrous oxidase accessory protein-like protein	Pcal_1931	gi|500176180	8	336	46797	10	26
glutamate dehydrogenase	glutamate dehydrogenase (NADP)	Pcal_1606	gi|500175856	2	388	46581	12	30
hypothetical protein	conserved hypothetical protein	Pcal_0801	gi|500175060	1	122	46259	8	13
hypothetical protein	conserved hypothetical protein	Pcal_0611	gi|500174871	14	827	46194	19	36
NADH dehydrogenase subunit D	NADH dehydrogenase subunit D	Pcal_0514	gi|500174775	2	665	46094	28	46
nodulation efficiency protein NfeD	Nodulation efficiency protein NfeD	Pcal_2167	gi|500176416	3	618	46004	13	20
hypothetical protein	conserved hypothetical protein	Pcal_0909	gi|500175167	8	269	45674	9	27
hypothetical protein	conserved hypothetical protein	Pcal_1374	gi|500175628	2	623	45643	16	34
hypothetical protein	conserved hypothetical protein	Pcal_0363	gi|500174630	8	1252	44535	39	47
sugar-binding protein	nucleoside-binding protein	Pcal_1844	gi|500176094	1	3434	44302	68	74
NADH-ubiquinone oxidoreductase	NADH dehydrogenase subunit N	Pcal_0521	gi|500174782	1	177	43675	2	5
hypothetical protein	conserved hypothetical protein	Pcal_1471	gi|500175723	4	299	43241	7	20
branched-chain amino acid ABC transporter substrate-binding protein	amino acid/amide ABC transporter substrate-binding protein, HAAT family	Pcal_0825	gi|500175084	1	1468	42707	44	73
hypothetical protein	hypothetical protein	Pcal_1130	gi|500175388	9	125	42370	6	12
serpin	proteinase inhibitor I4, serpin	Pcal_0455	gi|500174720	4	165	42139	5	18
hypothetical protein	hypothetical protein	Pcal_1999	gi|500176247	6	731	41923	17	37
ABC transporter substrate-binding protein	periplasmic solute binding protein	Pcal_1367	gi|500175621	1	861	41887	23	47

ATPase	ATPase associated with various cellular activities, AAA_5	Pcal_0360	gi|500174627	3	127	41579	4	13
hypothetical protein	hypothetical protein	Pcal_1490	gi|500175742	3	289	41380	5	20
ABC transporter	ABC-2 type transporter	Pcal_1407	gi|500175660	4	211	40958	4	10
ABC transporter substrate-binding protein	extracellular solute-binding protein, family 1	Pcal_0120	gi|500174390	2	1014	40956	23	60
pullulanase	pullulanase	Pcal_0976	gi|500175234	6	2902	40745	63	34
hypothetical protein	protein of unknown function DUF107	Pcal_1326	gi|500175581	1	370	40246	15	25
peptide ABC transporter permease	binding-protein-dependent transport systems inner membrane component	Pcal_0974	gi|500175232	2	672	39836	26	26
TRAP ABC transporter	TRAP transporter solute receptor, TAXI family	Pcal_0207	gi|500174476	2	1016	39762	26	46
nosD	Parallel beta-helix repeat containing protein	Pcal_1938	gi|500176187	6	327	39586	7	18
cytochrome c biogenesis protein	cytochrome c biogenesis protein, transmembrane region	Pcal_1482	gi|500175734	1	419	38901	8	18
excisionase	Phosphatidylserine/phosphatidylglycerophosphate/cardiolipin synthase and related enzyme-like protein	Pcal_0785	gi|500175044	3	209	38178	5	15
ABC transporter substrate-binding protein	ABC transporter, periplasmic binding protein, thiB subfamily	Pcal_0214	gi|500174483	2	1602	37000	69	46
primosomal protein	band 7 protein	Pcal_0235	gi|500174504	4	489	35979	15	34

hypothetical protein	conserved hypothetical protein	Pcal_0325	gi|500174592	7	562	35737	17	35
cytochrome c biogenesis protein	cytochrome c assembly protein	Pcal_1486	gi|500175738	2	149	34808	4	7
glycosyl transferase	glycosyl transferase, family 2	Pcal_0484	gi|500174748	5	147	34650	6	19
hypothetical protein	hypothetical protein	Pcal_0832	gi|500175091	14	467	34321	18	29
ABC transporter substrate-binding protein	conserved hypothetical protein	Pcal_1399	gi|500175652	1	801	33681	28	43
glycosyl transferase family 2	glycosyl transferase, family 2	Pcal_0487	gi|500174750	3	157	32721	6	12
hypothetical protein	hypothetical protein	Pcal_1613	gi|500175863	4	225	32058	7	20
nitrate reductase	nitrate reductase subunit	Pcal_1905	gi|500176154	1	445	31966	10	31
transcriptional regulator	transcriptional regulator, TrmB	Pcal_0346	gi|500174613	6	217	31944	6	18
hypothetical protein	hypothetical protein	Pcal_0352	gi|500174619	8	459	30355	11	33
4Fe-4S ferredoxin	4Fe-4S ferredoxin, iron-sulfur binding domain protein	Pcal_1496	gi|500175748	1	273	30004	9	32
amino acid ABC transporter substrate-binding protein	amino acid ABC transporter substrate-binding protein, PAAT family	Pcal_1148	gi|500175406	1	1899	29226	52	55
hypothetical protein	conserved hypothetical protein	Pcal_0044	gi|500174316	1	183	27931	10	28
senC	electron transport protein SCO1/SenC	Pcal_1948	gi|500176197	3	175	27717	9	22
hypothetical protein	hypothetical protein	Pcal_1550	gi|500175801	6	155	27086	5	16
RNA methyltransferase	conserved hypothetical protein	Pcal_1172	gi|500175430	1	146	26810	8	28
transcriptional regulator	Uncharacterized membrane-associated	Pcal_1547	gi|500175798	5	238	26210	10	21

	protein/domain-like protein							
hypothetical protein	hypothetical protein	Pcal_0082	gi|500174354	5	132	25602	2	13
hypothetical protein	hypothetical protein	Pcal_1940	gi|500176189	3	1491	25340	51	63
hypothetical protein	conserved hypothetical protein	Pcal_0340	gi|500174607	2	1182	24484	47	50
eight-cysteine-cluster	hypothetical protein	Pcal_1255	gi|500175512	3	244	24085	6	25
domain-containing protein								
30S ribosomal protein S2	SSU ribosomal protein S2P	Pcal_0027	gi|500174299	1	133	23955	7	16
protein-disulfide isomerase-like protein	Protein-disulfide isomerase-like protein	Pcal_0247	gi|500174516	4	371	23821	9	31
V-type ATPase subunit D	V-type ATPase, D subunit	Pcal_2162	gi|500176410	1	708	23300	26	60
ABC transporter	ABC transporter related protein	Pcal_1932	gi|500176181	2	226	23115	4	25
ABC transporter	ABC transporter related protein	Pcal_1937	gi|500176186	1	252	22859	6	19
V-type ATP synthase subunit E	H+-transporting two-sector ATPase, E	Pcal_2088	gi|500176336	1	397	21736	11	47
	subunit							
hypothetical protein	hypothetical protein	Pcal_0284	gi|500174553	3	325	21659	10	38
ABC transporter	ABC transporter related protein	Pcal_1484	gi|500175736	1	297	20891	11	48
hypothetical protein	hypothetical protein	Pcal_0148	gi|500174418	2	209	20418	4	19
hypothetical protein	hypothetical protein	Pcal_0083	gi|500174355	4	276	19905	7	25
hypothetical protein	hypothetical protein	Pcal_1694	gi|500175945	2	502	19753	23	28
FHA domain-containing protein	FHA domain containing protein	Pcal_0026	gi|500174298	1	113	16523	2	18
hypothetical protein	protein of unknown function DUF1641	Pcal_1437	gi|500175690	1	137	16191	5	29

hypothetical protein	hypothetical protein	Pcal_1552	gi|500175803	4	132	14064	4	26
hypothetical protein	hypothetical protein	Pcal_1485	gi|500175737	1	444	13440	18	50
hypothetical protein	hypothetical protein	Pcal_1245	gi|500175502	1	308	12484	6	13
hypothetical protein	hypothetical protein	Pcal_0309	gi|500174576	1	218	11680	5	36

Appendix table II. Glycopeptides identified by LC-MS/MS analysis of ConA enriched glycoproteins

Number	Protein	ORF	Peptide	Observed mass (m/z)	Glycan structure	
1	hypothetical protein	Pcal_0033	NVGN[198]WTGR	1351.54(2+)	$Hex(NAc)_2HexA(NAc)_2Hex_8$	(S10)
				1270.51(2+)	$Hex(NAc)_2HexA(NAc)_2Hex_7$	(S9)
2	hypothetical protein	Pcal_0194	LVVN[631]TTR	1301.05(2+)	$Hex(NAc)_2HexA(NAc)_2Hex_8$	(S10)
				1220.03(2+)	$Hex(NAc)_2HexA(NAc)_2Hex_7$	(S9)
				1139.00(2+)	$Hex(NAc)_2HexA(NAc)_2Hex_6$	(S8)
3	peptidase S49	Pcal_0153	VAN[56]ISASVDK	1401.58(2+)	$Hex(NAc)_2HexA(NAc)_2Hex_8$	(S10)
				1320.56(2+)	$Hex(NAc)_2HexA(NAc)_2Hex_7$	(S9)
4			LLN[66]ATR	1243.52(2+)	$Hex(NAc)_2HexA(NAc)_2Hex_8$	(S10)
				1162.49(2+)	$Hex(NAc)_2HexA(NAc)_2Hex_7$	(S9)
				1081.46(2+)	$Hex(NAc)_2HexA(NAc)_2Hex_6$	(S8)
5			LYSTIAGLN[146]K	1520.64(2+)	$Hex(NAc)_2HexA(NAc)_2Hex_9$	(S11)
				1439.61(2+)	$Hex(NAc)_2HexA(NAc)_2Hex_8$	(S10)
				1358.59(2+)	$Hex(NAc)_2HexA(NAc)_2Hex_7$	(S9)
				1277.56(2+)	$Hex(NAc)_2HexA(NAc)_2Hex_6$	(S8)
6	ABC transporter	Pcal_1519	N[79]ATAILLK	1402.60(2+)	$Hex(NAc)_2HexA(NAc)_2Hex_9$	(S11)

	substrate-binding protein			1321.57(2+)	$Hex(NAc)_2HexA(NAc)_2Hex_8$	(S10)
				1240.55(2+)	$Hex(NAc)_2HexA(NAc)_2Hex_7$	(S9)
				1159.52(2+)	$Hex(NAc)_2HexA(NAc)_2Hex_6$	(S8)
7			$ALEN^{203}GTIR$	1336.55(2+)	$Hex(NAc)_2HexA(NAc)_2Hex_8$	(S10)
				1255.52(2+)	$Hex(NAc)_2HexA(NAc)_2Hex_7$	(S9)
				1174.49(2+)	$Hex(NAc)_2HexA(NAc)_2Hex_6$	(S8)
8			$YAYSN^{343}YSR$	1411.53(2+)	$Hex(NAc)_2HexA(NAc)_2Hex_8$	(S10)
				1330.51(2+)	$Hex(NAc)_2HexA(NAc)_2Hex_7$	(S9)
9	hypothetical protein	Pcal_0611	$AVN^{145}TTVK$	1266.02(2+)	$Hex(NAc)_2HexA(NAc)_2Hex_8$	(S10)
				1184.99(2+)	$Hex(NAc)_2HexA(NAc)_2Hex_7$	(S9)
10			$IAN^{262}ESFRGR$	1424.57(2+)	$Hex(NAc)_2HexA(NAc)_2Hex_8$	(S10)
				1343.56(2+)	$Hex(NAc)_2HexA(NAc)_2Hex_7$	(S9)
11	pullulanase	Pcal_0976	$N^{51}GTVFDLTR$	1411.07(2+)	$Hex(NAc)_2HexA(NAc)_2Hex_8$	(S10)
				1330.04(2+)	$Hex(NAc)_2HexA(NAc)_2Hex_7$	(S9)
				1249.02(2+)	$Hex(NAc)_2HexA(NAc)_2Hex_6$	(S8)
12	hypothetical protein	Pcal_0034	$N^{75}FTLIIR$	1338.07(2+)	$Hex(NAc)_2HexA(NAc)_2Hex_8$	(S10)
				1257.05(2+)	$Hex(NAc)_2HexA(NAc)_2Hex_7$	(S9)