DEVELOPMENT OF A MODEL ORGANISM FOR BACTERIAL SURFACE LAYER PROTEIN *O*-GLYCOSYLATION

DOCTORAL THESIS PRESENTED TO THE

UNIVERSITÄT FÜR BODENKULTUR WIEN

(UNIVERSITY OF NATURAL RESOURCES AND APPLIED LIFE SCIENCES, VIENNA)

BY DIPL. BIOL. KRISTOF ZARSCHLER

TO OBTAIN THE DEGREE OF

DR. NAT. TECHN.

(DOCTOR RERUM NATURALIUM TECHNICARUM)

DEPARTMENT OF NANOBIOTECHNOLOGY

HEAD: O.UNIV.PROF. DIPL.-ING. DR.NAT.TECHN. UWE B. SLEYTR

SUPERVISOR: AO.UNIV.PROF. DIPL.-ING. DR.NAT.TECHN. PAUL MESSNER

AUGUST 2009

“Keine Schuld ist dringender, als die, Dank zu sagen.“
Marcus Tullius Cicero, (106 - 43 v. Chr.)

An erster Stelle möchte ich Herrn Prof. Dr. Uwe B. Sleytr für die Möglichkeit danken, dass ich meine Dissertation am Department für Nanobiotechnologie absolvieren konnte.

Für die hervorragende Betreuung meiner Dissertation, die tiefgründigen Einblicke in die Thematik und die faszinierenden Exkurse in die Welt der Glykobiologie möchte ich Herrn Prof. Dr. Paul Messner danken.

Weiterhin möchte ich Frau Prof. Dr. Christina Schäffer für die hilfreichen Diskussionen, kritischen Anmerkungen und kreativen Denkanstöße, die maßgeblich zum Gelingen dieser Arbeit beigetragen haben, meinen Dank aussprechen.

Ein großes Danke für das angenehme Arbeitklima, die gemeinsamen Freizeitaktivitäten, die ständige Hilfsbereitschaft, die konstruktiven Gepräche und angeregten Diskussionen geht an meine ehemaligen und derzeitigen Kolleginnen und Kollegen der Arbeitsgruppe Nanoglykobiologie und des gesamten Departments für Nanobiotechnologie. Besonders sind dabei Frau Andrea Scheberl, Frau Sonja Zayni, Frau Jacqueline Friedmann und Herr Harald Mayer zu erwähnen, denen ich für ihre ausgezeichnete technische Unterstützung danken möchte.

Mein herzlicher Dank gilt Frau Dipl. Ing. Birgit Kainz und Frau Dr. Judith Ferner-Ortner-Bleckmann für die zahlreichen offenen Gespräche und unterhaltsamen Unternehmungen.

Für das sorgfältige Korrekturlesen dieser Arbeit und die hilfreichen Bemerkungen möchte ich Frau Dr. Kerstin Steiner meinen Dank aussprechen.

Herrn Dr. Martin Pabst und Prof. Dr. Friedrich Altmann möchte ich für die entscheidenden Analysen meiner Proben danken.

Herrn Dieter Jäger gilt nicht nur großer Dank für die Hilfe bei allen administrativen, organisatorischen und bürokratischen Angelegenheiten, sondern auch für seine humorvolle Art, die maßgeblich zum positiven Arbeitsklima beigetragen hat.

Weiters möchte ich an dieser Stelle dem Fonds zur Förderung der wissenschaftlichen Forschung (FWF) und der Hochschuljubiläumsstiftung der Stadt Wien für die finanzielle Unterstützung danken.

Mein größter Dank gilt meiner Familie und meinen Eltern, die mich stets bei der Verwirklichung meiner Vorhaben und Träume unterstützt haben.

“Wer nicht gelegentlich auch einmal kausalwidrige Dinge zu denken vermag, wird seine Wissenschaft nie um eine neue Idee bereichern können.“

Max Planck (1858-1947), deutscher Physiker, Begründer der Quantentheorie

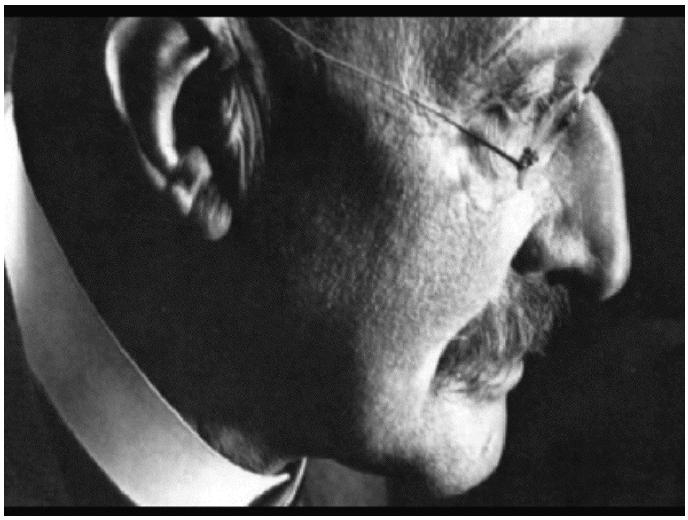


TABLE OF CONTENS

APPENDIX 5

Publication: Steiner, K., Novotny, R., Werz, D. B., **Zarschler, K.**, Seeberger, P. H., Hofinger, A., Kosma, P., Schäffer, C., Messner, P. (2008). Molecular basis of S-layer glycoprotein glycan biosynthesis in *Geobacillus stearothermophilus*. Journal of Biological Chemistry **283:**21120-21133.

APPENDIX 6

Abstract & Poster: **Zarschler, K.**, Janesch, B., Schäffer, C., Messner, P. (2009). Proposal of a pathway of the surface layer protein glycosylation in *Paenibacillus alvei* CCM 2051^{T} enabled by a glycan mutant approach. 15th European Carbohydrate Symposium, Vienna, Austria (19.-24.07.2009).

APPENDIX 7

Abstract & Poster: **Zarschler, K.**, Janesch, B., Schäffer, C., Messner, P. (2008). First insights into the surface layer glycosylation of *Paenibacillus alvei*. Life Sciences 2008 - Gemeinsame Jahrestagung der Österreichischen Gesellschaft für Biochemie und Molekularbiologie, der Österreichischen Gesellschaft für Genetik und Gentechnik, der Österreichischen Gesellschaft für Biotechnologie und des Austrian Networks für Gentherapie, Graz, Austria (22.-24.09.2008).

APPENDIX 8

Abstract & Poster: **Zarschler, K.**, Novotny, R., Steiner, K., Schäffer C., Messner, P. (2006). Is an ABC transporter responsible for the export of the cytoplasmatically synthesized surface layer (S-layer) glycan to the cell surface *Geobacillus stearothermophilus* NRS 2004/3a? FEBS Special Meeting: ATP-Binding Cassette (ABC) Proteins from Multidrug Resistance to Genetic Diseases, Innsbruck, Austria (04.-10.03.2006).

1. ZUSAMMENFASSUNG

Posttranslationale Glykosylierung ist sowohl in Eukaryoten als auch in Prokaryoten die häufigste und facettenreichste Modifikation von Proteinen. Neben verschiedenen Zellstrukturen wie Flagellen, Pili und Adhäsinen zählen glykosylierte S-Schicht-proteine zu den am besten untersuchten prokaryotischen Glykoproteinen. Die meisten S-Schicht(glyko)proteine besitzen die einzigartige Eigenschaft, von selbst zweidimensionale, kristalline Gitter auf der darunter liegenden Zellwandschicht auszubilden (Selbstorganisation). Diese geordneten, monomolekularen Gitter sind aus identen Untereinheiten aufgebaut und bedecken die bakterielle Zelloberfläche vollständig. Die Glykanketten der glykosylierten S-Schichtproteine ragen dabei wie die O-Antigene der Lipopolysaccharide Gram-negativer Bakterien von der Zellober-fläche weg.

Der mesophile, Gram-positive Organismus *Paenibacillus alvei* CCM 2051^T besitzt ein glykosyliertes S-Schichtprotein mit einer bekannten Glykanstruktur und ist, im Gegensatz zum thermophilen Bakterium *Geobacillus stearothermophilus* NRS 2004/3a, mit dem sich bedeutende Fortschritte in der Aufklärung der S-Schicht-glykan Biosynthese erzielen ließen, zugänglich für gentechnische Manipulationen. Seine *O*-Glykankette ist ein Heteropolymer aus [→3)-β-D-Gal*p*-(1[α-D-Glc*p*-(1→6)] →4)-β-D-Man*p*NAc-(1→] Wiederholungseinheiten, welches über einen Adapter aus -[GroA-2→OPO_2→4-β-D-Man*p*NAc-(1→4)] →3)-α-L-Rha*p*-(1→3)-α-L-Rha*p*-(1→3)-α-L-Rha*p*-(1→3)-β-D-Gal*p*-(1→ mit bestimmten Tyrosinen des S-Schichtproteins SpaA kovalent verbunden ist. Außer bei den S-Schichtglykoproteinen von *P. alvei, Thermoanaerobacter thermohydrosulfuricus* und *Thermoanaero-bacterium thermosaccharolyticum* sind *O*-glykosidische Bindungen an Tyrosine bei Prokaryoten weder bekannt noch näher untersucht und grundsätzlich nicht weit verbreitet in der Natur.

Um Einblicke in die Mechanismen der S-Schichtglykan Biosynthese zu erhalten, wurde ein S-Schichtglykan Biosynthesecluster (*slg*-Biosynthesecluster) identifiziert, vollständig sequenziert und die Funktion der einzelnen offenen Leserahmen (ORFs) im Datenbankvergleich analysiert. Der 24,3-kb große *slg*-Biosynthesecluster kodiert für 18 ORFs, deren Genprodukte sowohl in die Herstellung von nukleotidaktivierten Monosacchariden (UDP-Glukose, UDP-Galaktose, CDP-Glycerin, dTDP-L-Rhamnose) als auch in den Aufbau, Export und *en bloc* Transfer des

Polysaccharides auf das S-Schichtprotein involviert sind. Untersuchungen zur Regulation der S-Schichtglykan Biosynthese zeigten, dass 16 ORFs als eine Einheit (polycistronisch) transkribiert werden, während zwei ORFs, darunter die vermeintliche Oligosaccharyltransferase WsfB, unabhängig davon transkribiert werden.

Zur detaillierten Aufklärung der konkreten Funktion jedes einzelnen, im *slg*-Biosynthesecluster kodierten Enzyms wurde neben einem zuverlässigen Protokoll für die Transformation von Plasmid-DNS durch Elektroporation und einem Vektor für die heterologe Genexpression ein leistungsfähiges Knock-out System für die gezielte Inaktivierung von Genen entwickelt. Dieses System basiert auf dem handelsüblichen sogenannten *bacterial mobile group II intron* Ll.LtrB von *Lactococcus lactis* und wurde mit dem Promoter des S-Schichtgens *sgsE* von *G. stearothermophilus* NRS 2004/3a und dem *Geobacillus-Bacillus-E. coli* Vektor pNW33N kombiniert. Mit Hilfe des konstruierten Knock-out Systems wurde das Initiationsenzym der S-Schichtglykan Biosynthese in *P. alvei* CCM 2051^T, WsfP, identifiziert. Die Zerstörung des entsprechend Gens *wsfP* führte zur Erzeugung einer Mutante, welche die Eigenschaft, das S-Schicht Glykan herzustellen, verloren hatte. Durch die Expression von *wsfP* in der entsprechenden Mutante konnte die vollständige Glykosylierung wiederhergestellt werden, was dafür spricht, daß WsfP die S-Schicht-glykosylierung durch den Transfer von Galaktose von UDP-α-D-Gal auf einen Lipidanker initiiert. Bis auf diejenigen ORFs, die für Enzyme zur Biosynthese von Nukleotid-aktivierten Monosacchariden und für das Membranprotein eines ABC-Transporters kodieren, wurden neben *wsfP* alle ORFs des *slg*-Biosynthesecluster durch die Insertion von Ll.LtrB zerstört und die generierten Mutanten mittels Massen-spektrometrie analysiert. Zusammen mit den Ergebnissen des Datenbankvergleichs der entsprechenden Proteine erlauben diese Resultate eine beinahe vollständige Zuordnung der jeweiligen biologischen Funktion der Proteine im *slg* Cluster.

Die Identifikation eines ABC-Transportersystems (Wzt und Wzm) und das Fehlen der S-Schichtglykosylierung in der *wzt*::Ll.LtrB Mutante sprechen dafür, daß der Export der lipidgebundenen Glykankette unter ATP-Hydrolyse stattfindet, ähnlich dem ATP-Transporter-abhängigen LPS *O*-Polysaccharid Biosyntheseweg. Dementsprechend erfolgt der Aufbau der Glykankette durch das schrittweise Aneinanderfügen von Monosacchariden an das nicht-reduzierende Ende der wachsenden, lipid-gebundenen Glykankette. Nach dem Erreichen einer bestimmten Kettenlänge wird

das Polymer unter Verwendung eines ABC-Transporters durch die Zellmembran exportiert. Während des Exports kommt es durch die Bindung eines Glukoserests an *N*-Acetylmannosaminreste jeder Wiederholungseinheit zur Vollendung des Polysaccharids und zum abschließenden Transfer auf adäquate Tyrosine des Zielproteins SpaA durch die Oligosaccharyltransferase WsfB.

Neben dem *slg*-Biosynthesecluster wurde das Strukturgen *spaA*, das für das S-Schichtprotein von 983 Aminosäuren, inklusive des Signalpeptids von 24 Aminosäuren kodiert, sequenziert. Das reife S-Schichtprotein besitzt eine theoretische Molekülmasse von 105,95 kDa, einen berechneten isoelektrischen Punkt von 5,83 und verfügt über drei *S-layer homology* (SLH) Domänen, welche für die Verankerung der Proteinuntereinheiten an der Zelloberfläche verantwortlich sind.

Um die Eignung von SpaA für die Präsentation von chimären Glykoproteinen an der bakteriellen Zelloberfläche zu prüfen, wurden die beiden Fusionsproteine SpaA_6HIS und SpaA_eGFP hergestellt und erfolgreich in *P. alvei* CCM 2051^T exprimiert. Zusätzlich zu Immunoblot-Analysen ergaben Immunofluoreszenzfärbungen und fluoreszenzmikroskopische Analysen, daß beide Fussionsproteine ausreichend exprimiert, glykosyliert und erfolgreich an der Zelloberfläche von *P. alvei* CCM 2051^T präsentiert werden.

Die vielversprechenden Einblicke in die S-Schichtglykan Biosynthesemaschinerie und die Eignung des S-Schicht Systems für das kontrollierte Verankern von heterologen Glykoproteinen auf der Zelloberfläche machen *P. alvei* CCM 2051^T zu einem bevorzugten Kandidaten für das Design und die Präsentation von Peptiden und Proteinen in Kombination mit funktionellen, bioaktiven Glykanen. Derartige S-Schicht *Neo*glykoproteine könnten in verschiedenen Bereichen, in denen die Erkennung von Kohlenhydratmotiven eine wesentliche Rolle spielt, zum Beispiel bei der Rezeptor-Substrat-Interaktion, der Wirkstofffreisetzung am Zielort und der interzellulären Kommunikation, ihre biotechnologische Anwendung finden.

2. SUMMARY

Posttranslational glycosylation is the most prevalent and most diverse modification of proteins in eukaryotes as well as in prokaryotes. In addition to several cell structures such as flagella, pili, and adhesins, surface (S-) layer glycoproteins have become one of the best-studied prokaryotic glycoprotein species. Most S-layer (glyco)proteins have the unique feature to self-assemble into two-dimensional crystalline arrays on the supporting cell envelope. These regular monolayers are composed of identical subunits and cover the bacterial cells completely with the glycan moieties protruding from the cell surface, comparable to the lipopolysaccharide coating of Gram-negative bacteria.

The mesophilic Gram-positive organism *Paenibacillus alvei* CCM 2051^T possesses a glycosylated S-layer with an elucidated glycan structure and is, unlike the thermophilic bacterium *Geobacillus stearothermophilus* NRS 2004/3a, where significant advances in the elucidation of the S-layer protein glycosylation pathway have been achieved, amenable to genetic manipulation. Its *O*-glycan is a hetero-polymer of [→3)-β-D-Gal*p*-(1[α-D-Glc*p*-(1→6)] →4)-β-D-Man*p*NAc-(1→] repeating units that is linked by an adaptor of -[GroA-2→OPO_2→4-β-D-Man*p*NAc-(1→4)] →3)-α-L-Rha*p*-(1→3)-α-L-Rha*p*-(1→3)-α-L-Rha*p*-(1→3)-β-D-Gal*p*-(1→ to specific tyrosine residues of the S-layer protein SpaA. Apart from the S-layer glycoproteins of *P. alvei, Thermoanaerobacter thermohydrosulfuricus* and *Thermoanaerobacterium thermo-saccharolyticum*, *O*-glycosidic linkages to tyrosine residues have neither been identified nor investigated in prokaryotes and are generally rare in nature.

In order to gain insights into the mechanisms governing S-layer glycan biosynthesis, a gene cluster responsible for glycan biosynthesis (*slg* gene cluster) has been identified, entirely sequenced, and the genes have been assigned by protein database comparison. The ~24.3-kb gene cluster encodes 18 open reading frames (ORFs) whose translation products are involved in the biosynthesis of nucleotide-activated monosaccharides (UDP-glucose, UDP-galactose, CDP-glycerol, dTDP-L-rhamnose) and in assembly, export, as well as *en bloc* transfer of the polysaccharide to the S-layer protein. Transcriptional analysis of the *slg* gene cluster revealed that 16 ORFs are transcribed by a single polycistronic mRNA, whereas two ORFs, including the putative oligosaccharyltransferase WsfB (OTase), are transcribed independently.

To further investigate the function of individual enzymes encoded by the *slg* gene cluster in detail, a reliable protocol for introduction of plasmid-DNA into *P. alvei* CCM 2051^T by electroporation, a shuttle vector for heterologous gene expression, and an effective gene knockout system for insertional inactivation of desired target genes were developed. This system is based on the commercially available bacterial mobile group II intron Ll.LtrB of *Lactococcus lactis* in combination with the broad host-range S-layer gene promoter of *sgsE* from *G. stearothermophilus* NRS 2004/3a and the *Geobacillus-Bacillus-E. coli* shuttle vector pNW33N. By means of the constructed gene knockout system, the initiation enzyme of S-layer glycan biosynthesis in *P. alvei* CCM 2051^T, WsfP, was identified. Disruption of the corresponding gene *wsfP* resulted in a mutant, which had lost the ability to assemble the S-layer glycan chain. Glycosylation could be fully restored by expression of *wsfP* in the respective *P. alvei* mutant, confirming that WsfP initiates S-layer protein glycosylation by the transfer of a galactose residue from UDP-α-D-Gal to a lipid carrier. In addition to *wsfP*, all ORFs of the gene cluster, except those encoding nucleotide sugar biosynthesis enzymes and the ABC transporter integral transmembrane protein, were disrupted by insertion of the mobile group II intron, and glycosylated S-layer proteins produced in mutant backgrounds were analyzed by mass spectrometry. These results, combined with the observed similarity of the respective proteins, allowed an almost complete assignment of their biological function. The identification of an ABC transporter system (Wzm and Wzt) and the loss of S-layer glycosylation in the *wzt*::Ll.LtrB mutant corroborate the assumption of an ATP hydrolysis-driven export of the lipid-linked glycan chain comparable to the ABC-transporter-dependent pathway of the LPS *O*-antigen biosynthesis. According to this pathway, the glycan chain extension is accomplished by processive addition of monosaccharides to the nonreducing terminus of the lipid-linked growing chain. After extension to a certain chain length, the polymer is exported through the cytoplasmic membrane by the ABC transporter. The polysaccharide would be completed by the addition of a glucose residue to *N*-acetylmannosamine of each repeating unit and then transferred to appropriate tyrosine residues of the target protein SpaA by the oligosaccharyltransferase WsfB.

Besides the *slg* gene cluster, the S-layer structural gene *spaA* has been sequenced, encoding a protein of 983 amino acids, including a signal peptide of 24 amino acids. The mature S-layer protein has a theoretical molecular mass of 105.95 kDa, a

calculated pI of 5.83, and contains three aminoterminal S-layer homology domains involved in cell surface anchoring of the glycoprotein subunits.

To evaluate the applicability of SpaA for bacterial cell surface display of chimeric glycoproteins, full-length SpaA with a carboxyterminal hexahistidine tag (SpaA_6HIS) and a SpaA_eGFP fusion protein were created and successfully expressed. Immunoblot analysis together with immunofluorescence staining and fluorescence microscopy revealed that the fusion proteins SpaA_6HIS and SpaA_eGFP are efficiently expressed and glycosylated as well as successfully displayed on the surface of *P. alvei* CCM 2051^{T} cells.

Taking together the promising insights into the S-layer glycosylation machinery and the suitability of the S-layer system for the controlled surface display of heterologous glycoproteins, *P. alvei* CCM 2051^{T} has become our favorite candidate for the design and presentation of numerous peptides and proteins carrying functional, bioactive glycans. These S-layer *neo*glycoproteins may be useful in various areas of applications based on carbohydrate recognition, such as receptor-substrate interaction, drug delivery, or cell-cell communication.

3. INTRODUCTION

Carbohydrates are the most common biomolecules on earth, because of their substantial roles in all forms of life. They are used for storage and transport of energy, they are building blocks of the structural framework of RNA and DNA, and they are structural elements in connective tissues of animals and in the cell walls of bacteria, fungi, and plants. Furthermore, carbohydrates are attached to many proteins and lipids, where they play key roles in mediating interactions among cells and interactions between cells and other compounds in the cellular environment. Additionally, carbohydrates and their derivatives play major roles in the working processes of the immune system, in fertilization, pathogenesis, blood clotting, and development (1).

3.1 PROKARYOTIC GLYCOCONJUGATES

Glycoconjugates is the general classification for carbohydrates covalently linked to a non-carbohydrate moiety. Prokaryotes possess an extreme variety of glycosylated macromolecules as constituents of their cell wall (Figure 1). The major glyco-conjugates of bacteria are glycoproteins, peptidoglycans, glycolipids, lipopoly-saccharides (LPS), capsular polysaccharides and other cell wall-associated poly-saccharides, e.g., teichoic acids, teichuronic acids, lipoteichoic acids, and secondary cell wall polymers. In archaea, lacking both, peptidoglycan and a cell wall, sometimes S-layer glycoproteins and pseudomurein are the most frequent glycoconjugates (2).

3.1.1 NON-PROTEINACEOUS GLYCOCONJUGATES

Peptidoglycan or murein is an essential structural cell wall component of the bacterial cell wall on the outside of the cytoplasmic membrane of all bacteria. Its structure, architecture, function, biosynthesis, and degradation have been extensively investigated and excellently reviewed (3-6). Instead of murein, methanogenic Gram-positive archaea possess pseudomurein, a different type of polymer, that is involved in the formation of rigid cell envelope scaffolds and resembling peptidoglycan (2, 7).

Besides several layers of peptidoglycan, the cell wall of Gram-positive bacteria contains further cell wall-associated polysaccharides such as teichoic acids (8, 9) and teichuronic acids (10, 11). Whereas these two glycoconjugates are covalently linked

to *N*-acetylmuramic acid residues of the peptidoglycan, lipoteichoic acids are anchored to the cytoplasmic membrane via a glycerolipid (12, 13).

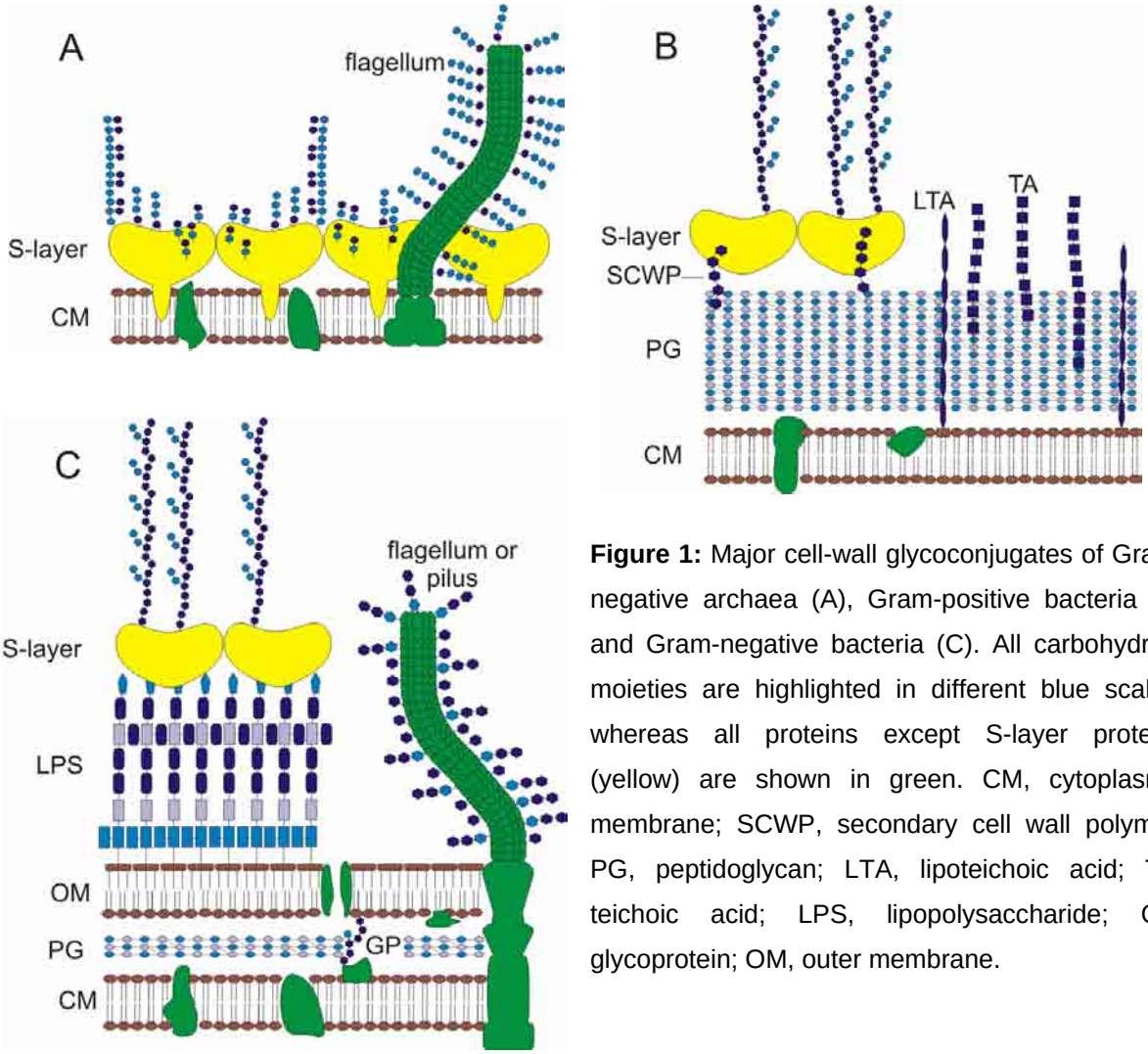


Figure 1: Major cell-wall glycoconjugates of Gram-negative archaea (A), Gram-positive bacteria (B) and Gram-negative bacteria (C). All carbohydrate moieties are highlighted in different blue scales, whereas all proteins except S-layer proteins (yellow) are shown in green. CM, cytoplasmic membrane; SCWP, secondary cell wall polymer; PG, peptidoglycan; LTA, lipoteichoic acid; TA, teichoic acid; LPS, lipopolysaccharide; GP, glycoprotein; OM, outer membrane.

Additionally to a relatively thin layer of peptidoglycan, amphipathic LPS molecules located in the outer membrane are essential and characteristic glycoconjugates in the cell envelope of Gram-negative bacteria stabilizing the overall membrane structure and protecting the membrane from certain kinds of chemical attack. The tripartite LPS molecule typically consists of a hydrophobic domain designated as lipid A or endotoxin, anchoring LPS in the outer membrane, a core oligosaccharide, and a distal polysaccharide or O-antigen extending from the cell surface and interacting with the environment. The structural and serological diversity of O-antigens derives from variations in the sugar composition and sequence, the position and stereochemistry of the *O*-glycosidic linkages, and the substitution of monomers with non-carbohydrate residues resulting in ~170 different O-antigen types. The

biochemistry, structure and biosynthesis of LPS has been analyzed in great detail over the last years (14-19).

In addition to LPS, the capsular polysaccharide K antigens are a second kind of serotype-specific surface polysaccharides isolated from different Gram-negative bacteria. These highly variable surface structures are virulence factors generally involved in resistance to phagocytosis and complement-mediated killing. The biosynthesis and assembly of capsular polysaccharides is a complex process summarized recently (20).

3.1.2 GLYCOPROTEINS

Carbohydrates linked to proteins or peptides have long thought to be absent from prokaryotes, but by now it is established knowledge that protein glycosylation is a frequent heterogenous posttranslational modification in all domains of life. Recent work of several research groups has revealed the great structural diversity of prokaryotic glycoproteins, and progress has been made in the identification of molecular mechanisms, immunological aspects and application potentials of prokaryotic glycosylation (21). Among the best investigated prokaryotic glycoproteins are archaeal and bacterial surface (S-) layer glycoproteins (22), flagella of *Helicobacter pylori* (23, 24) and *Campylobacter jejuni* (25-27), pili of *Neisseria meningitidis* (28-31) and *Pseudomonas aeruginosa* (32-34), the adhesins of *Escherichia coli* (35, 36) and *Haemophilus influenzae* (37, 38), and glycoproteins generated by the general glycosylation systems of *Bacteroides fragilis* (39, 40), *C. jejuni* (30, 41-46), and pathogenic *Neisseria* (47, 48).

In prokaryotes, the glycans, which often contain rare monosaccharides (e.g., di-*N*-acetylbacillosamine, *N*-acetylfucosamine, pseudaminic acid or legionaminic acid), are attached to the peptide backbone via the amide nitrogen of an asparagine residue (*N*-glycosylation) or via the functional hydroxy group of serine, threonine or tyrosine (*O*-glycosylation), and different carbohydrate-peptide linkage regions have been identified until now (44). An excellent overview of S-layer glycoproteins and non-S-layer glycoproteins including the linkage regions and method of detection was published recently (49).

The *O*-glycosidic linkage of a glycan chain via β-D-galactose or β-D-glucose to tyrosine was discovered as completely new types of linkage in the bacterial S-layer glycoproteins of *Paenibacillus alvei*, *Thermoanaerobacter thermohydrosulfuricus* and

Thermoanaerobacterium thermosaccharolyticum (50-56). In addition to these S-layer glycoproteins, *O*-glycosylated tyrosine has only been identified as tyrosine glucoside in insects (57, 58) as well as in glycogenin of glycogen-containing cells, where this enzyme initiates glycogen synthesis in an autocatalytic reaction linking individual α-D-glucose residues to tyrosine-194 in order to form a short priming chain of glucose residues as a substrate for the glycogen synthase (59-64).

In the following, the general glycosylation systems of *C. jejuni*, *B. fragilis*, *N. meningitidis* and *N. gonorrhoeae* will be described briefly, and particular attention to the glycosylation of S-layer proteins will be drawn in chapter 3.2. In Table 1, the four general glycosylation systems are compared.

The mucosal pathogen *C. jejuni* possesses a considerable repertoire of glyco-conjugates including lipooligosaccharides, capsular polysaccharides, and different secreted or surface located proteins being posttranslationally modified with *O*- or *N*-linked glycans (65). Besides the *O*-glycosylation of flagellin (25), up to 40 soluble and membrane proteins have been reported to be glycosylated in *C. jejuni* via the general *N*-glycosylation pathway encoded by the *p*rotein *g*lycosylation *l*ocus (*pgl*) (66). Several genes involved in protein glycosylation have been identified by site-specific mutation (67, 68), including *pglB* encoding an oligosaccharyltransferase mediating the *N*-linkage of a heptasaccharide to the peptide backbone (69). In the proposed model for *N*-linked glycosylation in *C. jejuni*, the assembly of the glycan chain proceeds through the sequential addition of monosaccharides from nucleotide-activated precursors onto a lipid carrier by the five glycosyltransferases PglA, PglC, PglH, PglI, and PglJ resulting in the formation of a branched heptasaccharide (Figure 2). This glycan is then transferred across the inner membrane into the periplasm by the ATP-binding cassette (ABC) transporter orthologue PglK and coupled by PglB to asparagine residues present in the sequon D/E-Y-N-X-S/T, where X and Y can be any amino acid except proline, to form the *N*-linked glycoproteins (46, 70, 71). The relaxed substrate specificity of PglB and the functional transfer of the *N*-glycosylation system into *E. coli* allow the production of engineered glycoconjugates for various biotechnological applications in this organism (72-74).

In addition to the general *N*-glycosylation system of *C. jejuni*, three different general prokaryotic protein *O*-glycosylation systems have been identified recently. Comstock and colleagues demonstrated that the major human intestinal symbiont *Bacteroides*

fragilis produces many glycoproteins that bind the fucose-specific *Aleuria aurantia* lectin (75). These fucosylated glycoproteins are located in the periplasm or outer membrane and have a variety of predicted functions. Detailed investigation of a selected glycoprotein showed, that the glycans are *O*-glycosidically linked to threonine residues and that the glycosylation requires the transport of the target protein to the periplasm. Furthermore, a region on the *B. fragilis* genome encoding a putative flippase, five putative glycosyltransferases, and other genes likely to be involved in oligosaccharide biosynthesis was identified to be required for the synthesis of all fucosylated glycoproteins. This region, named *lfg* for *l*ocus of *f*ragilis *g*lycosylation, appears to be part of a general glycosylation system not only in *B. fragilis*, but also in five other intestinal *Bacteroides* species (40). The mechanism of *O*-glycosylation in *B. fragilis* seems to have some similarities to protein glycosylation in other bacteria sharing the common features of assembly of the glycan on a lipid carrier by cytoplasmic glycosyltransferases using nucleotide-activated mono-saccharides, flipping of the glycan into the periplasm and subsequent *en bloc* transfer of the oligosaccharide to the protein (Figure 2) (46).

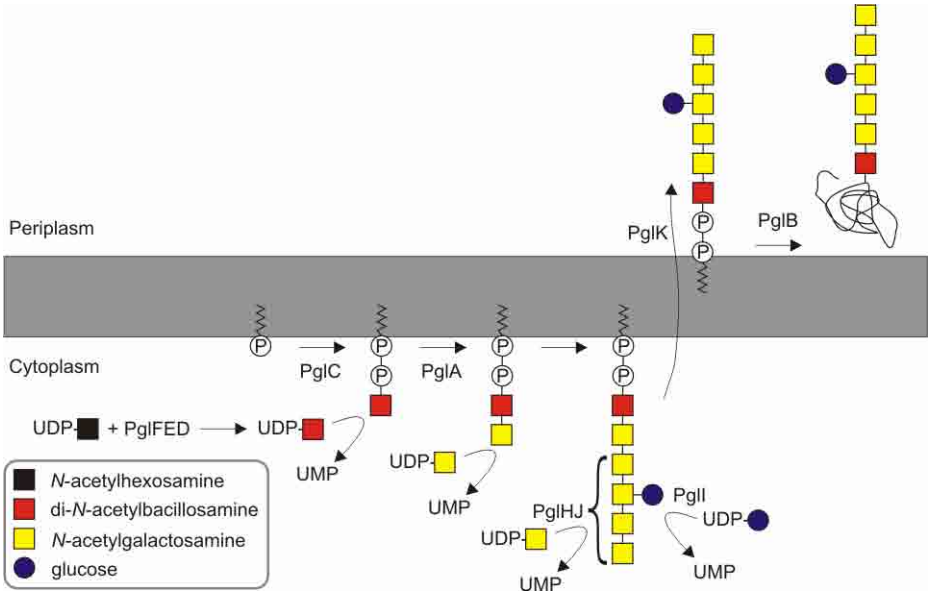


Figure 2: Current model of *N*-linked glycosylation in *Campylobacter jejuni* (76). In *C. jejuni*, PglF, PglE and PglD, acting as a dehydratase, an aminotransferase and an *N*-acetylase respectively, are required for the synthesis of UDP-Bac from UDP-GlcNAc. PglC then attaches the di-*N*-acetylbacillosamine residue to a lipid carrier, and PglA adds the α-1,3-linked GalNAc moiety. PglH and PglJ are involved in transferring the next four α-1,4-linked GalNAc moieties, and finally, the glycosyltransferase PglI adds the branching Glc moiety. Once assembled, the entire heptasaccharide is translocated across the inner membrane by PglK and transferred to the asparagine residue within the sequon D/E-Y-N-X-S/T, where X and Y can be any amino acid except proline, by the oligosaccharyltransferase PglB (46, 71).

The two remaining general prokaryotic protein *O*-glycosylation systems were identified in the pathogenic bacteria *Neisseria meningitidis* and *N. gonorrhoeae*. In both strains, an *O*-glycosylation pathway modifies a single abundant protein, pilin, representing the protein subunit forming pili. Several genes encoded by polymorphic *p*ilin *gl*ycosylation (*pgl*) clusters have been identified that are involved in glycosylation of this protein in pathogenic *Neisseria* (30, 77, 78). Besides pilin, one or more additional proteins are glycosylated by the neisserial pilin glycosylation pathway (47, 48). For *N. meningitidis*, the surface-exposed, flexible carboxyterminal domain of the outer membrane nitrite reductase AniA is decorated with the same phase-variable *O*-linked glycan modification as observed on pilin (48). Eleven functionally diverse membrane-associated proteins are targeted by the general *O*-glycosylation system of *N. gonorrhoeae* (47). In accordance with the findings in *B. fragilis*, trafficking of neisserial proteins to or through the periplasm is likely to be a prerequisite for glycosylation, because the corresponding *O*-OTases are predicted to act in this compartment (79, 80).

In all of the mentioned organisms it has been shown, that glycosylations are often found on proteins important in pathogenesis and/or in colonization of a particular environment. However, the biological significance of global protein glycosylation needs to be addressed in the future.

Table 1: Comparison between the different general glycosylation systems of *Campylobacter jejuni*, *Bacteroides fragilis*, *Neisseria meningitidis,* and *N. gonorrhoeae*.

	Campylobacter jejuni	*Bacteroides fragilis*	*Neisseria meningitidis*	*Neisseria gonorrhoeae*
Type of organism	Human gut mucosal pathogen causing gastroenteritis	Human intestinal symbiont	Human pathogen causing meningitis	Human pathogen causing gonorrhoea
Linkage type	*N*-glycosidic to asparagine	*O*-glycosidic to threonine	*O*-glycosidic to serine / threonine	
Linkage sugar	2,4-diacetamido-2,4,6-trideoxy-glucose (di-*N*-acetyl-bacillosamine)	Not determined	2,4-diacetamido-2,4,6-trideoxyhexose (DATDH)	

Function of glycosylated proteins	Diverse		Pilin protein PilE; nitrite reductase AniA	Pilin protein PilE and many functionally diverse membrane-associated proteins
Effect of glycosylation gene deletion	Decreased bacterial attachment and invasion of human epithelial cells; loss of colonization ability; reduced protein reactivity to antisera	No incorporation of fucose into glyco-proteins; *in vitro* growth defects; loss of colonization ability under competitive conditions	No dramatic phenotypic alterations	
Glycan assembly	Sequential addition of monosaccharides from nucleotide-activated precursors onto a lipid carrier by cytoplasmic glycosyltransferases			
Lipid carrier	Undecaprenylpyro-phosphate	Not determined	Undecaprenylpyrophosphate	
Membrane transport	Required for protein glycosylation			
Glycan transfer	*En bloc* transfer at the external face of the cytoplasmic membrane			

3.2 SURFACE (S-) LAYER PROTEIN GLYCOSYLATION

S-layers are common envelope components on prokaryotic cells with the unique feature of forming porous two-dimensional crystalline arrays on the supporting layer by an entropy-driven self-assembly process resulting in a complete coverage of the bacterial cells. High-resolution electron microscopy and scanning force microscopy studies revealed that S-layer lattices can have oblique (p1, p2), square (p4), or hexagonal (p3, p6) symmetry with morphological units composed of one, two, three, four, or six identical monomers, respectively (Figure 3). Since their first description in the early 1950s as regularly patterned monolayers (81, 82), an overwhelming body of knowledge on physicochemical and ultrastructural properties, functional aspects, genetics, biosynthesis, and application potentials has been accumulated and recently summarized (83).

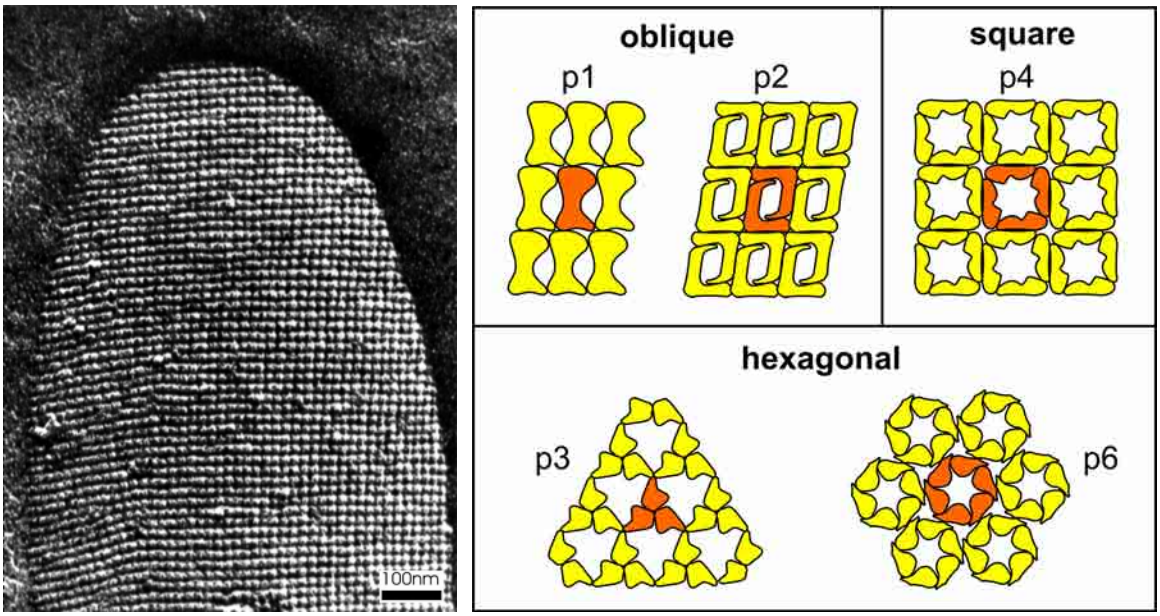


Figure 3: Electron micrograph of a freeze-etched and platinum/carbon-shadowed preparation of a Gram-positive organism exhibiting a square (p4) S-layer lattice (left). Schematic drawing showing different S-layer lattice types (right). In the oblique lattice, one morphological unit (red) consists of one (p1) or two (p2) identical subunits. Four subunits form one morphological unit in the square (p4) lattice type, whereas the hexagonal lattice type is either composed of three (p3) or six (p6) subunits.

More than 30 years ago, the S-layer proteins of the extreme halophilic archaeon *Halobacterium salinarum* and the two thermophilic bacteria *Thermoanaerobacter thermohydrosulfuricus* and *Thermoanaerobacterium thermosaccharolyticum* were the earliest examples of prokaryotic protein glycosylation (84, 85), and by now it is evident that glycosylation represents the major modification of S-layer proteins in archaea and bacteria. The S-layer glycan chains, which protrude from the cell surface up to 30-40 nm, create a polysaccharide coat similar to that of lipopoly-saccharide O-antigens from Gram-negative bacteria (19, 86). The long homo- or heteropolysaccharides representing bacterial S-layer glycans vary in chain length, resulting from different degrees of polymerization of identical repeating units (20-50 repeats), and are predominantly linked *O*-glycosidically via tyrosine, serine or threonine to two to six different glycosylation sites on the S-layer polypeptide (87).

Archaeal S-layer glycans consist almost exclusively of short heteropolysaccharides with up to ten sugar residues usually lacking any repeating unit or adaptor region. *N*-glycosidic linkages prevail in archaea, whereas short *O*-glycans, linked to threonine residues have also been identified on identical protein subunits, resulting overall in three to 25 glycan attachment sites (88). These sites are usually in accordance with

the tripeptide sequon N-X-S/T, where X is any amino acid except proline (89). However, replacement of the serine residue by valine, leucine or asparagine does not prevent *N*-glycosylation on a specific site of the S-layer glycoprotein from *Hb. salinarum* (90). In Table 2, the features of bacterial and archaeal S-layer glycoproteins are compared.

Table 2: Comparison between bacterial and archaeal S-layer glycoproteins (88).

S-layer glycan	Bacteria	Archaea
Type of glycan	*O*-glycan	*O*-glycan and/or *N*-glycan
Different glycans per S-layer protomer	1	1-3
Composition of glycan	Long homo- or heteropoly-saccharides (20–50 repeats)	Short heteropolysaccharides (up to 10 sugars)
Repeating unit	Usually 2-6 sugars	Not identified
Adaptor	Usually 1-3 sugars	Not identified
Attachment sites on protein	1-4	3-25

3.2.1 ARCHAEAL S-LAYER GLYCOSYLATION

Representing the main, in some cases sole, component of the layer surrounding archaeal cells, S-layer glycoproteins remain among the best-characterized archaeal glycoproteins (91). Although most archaeal S-layer proteins appear to be glycosylated, experimental evidence for this posttranslational modification is limited to some archaeal S-layer glycoproteins (83). In the following, the halobacterial and methanogenic S-layer *N*- and *O*-glycosylation is briefly summarized.

3.2.1.1 HALOBACTERIA

A hexagonally arranged S-layer composed of extremely acidic glycoprotein subunits is the sole cell wall component of *Hb. salinarum*, living in saturated sodium chloride solutions (92). The S-layer gene *csg* codes for a protein of 852 amino acids, including a leader sequence of 34 amino acids. Except a carboxyterminal region of 21 hydrophobic amino acids acting as a membrane anchor, the complete polypeptide chain mainly consists of polar and negatively charged amino acids (93). Each mature glycoprotein subunit has an apparent molecular mass of ~200 kDa with a carbo-

hydrate content of 10-12% (w/w) attached to the polypeptide via *N*- and *O*-glycosidic linkages (84). A sulfated high-molecular-mass saccharide composed of ten to 20 pentasaccharide repeating units is linked directly to a single asparagine residue within the typical acceptor sequence N-X-S/T via an *N*-acetylgalactosamine residue at the reducing end (94-96). Additionally, smaller sulfated oligosaccharides consisting of two to three glucuronic acids are *N*-glycosidically linked via glucose to ten different asparagine residues of the S-layer glycoprotein (94, 95, 97). The same species of sulfated oligosaccharides and identical linkage units have been detected in the glycosylated flagellins of this archaeon (98). Finally, neutral disaccharides of galactose and glucose are *O*-glycosidically attached to ~20 consecutive threonine residues located in close proximity to the carboxyterminal membrane spanning domain (94). This region, together with the cluster of threonine residues and the composition of the attached neutral disaccharides were also identified in the moderate halophile *Haloferax volcanii* (99). However, the composition of the *N*-glycosidically linked glycan chains is completely different compared to the extreme halophile. The sulfated amino sugars and glucuronic acids of *Hb. salinarum* are replaced by a pentasaccharide comprising two hexoses, two hexuronic acids, and a 190 Da species resulting in less negatively charged oligosaccharides in *Hf. volcanii* (100). Thus, the net surface charge of both organisms differs drastically, reflecting a possible adaptation from a moderately to an extremely halophilic environment (101).

3.2.1.2 METHANOGENS

Besides *Hb. salinarum* and *Hf. volcanii* other archaea such as the methanogens *Methanothermus fervidus* (102-104) and *Methanococcus voltae* (105, 106) were investigated in greater detail.

From the S-layer glycoprotein of the former organism, a heterosaccharide with the structure α-D-3-*O*-MeMan*p*-(1→6)-α-D-3-*O*-MeMan*p*-((1→2)-α-D-Man*p*)$_3$-(1→4)-D-GalNAc, which is *N*-linked to an asparagine residue of the peptide backbone, was isolated (104).

In the latter organism, a trisaccharide composed of 6-*O*-threonyl-β-Man*p*NAcA-(1→4)-β-Glc*p*NAc3NAcA-(1→3)-β-Glc*p*NAc linked to asparagine of the *N*-linked sequons N-X-S/T was observed to decorate the four flagellin proteins as well as the S-layer glycoprotein pointing to a common *N*-glycosylation mechanism in these organisms (105).

3.2.1.3 BIOSYNTHETIC ASPECTS OF ARCHAEAL S-LAYER GLYCOSYLATION

The identification of several *a*rchaeal *gl*ycosylation (*agl*) genes, encoding enzymes involved in the assembly of *N*-linked glycans in *Mc. voltae* and *Hf. volcanii* and their functional characterization by combined gene deletion and mass spectrometry approaches led to substantial progress in revealing the glycan biosynthesis during archaeal *N*-glycosylation (Figure 4) (106, 107).

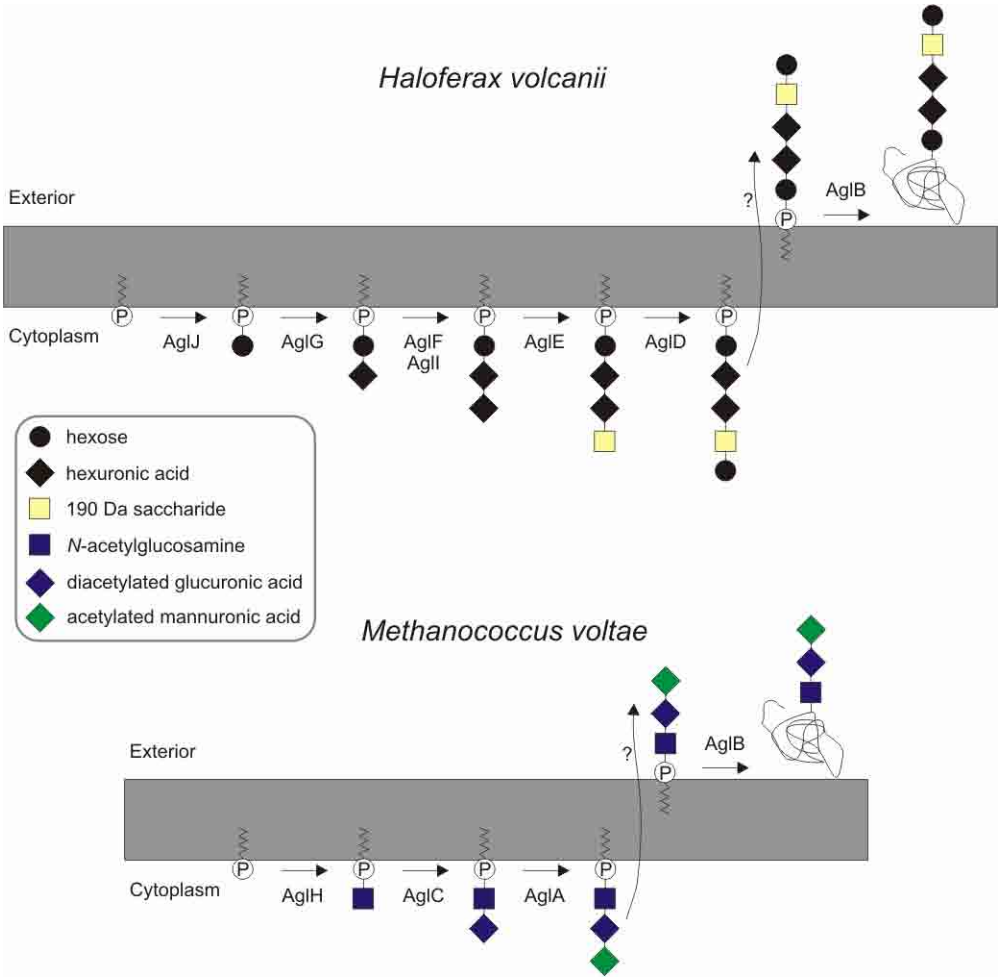


Figure 4: Schematic drawing of the archaeal *N*-glycosylation pathways observed in *Hf. volcanii* and *Mc. voltae* (108). In both organisms, monosaccharides are sequentially added to a dolichyl phosphate carrier on the cytoplasmic side of the membrane. After reorientation of the lipid-linked oligosaccharide to face the cell exterior, the archaeal oligosaccharyltransferase, AglB, transfers the polysaccharide to selected sequons of target proteins.

The *N*-linked trisaccharide of *Mc. voltae* is stepwise assembled in the cytoplasm by the actions of AglH, AglC, AglA, and AglK, whereas AglH is involved in the initial addition of the asparagine-bound linking sugar, *N*-acetylglucosamine to the dolichyl phosphate carrier, and AglA is responsible for the terminal sugar, a modified

mannuronic acid containing a covalently attached threonine residue (106). AglC and AglK are involved in the biosynthesis or transfer of the second sugar (2,3-diacetamido-2,3-dideoxy-β-glucuronic acid) within the glycan structure (109).

For *Hf. volcanii*, the glycosyltransferase AglJ was identified as a candidate for the initial transfer of the first hexose to the dolichyl phosphate carrier (110). However, AglD, AglE, AglF, AglG, and AglI participate in the cytoplasmic addition of the second to fifth sugar subunits of the pentasaccharide. AglD plays a role in the addition of the final hexose of the pentasaccharide, while AglE is involved in adding the 190 Da saccharide at the fourth position of the glycan (100, 111). AglF, AglG and AglI participate in transferring the hexuronic acids found at positions two and three of the pentasaccharide (112).

After the complete assembly, the archaeal lipid-linked glycans are reoriented to the periplasm, although no transporter responsible for translocating lipid-linked oligosaccharides across the plasma membrane has been identified so far (108). For both organisms, *Mc. voltae* and *Hf. volcanii*, an oligosaccharyltransferase AglB catalyzing the exterior *en bloc* transfer of an assembled glycan from its lipid carrier dolichyl phosphate onto asparagine residues of a nascent polypeptide chain found as part of the sufficient N-X-S/T consensus sequence was detected (89, 100, 106).

Since the reannotation of all genes located in the *agl* cluster responsible for *N*-glycosylation in *Hf. volcanii* was finished just recently, further analysis of the precise roles of each corresponding gene product is the main focus of future efforts (113).

3.2.2 Bacterial S-layer glycosylation

As mentioned above, the first convincing evidence for bacterial glycosylation was provided for S-layer glycoproteins of the thermophilics *Th. thermohydrosulfuricus* and *Ts. thermosaccharolyticum* (85) and since then, a considerable body of knowledge on structure, biosynthesis and molecular biology of S-layer glycoproteins has been accumulated and reviewed comprehensively (22). In addition to several members of the Gram-positive *Bacillaceae* including the species *Geobacillus stearothermophilus*, *G. tepidamans*, *P. alvei*, *Aneurinibacillus thermoaerophilus*, also Gram-negative species such as the oral anaerobe *Tannerella forsythia* (114, 115) and the enteric anaerobe *Bacteroides distasonis* (39) have been reported to possess glycosylated S-layers.

3.2.2.1 *GEOBACILLUS STEAROTHERMOPHILUS*

Currently, the moderately thermophilic *G. stearothermophilus* NRS 2004/3a is the best investigated model organism concerning bacterial S-layer glycosylation (116). An S-layer with an oblique lattice symmetry composed of glycoprotein subunits forms the outermost component of its cell wall (117, 118). The S-layer glycan is a poly-L-rhamnan chain which is *O*-glycosidically linked via galactose to a serine and two threonine residues of the S-layer protein SgsE (Figure 5) (119, 120).

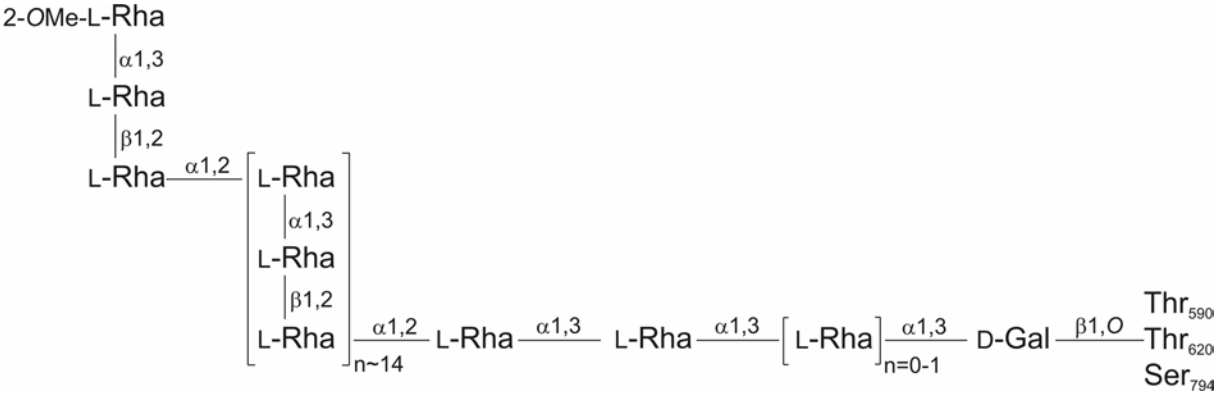


Figure 5: Structure of the S-layer glycan of *G. stearothermophilus* NRS 2004/3a (119). Rha, rhamnose; *O*Me, *O*-methyl; Thr, threonine; Ser, serine; n, number of repeats.

3.2.2.2 BIOSYNTHETIC ASPECTS OF BACTERIAL S-LAYER GLYCOSYLATION

For several S-layer glycoprotein possessing bacteria, polycistronic *S-l*ayer *g*lycosylation (*slg*) gene clusters with a size of ~16 to ~25 kb have been identified and sequenced. They include nucleotide sugar pathway genes that are arranged consecutively, glycosyltransferase genes, glycan processing genes, and transporter genes, all of them exhibiting high homology with components involved in the biosynthesis of different bacterial surface polysaccharides. As the G + C content at the *slg* locus is decreased in comparison to the respective bacterial genome, it is assumed that the investigated organisms have acquired their specific S-layer glycosylation potential by lateral gene transfer (121).

In *G. stearothermophilus* NRS 2004/3a, the ~17 kb *slg* gene clusters has been identified immediately downstream of the S-layer structural gene *sgsE*, and it encodes enzymes involved in dTDP-L-rhamnose and glycan biosynthesis, such as several glycosyltransferases, the components of the ABC transporter, and a putative oligosaccharyltransferase (Figure 6) (122).

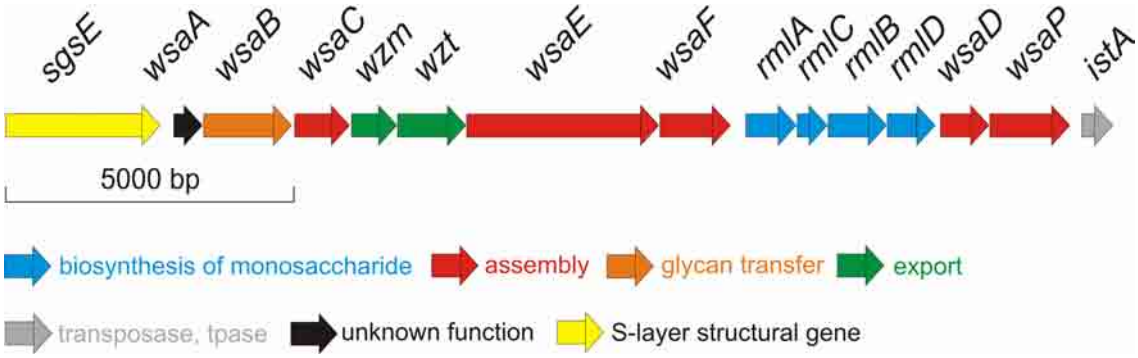


Figure 6: Genetic organization of the chromosomal *slg* gene of *G. stearothermophilus* NRS 2004/3a (122). Gene nomenclature: *sgsE*, the S-layer structural gene; *wsaA*, a gene of unknown function; *wsaB*, oligosaccharyltransferase; *wsaC*, rhamnosyltransferase; *wzm*, ABC transporter integral membrane protein; *wzt*, ATP-binding cassette (ABC) transporter nucleotide-binding protein; *wsaE*, trifunctional methyl-rhamnosyltransferase; *wsaF*, rhamnosyltransferase; *rmlA*, glucose-1-phosphate thymidylyltransferase; *rmlC*, dTDP-dehydrorhamnose 3,5-epimerase; *rmlB*, dTDP-D-glucose 4,6-dehydratase; *rmlD*, dTDP-dehydrorhamnose reductase; *wsaD*, rhamnosyltransferase; *wsaP*, UDP-galactose-lipid carrier transferase; *istA*, transposase .

The existence of a predicted ABC-2-type transporter system and the lack of a putative polymerase in this *slg* gene cluster indicate that the S-layer glycan chains are synthesized in a process comparable to the ABC transporter-dependent pathway of LPS *O*-polysaccharide biosynthesis (17). Based on homologies with proteins involved in the biosynthesis of different bacterial surface polysaccharides and on extensive *in vitro* characterization of selected glycosyltransferases, a model for S-layer glycosylation is proposed (Figure 7). The gene products from the *slg* gene cluster stepwise assemble a poly-L-rhamnan chain on the membrane-bound lipid carrier undecaprenol-pyrophosphate in the cytoplasm. Thereby WsaP, a homologue of WbaP responsible for initiation of capsular polysaccharides in *E. coli* (123), serves as the initiation enzyme catalyzing the transfer of galactose from its nucleotide-activated form (UDP-Gal) to a membrane-associated lipid carrier (124). The first adaptor rhamnose is α1,3-linked from dTDP-β-L-rhamnose to the lipid-bound galactose by WsaD. WsaC completes the adaptor by the transfer of one or two additional α1,3-linked rhamnoses to the first α1,3-linked rhamnose. Chain extension continues by the action of the two rhamnosyltransferases WsaE and WsaF, whereby WsaE catalyses the formation of the α1,2- and the α1,3-linkage, and WsaF is a β1,2-rhamnosyltransferase. The multifunctional enzyme WsfE is also responsible for S-layer glycan chain length termination by 2-*O*-methylation of terminal α1,3-linked

rhamnoses. Subsequently, the methylated glycan chain is recognized and transported across the cytoplasmic membrane by a process involving an ABC transporter and finally transferred to the appropriate serine or threonine residues of the target protein SgsE by an oligosaccharyltransferase WsaB (125). Similar reactions are described for PglL and PilO involved in protein *O*-glycosylation in *Neisseria meningitidis* and *Pseudomonas aeruginosa*, respectively (79), and for PglB involved in *N*-glycosylation in *C. jejuni* (126).

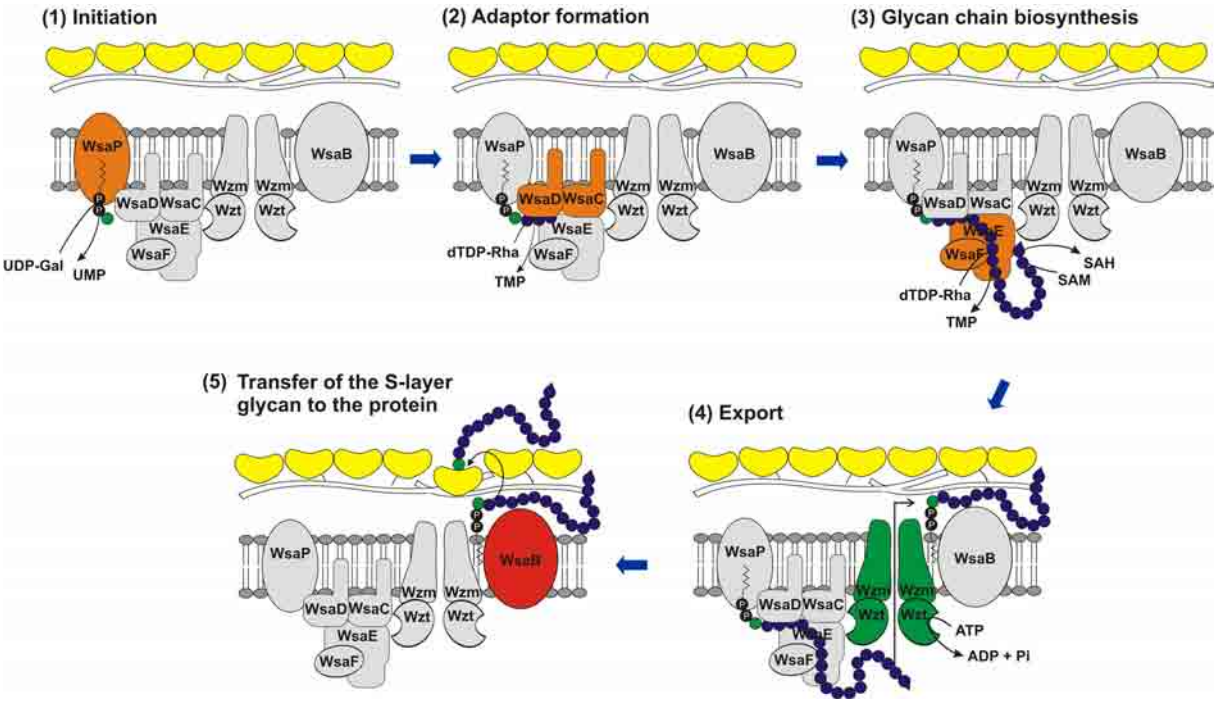


Figure 7: Proposed model of S-layer glycoprotein glycan biosynthesis in *Geobacillus stearothermophilus* NRS 2004/3a (125). The glycan biosynthesis is initiated by the transfer of a galactose residue from UDP-α-D-Gal to a lipid carrier catalyzed by WsaP (1, Initiation). The adaptor region is formed by the α1,3-linkage of a rhamnose residue from dTDP-β-L-Rha by WsaD, followed by the transfer of one or two additional α1,3-linked rhamnoses by WsaC (2, Adaptor formation). The repeating units are built up by the rhamnosyltransferases WsaE and WsaF, whereby WsaE is forming the α1,2- and the α1,3-linkages, and WsaF is forming the α1,2-linkage. The termination of chain growth is achieved by 2-*O*-methylation of the terminal repeating unit catalyzed by the *O*-methyltransferase domain of WsaE (3, Glycan chain biosynthesis). After binding of the Wzt component of the Wzm/Wzt ABC transporter system to the 2-*O*-methylated glycan chain, it is exported across the membrane (4, Export). The final transfer of the completed S-layer glycan to the S-layer protein is catalyzed by the oligosaccharyltransferase WsaB (5, Transfer of the S-layer glycan to the protein).

3.2.2.3 LIMITATIONS OF THE *GEOBACILLUS STEAROTHERMOPHILUS* S-LAYER GLYCOSYLATION SYSTEM

The creation and *in vivo* display of tailored *neo*glycoproteins requires the ability to manipulate an S-layer glycoprotein possessing organism by genetic engineering. This includes the introduction of foreign DNA by transformation, the expression of heterologous genes and the deletion or disruption of desired genes. Several attempts to transform the model organism *G. stearothermophilus* NRS 2004/3a have been made in the past and all of them failed. Within this context, diverse transformation protocols such as electroporation under variable conditions (127-129), transformation of protoplasts (130, 131) and the induction of natural competence (132) have been tried. Also using a wide range of plasmids exhibiting different antibiotic resistance genes and origins of replication did not result in transformands (133-139). To exclude that the plasmid DNA is degraded by host-specific nucleases, it was isolated from strains with or without a restriction modification system or enzymatically methylated *in vitro* and subsequently used for transformation (140). But again, no transformands of *G. stearothermophilus* NRS 2004/3a were obtained. Since introduction of foreign DNA into other S-layer carrying bacteria and archaea, including *Thermus thermophilus* (141, 142), *Caulobacter crescentus* (143), *Bacillus anthracis* (144), *Tannerella forsythia* (115), *Methanococcus voltae* (145), and *Haloferax volcanii* (146), can be done without major difficulties, it seems that, although described in the literature (130, 147, 148), transformation of thermophilic bacilli is a strain-specific trait and is not inhibited by the S-layer (149, 150).

However, transformation of the S-layer glycoprotein possessing mesophilic bacterium *Paenibacillus alvei* CCM 2051^T could be achieved easily by electroporation, so this organism was chosen for further examination of S-layer glycan biosynthesis and for *in vivo* display of S-layer *neo*glycoproteins (151).

3.2.2.4 WITH *PAENIBACILLUS ALVEI* TOWARDS S-LAYER *NEO*GLYCOPROTEINS

P. alvei is a mesophilic, Gram-positive organism, which was isolated and described for the first time in 1885 by Cheshire and Cheyne during analyzing an infection of bee larvae with European foulbrood and hence named *Bacillus alvei* (152). After comparison of the 16S rRNA the organism was reassigned to the new genus of *Paenibacillus* and is thus related to the rhizobacterium *P. polymyxa* and to *P. larvae,* the causative agent of American Foulbrood in honey bee *Apis mellifera* (153).

Besides the detection of *P. alvei* in insects (154, 155) and soil (156), there is a single report, where the organism was the source of bacteremia in human causing fever, swelling of joints and anemia (157).

The cell surface of the type strain *P. alvei* CCM 2051^T (= ATCC 6344^T; DSM 29^T) is completely covered with an oblique S-layer lattice composed of an identical glycoprotein species. Various aspects of its crystalline S-layer including ultra-structural characterization (158), glycosylation analysis (50, 51) and glycan bio-synthesis (159) have been investigated so far. In this strain, the S-layer *O*-glycan is a polymeric branched heterosaccharide linked to specific tyrosine residues of the protein backbone (Figure 8) (50, 51). Since *O*-glycosylation of tyrosine residues is not widely spread in nature, this organism is an excellent candidate for detailed examination of the S-layer glycan biosynthesis and identification of the involved enzyme machinery.

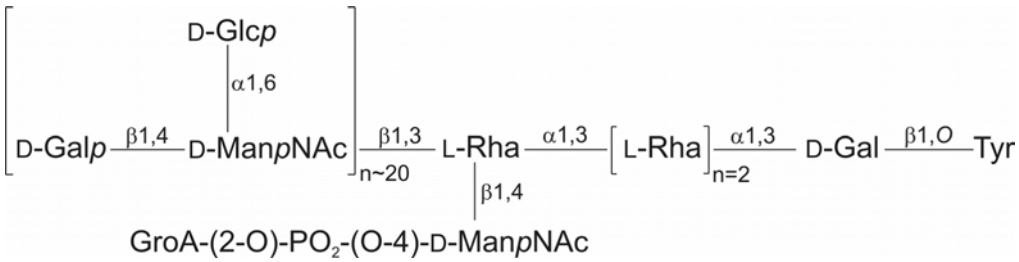


Figure 8: Structure of the S-layer glycan of *P. alvei* CCM 2051^T (50, 51). Glc, glucose; Gal, galactose; Rha, rhamnose; ManNAc, *N*-acetylmannosamine; Tyr, tyrosine; GroA, phosphoglyceric acid; n, number of repeats.

Once the S-layer glycoprotein biosynthesis in *P. alvei* CCM 2051^T is elucidated in detail, the protein portion as well as the glycan can be engineered to display tailor-made self-assembly S-layer glycoproteins at the bacterial cell surface or to re-crystallize these S-layer *neo*glycoproteins on various supports (Figure 9).

Alteration of the native S-layer glycan or assembly of completely new glycans on permissive sites of the S-layer protein portion can be achieved by deletion or replacement of endogenous enzymes involved in glycan biosynthesis and by heterologous expression of foreign glycosylation genes or complete gene cluster. The creation of a gene expression system and the construction of a gene knockout system for *P. alvei* CCM 2051^T based on bacterial mobile group II introns is reported in appendix 3 and 7. The identification of the *P. alvei* CCM 2051^T *slg* gene cluster and the proposal of a biosynthetic pathway for S-layer glycosylation enabled by a

glycan mutant approach using the newly developed gene knockout system are described in appendix 1 and 6. An introduction into mobile introns and their utilization as programmable gene-targeting vectors is given in chapter 3.4.

Manipulation of the S-layer protein portion includes the exchange or introduction of new glycosylation sites and the translational fusion and surface presentation of structural and / or functional peptide epitopes. A short summary about bacterial surface display is given in chapter 3.3., and the development of an S-layer based surface display system in *P. alvei* CCM 2051^{T} is described in appendix 2.

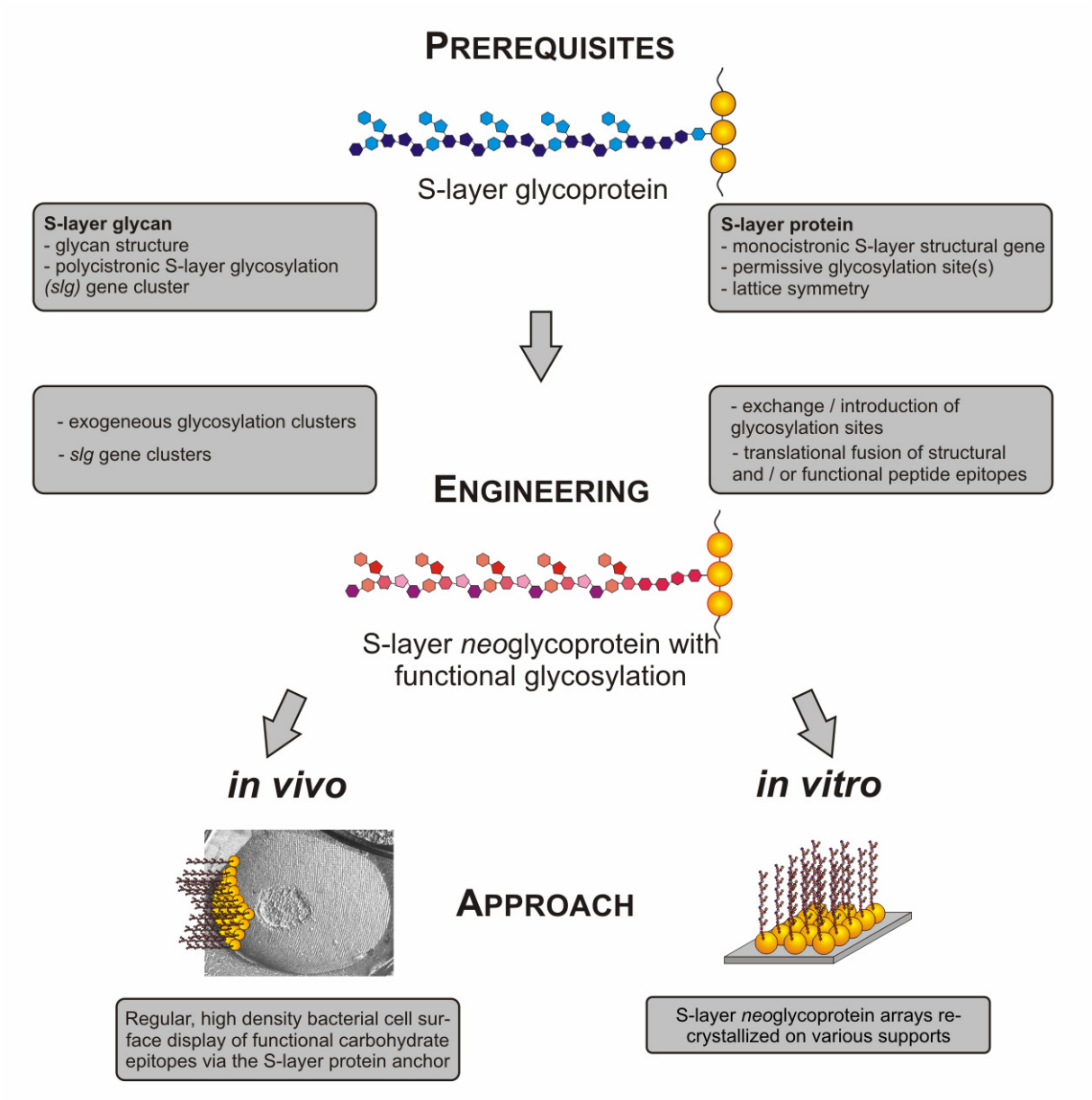


Figure 9: Strategies towards the production of S-layer *neo*glycoproteins (22). The first approach is the *in vivo* display of functional glycoproteins on the surface of bacteria enabled by means of recombinant DNA technology. The *in vitro* approach utilizes the recrystallization capability of the S-layer portion on a broad spectrum of supports.

3.3 BACTERIAL SURFACE DISPLAY

The presentation of heterologous proteins or peptide epitopes on the microbial cell surface by genetic engineering endows intact cells with new functionalities that have a myriad of biotechnological applications (Figure 10), including live vaccine development by presentation of epitopes on human commensal or attenuated pathogenic bacterial cells to trigger antigen-specific antibody responses (160), antibody production by expressing surface antigens for generation of polyclonal antibodies in animals (161), peptide library screening by sequential binding and elution or by fluorescence-activated cell sorting (162), whole-cell biocatalysis by immobilization of enzymes (163), biosensor development by anchoring enzymes, receptors or other signal-sensitive components (164, 165), and the removal of harmful chemicals and heavy metals (bioremediation) (166, 167). Both, bacteria and yeasts, have been investigated for this purpose. Here the cell surface display of bacteria using different carrier and passenger proteins is briefly summarized.

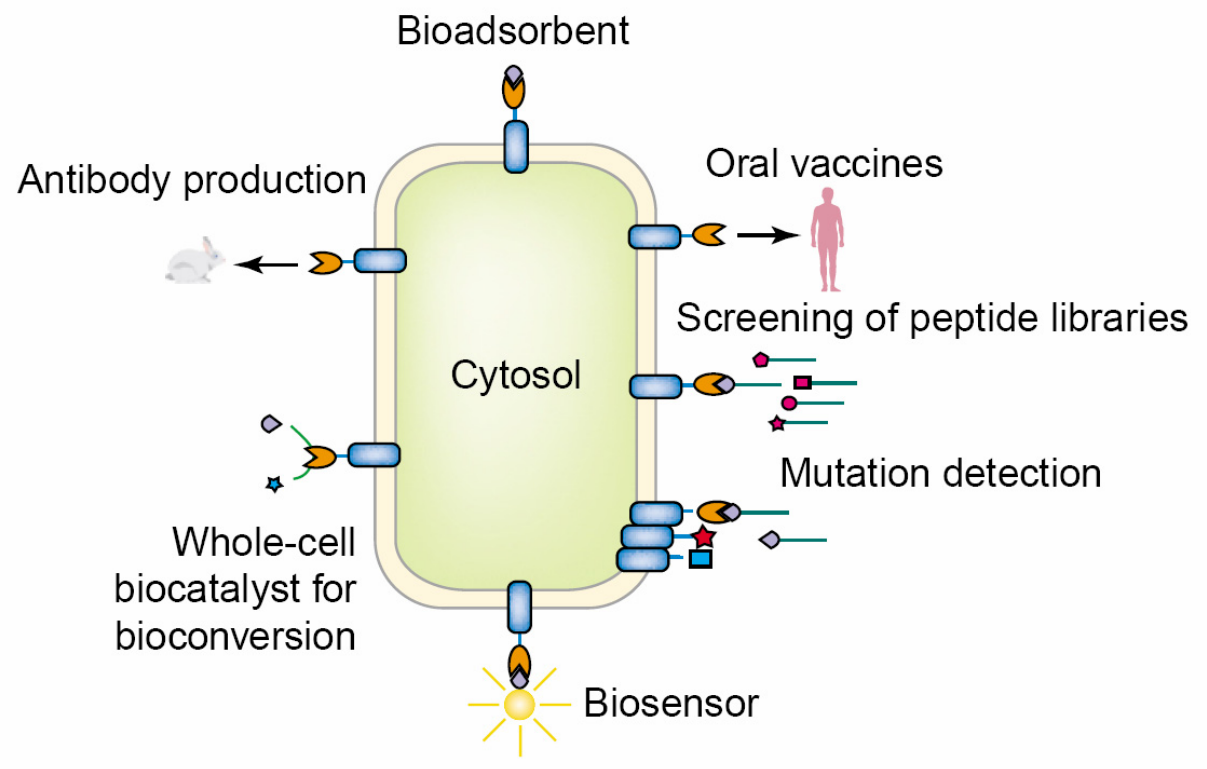


Figure 10: Biotechnological and industrial applications of microbial cell surface display (168). These systems have a versatile use as bioadsorbents, for the production of antibodies, as whole-cell biocatalysts, for biosensor development, for detection of single amino acid changes in peptides after random mutagenesis, for screening of peptide libraries, and for the development of live vaccines.

3.3.1 CARRIER PROTEINS

Carrier proteins are usually cell membrane-associated proteins used to attach functional target (passenger) proteins to the cell surface (166). An appropriate carrier should have an efficient signal peptide or transporting signal to allow the export of the fusion protein across the cell membrane and should offer a strong anchoring structure to avoid detachment of the fusion proteins from the cell surface. Furthermore, it should be amenable to insertion or fusion of heterologous sequences and it should be resistant to protease attack. Outer membrane proteins, lipoproteins, autotransporters, S-layer proteins, ice nucleation proteins, flagellae and fimbriae are the most popular carrier proteins of Gram-negative surface display systems, whereas for Gram-positive bacteria the staphylococcal protein A and different S-layer proteins have been utilized to display various polypeptides and proteins (168, 169).

Since S-layers provide a crystalline, regular immobilization matrix allowing the controlled and periodic surface display of functional domains at distinct positions with a high density while maintaining the intrinsic self-assembly property, these proteins are considered as optimal carriers for bacterial surface display (170, 171). Amongst others, the S-layer proteins RsaA of *Caulobacter crescentus* and EA1 / Sap of *Bacillus anthracis* have been used for anchoring heterologous target proteins on the cell surface of bacteria (170, 172-177). Besides the presentation of peptide epitopes, S-layer glycoproteins may be used for the *in vivo* display of functional glycosylation motifs enabled by means of recombinant DNA technology. These bioactive S-layer *neo*glycoproteins may be applicable in the areas of biomimetics, drug targeting, vaccine design, or diagnostics. However, for the alteration of the native S-layer glycan or assembly of completely new glycans, the detailed understanding of the S-layer protein glycosylation process is required (22, 171, 178).

3.3.2 PASSENGER PROTEINS

The passenger protein, which confers a selected functionality on the intact cell, is fused to or inserted into a carrier protein and is expressed on the cell exterior. The efficiency of the translocation process and final surface display is often significantly affected by the characteristics of the respective passenger protein, such as the folding structure, amino acid composition, size, and charge (168). Depending on the required application, a wide variety of passengers has been presented on different bacterial cell surfaces. Regarding S-layer proteins as carriers, several peptide

epitopes and enzymes have been carboxyterminally fused or inserted. This includes the repetitive insertion of a protein G IgG-binding domain into certain sites of full length RsaA, resulting in a functional, immunoreactive surface display at very high density (179). By translational fusion to the three S-layer homology (SLH) domains of EA1 and Sap, the active levansucrase of *Bacillus subtilis* and the immunogenic tetanus toxin fragment C of *Clostridium tetani* were targeted to the cell surface of *Bacillus anthracis*, retaining their enzymatic and antigenic properties (176, 177). Furthermore, the S-layer protein CTC of *Bacillus thuringiensis* was used for displaying *Mycoplasma gallisepticum* agglutinin and the avian influenza virus nucleoprotein on the cell surface of *B. thuringiensis*, respectively (180, 181). In both cases, agglutination assays showed the successful presentation of the recombinant proteins on the surface of vegetative cells. Additionally, a humoral response was triggered after oral immunization of chickens with heat-stable spores displaying the recombinant proteins, which demonstrates the possibility of developing oral vaccines using the S-layer surface display system.

3.4 MOBILE INTRONS FOR BACTERIAL GENE KNOCKOUT

The availability of complete genome sequences has revealed a very large number of genes with unknown function. Despite the existence of extensive databases for gene annotation by homology to known genes, the most straightforward approach to assign the function of a given gene is to create a targeted disruption of that gene and to characterize the phenotype of the mutant. Random mutations can be created by chemical or transposon mutagenesis, whereas specific mutations are usually generated by homologous recombination using double cross-over integration of heterologous non-replicating or conditionally replicating vectors in the genome. A different strategy is the targeted disruption of bacterial genes with mobile group II introns (targetrons), which was applied to inactivate certain genes of the *P. alvei* CCM 2051^{T} *slg* gene cluster.

3.4.1 MOBILE GROUP I AND GROUP II INTRONS

Catalytic RNA or ribozyme capable of self-splicing were discovered almost 30 years ago in the ciliate protozoa *Tetrahymena thermophila* (182, 183). In addition to the ability to excise themselves out of RNA transcripts and ligate their flanking RNA

sequences, these genetic elements are also mobile, coding for proteins which allow them to invade genomic sequences (184, 185). Group I and group II introns can either spread into homologous intron-less DNA sites (homing) or insert into novel genomic locations (transposition). Possessing a widespread occurrence, these elements have been identified in all domains of life. Group I introns are present in bacteria, bacteriophages, and in organellar and nuclear genomes of eukaryotes, whereas group II introns are found in bacteria, archaea and eukaryotic organelles (186-188). The molecular details of the RNA catalysis and insertion events of these introns have been intensively investigated and reviewed recently (189, 190). Group I and group II introns vary in sequence, structure, as well as in splicing and mobility mechanisms (191).

3.4.1.1 MOBILE GROUP I INTRONS

Group I introns, which have been found to interrupt genes for rRNA, tRNA and proteins, are variably distributed in the nuclei of protozoa, the mitochondria of fungi, the chloroplasts of algae, and in bacteria and their phages. Besides their common core secondary structure made up of usually nine or ten base-paired elements (P1-P10) that are organized into three domains at the tertiary structure level, they show a common mechanism of self-splicing (Figure 11).

Group I introns splice by a Mg^{2+}-dependent mechanism involving two trans-esterification reactions mediated by an exogenous guanosine cofactor. The 5' splice site is defined by the formation of the P1 helix by internal base pairing between an internal guide sequence (IGS) and the 5' exon. The splicing starts with the binding of guanosine to the cofactor pocket (G binding site) within P7 and the nucleophilic attack of guanosine at the 5' splice site. Cleavage of the 5' splice-site phosphate and covalent attachment of the guanosine to the 5' end of the intron leads to a conformational change within the RNA molecule and the release of the upstream 5' exon. The cleaved 5' exon, which is still connected to the intron by base pairing, subsequently cleaves the 3' splice site. Finally, the exons are ligated and the intron is excised (192, 193). Some group I introns may contain homing endonuclease genes coding for highly site-specific homing endonucleases, which are responsible for intron mobility and intron insertion into cognate target sites. However, intron movement of free intron RNAs directly into ectopic sites is achieved by reverse splicing (194).

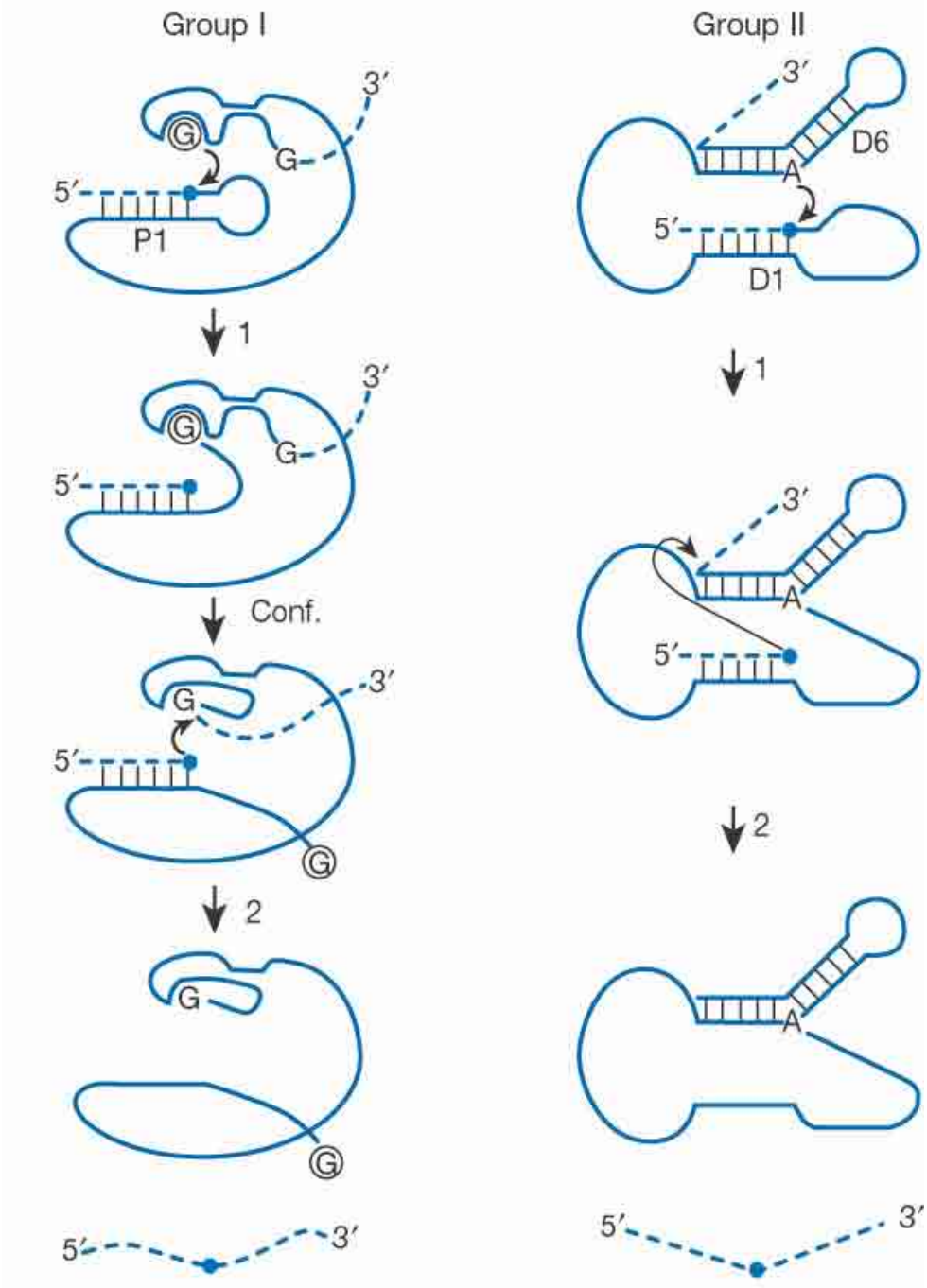


Figure 11: Pathways for group I and II intron self-splicing, with exons shown as dashed lines and introns as solid lines (190). For group I introns, an intron-bound guanosine (G) cleaves the 5' splice site while becoming covalently attached to the 5' end of the intron (Step 1). 'Conf.' indicates a conformational change whereby the G at the 3' end of the intron replaces the original G in the G-binding site. The cleaved 5' exon still held to the intron by base pairing (P1), cleaves the 3' splice site. As a result, the exons are ligated and the intron excised (Step 2).

For group II introns, an adenosine (A) within domain D6 attacks the 5' splice site, which is identified by base-pairing interactions involving domain D1 (Step 1). This results in a branched lariat RNA intermediate. The cleaved 5' exon then attacks the 3' splice site, ligating the exons and excising the lariat intron (Step 2).

3.4.1.2 MOBILE GROUP II INTRONS

Group II introns were discovered in mitochondrial and chloroplast genes of eukaryotic cells, later also in both Gram-negative and Gram-positive bacteria and have recently been found in archaea (195-197).

Like group I introns, group II introns have little sequence similarity, but share a conserved secondary structure required for catalytic activity (Figure 12). This structure, which forms an active site containing catalytically essential Mg^{2+} ions, consists of six double-helical domains (DI-DVI) emanating from a central wheel. Domain DI contains sequences involved in aligning the splice sites, including the exon binding sites (EBS1, EBS2, and δ). Other sequences in domain DI interact with the most highly conserved domain DV to form the catalytic core (198). Domain DIV, which is dispensable for ribozyme activity, may contain an intron ORF without disruption of the conserved structure. This intron encoded protein (IEP) assists RNA splicing and mediates efficient integration of the intron RNA into double-stranded DNA by reverse splicing and reverse transcription. The IEP consists of four conserved domains, an aminoterminal reverse transcriptase domain, a RNA-binding domain associated with RNA splicing or maturase activity, a carboxyterminal DNA-binding domain and a DNA-endonuclease domain Zn finger-like motif (188, 199).

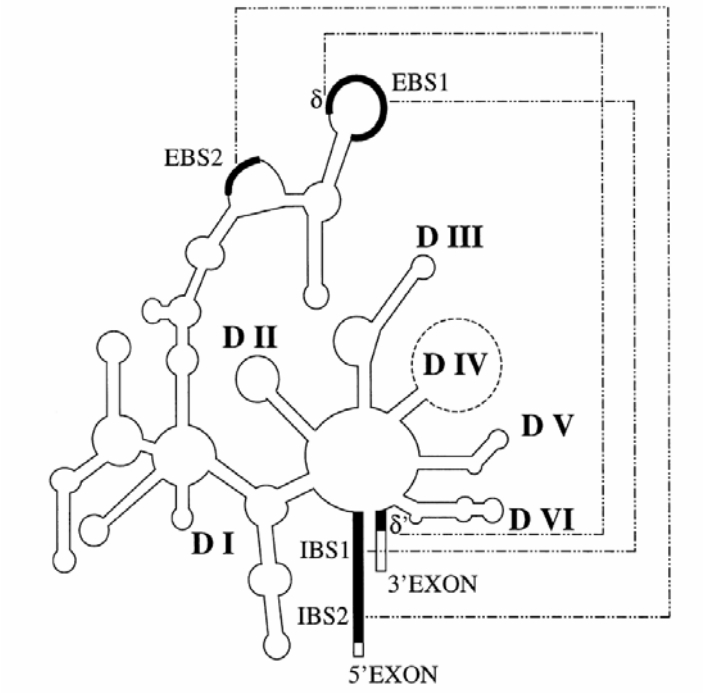


Figure 12: Secondary structure model of mobile group II introns showing the location of the EBS1, EBS2, and δ sequences in domain DI of the intron RNA and their basepairing with IBS1, IBS2, and δ' sequences in the 5' and 3' exons (200).

In a mechanism distinct from that of the group I introns, group II intron self-splicing involves two sequential transesterification reactions and does not require exogenous energy (Figure 11). The first reaction is triggered by the nucleophilic attack on the 5' splice junction by the 2'-OH of a bulged nucleotide, usually an adenosine residue termed bulging A, located in DVI domain near the 3' end of the intron, releasing the 5' exon and generating a lariat intermediate. The second reaction involves a nucleophilic attack on the 3' splice site by the liberated 3'-OH of the last nucleotide of the 5'-exon, resulting in ligated exons and an excised intron lariat with a 2'-5' phosphodiester bond (201). As in group I introns, the definition and binding of the 5' splice junction in group II introns depends on interaction between the intron and sequences in the 5' exon. The sequence elements EBS1 and EBS2 in domain DI each form five to six base pairs with the 5'-exon sequences IBS1 and IBS2 (intron-binding sites 1 and 2). Additional base-pairing interactions involving adjacent sequences (δ and EBS 3) also contribute to positioning of the 5' splice site (202-204).
Like group I introns, group II introns are mobile genetic elements, capable of inserting at intron-specific locations in intronless alleles (retrohoming) or at novel (ectopic) sites at low frequency (transposition). For the process of retrohoming, different mechanisms varying between group II introns originating from yeast mitochondria and bacteria are known and summarized elsewhere (188).
The major retrohoming mechanism used by the *Lactococcus lactis* Ll.LtrB intron is shown in Figure 13. After splicing, the IEP remains tight and specifically bound to the excised intron lariat RNA, forming a ribonucleoprotein particle (RNP) that initially binds DNA nonspecifically and then searches for DNA target sites. The RNPs of group II introns recognize DNA target sequences by using both the IEP and base pairing of the intron RNA. After recognition of a small number of nucleotide residues in the double-stranded DNA by the IEP, which causes partial DNA unwinding, the EBS1, EBS2, and δ sequences of the intron RNA interact with the 5'- and 3'-exon sequences IBS1, IBS2, and δ' (205, 206). Once the DNA target site is identified by the RNP, the RNA component cleaves the sense-strand and integrates at the target site, whereas the IEP uses its endonuclease activity to cleave the antisense-strand. The 3' end of the cleaved antisense strand is used as a primer for reverse transcription of the inserted intron RNA by the reverse transcriptase activity of the IEP, generating a cDNA copy of the intron that is subsequently integrated into its new location by RecA-independent repair mechanisms (207).

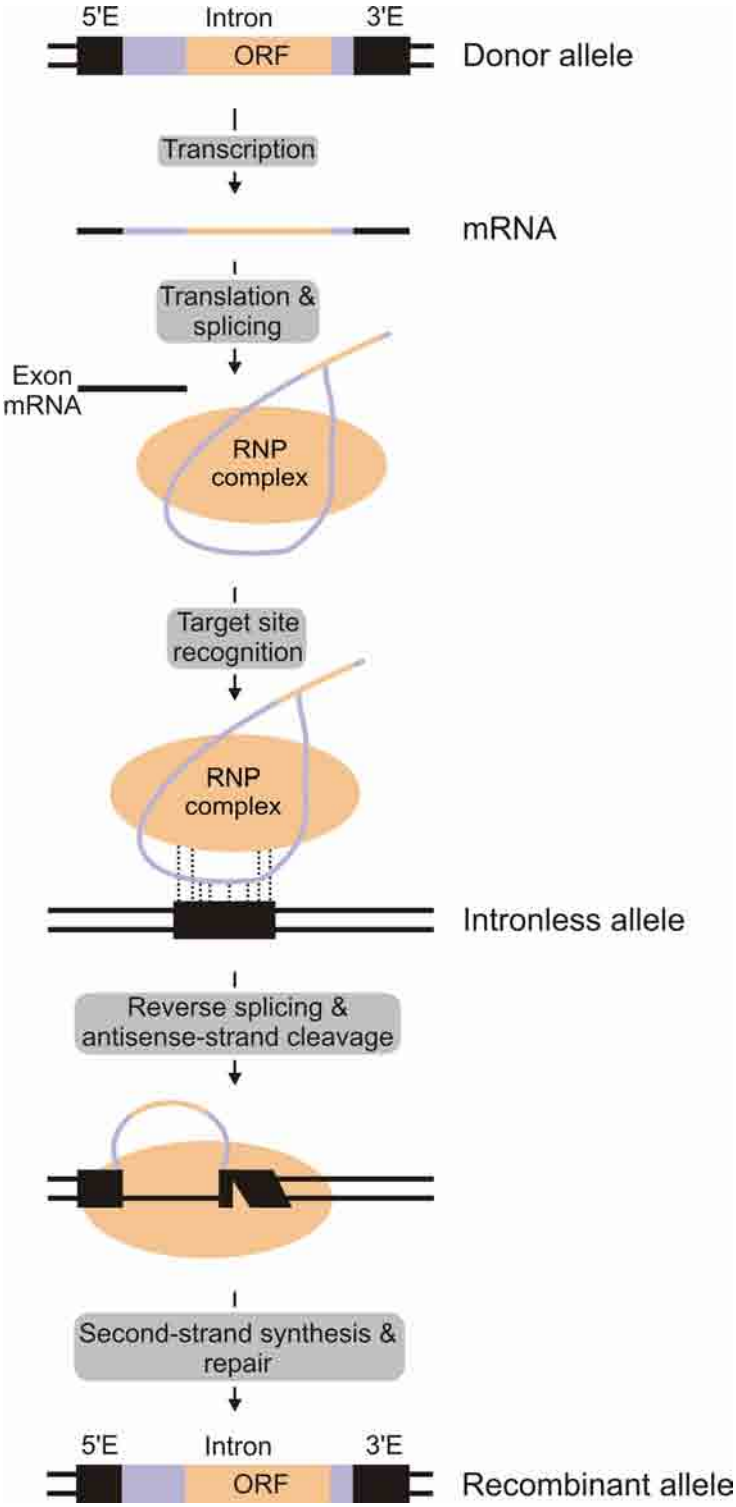


Figure 13: Retrohoming pathway of bacterial group II introns (199, 200). *L. lactis* Ll.LtrB intron retrohoming is initiated by the RNP complex containing the intron-encoded protein and the intron lariat RNA. After translation, the IEP assists in RNA splicing. It then remains tightly bound to the intron RNA to form the RNP complex. The IEP recognizes a small number of nucleotide residues of the DNA target site in the intronless allele. This leads to DNA unwinding, enabling the intron RNA to basepair with the IBS and δ' sequences for reverse splicing at the intron insertion site. After antisense-strand cleavage by the IEP the 3' end of the cleaved antisense strand is used as a primer for reverse transcription of the inserted intron RNA. Finally, this cDNA is integrated by a RecA-independent repair mechanism, resulting in the recombinant allele.

In addition to retrohoming, group II introns retrotranspose at low frequency to ectopic sites, favorably into transmissible genetic elements, such as plasmids, leading to intron spread within and between species (208, 209).

3.4.1.3 GROUP II INTRONS AS GENE-TARGETING VECTORS

The recognition of the DNA target sites mainly by base pairing of the intron RNA, the possible mobilization of foreign genetic information inserted within the intron, the requirements of minimal host functions combined with their very high insertion frequency and specificity by a homologous recombination-independent process make group II introns suitable for use as biotechnological tools (Figure 14). The *L. lactis* LI.LtrB intron has been engineered to increase the frequency and efficiency of mobility and used for efficient targeted gene disruption in pro- and eukaryotic cells (210-214). By removing most of the LtrA ORF from the domain DIV loop and cloning of the intact LtrA ORF downstream from the 3' exon, the frequency of mobility was increased to almost 100%, being probably a result of greater stability of the intron RNA (212). It is currently possible to reprogram group II introns to insert into any desired target sequence. Retargeting of group II introns, which are generally expressed from a donor plasmid, is performed by a computer algorithm, scanning the target sequence, selecting the optimal positions for IEP recognition and designing primers to modify the intron sequences EBS1, EBS2, and δ as well as the complementary IBS1 and IBS2 sequences in the 5' exon (215). Selection of integrated introns into the target DNA is achieved by *in situ* PCR or by using selectable markers, such as antibiotic resistance genes. This heterologous DNA is inserted into the domain DIV of the group II intron after removal of the IEP sequence, which generally has a negative effect on homing, but the modified introns nonetheless display detectable homing (216, 217).

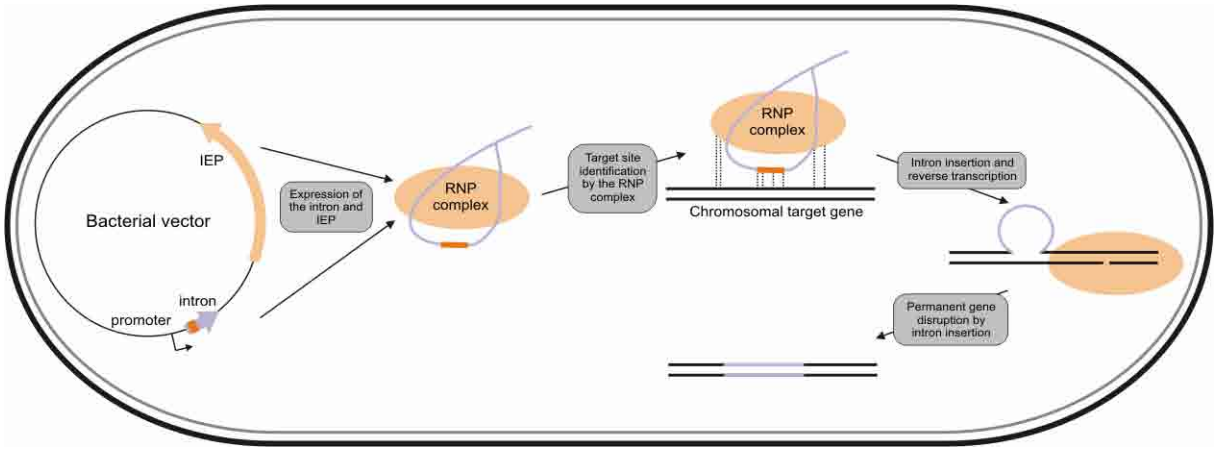


Figure 14: Use of targeted mobile group II introns (targetrons) for gene disruption (188). The DNA binding specificity of the LI.LtrB group II intron is determined primarily by the EBS regions encoded on the RNA portion of the ribonucleoprotein complex (RNP). This allows rapid mutation of the EBS regions to target specific chromosomal regions for insertional mutagenesis.

Mobile group II introns as programmable gene-targeting vectors termed TargeTron™ Gene Knockout System are available from Sigma-Aldrich. This commercial system was the basis for the construction of an effective gene knockout system for insertional inactivation of target genes located on the bacterial chromosome of *P. alvei* CCM 2051^{T}.

4. AIMS OF THE WORK

Substantial progress in revealing of the S-layer glycan biosynthesis pathway has been made in the Gram-positive, moderately thermophilic organism *Geobacillus stearothermophilus* NRS 2004/3a. However, a severe drawback of this organism is its resistance to take up foreign DNA, preventing the creation of mutant strains with abolished or modified glycosylation patterns and, thus, eliminating the possibility of an *in vivo* analysis of the glycosylation process. So far, unraveling of the glycan biosynthesis mechanism is limited to *in vitro* testing of individual enzymes and to heterologous carbohydrate-engineering approaches (124, 125, 218).

In order to establish a new model organism for genetic manipulation of bacterial S-layer protein glycosylation *in vivo*, the mesophilic bacterium *Paenibacillus alvei* CCM 2051^T turned out to be an optimal candidate for this work.

- One aim is the development of a reliable method to transform plasmid-DNA into this organism and the creation of a suitable vector for the expression of heterologous genes. For this reason, several protocols for different types of transformation, such as electroporation and natural competence, as well as various available plasmids will be investigated.
- The construction of an effective gene knockout system for insertional inactivation of desired target genes is the next aim of this study. For this purpose, the bacterial mobile group II intron Ll.LtrB of *L. lactis* together with a constitutive promoter will be cloned into a suitable plasmid replicating in *P. alvei* CCM 2051^T. Subsequently, the intron will be retargeted to insert into a gene of interest and the resulting mutant will be examined by SDS-PAGE followed by PAS staining and carbohydrate analysis.
- The most important aim of this work is the detection of a gene locus coding for the S-layer glycan biosynthesis (*slg* gene cluster) in *P. alvei* CCM 2051^T, accompanied by the assignment of the proper biological function of each encoded protein. Therefore, by successive chromosome walking the complete sequence of the *slg* gene cluster will be determined and putative genes will be assigned by protein database comparison. In order to determine the role of different enzymes in the process of S-layer glycosylation and to analyze the glycosylation process *in vivo*, corresponding genes will be disrupted by applying the newly-created gene knockout system.

- In addition to the identification of the structural gene encoding the S-layer protein SpaA of *P. alvei* CCM 2051^T, a further aim is its utilization for the presentation of chimeric glycoproteins on the cell surface representing an essential requirement for the future construction of bioactive S-layer *neo*glycoproteins. For this purpose, the easily detectable S-layer fusion proteins SpaA_6HIS and SpaA_eGFP will be constructed and heterologously expressed in *P. alvei* CCM 2051^T. Their glycosylation, subcellular localization, accessibility, and functionality will be investigated by immunoblot detection and immunofluorescence staining followed by fluorescence microscopical analysis.

5. REFERENCES

1. **Lehninger, A. L., D. L. Nelson, and M. M. Cox.** 1993. Principles of Biochemistry, vol. 2. Worth Publishers, Inc., New York.
2. **Kandler, O., and H. König.** 1998. Cell wall polymers in Archaea (Archaebacteria). Cellular and Molecular Life Sciences **54:**305-308.
3. **Sharon, N.** 1975. Complex carbohydrates: Their chemistry, biosynthesis and functions, Addison-Wesley Pub. Co., Reading, MA.
4. **van Heijenoort, J.** 2001. Formation of the glycan chains in the synthesis of bacterial peptidoglycan. Glycobiology **11:**25R-36R.
5. **James, D. W., J. Preiss, and A. D. Elbein.** 1985. Biosynthesis of polysaccharides, p. 107-207. *In* G. O. Aspinall (ed.), The polysaccharides, vol. 3. Academic Press, New York.
6. **Höltje, J. V.** 1998. Growth of the stress-bearing and shape-maintaining murein sacculus of *Escherichia coli*. Microbiology and Molecular Biology Reviews **62:**181-203.
7. **Kandler, O., and H. König.** 1993. Cell envelopes of archaea: structure and chemistry., p. 223–259. *In* M. Kates, D. J. Kushner, and A. T. Matheson (ed.), The Biochemistry of Archaea (Archaebacteria). Elsevier, Amsterdam.
8. **Archibald, A. R., J. Baddiley, and N. L. Blumsom.** 1968. The teichoic acids. Advances in Enzymology and Related Areas of Molecular Biology **30:**223-253.
9. **Heptinstall, S., A. R. Archibald, and J. Baddiley.** 1970. Teichoic acids and membrane function in bacteria. Nature **225:**519-521.
10. **Ward, J. B.** 1981. Teichoic and teichuronic acids: biosynthesis, assembly, and location. Microbiological Reviews **45:**211-243.
11. **Hancock, I. C., and J. Baddiley.** 1985. Biosynthesis of the bacterial envelope polymers teichoic acid and teichuronic acid, p. 279–307. *In* A. N. Martonosi (ed.), The Enzymes of Biological Membranes, 2 ed, vol. 2. Plenum, New York.
12. **Fischer, W.** 1994. Lipoteichoic acids and lipoglycans., p. 199–215. *In* J.-M. Ghuysen and R. Hakenbeck (ed.), Bacterial Cell Wall. Elsevier, Amsterdam.
13. **Ward, J. B., and C. A. Curtis.** 1982. The biosynthesis and linkage of teichuronic acid to peptidoglycan in *Bacillus licheniformis*. European Journal of Biochemistry **122:**125-132.

14. **Raetz, C. R. H.** 1990. Biochemistry of endotoxins. Annual Review of Biochemistry **59:**129-170.
15. **Raetz, C. R. H.** 1998. Enzymes of lipid A biosynthesis: targets for the design of new antibiotics. Progress in Clinical Biological Research **397:**1-14.
16. **Raetz, C. R. H., Z. Guan, B. O. Ingram, D. A. Six, F. Song, X. Wang, and J. Zhao.** 2009. Discovery of new biosynthetic pathways: the lipid A story. Journal of Lipid Research **50 Suppl:**S103-108.
17. **Raetz, C. R. H., and C. Whitfield.** 2002. Lipopolysaccharide endotoxins. Annual Review of Biochemistry **71:**635-700.
18. **Reeves, P. R., M. Hobbs, M. A. Valvano, M. Skurnik, C. Whitfield, D. Coplin, N. Kido, J. Klena, D. Maskell, C. R. H. Raetz, and P. D. Rick.** 1996. Bacterial polysaccharide synthesis and gene nomenclature. Trends in Microbiology **4:**495-503.
19. **Whitfield, C.** 1995. Biosynthesis of lipopolysaccharide O antigens. Trends in Microbiology **3:**178-185.
20. **Whitfield, C.** 2006. Biosynthesis and assembly of capsular polysaccharides in *Escherichia coli*. Annual Review of Biochemistry **75:**39-68.
21. **Messner, P.** 2009. Prokaryotic protein glycosylation is rapidly expanding from "curiosity" to "ubiquity". ChemBioChem, DOI:10.1002/cbic.200900388.
22. **Messner, P., K. Steiner, K. Zarschler, and C. Schäffer.** 2008. S-layer nanoglycobiology of bacteria. Carbohydrate Research **343:**1934-1951.
23. **Josenhans, C., L. Vossebein, S. Friedrich, and S. Suerbaum.** 2002. The *neuA*/*flmD* gene cluster of *Helicobacter pylori* is involved in flagellar biosynthesis and flagellin glycosylation. FEMS Microbiology Letters **210:**165-172.
24. **Schirm, M., E. C. Soo, A. J. Aubry, J. Austin, P. Thibault, and S. M. Logan.** 2003. Structural, genetic and functional characterization of the flagellin glycosylation process in *Helicobacter pylori*. Molecular Microbiology **48:**1579-1592.
25. **Thibault, P., S. M. Logan, J. F. Kelly, J. R. Brisson, C. P. Ewing, T. J. Trust, and P. Guerry.** 2001. Identification of the carbohydrate moieties and glycosylation motifs in *Campylobacter jejuni* flagellin. Journal of Biological Chemistry **276:**34862-34870.

26. **Guerry, P., P. Doig, R. A. Alm, D. H. Burr, N. Kinsella, and T. J. Trust.** 1996. Identification and characterization of genes required for post-translational modification of *Campylobacter coli* VC167 flagellin. Molecular Microbiology **19:**369-378.

27. **Doig, P., N. Kinsella, P. Guerry, and T. J. Trust.** 1996. Characterization of a posttranslational modification of *Campylobacter* flagellin: identification of a sero-specific glycosyl moiety. Molecular Microbiology **19:**379-387.

28. **Jennings, M. P., M. Virji, D. Evans, V. Foster, Y. N. Srikhanta, L. Steeghs, P. van der Ley, and E. R. Moxon.** 1998. Identification of a novel gene involved in pilin glycosylation in *Neisseria meningitidis*. Molecular Microbiology **29:**975-984.

29. **Marceau, M., K. Forest, J. L. Beretti, J. Tainer, and X. Nassif.** 1998. Consequences of the loss of *O*-linked glycosylation of meningococcal type IV pilin on piliation and pilus-mediated adhesion. Molecular Microbiology **27:**705-715.

30. **Power, P. M., L. F. Roddam, M. Dieckelmann, Y. N. Srikhanta, Y. C. Tan, A. W. Berrington, and M. P. Jennings.** 2000. Genetic characterization of pilin glycosylation in *Neisseria meningitidis*. Microbiology **146:**967-979.

31. **Virji, M., E. Stimson, K. Makepeace, A. Dell, H. R. Morris, G. Payne, J. R. Saunders, and E. R. Moxon.** 1996. Posttranslational modifications of meningococcal pili. Identification of a common trisaccharide substitution on variant pilins of strain C311. Annals of the New York Academy of Sciences **797:**53-64.

32. **Castric, P.** 1995. *pilO*, a gene required for glycosylation of *Pseudomonas aeruginosa* 1244 pilin. Microbiology **141:**1247-1254.

33. **Castric, P., F. J. Cassels, and R. W. Carlson.** 2001. Structural characterization of the *Pseudomonas aeruginosa* 1244 pilin glycan. Journal of Biological Chemistry **276:**26479-26485.

34. **DiGiandomenico, A., M. J. Matewish, A. Bisaillon, J. R. Stehle, J. S. Lam, and P. Castric.** 2002. Glycosylation of *Pseudomonas aeruginosa* 1244 pilin: glycan substrate specificity. Molecular Microbiology **46:**519-530.

35. **Benz, I., and M. A. Schmidt.** 2001. Glycosylation with heptose residues mediated by the *aah* gene product is essential for adherence of the AIDA-I adhesin. Molecular Microbiology **40:**1403-1413.

36. **Moormann, C., I. Benz, and M. A. Schmidt.** 2002. Functional substitution of the TibC protein of enterotoxigenic *Escherichia coli* strains for the autotransporter adhesin heptosyltransferase of the AIDA system. Infection and Immunity **70:**2264-2270.

37. **Grass, S., A. Z. Buscher, W. E. Swords, M. A. Apicella, S. J. Barenkamp, N. Ozchlewski, and J. W. St Geme, 3rd.** 2003. The *Haemophilus influenzae* HMW1 adhesin is glycosylated in a process that requires HMW1C and phosphoglucomutase, an enzyme involved in lipooligosaccharide biosynthesis. Molecular Microbiology **48:**737-751.

38. **Gross, J., S. Grass, A. E. Davis, P. Gilmore-Erdmann, R. R. Townsend, and J. W. St Geme, 3rd.** 2008. The *Haemophilus influenzae* HMW1 adhesin is a glycoprotein with an unusual *N*-linked carbohydrate modification. Journal of Biological Chemistry **283:**26010-26015.

39. **Fletcher, C. M., M. J. Coyne, D. L. Bentley, O. F. Villa, and L. E. Comstock.** 2007. Phase-variable expression of a family of glycoproteins imparts a dynamic surface to a symbiont in its human intestinal ecosystem. Proceedings of the National Academy of Sciences of the United States of America **104:**2413-2418.

40. **Fletcher, C. M., M. J. Coyne, O. F. Villa, M. Chatzidaki-Livanis, and L. E. Comstock.** 2009. A general *O*-glycosylation system important to the physiology of a major human intestinal symbiont. Cell **137:**321-331.

41. **Logan, S. M.** 2006. Flagellar glycosylation - a new component of the motility repertoire? Microbiology **152:**1249-1262.

42. **Power, P. M., and M. P. Jennings.** 2003. The genetics of glycosylation in Gram-negative bacteria. FEMS Microbiology Letters **218:**211-222.

43. **Schmidt, M. A., L. W. Riley, and I. Benz.** 2003. Sweet new world: glycoproteins in bacterial pathogens. Trends in Microbiology **11:**554-561.

44. **Abu-Qarn, M., J. Eichler, and N. Sharon.** 2008. Not just for Eukarya anymore: protein glycosylation in Bacteria and Archaea. Current Opinion in Structural Biology **18:**544-550.

45. **Szymanski, C. M., S. M. Logan, D. Linton, and B. W. Wren.** 2003. *Campylobacter* - a tale of two protein glycosylation systems. Trends in Microbiology **11:**233-238.

46. **Szymanski, C. M., and B. W. Wren.** 2005. Protein glycosylation in bacterial mucosal pathogens. Nature Reviews Microbiology **3:**225-237.

47. **Vik, A., F. E. Åas, J. H. Anonsen, S. Bilsborough, A. Schneider, W. Egge-Jacobsen, and M. Koomey.** 2009. Broad spectrum *O*-linked protein glycosylation in the human pathogen *Neisseria gonorrhoeae*. Proceedings of the National Academy of Sciences of the United States of America **106:**4447-4452.

48. **Ku, S. C., B. L. Schulz, P. M. Power, and M. P. Jennings.** 2009. The pilin *O*-glycosylation pathway of pathogenic *Neisseria* is a general system that glycosylates AniA, an outer membrane nitrite reductase. Biochemical and Biophysical Research Communications **378:**84-89.

49. **Messner, P., and C. Schäffer.** 2003. Prokaryotic glycoproteins, p. 51-124. *In* W. Herz, H. Falk, and G. W. Kirby (ed.), Progress in the Chemistry of Organic Natural Products, vol. 85. Springer, Wien.

50. **Altman, E., J. R. Brisson, P. Messner, and U. B. Sleytr.** 1991. Structure of the glycan chain from the surface layer glycoprotein of *Bacillus alvei* CCM 2051. Biochemistry and Cell Biology **69:**72-78.

51. **Messner, P., R. Christian, C. Neuninger, and G. Schulz.** 1995. Similarity of "core" structures in two different glycans of tyrosine-linked eubacterial S-layer glycoproteins. Journal of Bacteriology **177:**2188-2193.

52. **Christian, R., P. Messner, C. Weiner, U. B. Sleytr, and G. Schulz.** 1988. Structure of a glycan from the surface-layer glycoprotein of *Clostridium thermohydrosulfuricum* strain L111-69. Carbohydrate Research **176:**160-163.

53. **Bock, K., J. Schuster-Kolbe, E. Altman, G. Allmaier, B. Stahl, R. Christian, U. B. Sleytr, and P. Messner.** 1994. Primary structure of the *O*-glycosidically linked glycan chain of the crystalline surface layer glycoprotein of *Thermoanaerobacter thermohydrosulfuricus* L111-69. Galactosyl tyrosine as a novel linkage unit. Journal of Biological Chemistry **269:**7137-7144.

54. **Messner, P., R. Christian, J. Kolbe, G. Schulz, and U. B. Sleytr.** 1992. Analysis of a novel linkage unit of *O*-linked carbohydrates from the crystalline surface layer glycoprotein of *Clostridium thermohydrosulfuricum* S102-70. Journal of Bacteriology **174:**2236-2240.

55. **Altman, E., C. Schäffer, J. R. Brisson, and P. Messner.** 1995. Characterization of the glycan structure of a major glycopeptide from the

surface layer glycoprotein of *Clostridium thermosaccharolyticum* E207-71. European Journal of Biochemistry **229:**308-315.

56. **Schäffer, C., K. Dietrich, B. Unger, A. Scheberl, F. A. Rainey, H. Kählig, and P. Messner.** 2000. A novel type of carbohydrate-protein linkage region in the tyrosine-bound S-layer glycan of *Thermoanaerobacterium thermosaccharolyticum* D120-70. European Journal of Biochemistry **267:**5482-5492.
57. **Chen, P. S., H. K. Mitchell, and M. Neuweg.** 1978. Tyrosine glucoside in *Drosophila busckii*. Insect Biochemistry **8:**279-286.
58. **Kramer, K. J., T. L. Hopkins, R. F. Ahmed, D. Mueller, and G. Lookhart.** 1980. Tyrosine metabolism for cuticle tanning in the tobacco hornworm, *Manduca sexta* (L.) and other *Lepidoptera*: identification of β-D-glucopyranosyl-*O*-L-tyrosine and other metabolites. Archives of Biochemistry and Biophysics **205:**146-155.
59. **Aon, M. A., and J. A. Curtino.** 1985. Protein-bound glycogen is linked to tyrosine residues. Biochemical Journal **229:**269-272.
60. **Rodriguez, I. R., and W. J. Whelan.** 1985. A novel glycosyl-amino acid linkage: rabbit-muscle glycogen is covalently linked to a protein via tyrosine. Biochemical and Biophysical Research Communications **132:**829-836.
61. **Smythe, C., and P. Cohen.** 1991. The discovery of glycogenin and the priming mechanism for glycogen biogenesis. European Journal of Biochemistry **200:**625-631.
62. **Pitcher, J., C. Smythe, and P. Cohen.** 1988. Glycogenin is the priming glucosyltransferase required for the initiation of glycogen biogenesis in rabbit skeletal muscle. European Journal of Biochemistry **176:**391-395.
63. **Lomako, J., W. M. Lomako, and W. J. Whelan.** 1988. A self-glucosylating protein is the primer for rabbit muscle glycogen biosynthesis. FASEB Journal **2:**3097-3103.
64. **Lomako, J., W. M. Lomako, and W. J. Whelan.** 2004. Glycogenin: the primer for mammalian and yeast glycogen synthesis. Biochimica et Biophysica Acta **1673:**45-55.
65. **Guerry, P., and C. M. Szymanski.** 2008. *Campylobacter* sugars sticking out. Trends in Microbiology **16:**428-435.
66. **Fry, B. N., V. Korolik, J. A. ten Brinke, M. T. T. Pennings, R. Zalm, B. J. J. Teunis, P. J. Coloe, and B. A. M. van der Zeijst.** 1998. The

lipopolysaccharide biosynthesis locus of *Campylobacter jejuni* 81116. Microbiology **144:**2049-2061.

67. **Szymanski, C. M., R. Yao, C. P. Ewing, T. J. Trust, and P. Guerry.** 1999. Evidence for a system of general protein glycosylation in *Campylobacter jejuni*. Molecular Microbiology **32:**1022-1030.
68. **Linton, D., E. Allan, A. V. Karlyshev, A. D. Cronshaw, and B. W. Wren.** 2002. Identification of *N*-acetylgalactosamine-containing glycoproteins PEB3 and CgpA in *Campylobacter jejuni*. Molecular Microbiology **43:**497-508.
69. **Young, N. M., J. R. Brisson, J. Kelly, D. C. Watson, L. Tessier, P. H. Lanthier, H. C. Jarrell, N. Cadotte, F. St Michael, E. Aberg, and C. M. Szymanski.** 2002. Structure of the *N*-linked glycan present on multiple glycoproteins in the Gram-negative bacterium, *Campylobacter jejuni*. Journal of Biological Chemistry **277:**42530-42539.
70. **Nita-Lazar, M., M. Wacker, B. Schegg, S. Amber, and M. Aebi.** 2005. The N-X-S/T consensus sequence is required but not sufficient for bacterial *N*-linked protein glycosylation. Glycobiology **15:**361-367.
71. **Kowarik, M., N. M. Young, S. Numao, B. L. Schulz, I. Hug, N. Callewaert, D. C. Mills, D. C. Watson, M. Hernandez, J. F. Kelly, M. Wacker, and M. Aebi.** 2006. Definition of the bacterial *N*-glycosylation site consensus sequence. EMBO Journal **25:**1957-1966.
72. **Feldman, M. F., M. Wacker, M. Hernandez, P. G. Hitchen, C. L. Marolda, M. Kowarik, H. R. Morris, A. Dell, M. A. Valvano, and M. Aebi.** 2005. Engineering *N*-linked protein glycosylation with diverse O antigen lipopolysaccharide structures in *Escherichia coli*. Proceedings of the National Academy of Sciences of the United States of America **102:**3016-3021.
73. **Wacker, M., M. F. Feldman, N. Callewaert, M. Kowarik, B. R. Clarke, N. L. Pohl, M. Hernandez, E. D. Vines, M. A. Valvano, C. Whitfield, and M. Aebi.** 2006. Substrate specificity of bacterial oligosaccharyltransferase suggests a common transfer mechanism for the bacterial and eukaryotic systems. Proceedings of the National Academy of Sciences of the United States of America **103:**7088-7093.
74. **Wacker, M., D. Linton, P. G. Hitchen, M. Nita-Lazar, S. M. Haslam, S. J. North, M. Panico, H. R. Morris, A. Dell, B. W. Wren, and M. Aebi.** 2002. *N*-

linked glycosylation in *Campylobacter jejuni* and its functional transfer into *E. coli.* Science **298:**1790-1793.

75. **Coyne, M. J., B. Reinap, M. M. Lee, and L. E. Comstock.** 2005. Human symbionts use a host-like pathway for surface fucosylation. Science **307:**1778-1781.
76. **Glover, K. J., E. Weerapana, and B. Imperiali.** 2005. *In vitro* assembly of the undecaprenylpyrophosphate-linked heptasaccharide for prokaryotic *N*-linked glycosylation. Proceedings of the National Academy of Sciences of the United States of America **102:**14255-14259.
77. **Power, P. M., L. F. Roddam, K. Rutter, S. Z. Fitzpatrick, Y. N. Srikhanta, and M. P. Jennings.** 2003. Genetic characterization of pilin glycosylation and phase variation in *Neisseria meningitidis*. Molecular Microbiology **49:**833-847.
78. **Kahler, C. M., L. E. Martin, Y. L. Tzeng, Y. K. Miller, K. Sharkey, D. S. Stephens, and J. K. Davies.** 2001. Polymorphisms in pilin glycosylation locus of *Neisseria meningitidis* expressing class II pili. Infection and Immunity **69:**3597-3604.
79. **Faridmoayer, A., M. A. Fentabil, D. C. Mills, J. S. Klassen, and M. F. Feldman.** 2007. Functional characterization of bacterial oligosaccharyl-transferases involved in *O*-linked protein glycosylation. Journal of Bacteriology **189:**8088-8098.
80. **Åas, F. E., A. Vik, J. Vedde, M. Koomey, and W. Egge-Jacobsen.** 2007. *Neisseria gonorrhoeae O*-linked pilin glycosylation: functional analyses define both the biosynthetic pathway and glycan structure. Molecular Microbiology **65:**607-624.
81. **Houwink, A. L.** 1953. A macromolecular mono-layer in the cell wall of *Spirillum spec*. Biochimica et Biophysica Acta **10:**360-366.
82. **Labaw, L. W., and V. M. Mosley.** 1954. Periodic structure in the flagella and cell walls of a bacterium. Biochimica et Biophysica Acta **15:**325-331.
83. **Sleytr, U. B., and P. Messner.** 2009. Crystalline Cell Surface Layers (S Layers), p. 89-98. *In* M. Schaechter (ed.), Encyclopedia of Microbiology, 3 ed, vol. 1. Academic Press, San Diego.
84. **Mescher, M. F., and J. L. Strominger.** 1976. Purification and characterization of a prokaryotic glycoprotein from cell-envelope of *Halobacterium salinarium*. Journal of Biological Chemistry **251:**2005-2014.

85. **Sleytr, U. B., and K. J. Thorne.** 1976. Chemical characterization of the regularly arranged surface layers of *Clostridium thermosaccharolyticum* and *Clostridium thermohydrosulfuricum*. Journal of Bacteriology **126:**377-383.

86. **Schäffer, C., T. Wugeditsch, C. Neuninger, and P. Messner.** 1996. Are S-layer glycoproteins and lipopolysaccharides related? Microbial Drug Resistance **2:**17-23.

87. **Schäffer, C., and P. Messner.** 2001. Glycobiology of surface layer proteins. Biochimie **83:**591-599.

88. **Schäffer, C., M. Graninger, and P. Messner.** 2001. Prokaryotic glycosylation. Proteomics **1:**248-261.

89. **Abu-Qarn, M., and J. Eichler.** 2007. An analysis of amino acid sequences surrounding archaeal glycoprotein sequons. Archaea **2:**73-81.

90. **Zeitler, R., E. Hochmuth, R. Deutzmann, and M. Sumper.** 1998. Exchange of Ser-4 for Val, Leu or Asn in the sequon Asn-Ala-Ser does not prevent *N*-glycosylation of the cell surface glycoprotein from *Halobacterium halobium*. Glycobiology **8:**1157-1164.

91. **Eichler, J., and M. W. W. Adams.** 2005. Posttranslational protein modification in Archaea. Microbiology and Molecular Biology Reviews **69:**393-425.

92. **Wieland, F., W. Dompert, G. Bernhardt, and M. Sumper.** 1980. Halobacterial glycoprotein saccharides contain covalently linked sulfate. FEBS Letters **120:**110-114.

93. **Lechner, J., and M. Sumper.** 1987. The primary structure of a prokaryotic glycoprotein - cloning and sequencing of the cell-surface glycoprotein gene of *Halobacteria*. Journal of Biological Chemistry **262:**9724-9729.

94. **Wieland, F., J. Lechner, and M. Sumper.** 1982. The cell-wall glycoprotein of halobacteria - structural, functional and biosynthetic aspects. Zentralblatt für Bakteriologie, Mikrobiologie und Hygiene, I. Abteilung, Originale C - Allgemeine, Angewandte und Ökologische Mikrobiologie **3:**161-170.

95. **Paul, G., F. Lottspeich, and F. Wieland.** 1986. Asparaginyl-*N*-acetylgalactosamine - linkage unit of halobacterial glycosaminoglycan. Journal of Biological Chemistry **261:**1020-1024.

96. **Paul, G., and F. Wieland.** 1987. Sequence of the halobacterial glycosaminoglycan. Journal of Biological Chemistry **262:**9587-9593.

97. **Wieland, F., R. Heitzer, and W. Schaefer.** 1983. Asparaginylglucose - novel type of carbohydrate linkage. Proceedings of the National Academy of Sciences of the United States of America-Biological Sciences **80:**5470-5474.

98. **Wieland, F., G. Paul, and M. Sumper.** 1985. Halobacterial flagellins are sulfated glycoproteins. Journal of Biological Chemistry **260:**5180-5185.

99. **Sumper, M., E. Berg, R. Mengele, and I. Strobel.** 1990. Primary structure and glycosylation of the S-layer protein of *Haloferax volcanii*. Journal of Bacteriology **172:**7111-7118.

100. **Abu-Qarn, M., S. Yurist-Doutsch, A. Giordano, A. Trauner, H. R. Morris, P. Hitchen, O. Medalia, A. Dell, and J. Eichler.** 2007. *Haloferax volcanii* AglB and AglD are involved in *N*-glycosylation of the S-layer glycoprotein and proper assembly of the surface layer. Journal of Molecular Biology **374:**1224-1236.

101. **Mengele, R., and M. Sumper.** 1992. Drastic differences in glycosylation of related S-layer glycoproteins from moderate and extreme halophiles. Journal of Biological Chemistry **267:**8182-8185.

102. **Bröckl, G., M. Behr, S. Fabry, R. Hensel, H. Kaudewitz, E. Biendl, and H. König.** 1991. Analysis and nucleotide sequence of the genes encoding the surface layer glycoproteins of the hyperthermophilic methanogens *Methanothermus fervidus* and *Methanothermus sociabilis*. European Journal of Biochemistry **199:**147-152.

103. **Kalmokoff, M. L., S. F. Koval, and K. F. Jarrell.** 1992. Relatedness of the flagellins from methanogens. Archives of Microbiology **157:**481-487.

104. **Kärcher, U., H. Schröder, E. Haslinger, G. Allmaier, R. Schreiner, F. Wieland, A. Haselbeck, and H. König.** 1993. Primary structure of the heterosaccharide of the surface glycoprotein of *Methanothermus fervidus*. Journal of Biological Chemistry **268:**26821-26826.

105. **Voisin, S., R. S. Houliston, J. Kelly, J. R. Brisson, D. Watson, S. L. Bardy, K. F. Jarrell, and S. M. Logan.** 2005. Identification and characterization of the unique *N*-linked glycan common to the flagellins and S-layer glycoprotein of *Methanococcus voltae*. Journal of Biological Chemistry **280:**16586-16593.

106. **Chaban, B., S. Voisin, J. Kelly, S. M. Logan, and K. F. Jarrell.** 2006. Identification of genes involved in the biosynthesis and attachment of

Methanococcus voltae *N*-linked glycans: insight into *N*-linked glycosylation pathways in Archaea. Molecular Microbiology **61:**259-268.

107. **Abu-Qarn, M., and J. Eichler.** 2006. Protein *N*-glycosylation in Archaea: defining *Haloferax volcanii* genes involved in S-layer glycoprotein glycosylation. Molecular Microbiology **61:**511-525.
108. **Yurist-Doutsch, S., B. Chaban, D. J. VanDyke, K. F. Jarrell, and J. Eichler.** 2008. Sweet to the extreme: Protein glycosylation in Archaea. Molecular Microbiology **68:**1079-1084.
109. **Chaban, B., S. M. Logan, J. F. Kelly, and K. F. Jarrell.** 2009. AglC and AglK are involved in biosynthesis and attachment of diacetylated glucuronic acid to the *N*-glycan in *Methanococcus voltae*. Journal of Bacteriology **191:**187-195.
110. **Plavner, N., and J. Eichler.** 2008. Defining the topology of the *N*-glycosylation pathway in the halophilic archaeon *Haloferax volcanii*. Journal of Bacteriology **190:**8045-8052.
111. **Abu-Qarn, M., A. Giordano, F. Battaglia, A. Trauner, P. G. Hitchen, H. R. Morris, A. Dell, and J. Eichler.** 2008. Identification of AglE, a second glycosyltransferase involved in *N*-glycosylation of the *Haloferax volcanii* S-layer glycoprotein. Journal of Bacteriology **190:**3140-3146.
112. **Yurist-Doutsch, S., M. Abu-Qarn, F. Battaglia, H. R. Morris, P. G. Hitchen, A. Dell, and J. Eichler.** 2008. *AglF*, *aglG* and *aglI*, novel members of a gene island involved in the *N*-glycosylation of the *Haloferax volcanii* S-layer glycoprotein. Molecular Microbiology **69:**1234-1245.
113. **Yurist-Doutsch, S., and J. Eichler.** 2009. Manual annotation, transcriptional analysis, and protein expression studies reveal novel genes in the *agl* cluster responsible for *N*-glycosylation in the halophilic archaeon *Haloferax volcanii*. Journal of Bacteriology **191:**3068-3075.
114. **Lee, S. W., M. Sabet, H. S. Um, J. Yang, H. C. Kim, and W. Zhu.** 2006. Identification and characterization of the genes encoding a unique surface (S-) layer of *Tannerella forsythia*. Gene **371:**102-111.
115. **Sakakibara, J., K. Nagano, Y. Murakami, N. Higuchi, H. Nakamura, K. Shimozato, and F. Yoshimura.** 2007. Loss of adherence ability to human gingival epithelial cells in S-layer protein-deficient mutants of *Tannerella forsythensis*. Microbiology **153:**866-876.

116. **Nazina, T. N., T. P. Tourova, A. B. Poltaraus, E. V. Novikova, A. A. Grigoryan, A. E. Ivanova, A. M. Lysenko, V. V. Petrunyaka, G. A. Osipov, S. S. Belyaev, and M. V. Ivanov.** 2001. Taxonomic study of aerobic thermophilic bacilli: descriptions of *Geobacillus subterraneus* gen. nov., sp. nov. and *Geobacillus uzenensis* sp. nov. from petroleum reservoirs and transfer of *Bacillus stearothermophilus*, *Bacillus thermocatenulatus*, *Bacillus thermoleovorans*, *Bacillus kaustophilus*, *Bacillus thermodenitrificans* to *Geobacillus* as the new combinations *G. stearothermophilus*, *G. thermocatenulatus*, *G. thermoleovorans*, *G. kaustophilus*, *G. thermoglucosidasius* and *G. thermodenitrificans*. International Journal of Systematic and Evolutionary Microbiology **51:**433-446.

117. **Küpcü, Z., L. März, P. Messner, and U. B. Sleytr.** 1984. Evidence for the glycoprotein nature of the crystalline cell wall surface layer of *Bacillus stearothermophilus* strain NRS 2004/3a. FEBS Letters **173:**185-190.

118. **Messner, P., D. Pum, and U. B. Sleytr.** 1986. Characterization of the ultrastructure and the self-assembly of the surface layer of *Bacillus stearothermophilus* strain NRS 2004/3a. Journal of Ultrastructure and Molecular Structure Research **97:**73-88.

119. **Schäffer, C., T. Wugeditsch, H. Kählig, A. Scheberl, S. Zayni, and P. Messner.** 2002. The surface layer (S-layer) glycoprotein of *Geobacillus stearothermophilus* NRS 2004/3a. Analysis of its glycosylation. Journal of Biological Chemistry **277:**6230-6239.

120. **Steiner, K., G. Pohlentz, K. Dreisewerd, S. Berkenkamp, P. Messner, J. Peter-Katalinic, and C. Schäffer.** 2006. New insights into the glycosylation of the surface layer protein SgsE from *Geobacillus stearothermophilus* NRS 2004/3a. Journal of Bacteriology **188:**7914-7921.

121. **Novotny, R., A. Pföstl, P. Messner, and C. Schäffer.** 2004. Genetic organization of chromosomal S-layer glycan biosynthesis loci of *Bacillaceae*. Glycoconjugate Journal **20:**435-447.

122. **Novotny, R., C. Schäffer, J. Strauss, and P. Messner.** 2004. S-layer glycan-specific loci on the chromosome of *Geobacillus stearothermophilus* NRS 2004/3a and dTDP-L-rhamnose biosynthesis potential of *G. stearothermophilus* strains. Microbiology **150:**953-965.

123. **Drummelsmith, J., and C. Whitfield.** 1999. Gene products required for surface expression of the capsular form of the group 1 K antigen in *Escherichia coli* (O9a:K30). Molecular Microbiology **31:**1321-1332.

124. **Steiner, K., R. Novotny, K. Patel, E. Vinogradov, C. Whitfield, M. A. Valvano, P. Messner, and C. Schäffer.** 2007. Functional characterization of the initiation enzyme of S-layer glycoprotein glycan biosynthesis in *Geobacillus stearothermophilus* NRS 2004/3a. Journal of Bacteriology **189:**2590-2598.

125. **Steiner, K., R. Novotny, D. B. Werz, K. Zarschler, P. H. Seeberger, A. Hofinger, P. Kosma, C. Schäffer, and P. Messner.** 2008. Molecular basis of S-layer glycoprotein glycan biosynthesis in *Geobacillus stearothermophilus*. Journal of Biological Chemistry **283:**21120-21133.

126. **Kowarik, M., S. Numao, M. F. Feldman, B. L. Schulz, N. Callewaert, E. Kiermaier, I. Catrein, and M. Aebi.** 2006. *N*-linked glycosylation of folded proteins by the bacterial oligosaccharyltransferase. Science **314:**1148-1150.

127. **Vehmaanperä, J.** 1989. Transformation of *Bacillus amyloliquefaciens* by electroporation. FEMS Microbiology Letters **52:**165-169.

128. **Brigidi, P., E. De Rossi, M. L. Bertarini, G. Riccardi, and D. Matteuzzi.** 1990. Genetic transformation of intact cells of *Bacillus subtilis* by electroporation. FEMS Microbiology Letters **55:**135-138.

129. **Narumi, I., K. Sawakami, S. Nakamoto, N. Nakayama, T. Yanagisawa, N. Takahashi, and H. Kihara.** 1992. A newly isolated *Bacillus stearo-thermophilus* K1041 and its transformation by electroporation. Biotechnology Techniques **6:**83-86.

130. **Wu, L. J., and N. E. Welker.** 1989. Protoplast transformation of *Bacillus stearothermophilus* NUB36 by plasmid DNA. Journal of General Microbiology **135:**1315-1324.

131. **Imanaka, T., M. Fujii, I. Aramori, and S. Aiba.** 1982. Transformation of *Bacillus stearothermophilus* with plasmid DNA and characterization of shuttle vector plasmids between *Bacillus stearothermophilus* and *Bacillus subtilis*. Journal of Bacteriology **149:**824-830.

132. **Spizizen, J.** 1958. Transformation of biochemically deficient strains of *Bacillus subtilis* by deoxyribonucleate. Proceedings of the National Academy of Sciences of the United States of America **44:**1072-1078.

133. **De Rossi, E., P. Brigidi, M. Rossi, D. Matteuzzi, and G. Riccardi.** 1991. Characterization of Gram-positive broad host-range plasmids carrying a thermophilic replicon. Research in Microbiology **142:**389-396.

134. **Soutschek-Bauer, E., W. Scholz, E. Grill, and W. L. Staudenbauer.** 1987. Thermostability and superhelicity of plasmid DNA in *Bacillus stearothermophilus*. Molecular Genetics and Genomics **209:**575-579.

135. **Hoshino, T., T. Ikeda, K. Furukawa, and N. Tomizuka.** 1985. Genetic relationship between pUB110 and antibiotic-resistant plasmids obtained from thermophilic bacilli. Canadian Journal of Microbiology **31:**614-619.

136. **Matsumura, M., Y. Katakura, T. Imanaka, and S. Aiba.** 1984. Enzymatic and nucleotide sequence studies of a kanamycin-inactivating enzyme encoded by a plasmid from thermophilic bacilli in comparison with that encoded by plasmid pUB110. Journal of Bacteriology **160:**413-420.

137. **De Rossi, E., P. Brigidi, N. E. Welker, G. Riccardi, and D. Matteuzzi.** 1994. New shuttle vector for cloning in *Bacillus stearothermophilus*. Research in Microbiology **145:**579-583.

138. **Narumi, I., N. Nakayama, S. Nakamoto, T. Kimura, T. Yanagisawa, and H. Kihara.** 1993. Construction of a new shuttle vector pSTE33 and its stabilities in *Bacillus stearothermophilus*, *Bacillus subtilis*, and *Escherichia coli*. Biotechnology Letters **15:**815-820.

139. **Sullivan, M. A., R. E. Yasbin, and F. E. Young.** 1984. New shuttle vectors for *Bacillus subtilis* and *Escherichia coli* which allow rapid detection of inserted fragments. Gene **29:**21-26.

140. **Alegre, M. T., M. C. Rodriguez, and J. M. Mesas.** 2004. Transformation of *Lactobacillus plantarum* by electroporation with *in vitro* modified plasmid DNA. FEMS Microbiology Letters **241:**73-77.

141. **Lasa, I., J. R. Caston, L. A. Fernandez-Herrero, M. A. de Pedro, and J. Berenguer.** 1992. Insertional mutagenesis in the extreme thermophilic eubacteria *Thermus thermophilus* HB8. Molecular Microbiology **6:**1555-1564.

142. **de Grado, M., P. Castan, and J. Berenguer.** 1999. A high-transformation-efficiency cloning vector for *Thermus thermophilus*. Plasmid **42:**241-245.

143. **Gilchrist, A., and J. Smit.** 1991. Transformation of freshwater and marine caulobacters by electroporation. Journal of Bacteriology **173:**921-925.

144. **Trieu-Cuot, P., C. Carlier, P. Martin, and P. Courvalin.** 1987. Plasmid transfer by conjugation from *Escherichia coli* to Gram-positive bacteria. FEMS Microbiology Letters **48:**289-294.

145. **Thomas, N. A., S. Mueller, A. Klein, and K. F. Jarrell.** 2002. Mutants in *flaI* and *flaJ* of the archaeon *Methanococcus voltae* are deficient in flagellum assembly. Molecular Microbiology **46:**879-887.

146. **Cline, S. W., W. L. Lam, R. L. Charlebois, L. C. Schalkwyk, and W. F. Doolittle.** 1989. Transformation methods for halophilic archaebacteria. Canadian Journal of Microbiology **35:**148-152.

147. **Zhang, M., H. Nakai, and T. Imanaka.** 1988. Useful host-vector systems in *Bacillus stearothermophilus*. Applied and Environmental Microbiology **54:**3162-3164.

148. **Aiba, S., T. Imanaka, and J. Koizumi.** 1983. Transformation of *Bacillus stearothermophilus* with plasmid DNA and its application to molecular cloning. Annals of the New York Academy of Sciences **413:**57-70.

149. **Oskam, L., G. Venema, and S. Bron.** 1992. Plasmid maintenance in *Bacillus stearothermophilus* is strain-dependent. FEMS Microbiology Letters **72:**203-208.

150. **Studholme, D. J., R. A. Jackson, and D. J. Leak.** 1999. Phylogenetic analysis of transformable strains of thermophilic *Bacillus* species. FEMS Microbiology Letters **172:**85-90.

151. **Zarschler, K., B. Janesch, S. Zayni, C. Schäffer, and P. Messner.** 2009. Construction of a gene knockout system for application in *Paenibacillus alvei* CCM 2051^{T}, exemplified by the S-layer glycan biosynthesis initiation enzyme WsfP. Applied and Environmental Microbiology **75:**3077-3085.

152. **Cheshire, F. R., and W. W. Cheyne.** 1885. The pathogenic history and the history under cultivation of a new *Bacillus* (*B. alvei*), the cause of a disease of the hive bee hitherto known as foul brood. Journal of the Royal Microscopical Society **2:**581-601.

153. **Ash, C., F. G. Priest, and M. D. Collins.** 1993. Molecular identification of rRNA group 3 bacilli (Ash, Farrow, Wallbanks and Collins) using a PCR probe test. Proposal for the creation of a new genus *Paenibacillus*. Antonie van Leeuwenhoek **64:**253-260.

154. **Balaraman, K., U. S. Bheema Rao, and P. K. Rajagopalan.** 1979. Bacterial pathogens of mosquito larvae - *Bacillus alvei* (Cheshire and Cheyene) and *Bacillus brevis* (Migula) - isolated in Pondicherry. The Indian Journal of Medical Research **70:**615-619.

155. **Gilliam, M.** 1985. Microbes from apiarian sources - *Bacillus spp* in frass of the greater wax moth. Journal of Invertebrate Pathology **45:**218-224.

156. **Snirnov, A. M.** 1976. Disinfection of apiary soil in foulbrood of bees. Veterinariia**:**32-37.

157. **Reboli, A. C., C. S. Bryan, and W. E. Farrar.** 1989. Bacteremia and infection of a hip prosthesis caused by *Bacillus alvei*. Journal of Clinical Microbiology **27:**1395-1396.

158. **Sára, M., K. Moser-Thier, U. Kainz, and U. B. Sleytr.** 1990. Characterization of S-layers from mesophilic *Bacillaceae* and studies on their protective role towards muramidases. Archives of Microbiology **153:**209-214.

159. **Hartmann, E., P. Messner, G. Allmeier, and H. König.** 1993. Proposed pathway for biosynthesis of the S-layer glycoprotein of *Bacillus alvei*. Journal of Bacteriology **175:**4515-4519.

160. **Georgiou, G., C. Stathopoulos, P. S. Daugherty, A. R. Nayak, B. L. Iverson, and R. Curtiss, 3rd.** 1997. Display of heterologous proteins on the surface of microorganisms: from the screening of combinatorial libraries to live recombinant vaccines. Nature Biotechnology **15:**29-34.

161. **Miller, K. D., N. B. Pefaur, and C. L. Baird.** 2008. Construction and screening of antigen targeted immune yeast surface display antibody libraries. Current Protocols in Cytometry **45:**1-30.

162. **Daugherty, P. S.** 2007. Protein engineering with bacterial display. Current Opinion in Structural Biology **17:**474-480.

163. **Ishige, T., K. Honda, and S. Shimizu.** 2005. Whole organism biocatalysis. Current Opinion in Chemical Biology **9:**174-180.

164. **Shibasaki, S., M. Ueda, K. Ye, K. Shimizu, N. Kamasawa, M. Osumi, and A. Tanaka.** 2001. Creation of cell surface-engineered yeast that display different fluorescent proteins in response to the glucose concentration. Applied Microbiology and Biotechnology **57:**528-533.

165. **Yasui, M., S. Shibasaki, K. Kuroda, M. Ueda, N. Kawada, J. Nishikawa, T. Nishihara, and A. Tanaka.** 2002. An arming yeast with the ability to entrap

fluorescent 17β-estradiol on the cell surface. Applied Microbiology and Biotechnology **59:**329-331.

166. **Wu, C. H., A. Mulchandani, and W. Chen.** 2008. Versatile microbial surface-display for environmental remediation and biofuels production. Trends in Microbiology **16:**181-188.
167. **Saleem, M., H. Brim, S. Hussain, M. Arshad, M. B. Leigh, and H. Zia ul.** 2008. Perspectives on microbial cell surface display in bioremediation. Biotechnology Advances **26:**151-161.
168. **Lee, S. Y., J. H. Choi, and Z. Xu.** 2003. Microbial cell-surface display. Trends in Biotechnology **21:**45-52.
169. **Benhar, I.** 2001. Biotechnological applications of phage and cell display. Biotechnology Advances **19:**1-33.
170. **Smit, J.** 2008. Heads up S-layer display: the power of many. Structure **16:**1151-1153.
171. **Sleytr, U. B., C. Huber, N. Ilk, D. Pum, B. Schuster, and E. M. Egelseer.** 2007. S-layers as a tool kit for nanobiotechnological applications. FEMS Microbiology Letters **267:**131-144.
172. **Bingle, W. H., J. F. Nomellini, and J. Smit.** 1997. Cell-surface display of a *Pseudomonas aeruginosa* strain K pilin peptide within the paracrystalline S-layer of *Caulobacter crescentus*. Molecular Microbiology **26:**277-288.
173. **Bingle, W. H., J. F. Nomellini, and J. Smit.** 1997. Linker mutagenesis of the *Caulobacter crescentus* S-layer protein: toward a definition of an N-terminal anchoring region and a C-terminal secretion signal and the potential for heterologous protein secretion. Journal of Bacteriology **179:**601-611.
174. **Bingle, W. H., J. F. Nomellini, and J. Smit.** 2000. Secretion of the *Caulobacter crescentus* S-layer protein: further localization of the C-terminal secretion signal and its use for secretion of recombinant proteins. Journal of Bacteriology **182:**3298-3301.
175. **Mesnage, S., E. Tosi-Couture, M. Mock, and A. Fouet.** 1999. The S-layer homology domain as a means for anchoring heterologous proteins on the cell surface of *Bacillus anthracis*. Journal of Applied Microbiology **87:**256-260.
176. **Mesnage, S., M. Weber-Levy, M. Haustant, M. Mock, and A. Fouet.** 1999. Cell surface-exposed tetanus toxin fragment C produced by recombinant

Bacillus anthracis protects against tetanus toxin. Infection and Immunity **67:**4847-4850.

177. **Mesnage, S., E. Tosi-Couture, and A. Fouet.** 1999. Production and cell surface anchoring of functional fusions between the SLH motifs of the *Bacillus anthracis* S-layer proteins and the *Bacillus subtilis* levansucrase. Molecular Microbiology **31:**927-936.
178. **Sleytr, U. B., E. M. Egelseer, N. Ilk, D. Pum, and B. Schuster.** 2007. S-Layers as a basic building block in a molecular construction kit. Febs Journal **274:**323-334.
179. **Nomellini, J. F., G. Duncan, I. R. Dorocicz, and J. Smit.** 2007. S-layer-mediated display of the immunoglobulin G-binding domain of streptococcal protein G on the surface of *Caulobacter crescentus*: development of an immunoactive reagent. Applied and Environmental Microbiology **73:**3245-3253.
180. **Liu, M., S. Guo, S. Hu, Y. Xiao, Q. Xu, Z. Li, D. Bi, and M. Sun.** 2008. Displaying the protein of *Mycoplasma gallisepticum* agglutinin on the cell surface of *Bacillus thuringiensis* with the S-layer protein. Veterinary Microbiology **130:**99-106.
181. **Liu, M., S. Li, S. Hu, C. Zhao, D. Bi, and M. Sun.** 2008. Display of avian influenza virus nucleoprotein on *Bacillus thuringiensis* cell surface using CTC as a fusion partner. Applied Microbiology and Biotechnology **78:**669-676.
182. **Zaug, A. J., and T. R. Cech.** 1982. The intervening sequence excised from the ribosomal RNA precursor of *Tetrahymena* contains a 5-terminal guanosine residue not encoded by the DNA. Nucleic Acids Research **10:**2823-2838.
183. **Krüger, K., P. J. Grabowski, A. J. Zaug, J. Sands, D. E. Gottschling, and T. R. Cech.** 1982. Self-splicing RNA: autoexcision and autocyclization of the ribosomal RNA intervening sequence of *Tetrahymena*. Cell **31:**147-157.
184. **Hougland, J. L., J. A. Piccirilli, M. Forconi, J. Lee, and D. Herschlag.** 2006. How the group I intron works: A case study of RNA structure and function, p. 133–205. *In* R. F. Gesteland, T. R. Cech, and J. F. Atkins (ed.), The RNA World, 3 ed, vol. 43. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
185. **Pyle, A. M., and A. M. Lambowitz.** 2006. Group II introns: ribozymes that splice RNA and invade DNA, p. 469–505. *In* R. F. Gesteland, T. R. Cech, and

J. F. Atkins (ed.), The RNA World, 3 ed, vol. 43. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

186. **Cech, T. R.** 2002. Ribozymes, the first 20 years. Biochemical Society Transactions **30:**1162-1166.
187. **Woodson, S. A.** 2005. Structure and assembly of group I introns. Current Opinion in Structural Biology **15:**324-330.
188. **Lambowitz, A. M., and S. Zimmerly.** 2004. Mobile group II introns. Annual Review of Genetics **38:**1-35.
189. **Saldanha, R., G. Mohr, M. Belfort, and A. M. Lambowitz.** 1993. Group I and group II introns. FASEB Journal **7:**15-24.
190. **Doudna, J. A., and T. R. Cech.** 2002. The chemical repertoire of natural ribozymes. Nature **418:**222-228.
191. **Lambowitz, A. M., and M. Belfort.** 1993. Introns as mobile genetic elements. Annual Review of Biochemistry **62:**587-622.
192. **Cech, T. R.** 1988. Conserved sequences and structures of group I introns: building an active site for RNA catalysis - a review. Gene **73:**259-271.
193. **Cech, T. R.** 1990. Self-splicing of group I introns. Annual Review of Biochemistry **59:**543-568.
194. **Haugen, P., D. M. Simon, and D. Bhattacharya.** 2005. The natural history of group I introns. Trends in Genetics **21:**111-119.
195. **Ferat, J. L., and F. Michel.** 1993. Group II self-splicing introns in bacteria. Nature **364:**358-361.
196. **Michel, F., K. Umesono, and H. Ozeki.** 1989. Comparative and functional anatomy of group II catalytic introns - a review. Gene **82:**5-30.
197. **Toro, N.** 2003. Bacteria and Archaea Group II introns: additional mobile genetic elements in the environment. Environmental Microbiology **5:**143-151.
198. **Costa, M., and F. Michel.** 1999. Tight binding of the 5' exon to domain I of a group II self-splicing intron requires completion of the intron active site. EMBO Journal **18:**1025-1037.
199. **Toro, N., J. I. Jimenez-Zurdo, and F. M. Garcia-Rodriguez.** 2007. Bacterial group II introns: not just splicing. FEMS Microbiology Reviews **31:**342-358.
200. **Martinez-Abarca, F., and N. Toro.** 2000. Group II introns in the bacterial world. Molecular Microbiology **38:**917-926.

201. **Jacquier, A.** 1996. Group II introns: elaborate ribozymes. Biochimie **78:**474-487.

202. **Chanfreau, G., and A. Jacquier.** 1993. Interaction of intronic boundaries is required for the second splicing step efficiency of a group II intron. EMBO Journal **12:**5173-5180.

203. **Chanfreau, G., and A. Jacquier.** 1994. Catalytic site components common to both splicing steps of a group II intron. Science **266:**1383-1387.

204. **Jacquier, A., and N. Jacquesson-Breuleux.** 1991. Splice site selection and role of the lariat in a group II intron. Journal of Molecular Biology **219:**415-428.

205. **Guo, H., S. Zimmerly, P. S. Perlman, and A. M. Lambowitz.** 1997. Group II intron endonucleases use both RNA and protein subunits for recognition of specific sequences in double-stranded DNA. EMBO Journal **16:**6835-6848.

206. **Mohr, G., D. Smith, M. Belfort, and A. M. Lambowitz.** 2000. Rules for DNA target-site recognition by a lactococcal group II intron enable retargeting of the intron to specific DNA sequences. Genes and Development **14:**559-573.

207. **Cousineau, B., D. Smith, S. Lawrence-Cavanagh, J. E. Mueller, J. Yang, D. Mills, D. Manias, G. Dunny, A. M. Lambowitz, and M. Belfort.** 1998. Retrohoming of a bacterial group II intron: mobility via complete reverse splicing, independent of homologous DNA recombination. Cell **94:**451-462.

208. **Cousineau, B., S. Lawrence, D. Smith, and M. Belfort.** 2000. Retrotransposition of a bacterial group II intron. Nature **404:**1018-1021.

209. **Ichiyanagi, K., A. Beauregard, and M. Belfort.** 2003. A bacterial group II intron favors retrotransposition into plasmid targets. Proceedings of the National Academy of Sciences of the United States of America **100:**15742-15747.

210. **Yao, J., and A. M. Lambowitz.** 2007. Gene targeting in gram-negative bacteria by use of a mobile group II intron ("Targetron") expressed from a broad-host-range vector. Applied and Environmental Microbiology **73:**2735-2743.

211. **Heap, J. T., O. J. Pennington, S. T. Cartman, G. P. Carter, and N. P. Minton.** 2007. The ClosTron: a universal gene knock-out system for the genus *Clostridium*. Journal of Microbiological Methods **70:**452-464.

212. **Guo, H., M. Karberg, M. Long, J. P. Jones, 3rd, B. Sullenger, and A. M. Lambowitz.** 2000. Group II introns designed to insert into therapeutically relevant DNA target sites in human cells. Science **289:**452-457.

213. **Rodriguez, S. A., G. Davis, and K. E. Klose.** 2009. Targeted gene disruption in *Francisella tularensis* by group II introns. Methods.

214. **Malhotra, M., and S. Srivastava.** 2008. An *ipdC* gene knock-out of *Azospirillum brasilense* strain SM and its implications on indole-3-acetic acid biosynthesis and plant growth promotion. Antonie van Leeuwenhoek **93:**425-433.

215. **Perutka, J., W. Wang, D. Goerlitz, and A. M. Lambowitz.** 2004. Use of computer-designed group II introns to disrupt *Escherichia coli* DExH/D-box protein and DNA helicase genes. Journal of Molecular Biology **336:**421-439.

216. **Frazier, C. L., J. San Filippo, A. M. Lambowitz, and D. A. Mills.** 2003. Genetic manipulation of *Lactococcus lactis* by using targeted group II introns: generation of stable insertions without selection. Applied and Environmental Microbiology **69:**1121-1128.

217. **Zhong, J., M. Karberg, and A. M. Lambowitz.** 2003. Targeted and random bacterial gene disruption using a group II intron (targetron) vector containing a retrotransposition-activated selectable marker. Nucleic Acids Research **31:**1656-1664.

218. **Steiner, K., A. Hanreich, B. Kainz, P. G. Hitchen, A. Dell, P. Messner, and C. Schäffer.** 2008. Recombinant glycans on an S-layer self-assembly protein: a new dimension for nanopatterned biomaterials. Small **4:**1728-1740.

6. MANUSCRIPTS AND PUBLICATIONS

Zarschler, K., Janesch, B., Pabst, M., Altmann, F., Messner, P., Schäffer, C., Protein tyrosine *O*-glycosylation: molecular insights into a rather unexplored prokaryotic glycosylation system. (Manuscript in preparation).

Zarschler, K., Kainz, B., Janesch, B., Messner, P., Schäffer, C., Surface display of chimeric glycoproteins by the surface (S-) layer system of *Paenibacillus alvei* CCM 2051^T. (Manuscript in preparation).

Zarschler, K., Janesch, B., Zayni, S., Schäffer, C., Messner, P. (2009). Construction of a gene knockout system for application in *Paenibacillus alvei* CCM 2051^T, exemplified by the S-layer glycan biosynthesis initiation enzyme WsfP. Applied and Environmental Microbiology **75:**3077-3085.

Messner, P., Steiner, K., **Zarschler, K.**, Schäffer, C. (2008). S-layer nano-glycobiology of bacteria. Carbohydrate Research **343:**1934-1951.

Steiner, K., Novotny, R., Werz, D. B., **Zarschler, K.**, Seeberger, P. H., Hofinger, A., Kosma, P., Schäffer, C., Messner, P. (2008). Molecular basis of S-layer glycoprotein glycan biosynthesis in *Geobacillus stearothermophilus*. Journal of Biological Chemistry **283:**21120-21133.

7. PUBLISHED CONTRIBUTIONS TO SCIENTIFIC CONFERENCES

Zarschler, K., Janesch, B., Schäffer, C., Messner, P. (2009). Proposal of a pathway of the surface layer protein glycosylation in *Paenibacillus alvei* CCM 2051^T enabled by a glycan mutant approach. 15th European Carbohydrate Symposium, Vienna, Austria (19.-24.07.2009).

Schäffer, C., **Zarschler, K.**, Ristl, R., Posch, G., Janesch, B., Sekot, G., Scheberl, A., Zayni, S., Messner, P. (2008). S-layer protein based glycoengineering and nanobiotechnology. Nanosens 2008, Vienna, Austria (29.-30.09.2008).

Zarschler, K., Janesch, B., Schäffer, C., Messner, P. (2008). First insights into the surface layer glycosylation of *Paenibacillus alvei*. Life Sciences 2008 - Gemeinsame Jahrestagung der Österreichischen Gesellschaft für Biochemie und Molekular-biologie, der Österreichischen Gesellschaft für Genetik und Gentechnik, der Österreichischen Gesellschaft für Biotechnologie und des Austrian Networks für Gentherapie, Graz, Austria (22.-24.09.2008).

Messner, P., Steiner, K., **Zarschler, K.**, Ristl, R., Schäffer, C. (2008). S-layer *neo*glycoproteins. Gordon Research Conference on Bacterial Cell Surfaces, New London, NH, USA (22.-27.06.2008).

Steiner, K., Werz, D.B., **Zarschler, K.**, Seeberger, P.S., Kosma, P., Messner, P., Schäffer, C. (2008). Functional characterization of S layer glycoprotein glycan biosynthesis in *Geobacillus stearothermophilus* NRS 2004/3a. Proc., FASEB Conference on Microbial Polysaccharides of Medical, Agricultural & Industrial Importance, Carefree, AZ, USA (15.-20.06.2008).

Messner, P., Steiner, K., **Zarschler, K.**, Prchal, M., Zayni, S., Scheberl, A., Sleytr, U.B., and Schäffer, C. (2007). „Intelligent" glycoprotein self-assembly systems: novel strategies for nanobiotechnology. Proc. 14th Eur. Carbohydrate Symposium, Lübeck, (02.-07.09.2007).

Messner, P., Steiner, K., **Zarschler, K.**, Prchal, M., Hanreich, A., Zayni, S., Scheberl, A., Sleytr, U.B., Schäffer, C. (2007). S-layer glycoproteins - a versatile tool in nanobiotechnology. Proc. 19th Int. Glycoconjugate Symposium, Cairns, Australia (15.-20.07.2007).

Steiner, K., **Zarschler, K.**, Messner, P., Schäffer, C. (2007). Unraveling details of the S layer glycoprotein glycan biosynthesis of *Geobacillus stearothermophilus* NRS 2004/3a. 7th Carbohydrate Bioengineering Meeting, Braunschweig, Germany (22.-25.04.2007).

Zarschler, K., Steiner, K., Zayni, S., Scheberl, A., Schäffer, C., Messner, P. (2007). S-layer *O*-glycosylation in *Paenibacillus alvei*. 11th Austrian Carbohydrate Workshop, Graz, Austria (15.02.2007).

Schäffer, C., Steiner, K., **Zarschler, K.**, Giry-Laterriere, M., Novotny, R., Messner, P. (2006). Glycosylation of bacterial surface layer proteins: new aspects. Proc., FASEB Conference on Microbial Polysaccharides of Medical, Agricultural & Industrial Importance. Tucson, AZ, USA (17.-22.06.2006).

Zarschler, K., Novotny, R., Steiner, K., Schäffer C., Messner, P. (2006). Is an ABC transporter responsible for the export of the cytoplasmatically synthesized surface layer (S-layer) glycan to the cell surface *Geobacillus stearothermophilus* NRS 2004/3a? FEBS Special Meeting: ATP-Binding Cassette (ABC) Proteins from Multidrug Resistance to Genetic Diseases, Innsbruck, Austria (04.-10.03.2006).

Novotny, R., Steiner, K., **Zarschler, K.**, Giry-Laterriere, M., Zayni, S., Scheberl, A., Schäffer, C., Messner, P. (2005). S-layer glycoproteins. Life Sciences 2005 - Gemeinsame Jahrestagung der Österreichischen Gesellschaften für Biochemie und Molekularbiologie, Genetik und Gentechnologie und Biotechnologie und des Austrian Network for Gene Therapy, Vienna, Austria (26.-28.09.2005).

8. *CURRICULUM VITAE*

PERSONAL INFORMATION

NATIONALITY:	German
DATE OF BIRTH:	04.11. 1979
PLACE OF BIRTH:	Dresden, Germany
ADDRESS:	Sieveringer Straße 21/2/1, 1190 Vienna, Austria
TELEPHONE:	0043 1 9681559
E-MAIL:	kristof.zarschler@gmail.com

EDUCATION

1986 -1992	Primary school Bannewitz, Germany
1992 -1998	Secondary school Manfred-von-Ardenne Gymnasium Freital, Germany
1998	General Higher Education Entrance Qualification (Abitur)
1998 - 1999	Military service
1999 - 2005	University studies of Biology at the Dresden University of Technology, Germany
2004	Final Examination
2004 - 2005	Diploma thesis
2005 - today	PhD thesis at the Department of Nanobiotechnology, University of Natural Resources and Applied Life Sciences, Vienna, Austria

WORK EXPERIENCE

03/2001	Practical training at the laboratory of Prof. Gutzeit, Institute for Zoology, Dresden University of Technology
11/2001 - 02/2003	Part-time research assistant in the laboratory of Prof. Rödel, Institute for Genetics, Dresden University of Technology, Germany
07/2002 - 10/2002	Internship Elbion AG, Radebeul, Germany

10/2003 - 02/2004 Practical training in the Apoptosis laboratory of Dr. Alfons Lawen at the Monash University, Clayton Campus, Victoria, Australia

TEACHING ACTIVITY

04/2007 Lecturing tutor “Methods in Ultrastructure Research”, Department of Nanobiotechnology, University of Natural Resources and Applied Life Sciences, Vienna

RESEARCH PROJECTS

Paenibacillus alvei CCM 2051^{T} for creation of tailored S-layer *neo*glycoproteins.
Hochschuljubiläumsstiftung der Stadt Wien
Project number: H-02229-2007
Project leader: Kristof Zarschler

SOFTWARE QUALIFICATION

Microsoft Office (Word, Excel, PowerPoint)
Adobe Photoshop
Adobe GO LIVE
Corel DRAW and Photo-Paint
SECentral CloneManager
Invitrogen Vector NTI

DRIVING LICENCE

B, C1, BE, CE1, M, L

FOREIGN LANGUAGES

English (excellent)
Russian (basic)
Italian (basic)

Manuscript in preparation

Protein tyrosine *O*-glycosylation: molecular insights into a rather unexplored prokaryotic glycosylation system

Kristof Zarschler[a], Bettina Janesch[a], Martin Pabst[b], Friedrich Altmann[b], Paul Messner[a]*, and Christina Schäffer[a]*

Departments of [a]NanoBiotechnology and [b]Chemistry, University of Natural Resources and Applied Life Sciences (BOKU), A-1190 Vienna, Austria.

Running title: Protein tyrosine *O*-glycosylation in *Paenibacillus alvei*

* Corresponding authors

Mailing address: Department of NanoBiotechnology, University of Natural Resources and Applied Life Sciences Vienna, Muthgasse 11, A-1190 Vienna, Austria

Phone: +43-1-47654 ext. 2200

Fax: +43-1-4789112

E-mail: paul.messner@boku.ac.at

christina.schaeffer@boku.ac.at

ABSTRACT

Protein glycosylation is a frequent and heterogeneous posttranslational modification in all domains of life. Whereas *N*-glycosylation to asparagine and *O*-glycosylation to serine or threonine residues has been studied in great detail, only few data is available about *O*-glycosidic linkages to tyrosine. In this study, we describe the identification and characterization of a prokaryotic protein tyrosine *O*-glycosylation system. In the Gram-positive, mesophilic bacterium *Paenibacillus alvei* CCM 2051^T, a polysaccharide consisting of [→3)-β-D-Gal*p*-(1[α-D-Glc*p*-(1→6)] →4)-β-D-Man*p*NAc-(1→] repeating units is *O*-glycosidically linked via an adaptor with the structure -[GroA-2→OPO_2→4-β-D-Man*p*NAc-(1→4)] →3)-α-L-Rha*p*-(1→3)-α-L-Rha*p*-(1→3)-α-L-Rha*p*-(1→3)-β-D-Gal*p*-(1→ to specific tyrosine residues of the S-layer protein SpaA. A ~24.3-kb S-layer glycosylation (*slg*) gene cluster, which is required for the biosynthesis of this glycan chain, encodes 18 open reading frames whose translation products are involved in the biosynthesis of nucleotide-activated monosaccharides, in assembly, in export as well as in transfer of the polysaccharide to the S-layer target protein. All genes of the cluster, except those encoding the nucleotide sugar biosynthesis enzymes and the ABC transporter integral transmembrane protein, were disrupted by insertion of the mobile group II intron LI.LtrB, and S-layer glycoproteins produced in mutant backgrounds were analyzed by mass spectrometry. There is evidence that the glycan chain is synthesized in a process comparable to the ABC transporter-dependent pathway of lipopolysaccharide O-polysaccharide biosynthesis. Furthermore, with WsfB, we have identified an oligosaccharyl:protein transferase required for the covalent β-D-Gal→Tyr linkage of the complete glycan chain.

INTRODUCTION

Covalent attachment of glycans to the protein backbone via the amide nitrogen of an asparagine residue (*N*-glycosylation) or via the hydroxyl group of serine or threonine (*O*-glycosylation) has been reported for many natural glycoproteins (1). In contrast, as a rare event, in insect larvae (2, 3) as well as in glycogenin of glycogen-containing eukaryotic cells (4, 5), an *O*-glycosidic linkage between a tyrosine residue and α-D-glucose has been observed. In prokaryotes, *O*-glycosidic linkages of glycans via β-D-galactose or β-D-glucose residues to the phenolic OH-group of tyrosine were discovered as completely new types of linkage in the S-layer glycoproteins of

Paenibacillus alvei, *Thermoanaerobacter thermohydrosulfuricus* and *Thermoanaerobacterium thermosaccharolyticum* strains, respectively (6-12). As a specific property of S-layer constituents, these S-layer glycoproteins self-assemble into two-dimensional crystalline arrays on the supporting cell envelope (13), covering the bacterium completely. The glycan chains protrude from the cell surface, comparable to the LPS coating of Gram-negative bacteria (14). For several S-layer glycoprotein carrying bacteria, polycistronic *S*-layer *g*lycosylation (*slg*) gene clusters with a size of ~16 to ~25 kb have been identified and sequenced (15). They include nucleotide sugar pathway genes that are arranged consecutively, glycosyltransferase genes, glycan processing genes, and transporter genes, all of them exhibiting high homology with components involved in the biosynthesis of different bacterial surface polysaccharides (16).

In the mesophilic, Gram-positive bacterium *Paenibacillus alvei* CCM 2051^T, the S-layer *O*-glycan is a polymeric branched heterosaccharide of, on average, 22 [→3)-β-D-Gal*p*-(1[α-D-Glc*p*-(1→6)]→4)-β-D-Man*p*NAc-(1→] repeating units linked via an adaptor with the structure -[GroA-2→OPO_2→4-β-D-Man*p*NAc-(1→4)]→3)-α-L-Rha*p*-(1→3)-α-L-Rha*p*-(1→3)-α-L-Rha*p*-(1→3)-β-D-Gal*p*-(1→ to specific tyrosine residues of the S-layer protein backbone (6, 7).

Recently, we have identified the initiation enzyme of S-layer glycan biosynthesis, WsfP, as part of a *slg* gene cluster of *P. alvei* CCM 2051^T and we have developed a gene disruption system for this organism (17).

In the present study, we report the determination of the nucleotide sequence of the complete *slg* gene cluster of *P. alvei* CCM 2051^T and the genetic assignment of these genes. Based on the developed gene disruption system, we created insertional mutants for nine of the 18 genes located in the *slg* gene cluster, thereby proposing a transport mechanism for the glycan across the cytoplasmic membrane and identifying the oligosaccharyl:protein transferase (OTase) responsible for the transfer of the glycan chain to specific tyrosine residues of the target protein SpaA. Furthermore, we propose a working model for the S-layer glycan biosynthesis route in *P. alvei* CCM 2051^T based on the observed similarity between the proteins encoded by genes in the *slg* gene cluster and database entries, as well as on the effects of the disruption of selected genes of the gene cluster on S-layer glycan composition as determined by mass spectrometry.

MATERIALS AND METHODS

Bacterial strains and growth conditions. Bacterial strains, plasmids, and primers are listed in Tables 1 and 2. *P. alvei* CCM 2051^T was obtained from the Czech Collection of Microorganisms (CCM, Brno, Czech Republic) and was cultivated at 37°C and 200 rpm in Luria-Bertani (LB) broth or on LB agar plates supplemented with 10 µg/ml chloramphenicol (Cm), when appropriate. *Escherichia coli* DH5α (Invitrogen, Lofer, Austria) was grown in LB broth at 37°C supplemented with 30 µg/ml Cm, when appropriate.

Analytical and general methods. Genomic DNA of *P. alvei* CCM 2051^T was isolated as described recently (17). Restriction and cloning enzymes were purchased from Invitrogen. The MinElute gel extraction kit (Qiagen, Vienna, Austria) was used to purify DNA fragments from agarose gels, and the MinElute reaction cleanup kit (Qiagen) was used to purify digested oligonucleotides and plasmids. Plasmid DNA from transformed cells was isolated with the Plasmid Miniprep kit (Qiagen). Agarose gel electrophoresis was performed as described elsewhere (18). Transformation of *E. coli* DH5α was done according to the manufacturer's protocol (Invitrogen). Transformands were screened by *in situ* PCR using RedTaq ReadyMix PCR mix (Sigma-Aldrich, Vienna, Austria) and recombinant clones were analyzed by restriction mapping and confirmed by sequencing (Agowa, Berlin, Germany). Transformation of *P. alvei* CCM 2051^T was performed as described recently (17). Sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) was carried out according to a standard protocol (19) using a Protean II electrophoresis apparatus (Bio-Rad, Vienna, Austria). Protein bands were visualized with Coomassie Brilliant Blue G 250 staining reagent. The isolation and purification of S-layer glycoprotein essentially followed published methods (20).

PCR and DNA sequencing. Primers for PCR and DNA sequencing were purchased from Invitrogen and PCR conditions were optimized for each primer pair. PCR was performed using the Herculase® II Fusion DNA Polymerase (Stratagene, La Jolla, CA) and the thermal cycler My Cycler^{TM} (Bio-Rad). Amplification products were purified using the MinElute PCR purification kit (Qiagen). For the identification of the genes responsible for dTDP-L-rhamnose biosynthesis, the highly conserved seven

amino acid stretch TDYVFDG of RmlD was used for the design of the degenerate oligonucleotide primer proof_RmlD_for. For sequence determination of the *slg* gene cluster chromosome walking was applied as previously described (21, 22).

RT-PCR. Total RNA was extracted from *P. alvei* CCM 2051^{T} using the RNeasy Protect Bacteria Mini Kit (Qiagen) and subsequently treated with RNase-free DNase I (Fermentas, St. Leon-Rot, Germany) to remove DNA contamination. First strand cDNA was synthesized utilizing the Revert Aid^{TM} Premium Reverse Transcriptase (Fermentas) according to the manufacturer's instructions using a reverse primer specific for *wsfB* (1f), *wsfA* (4r), *wsfF* (8r) or *pcrB* (9r). After termination of the reaction by heating at 85°C for 5 min, one tenth of each cDNA reaction mixture was used as template for PCR using the Phusion™ High-Fidelity DNA Polymerase (New England Biolabs, Frankfurt/Main, Germany). PCR reactions were carried out with primer pairs annealing to *wsfB*/*galE* (1f/1r), *galU*/*wzm* (2f/2r), *wzt*/*wsfA* (3f/3r), *wsfA* (4f/4r), *wsfC*/*wsfD* (5f/5r), *wsfD*/*wsfE* (6f/6r), *rmlB*/*wsfF* (8f/8r), and *wsfH*/*pcrB* (9f/9r). As a positive control, genomic DNA was used, whereas DNase I-treated RNA without the cDNA-generating step served as a control for contamination of total RNA with chromosomal DNA. PCR products were analyzed by agarose gel electrophoresis.

Sequence analysis. Nucleotide and protein sequences were analyzed using the BLASTN and BLASTP sequence homology analysis tools (National Center for Biotechnology Information, Bethesda, MD). Open reading frames in the DNA sequence were identified by using the Clone Manager Professional Suite (SECentral, Cary, NC) and the ORF Finder analysis tool (National Center for Biotechnology Information). The TMHMM Server v. 2.0 transmembrane prediction program and the SignalP 3.0 Server (Center for Biological Sequence Analysis, Technical University of Denmark, Lyngby, Denmark) were used to identify putative protein transmembrane-spanning domains and the presence and location of signal peptide cleavage sites, respectively. The GC content of the entire *slg* gene cluster was determined using the GC Content and GC Skew program (Nano+Bio-Center, University of Kaiserslautern, Germany). For *in silico* reverse translation, the Sequence Manipulation Suite was used (23). Bacterial promoters, transcriptional terminators, operons and genes were predicted by the BProm and FindTerm modules of the FGenesB gene prediction program in Molquest software (SoftBerry Inc., Mount Kisco, NY). The presence of

conserved motifs in a given protein sequence was analyzed by the Pfam protein families database (24). Physical and chemical parameters for a given protein were calculated using the ProtParam tool (25).

Gene knockout. Specific disruption of nine genes located in the *slg* gene cluster of *P. alvei* CCM 2051^T was performed as described recently (17). The Ll.LtrB targetron of pTT_*wsfP*1176 was retargeted prior to transformation into *P. alvei* CCM 2051^T. Therefore, identification of potential insertion sites and design of PCR primers for the modification of the intron RNA was accomplished by a computer algorithm (www.Sigma-Aldrich.com/Targetronaccess). For each gene, insertion sites were chosen based on their location and intron insertion efficiency and modifications of the intron RNA sequences were introduced via PCR by primer-mediated mutation (Table 3). The retargeted Ll.LtrB targetron was subsequently digested with *Hin*dIII and *Bsr*GI and ligated into pTT_*wsfP*1176 digested with the same restriction enzymes, thereby replacing the *wsfP* targetron. Creation of *P. alvei* CCM 2051^T gene knockout mutants and confirmation of intron insertion was achieved as described recently (17). All mutant strains were analyzed for the migration behavior of SpaA by Coomassie Blue staining after electrophoresis of biomass on SDS-PA gels (11).

Protein elution from SDS-PAGE and pronase digest. Glycan structures linked to the S-layer protein SpaA of *P. alvei* CCM 2051^T wild-type and mutant cells were analyzed by mass spectrometry. Aliquots of biomass from wild-type and mutant cells were run on a 10% PA gel and stained with Coomassie Brilliant Blue G 250. Relevant SDS-PA gel bands were excised from the gel and destained (26). Destained gel pieces were minced into small particles and loaded onto a BioRad Model 422 electro eluter. After elution from gel, the protein was dialyzed against 10 mM ammonia formate puffer and precipitated using 5 volumes of acetone (-20°C, 1 h). 200 µl of 0.15 M Tris-HCl buffer, pH 7.8 (containing 1 mM $CaCl_2$ and 0.02% NaN_3) was added to the dried protein, followed by proteolytic digestion using 1 µg pronase for 24 h at 37°C. Subsequently, an additional amount of 1 µg pronase was added and incubation was continued for 24 h.

Glycopeptides were enriched using a PGC-SPE cartridge (Thermo Scientific, Vienna, Austria, 10 mg) according to (27), except using 150 mM ammonia formate buffer, pH

9.0, in 60% acetonitrile for elution. Prior to LC-MS analysis the samples were dried under vacuum and dissolved in 20 µl of distilled water.

Glycopeptide LC-ESI-MS/MS. Analysis of glycopeptides was performed by positive-ion LC-ESI-MS/MS using a 50 x 0.32 mm porous graphitic carbon column (Thermo Scientific) (28). A flow rate of 5 µl was maintained with a Dionex Ultimate 3000 cap flow system using 150 mM ammonia formate buffer (pH 9) as solvent A and acetonitrile as solvent B.

Glycopeptides were eluted from the separation column using a gradient form 5 to 45% of solvent B over 25 min. Analysis was carried out with a Waters Q-TOF Ultima Global mass spectrometer with standard ESI-source and a MassLynx V4.0 SP4 software for evaluation of obtained peaks. For online MSMS experiments the CE was set to 72 with a LM Res of 8 and a HM Res of 11. Capillary voltage, MS profile, cone voltage, RF LENS 1 setting and ESI probe adjustment was optimized to gain maximum signal intensity.

RESULTS

Identification of the *P. alvei* CCM 2051^T S-layer glycosylation (*slg*) gene cluster. Since L-rhamnose represents the main constituent of the adaptor region of the S-layer glycan of *P. alvei* CCM 2051^T, we have chosen the genes responsible for dTDP-L-rhamnose biosynthesis as suitable candidates for the design of degenerate primers to achieve entry into a putative *slg* gene cluster. Using a genomic DNA bank of *P. alvei* CCM 2051^T as a template for the PCR amplification reaction and the primer proof_RmlD_for, a ~1.4 kb DNA fragment was obtained. After confirmation of the presence of the *rmlD* gene, sequencing of upstream and downstream regions by chromosome walking revealed the presence of 18 genes contained in a 24.3 kb *slg* gene cluster, coding for components of the putative S-layer protein glycosylation machinery (Figure 1). Most of the putative gene products encoded by the assigned ORFs showed high homology to proteins involved in the biosynthesis of bacterial surface polysaccharides. Based on these sequence similarities, putative biological functions have been assigned to almost all of the genes of the *slg* gene cluster (Table 4).

Genetic characterization of the *slg* gene cluster. The *slg* gene cluster of *P. alvei* CCM 2051^T is flanked by genes coding for enzymes involved in lantibiotic biosynthesis (29) and lipid / lipoteichoic acid biosynthesis (30), respectively. Using the *E. coli* σ^{70} promoter consensus sequence, putative promoter sites were recognized along the sequence, which were located upstream of the *wsfB*, the *galE*, the *galU*, the *wzm*, the *wsfC*, the *wsfD*, and the *wsfE* genes. Downstream of *wsf*B, *galE*, *galU*, and *wsfH* putative ρ-independent bacterial terminators were identified. The prediction of transcription units resulted in a total number of seven, as indicated in Figure 1. This is in clear contrast to the closed spacing from ORF *wzm* to *wsfH* and their identical transcriptional direction, indicating that these ORFs are transcribed as a single operon. To identify specific mRNA(s) of the gene cluster, total RNA of *P. alvei* CCM 2051^T was isolated and reverse transcribed into cDNA. The amplification of cDNA using primer combinations spanning the regions *wsfB*/*galE*, *galU*/*wzm*, *wzt*/*wsfA*, *wsfA*, *wsfC*/*wsfD*, *wsfD*/*wsfE*, *rmlB*/*wsfF*, and *wsfH*/*pcrB* revealed that the *slg* gene cluster is transcribed as a polycistronic unit starting with *galU* and ending with *wsfH* (Figure 2). No PCR products were obtained, when primers annealing to *wsfB*/*galE* (1f/1r) and cDNA reverse transcribed with primer 1f and when primers annealing to *wsfH*/*pcrB* (9f/9r) and cDNA reverse transcribed with primer 1f were used (data not shown). This observation indicates that *wsfB* and *galE* are transcribed separately and that *pcrB* is not part of the *slg* gene cluster.

The characteristics of the genes within the analyzed region will be discussed in the order in which they are present, starting with *wsfB*.

wsfB. The gene product of *wsfB* (Figure 3) contains 12 potential transmembrane domains and a conserved Wzy_C motif characteristic for O-antigen ligases and oligosaccharyltransferases (OTases), responsible for the transfer of undecaprenyl-pyrophosphate-linked sugars to the LPS core or a target protein, respectively (31). OTases, such as PglL of *Neisseria meningitidis* and PilO of *Pseudo-monas aeruginosa*, show only low levels of amino acid similarity, but possess similar transmembrane topology and small regions of high homology (32). Both OTases exhibit relaxed glycan specificity, but require the translocation of the corresponding undecaprenyl-pyrophosphate-linked oligosaccharide substrates into the periplasm (33). As PglL, WsfB possesses in its carboxyterminal part a tetratricopeptide repeat

(TPR) described as mediator for protein-protein interactions and the assembly of multiprotein complexes (34).
The similar membrane spanning topology and the presence of the Wzy_C motif in WsfB supports the assumption that this enzyme belongs to the family of OTases transferring the S-layer glycan chain from the lipid carrier to certain tyrosine residues in SpaA.

galE and *galU*. The translation products of these two genes are homologous to the UDP-glucose 4-epimerase (GalE) and the UTP-glucose-1-phosphate uridylyltransferase (GalU), respectively. GalE catalyses the interconversion between UDP-glucose and UDP-galactose, and GalU mediates for the transfer of UTP to glucose-1-phosphate resulting in UDP-glucose. The existence of galactose and glucose in the S-layer *O*-glycan of *P. alvei* CCM 2051^T suggests that GalE and GalU are involved in the biosynthesis of the sugar precursors UDP-galactose and UDP-glucose.

wzm and *wzt*. The deduced 232 and 434 amino acid proteins encoded by *wzm* and *wzt*, respectively, reveal high similarity to proteins of the ABC-2 transporter family, involved in the transport of bacterial surface polysaccharides to the cell surface (35). The presence of six transmembrane domains in the putative translation product of *wzm* suggests that this protein is the integral membrane component of the transporter. An ATP binding site and an ATP transporter signature motif identified in the putative translation product of *wzt* indicate the involvement of this protein in the transport of a substrate across the cytoplasmic membrane (36, 37). Its extended carboxyterminal part obviously contains an O-polysaccharide binding domain determining the transporter's substrate specificity as observed for the polymannan O-antigenic polysaccharides of *E. coli* O8 and O9a. Furthermore, several amino acids identified in *wzt* of *E. coli* O9a to be critical for binding and export of O-antigenic polysaccharide were also found in the homologous protein of *P. alvei* CCM 2051^T (38, 39). For example, G333 and G387 located in the carbohydrate-binding pocket of the *E. coli* protein correspond to G334 and G387 of Wzt of *P. alvei* CCM 2051^T (Figure 4).

wsfA. Throughout the whole gene, *wsfA* is homologous to genes coding for asparagine synthetase B (AsnB). This enzyme acts as a homodimer with each

monomer composed of a glutaminase domain, hydrolyzing glutamine to glutamic acid, and an ammonia and a asparagine synthetase domain, catalyzing the ATP-dependent conversion of aspartate to asparagine (40, 41). An *asnB* mutant of *Corynebacterium glutamicum* was isolated as a lysozyme- and temperature-sensitive mutant (42), and in *Mycobacterium smegmatis* AsnB is involved in natural resistance to rifampin, erythromycin, and novobiocin (43). Up to date, no specific function could be assigned to the gene product of *wsfA* in *P. alvei* CCM 2051^T.

The stop codon of *wsfA* overlaps with the putative start codon of *tagD*. Comparison of the translation product of *tagD* with proteins in the database showed a high degree of amino acid homology to the glycerol-3-phosphate cytidyltransferase (GCT) involved in the formation of CDP-glycerol and pyrophosphate from CTP and glycerol-3-phosphate (44, 45). The existence of a glyceric acid phosphate residue in the adapter saccharide of the S-layer *O*-glycan of *P. alvei* CCM 2051^T suggests, that TagD catalyzes the synthesis of the building block CDP-glycerol.

wsfC. The protein encoded by *wsfC* is the largest found in the *slg* gene cluster, coding for a tripartite transferase of 147.34 kDa. Two glycosyltransferase family 2 motifs were identified in the central and carboxyterminal part, while a single CDP-glycerol:poly(glycerophosphate) glycerophosphotransferase motif was found at the aminoterminal part of the protein. The aminoterminal region shows similarity to the TagB protein of *Bacillus subtilis* catalyzing, there, the incorporation of a single glycerol phosphate residue from CDP-glycerol to the nonreducing end of the membrane-bound undecaprenyl-phosphate-linked *N*-acetylmannosamine-β-(1,4)-*N*-acetylglucosamine-1-phosphate (46). The central part of WsfC is homologous to the glycosyltransferase LgtD of different *Rickettsia* strains. In *Haemophilus influenzae* and *Neisseria gonorrhoeae*, LgtD is involved in LPS and LOS biosynthesis, respectively, possessing acetylgalactosaminyltransferase and galactosyltransferase activity (47-49). Sequence homology searches for the predicted amino acid sequences of the carboxyterminal part of WsfC showed homology with cyanobacterial and archaeal glycosyltransferases, with AglG being involved in the *N*-glycosylation of the *Haloferax volcanii* S-layer glycoprotein, showing hexuronic acid transferase activity (50).

wsfD. The *wsfD* gene product contains nine transmembrane domains and is similar to uncharacterized transmembrane proteins of various Gram-positive bacteria. In the aminoterminal part, a dolichyl-phosphate-mannose-protein mannosyltransferase domain spanning seven transmembrane domains was identified. In fungi, dolichyl-phosphate-mannose-protein mannosyltransferases (PMTs) are integral endoplasmic reticulum (ER) membrane proteins being responsible for initiation of protein *O*-mannosylation in the ER. A PMT of *Saccharomyces cerevisiae*, ScPmt1p, is an integral membrane glycoprotein of 817 amino acids, located in the ER and catalyzing the transfer of mannose from the lipid carrier Dol-P-β-D-mannose to serine/threonine residues of specific protein acceptors (51-53). As ScPmt1p, WsfD possesses an aminoterminal loop, a large hydrophilic loop and a carboxyterminal region all facing the ER or the external face of the cytoplasmic membrane, respectively (54, 55). Although no mannose residue was found in the S-layer *O*-glycan of *P. alvei* CCM 2051^{T}, WsfD could be involved in the transfer of a hexose residue from a lipid carrier to the glycan chain.

wsfE. The 364-amino acid translation product of *wsfE* shows high similarity to several glycosyltransferases of various *Clostridium*, *Pseudomonas* and *Vibrio* strains (56-58), and it contains a glycosyltransferases group 1 motif (GT1_wbuB_like). In *E. coli*, WbuB is involved in the biosynthesis of the O26 O-antigen, thereby acting as an *N*-acetyl-L-fucosamine (L-FucNAc) transferase (59).

rmlA, rmlC, rmlB, and rmlD. The *rmlACBD* gene products show a high degree of amino acid homology to the RmlACBD proteins involved in the biosynthesis of dTDP-L-rhamnose in different *Geobacillus* strains (16, 60) and in other bacteria (61). Since the adaptor oligosaccharide of the S-layer *O*-glycan of *P. alvei* CCM 2051^{T} contains three L-rhamnose residues, it is conceivable that RmlACBD are providing the nucleotide-activated building block dTDP-L-rhamnose. The putative start codon of *rmlD* overlaps with the stop codon of *rmlB*.

wsfF and *wsfG*. The stop codon of *wsfF* overlaps with the putative start codon of *wsfG*. The protein products of these two genes possess a glycosyltransferase family 2 motif in their aminoterminal parts. The deduced protein sequence of WsfF shows significant homology to the putative sugar transferase WsdG of *Aneurini-*

bacillus thermoaerophilus DSM 10155/G^{+} and to certain rhamnosyltransferases (15). The protein encoded by *wsfG* is highly similar to the β1,2-rhamnosyltransferase WsaF of *Geobacillus stearothermophilus* NRS 2004/3a, transferring an L-rhamnose residue to the linkage sugar galactose during S-layer glycan biosynthesis in this strain (62). For WsfG, a single transmembrane-spanning domain was found at the carboxyterminal part of the protein.

wsfP. The UDP-Gal:phosphoryl-polyprenol Gal-1-phosphate transferase WsfP was recently identified as the initiation enzyme of S-layer glycan biosynthesis in *P. alvei* CCM 2051^{T} (17).

wsfH. The predicted protein named WsfH shows high similarity to several glycosyltransferases found in different cyanobacteria (63, 64). A cytoplasmic glycosyltransferase family 2 motif and a single transmembrane-spanning domain were found at the amino- and carboxyterminal part of the protein, respectively. Since WsfH also shares amino acid similarity with several polyprenyl-phosphate β-D-glucosyltransferases it is likely that it is responsible for the intracellular transfer of a glucose residue to the membrane-associated lipid carrier undecaprenyl-pyrophosphate.
The region downstream of *wsfH* contains a putative ρ-independent bacterial terminator followed by a putative promoter site allowing the transcription of a gene coding for PcrB, an enzyme involved in lipid / lipoteichoic acid biosynthesis (30).

S-layer protein glycosylation in *slg* gene cluster mutant backgrounds. To investigate the role of individual genes from the *P. alvei* CCM 2051^{T} *slg* gene cluster, insertional knockout mutants in nine genes of the gene cluster were constructed. All mutants still produced the S-layer protein SpaA as detected by SDS-PAGE followed by Coomassie staining (Figure 5). However, for SpaA produced in–mutants with an insertion of Ll.LtrB in the *wsfA*, *wsfB*, *wsfC*, *wsfE*, *wsfF*, *wsfG* or *wzt* gene, only a single protein band corresponding to the non-glycosylated S-layer protein could be detected by SDS-PAGE. These findings suggest that mutations of *wsfA*, *wsfB*, *wsfC*, *wsfE*, *wsfF*, *wsfG* and *wzt* genes had significant effects on SpaA glycosylation in *P. alvei* CCM 2051^{T}. In contrast, knockout mutants in *wsfD* and *wsfH* genes resulted

in production of glycosylated SpaA protein of similar mobility to that produced from cells containing the wild-type *slg* gene cluster.

To ensure that no polar effects on downstream gene expression caused by the intron insertion as described by Rodriguez *et al.* (65) had occurred, but rather the inactivation of the target gene itself is responsible for the lack of S-layer glycosylation in the *wsfA*, *wsfB*, *wsfC*, *wsfE*, *wsfF*, *wsfG* or *wzt* mutants, we analyzed the mRNA of the *wsfE*::Ll.LtrB mutant for these effects. Using the RT-PCR approach described above, no differences between wild-type and *wsfE*::Ll.LtrB cells could be observed, indicating, that despite the intron insertion the polycistronic mRNA is not interrupted (data not shown).

Structural characterization of S-layer glycans produced in *slg* gene cluster mutant backgrounds. The identification of the glycan structures linked to the S-layer protein SpaA was accomplished by mass spectrometry of peptides derived from pronase digestion. While in the S-layer protein of the *wsfA*, *wsfB*, *wsfC*, *wsfE*, *wsfF*, *wsfG* or *wzt* mutants no such glycopeptides could be detected at all, the glycopeptides of the SpaA protein produced in either *wsfD* or *wsfH* mutant strains yielded identical MSMS results (Figure 6). The on-line MSMS spectrum of the SpaA glycopeptides of wild-type *P. alvei* CCM 2051^{T} confirms the known branched structure of the repeating units (Figure 6A). The three peaks at 528 Da, 1055 Da, and 1582 Da represent one, two or three repeating units of two hexoses and one *N*-acetylhexosamine residue, each.

The corresponding on-line MSMS spectrum of the *wsfD*::Ll.LtrB (and *wsfH*::Ll.LtrB) mutant strain shows four peaks at 366 Da, 731 Da, 1097 Da and 1462 Da, which are consistent with one to four repeating units containing one hexose and one *N*-acetylhexosamine residue.

This observation shows the absence of the α1,6-linked glucose residues of the repeating units in the glycosylated peptides of SpaA protein produced in *wsfD*::Ll.LtrB and *wsfH*::Ll.LtrB mutant strains, thus indicating the involvement of WsfD and WsfH in the glucosylation of the *N*-acetylmannosamine residues of the repeating units. However, no changes in the structure of the adaptor region were detected (data not shown).

DISCUSSION

In this study, we identified the *slg* gene cluster of the mesophilic, Gram-positive bacterium *P. alvei* CCM 2051^T encoding the *O*-glycosylation of tyrosine residues of the S-layer protein of this organism. The sequenced ~24.3 kb region contains 18 genes, of which the derived protein sequences show homology to proteins involved in the biosynthesis of different bacterial surface polysaccharides, such as lipopolysaccharides, exopolysaccharides, and capsule polysaccharides. Both, the observed similarity of the putative Wsf proteins with database entries of enzymes involved in bacterial polysaccharide biosyntheses, as well as the disruption of their corresponding genes gave first insights into the S-layer glycan biosynthesis pathway of *P. alvei* CCM 2051^T (Figure 7).

Seven ORFs located in the *slg* gene cluster are involved in the biosynthesis of nucleotide-activated monosaccharides. Next to *galE*, converting UDP-glucose to UDP-galactose, *galU* transferring UTP to glucose-1-phosphate, resulting in UDP-glucose, is present. While *tagD* is involved in the formation of CDP-glycerol, the four *rml* genes code for the biosynthesis of dTDP-L-rhamnose. Since the S-layer glycan contains glucose, galactose, phosphoglyceric acid, and rhamnose, the presence of these genes in the *slg* gene cluster is not surprising. However, no genes for the biosynthesis of nucleotide-activated *N*-acetylmannosamine are located in the *slg* gene cluster. This observation confirms the assumption, that housekeeping genes are additionally required for S-layer glycan biosynthesis (16).

As recently described and depicted in Figure 7A, the UDP-Gal:phosphoryl-polyprenol Gal-1-phosphate transferase WsfP acts as the initiation enzyme of S-layer glycan biosynthesis in *P. alvei* CCM 2051^T (17).

The *slg* gene cluster encodes two rhamnosyltransferases (WsfF and WsfG) which are obviously involved in the assembly of the L-rhamnose containing adaptor saccharide of the S-layer *O*-glycan, with WsfG transferring an L-rhamnose residue onto the linkage sugar galactose and WsfF acting as an α-L-rhamnose α-1,3-L-rhamnosyltransferase (Figure 7B). Since WsfE is related to the aminosugar transferase WbuB, it might be responsible for the addition of an *N*-acetylmannosamine residue to the growing glycan chain (Figure 7C).

On the basis of the observed similarity in the database, the proposed function of the tripartite transferase WsfC seems to be the transfer of a single glycerol phosphate from CDP-glycerol to the *N*-acetylmannosamine residue of the adaptor

oligosaccharide and of a galactose residue to the *N*-acetylmannosamine residues of the repeating units. Since this enzyme exhibits a third transferase domain, it may also catalyze the incorporation of another sugar into the growing glycan chain (Figure 7C).

The S-layer glycoproteins of the mutant strains *wsfD*::Ll.LtrB and *wsfH*::Ll.LtrB show identical migration behavior in SDS-PAGE compared to wild-type cells. MS analysis of the S-layer *O*-glycan of both mutants showed the lack of glucose residues, being part of the each repeating unit of the mature glycan, suggesting that both enzymes, WsfD and WsfH, are involved in the process of glucose addition to the glycan chain. Due to the similarity of WsfD to fungal Pmts and of WsfH to several polyprenyl-phosphate β-D-glucosyltransferases, we assume, that WsfH transfers a glucose residue to undecaprenyl-pyrophosphate in the cytoplasm, the lipid carrier is then re-orientated to the external face of the cytoplasmic membrane, and WsfD adds the glucose residue to the exported glycan chain (Figure 7D).

The identification of an ABC transporter system (Wzm and Wzt) and the loss of S-layer glycosylation in the *wzt*::Ll.LtrB mutant corroborate the assumption of the ATP hydrolysis-driven export of the undecaprenyl-pyrophosphate-linked glycan chain to the external face of the cytoplasmic membrane comparable to the ABC-transporter-dependent pathway of the LPS O-polysaccharide biosynthesis. According to this pathway, glycan chain extension is achieved by processive addition of sugar residues to the nonreducing terminus of the undecaprenyl-pyrophosphate-linked growing chain. Although not yet detected in *P. alvei* CCM 2051^{T}, nonreducing terminal modifications, such as 2-*O*-methyl groups, were described as chain length termination signal recognized by the carboxyterminal domain of Wzt (38). The polymer is then exported through the cytoplasmic membrane by the ABC transporter for ligation, independent of the existence of a Wzx flippase or Wzy polymerase homolog (66, 67). After the addition of glucose to the *N*-acetylmannosamine residue of each repeating unit by WsfD, the glycan chain is transferred *en bloc* from the lipid carrier to specific tyrosine residues of SpaA by the OTase WsfB (Figure 7E). This conclusion is based on the presence of the conserved Wzy_C motif and the similar transmembrane topology of the excellently characterized OTases WaaL, PglL and PilO, as well as on the loss of S-layer glycosylation in the *wsfB*::Ll.LtrB mutant.

The role of WsfA in S-layer glycosylation biosynthesis remains still unclear. Since the *wsfA*::Ll.LtrB mutant shows an altered S-layer migration due to the loss of the SpaA-

linked glycan chain, its involvement in the glycosylation process is evident, but needs to be further investigated.

Although several putative bacterial promoters and terminators have been identified by different prediction programs, most of the genes coding for GalU, Wzm, Wzt, WsfA, TagD, WsfCDE, RmlACBD and WsfFGPH are transcribed by a single polycistronic mRNA. However, the UDP-glucose 4-epimerase GalE and the OTase WsfB are transcribed independently. A common phenomenon of bacterial polysaccharide biosynthesis gene clusters is the low % G+C content compared to the respective bacterial genome as a whole (14, 15). For individual genes of the described *slg* gene cluster the % G+C content ranges between 25 and 43%, whereas the genome of *P. alvei* CCM 2051^T has a DNA base composition of 44.6 or 46.2%, depending on the method of its determination (68).

In conclusion, the current report describes the identification, annotation and characterization of the *slg* gene cluster of *P. alvei* CCM 2051^T involved in tyrosine *O*-glycosylation of the S-layer protein including identification of the corresponding ABC transporter and the OTase system.

ACKNOWLEDGEMENTS

We thank Andrea Scheberl and Sonja Zayni for excellent technical assistance. Financial support came from the Austrian Science Fund, project P20745-B11 (to P.M.) and projects P19047-B12 and P20605-12 (to C.S.), and the Hochschul-jubiläumsstiftung der Stadt Wien, project H-02229-2007 (to K.Z.).

REFERENCES

1. **Varki, A., R. Cummings, J. Esko, H. Freeze, G. Hart, and J. Marth (ed.).** 1999. Essentials of Glycobiology, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
2. **Chen, P. S., H. K. Mitchell, and M. Neuweg.** 1978. Tyrosine glucoside in *Drosophila busckii*. Insect Biochemistry **8:**279-286.
3. **Kramer, K. J., T. L. Hopkins, R. F. Ahmed, D. Mueller, and G. Lookhart.** 1980. Tyrosine metabolism for cuticle tanning in the tobacco hornworm,

Manduca sexta (L.) and other *Lepidoptera*: identification of β-D-glucopyranosyl-*O*-L-tyrosine and other metabolites. Archives of Biochemistry and Biophysics **205:**146-155.

4. **Aon, M. A., and J. A. Curtino.** 1985. Protein-bound glycogen is linked to tyrosine residues. Biochemical Journal **229:**269-272.
5. **Rodriguez, I. R., and W. J. Whelan.** 1985. A novel glycosyl-amino acid linkage: rabbit-muscle glycogen is covalently linked to a protein via tyrosine. Biochemical and Biophysical Research Communications **132:**829-836.
6. **Altman, E., J. R. Brisson, P. Messner, and U. B. Sleytr.** 1991. Structure of the glycan chain from the surface layer glycoprotein of *Bacillus alvei* CCM 2051. Biochemistry and Cell Biology **69:**72-78.
7. **Messner, P., R. Christian, C. Neuninger, and G. Schulz.** 1995. Similarity of "core" structures in two different glycans of tyrosine-linked eubacterial S-layer glycoproteins. Journal of Bacteriology **177:**2188-2193.
8. **Christian, R., P. Messner, C. Weiner, U. B. Sleytr, and G. Schulz.** 1988. Structure of a glycan from the surface-layer glycoprotein of *Clostridium thermohydrosulfuricum* strain L111-69. Carbohydrate Research **176:**160-163.
9. **Bock, K., J. Schuster-Kolbe, E. Altman, G. Allmaier, B. Stahl, R. Christian, U. B. Sleytr, and P. Messner.** 1994. Primary structure of the *O*-glycosidically linked glycan chain of the crystalline surface layer glycoprotein of *Thermoanaerobacter thermohydrosulfuricus* L111-69. Galactosyl tyrosine as a novel linkage unit. Journal of Biological Chemistry **269:**7137-7144.
10. **Messner, P., R. Christian, J. Kolbe, G. Schulz, and U. B. Sleytr.** 1992. Analysis of a novel linkage unit of *O*-linked carbohydrates from the crystalline surface layer glycoprotein of *Clostridium thermohydrosulfuricum* S102-70. Journal of Bacteriology **174:**2236-2240.
11. **Altman, E., C. Schäffer, J. R. Brisson, and P. Messner.** 1995. Characterization of the glycan structure of a major glycopeptide from the surface layer glycoprotein of *Clostridium thermosaccharolyticum* E207-71. European Journal of Biochemistry **229:**308-315.
12. **Schäffer, C., K. Dietrich, B. Unger, A. Scheberl, F. A. Rainey, H. Kählig, and P. Messner.** 2000. A novel type of carbohydrate-protein linkage region in the tyrosine-bound S-layer glycan of *Thermoanaerobacterium thermosaccharolyticum* D120-70. European Journal of Biochemistry **267:**5482-5492.

13. **Sleytr, U. B., and P. Messner.** 2009. Crystalline Cell Surface Layers (S Layers), p. 89-98. *In* M. Schaechter (ed.), Encyclopedia of Microbiology, 3 ed, vol. 1. Academic Press, San Diego.
14. **Messner, P., K. Steiner, K. Zarschler, and C. Schäffer.** 2008. S-layer nanoglycobiology of bacteria. Carbohydrate Research **343:**1934-1951.
15. **Novotny, R., A. Pföstl, P. Messner, and C. Schäffer.** 2004. Genetic organization of chromosomal S-layer glycan biosynthesis loci of *Bacillaceae*. Glycoconjugate Journal **20:**435-447.
16. **Novotny, R., C. Schäffer, J. Strauss, and P. Messner.** 2004. S-layer glycan-specific loci on the chromosome of *Geobacillus stearothermophilus* NRS 2004/3a and dTDP-L-rhamnose biosynthesis potential of *G. stearo-thermophilus* strains. Microbiology **150:**953-965.
17. **Zarschler, K., B. Janesch, S. Zayni, C. Schäffer, and P. Messner.** 2009. Construction of a gene knockout system for application in *Paenibacillus alvei* CCM 2051^{T}, exemplified by the S-layer glycan biosynthesis initiation enzyme WsfP. Applied and Environmental Microbiology **75:**3077-3085.
18. **Sambrook, J., E. F. Fritsch, and T. Maniatis.** 1989. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor,NY.
19. **Laemmli, U. K.** 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature **227:**680-685.
20. **Messner, P., and U. B. Sleytr.** 1988. Separation and purification of S-layers from Gram-positive and Gram-negative bacteria, p. 97-104. *In* I. C. Hancock and I. R. Poxton (ed.), Bacterial cell surface techniques. John Wiley & Sons, Chichester.
21. **Kneidinger, B., M. Graninger, and P. Messner.** 2001. Chromosome walking by cloning of distinct PCR fragments. Biotechniques **30:**248-249.
22. **Pilhofer, M., A. P. Bauer, M. Schrallhammer, L. Richter, W. Ludwig, K. H. Schleifer, and G. Petroni.** 2007. Characterization of bacterial operons consisting of two tubulins and a kinesin-like gene by the novel two-step gene walking method. Nucleic Acids Research **35:**e135.
23. **Stothard, P.** 2000. The sequence manipulation suite: JavaScript programs for analyzing and formatting protein and DNA sequences. Biotechniques **28:**1102-1104.

24. **Finn, R. D., J. Tate, J. Mistry, P. C. Coggill, S. J. Sammut, H. R. Hotz, G. Ceric, K. Forslund, S. R. Eddy, E. L. Sonnhammer, and A. Bateman.** 2008. The Pfam protein families database. Nucleic Acids Research **36:**D281-D288.

25. **Gasteiger, E., C. Hoogland, A. Gattiker, S. Duvaud, M. R. Wilkins, R. D. Appel, and A. Bairoch.** 2005. Protein identification and analysis tools on the ExPASy server, p. 571-607. *In* J. M. Walker (ed.), The Proteomics Protocols Handbook. Humana Press, Totowa, NJ.

26. **Stadlmann, J., M. Pabst, D. Kolarich, R. Kunert, and F. Altmann.** 2008. Analysis of immunoglobulin glycosylation by LC-ESI-MS of glycopeptides and oligosaccharides. Proteomics **8:**2858-2871.

27. **Packer, N. H., M. A. Lawson, D. R. Jardine, and J. W. Redmond.** 1998. A general approach to desalting oligosaccharides released from glycoproteins. Glycoconjugate Journal **15:**737-747.

28. **Pabst, M., and F. Altmann.** 2008. Influence of electrosorption, solvent, temperature, and ion polarity on the performance of LC-ESI-MS using graphitic carbon for acidic oligosaccharides. Analytical Chemistry **80:**7534-7542.

29. **Bierbaum, G., and H. G. Sahl.** 2009. Lantibiotics: mode of action, biosynthesis and bioengineering. Current Pharmaceutical Biotechnology **10:**2-18.

30. **Chen, A., D. Zhang, and C. D. Poulter.** 1993. (S)-geranylgeranylglyceryl phosphate synthase. Purification and characterization of the first pathway-specific enzyme in archaebacterial membrane lipid biosynthesis. Journal of Biological Chemistry **268:**21701-21705.

31. **Power, P. M., K. L. Seib, and M. P. Jennings.** 2006. Pilin glycosylation in *Neisseria meningitidis* occurs by a similar pathway to *wzy*-dependent O-antigen biosynthesis in *Escherichia coli*. Biochemical and Biophysical Research Communications **347:**904-908.

32. **Abeyrathne, P. D., C. Daniels, K. K. Poon, M. J. Matewish, and J. S. Lam.** 2005. Functional characterization of WaaL, a ligase associated with linking O-antigen polysaccharide to the core of *Pseudomonas aeruginosa* lipopolysaccharide. Journal of Bacteriology **187:**3002-3012.

33. **Faridmoayer, A., M. A. Fentabil, D. C. Mills, J. S. Klassen, and M. F. Feldman.** 2007. Functional characterization of bacterial oligosaccharyl-

transferases involved in *O*-linked protein glycosylation. Journal of Bacteriology **189:**8088-8098.

34. **D'Andrea, L. D., and L. Regan.** 2003. TPR proteins: the versatile helix. Trends in Biochemical Sciences **28:**655-662.
35. **Bronner, D., B. R. Clarke, and C. Whitfield.** 1994. Identification of an ATP-binding cassette transport system required for translocation of lipopoly-saccharide O-antigen side-chains across the cytoplasmic membrane of *Klebsiella pneumoniae* serotype O1. Molecular Microbiology **14:**505-519.
36. **Walker, J. E., M. Saraste, M. J. Runswick, and N. J. Gay.** 1982. Distantly related sequences in the alpha- and beta-subunits of ATP synthase, myosin, kinases and other ATP-requiring enzymes and a common nucleotide binding fold. EMBO Journal **1:**945-951.
37. **Rocchetta, H. L., and J. S. Lam.** 1997. Identification and functional characterization of an ABC transport system involved in polysaccharide export of A-band lipopolysaccharide in *Pseudomonas aeruginosa*. Journal of Bacteriology **179:**4713-4724.
38. **Cuthbertson, L., M. S. Kimber, and C. Whitfield.** 2007. Substrate binding by a bacterial ABC transporter involved in polysaccharide export. Proceedings of the National Academy of Sciences of the United States of America **104:**19529-19534.
39. **Cuthbertson, L., J. Powers, and C. Whitfield.** 2005. The C-terminal domain of the nucleotide-binding domain protein Wzt determines substrate specificity in the ATP-binding cassette transporter for the lipopolysaccharide O-antigens in *Escherichia coli* serotypes O8 and O9a. Journal of Biological Chemistry **280:**30310-30319.
40. **Milman, H. A., and D. A. Cooney.** 1979. Partial purification and properties of L-asparagine synthetase from mouse pancreas. Biochemical Journal **181:**51-59.
41. **Scofield, M. A., W. S. Lewis, and S. M. Schuster.** 1990. Nucleotide sequence of *Escherichia coli asnB* and deduced amino acid sequence of asparagine synthetase B. Journal of Biological Chemistry **265:**12895-12902.
42. **Hirasawa, T., M. Wachi, and K. Nagai.** 2000. A mutation in the *Corynebacterium glutamicum ltsA* gene causes susceptibility to lysozyme,

temperature-sensitive growth, and L-glutamate production. Journal of Bacteriology **182:**2696-2701.

43. **Ren, H., and J. Liu.** 2006. AsnB is involved in natural resistance of *Mycobacterium smegmatis* to multiple drugs. Antimicrobial Agents and Chemotherapy **50:**250-255.
44. **Park, Y. S., T. D. Sweitzer, J. E. Dixon, and C. Kent.** 1993. Expression, purification, and characterization of CTP:glycerol-3-phosphate cytidylyltransferase from *Bacillus subtilis*. Journal of Biological Chemistry **268:**16648-16654.
45. **Mauel, C., M. Young, and D. Karamata.** 1991. Genes concerned with synthesis of poly(glycerol phosphate), the essential teichoic acid in *Bacillus subtilis* strain 168, are organized in two divergent transcription units. Journal of General Microbiology **137:**929-941.
46. **Bhavsar, A. P., R. Truant, and E. D. Brown.** 2005. The TagB protein in *Bacillus subtilis* 168 is an intracellular peripheral membrane protein that can incorporate glycerol phosphate onto a membrane-bound acceptor *in vitro*. Journal of Biological Chemistry **280:**36691-36700.
47. **Shao, J., J. Zhang, P. Kowal, Y. Lu, and P. G. Wang.** 2002. Overexpression and biochemical characterization of β-1,3-*N*-acetylgalactosaminyltransferase LgtD from *Haemophilus influenzae* strain Rd. Biochemical and Biophysical Research Communications **295:**1-8.
48. **Randriantsoa, M., S. Drouillard, C. Breton, and E. Samain.** 2007. Synthesis of globopentaose using a novel β-1,3-galactosyltransferase activity of the *Haemophilus influenzae* β-1,3-*N*-acetylgalactosaminyltransferase LgtD. FEBS Letters **581:**2652-2656.
49. **Gotschlich, E. C.** 1994. Genetic locus for the biosynthesis of the variable portion of *Neisseria gonorrhoeae* lipooligosaccharide. Journal of Experimental Medicine **180:**2181-2190.
50. **Yurist-Doutsch, S., M. Abu-Qarn, F. Battaglia, H. R. Morris, P. G. Hitchen, A. Dell, and J. Eichler.** 2008. *AglF*, *aglG* and *aglI*, novel members of a gene island involved in the *N*-glycosylation of the *Haloferax volcanii* S-layer glycoprotein. Molecular Microbiology **69:**1234-1245.
51. **Strahl-Bolsinger, S., and W. Tanner.** 1991. Protein *O*-glycosylation in *Saccharomyces cerevisiae*. Purification and characterization of the dolichyl-

phosphate-D-mannose-protein *O*-D-mannosyltransferase. European Journal of Biochemistry **196:**185-190.

52. **Strahl-Bolsinger, S., T. Immervoll, R. Deutzmann, and W. Tanner.** 1993. PMT1, the gene for a key enzyme of protein *O*-glycosylation in *Saccharomyces cerevisiae*. Proceedings of the National Academy of Sciences of the United States of America **90:**8164-8168.
53. **Gentzsch, M., S. Strahl-Bolsinger, and W. Tanner.** 1995. A new Dol-P-Man:protein *O*-D-mannosyltransferase activity from *Saccharomyces cerevisiae*. Glycobiology **5:**77-82.
54. **Girrbach, V., T. Zeller, M. Priesmeier, and S. Strahl-Bolsinger.** 2000. Structure-function analysis of the dolichyl phosphate-mannose: protein *O*-mannosyltransferase ScPmt1p. Journal of Biological Chemistry **275:**19288-19296.
55. **Strahl-Bolsinger, S., and A. Scheinost.** 1999. Transmembrane topology of Pmt1p, a member of an evolutionarily conserved family of protein *O*-mannosyltransferases. Journal of Biological Chemistry **274:**9068-9075.
56. **Nölling, J., G. Breton, M. V. Omelchenko, K. S. Makarova, Q. Zeng, R. Gibson, H. M. Lee, J. Dubois, D. Qiu, J. Hitti, Y. I. Wolf, R. L. Tatusov, F. Sabathe, L. Doucette-Stamm, P. Soucaille, M. J. Daly, G. N. Bennett, E. V. Koonin, and D. R. Smith.** 2001. Genome sequence and comparative analysis of the solvent-producing bacterium *Clostridium acetobutylicum*. Journal of Bacteriology **183:**4823-4838.
57. **Gross, H., V. O. Stockwell, M. D. Henkels, B. Nowak-Thompson, J. E. Loper, and W. H. Gerwick.** 2007. The genomisotopic approach: a systematic method to isolate products of orphan biosynthetic gene clusters. Chemistry & Biology **14:**53-63.
58. **Chen, C. Y., K. M. Wu, Y. C. Chang, C. H. Chang, H. C. Tsai, T. L. Liao, Y. M. Liu, H. J. Chen, A. B. Shen, J. C. Li, T. L. Su, C. P. Shao, C. T. Lee, L. I. Hor, and S. F. Tsai.** 2003. Comparative genome analysis of *Vibrio vulnificus*, a marine pathogen. Genome Research **13:**2577-2587.
59. **D'Souza, J. M., L. Wang, and P. Reeves.** 2002. Sequence of the *Escherichia coli* O26 O antigen gene cluster and identification of O26 specific genes. Gene **297:**123-127.

60. **Zayni, S., K. Steiner, A. Pföstl, A. Hofinger, P. Kosma, C. Schäffer, and P. Messner.** 2007. The dTDP-4-dehydro-6-deoxyglucose reductase encoding *fcd* gene is part of the surface layer glycoprotein glycosylation gene cluster of *Geobacillus tepidamans* GS5-97^T. Glycobiology**:**433-443.

61. **Graninger, M., B. Kneidinger, K. Bruno, A. Scheberl, and P. Messner.** 2002. Homologs of the Rml enzymes from *Salmonella enterica* are responsible for dTDP-β-L-rhamnose biosynthesis in the Gram-positive thermophile *Aneurinibacillus thermoaerophilus* DSM 10155. Applied and Environment Microbiology **68:**3708-3715.

62. **Steiner, K., R. Novotny, D. B. Werz, K. Zarschler, P. H. Seeberger, A. Hofinger, P. Kosma, C. Schäffer, and P. Messner.** 2008. Molecular basis of S-layer glycoprotein glycan biosynthesis in *Geobacillus stearothermophilus*. Journal of Biological Chemistry **283:**21120-21133.

63. **Kaneko, T., N. Nakajima, S. Okamoto, I. Suzuki, Y. Tanabe, M. Tamaoki, Y. Nakamura, F. Kasai, A. Watanabe, K. Kawashima, Y. Kishida, A. Ono, Y. Shimizu, C. Takahashi, C. Minami, T. Fujishiro, M. Kohara, M. Katoh, N. Nakazaki, S. Nakayama, M. Yamada, S. Tabata, and M. M. Watanabe.** 2007. Complete genomic structure of the bloom-forming toxic cyanobacterium *Microcystis aeruginosa* NIES-843. DNA Research **14:**247-256.

64. **Welsh, E. A., M. Liberton, J. Stockel, T. Loh, T. Elvitigala, C. Wang, A. Wollam, R. S. Fulton, S. W. Clifton, J. M. Jacobs, R. Aurora, B. K. Ghosh, L. A. Sherman, R. D. Smith, R. K. Wilson, and H. B. Pakrasi.** 2008. The genome of *Cyanothece* 51142, a unicellular diazotrophic cyanobacterium important in the marine nitrogen cycle. Proceedings of the National Academy of Sciences of the United States of America **105:**15094-15099.

65. **Rodriguez, S. A., J. J. Yu, G. Davis, B. P. Arulanandam, and K. E. Klose.** 2008. Targeted inactivation of *Francisella tularensis* genes by group II introns. Applied and Environmental Microbiology **74:**2619-2626.

66. **Whitfield, C.** 1995. Biosynthesis of lipopolysaccharide O antigens. Trends in Microbiology **3:**178-185.

67. **Raetz, C. R. H., and C. Whitfield.** 2002. Lipopolysaccharide endotoxins. Annual Review of Biochemistry **71:**635-700.

68. **Claus, D., and R. C. W. Berkeley.** 1986. *Bacillus* Cohn 1872., p. 1105-1139. *In* P. H. A. Sneath (ed.), Bergey's Manual of Systematic Bacteriology, Vol 2. The Williams and Wilkins Co., Baltimore.

Table 1. Bacterial strains and plasmids used in this study.

Strain or plasmid	Genotype and/or relevant characteristics	Source or reference
Strains		
P. alvei CCM 2051^T	wild-type isolate, Km^r	Czech Collection of Micro-organisms
P. alvei CCM 2051^T Δ*wsfP*	*P. alvei* CCM 2051^T carrying a targetron insertion at the *wsfP* locus, Km^r	(17)
P. alvei CCM 2051^T Δ*wsfA*	*P. alvei* CCM 2051^T carrying a targetron insertion at the *wsfA* locus, Km^r	This study
P. alvei CCM 2051^T Δ*wsfB*	*P. alvei* CCM 2051^T carrying a targetron insertion at the *wsfB* locus, Km^r	This study
P. alvei CCM 2051^T Δ*wsfC*	*P. alvei* CCM 2051^T carrying a targetron insertion at the *wsfC* locus, Km^r	This study
P. alvei CCM 2051^T Δ*wsfD*	*P. alvei* CCM 2051^T carrying a targetron insertion at the *wsfD* locus, Km^r	This study
P. alvei CCM 2051^T Δ*wsfE*	*P. alvei* CCM 2051^T carrying a targetron insertion at the *wsfE* locus, Km^r	This study
P. alvei CCM 2051^T Δ*wsfF*	*P. alvei* CCM 2051^T carrying a targetron insertion at the *wsfF* locus, Km^r	This study
P. alvei CCM 2051^T Δ*wsfG*	*P. alvei* CCM 2051^T carrying a targetron insertion at the *wsfG* locus, Km^r	This study
P. alvei CCM 2051^T Δ*wsfH*	*P. alvei* CCM 2051^T carrying a targetron insertion at the *wsfH* locus, Km^r	This study
P. alvei CCM 2051^T Δ*wzt*	*P. alvei* CCM 2051^T carrying a targetron insertion at the *wzt* locus, Km^r	This study

Escherichia coli DH5α	F⁻ ϕ80d*lacZ*Δ*M15* Δ(*lacZYA-argF*)*U169* *deoR recA1 endA1 hsdR17* (rK⁻ mK⁻) *phoA supE44 thi-1 gyrA96 relA1* λ⁻	Invitrogen
Plasmids		
pEXALV	*P. alvei* expression vector	(17)
pTT_*wsfP*1176	pTT_*plc* targeted for insertion at position 1176/1177 from the initial ATG of *wsfP*	(17)
pTT_*wsfA*243	pTT_*wsfP*1176 targeted for insertion at position 243/244 from the initial ATG of *wsfA*	This study
pTT_*wsfB*587	pTT_*wsfP*1176 targeted for insertion at position 587/588 from the initial ATG of *wsfB*	This study
pTT_*wsfC*522	pTT_*wsfP*1176 targeted for insertion at position 522/523 from the initial ATG of *wsfC*	This study
pTT_*wsfD*136	pTT_*wsfP*1176 targeted for insertion at position 136/137 from the initial ATG of *wsfD*	This study
pTT_*wsfE*101	pTT_*wsfP*1176 targeted for insertion at position 101/102 from the initial ATG of *wsfE*	This study
pTT_*wsfF*116	pTT_*wsfP*1176 targeted for insertion at position 116/117 from the initial ATG of *wsfF*	This study
pTT_*wsfG*93	pTT_*wsfP*1176 targeted for insertion at position 93/94 from the initial ATG of *wsfG*	This study
pTT_*wsfH*418	pTT_*wsfP*1176 targeted for insertion at position 418/419 from the initial ATG of *wsfH*	This study
pTT_*wzt*534	pTT_*wsfP*1176 targeted for insertion at position 534/535 from the initial ATG of *wzt*	This study

Table 2. Oligonucleotide primers used for PCR amplification reactions.

Oligonucleotide	Sequence (5' → 3')*
proof_RmlD_for	ACIGAYTAYGTITTYGAYGG
proof_wSpa_for	GCIGAYGCIGCIAARACIACICARG
1f	CATAGATAACGCGAAGCTGGCAGGCAATAA
1r	ACTGCTGGAACAAGAGGAGCAAGTCATA
2f	GGTACATTAGAAGCCATTATTGCATATCA
2r	CGGAATATTGTCTACAGGTACTGTCTTA
3f	GAACCAGAGATTCTTATTGTGGATGAG
3r	TCTGTTGCACATAGTATAACGGCTTAATTC
4f	GTGGAGGACTGGATAGCACCATAGTTAC
4r	CGATCTTGCCTCAACGGAACACAT
5f	ATGATGATGCGCTTGCGGATGATAATCT
5r	TCTTGCTCATACCACATATGAGTAGGTCTT
6f	TGCCGACTATTGATAACAATTAC
6r	TTAATCCAGTTCCAATCCACATGCATTAAG
8f	GAAGATCGACTTGGACATGACCGTCGATA
8r	CGCCATCCAGACATCATCTTGATCACA
9f	ACCGCTCTATCTTGTACGCGAGGT
9r	CGTGCCGCCAGTCATCTATTGGTTGTAA

Table 3. Insertion position, primer set for retargeting of the Ll.LtrB targetron and primers used for confirmation of intron insertion.

ORF	insertion position	E-value	Primer sets for retargeting of Ll.LtrB	Primer pair for confirmation of intron insertion
wsfB	587|588s	0.205	B_587|588s-IBS: AAAAAAGCTTATAATTATCCTTACATTACGTTATCGTGCGCCCAGATAGGGTG B_587|588s-EBS1d: CAGATTGTACAAATGTGGTGATAACAGATAAGTCGTTATCCATAACTTACCTTTCTTTGT B_587|588s-EBS2: TGAACGCAAGTTTCTAATTTCGGTTTAATGTCGATAGAGGAAAGTGTCT	B_KO_for: TGGCAGATTTCGATCAACCAAT B_KO_rev: GCTATCAGAATAAGCATCCCGATAA
galE	n.d.			
galU	n.d.			
wzm	n.d.			
wzt	534|535s	0.002	wzt_534|535s-IBS: AAAAAAGCTTATAATTATCCTTACTGTCCGTTGGGGTGCGCCCAGATAGGGTG wzt_534|535s-EBS1d: CAGATTGTACAAATGTGGTGATAACAGATAAGTCGTTGGGGATAACTTACCTTTCTTTGT wzt_534|535s-EBS2: TGAACGCAAGTTTCTAATTTCGGTTGACAGTCGATAGAGGAAAGTGTCT	wzt_KO_for: ATCAGTGCTAGAAGTCAAGAATGT wzt_KO_rev: CGAATGCGATACGAATAGTATTGT
wsfA	243|244a	0.125	A_243|244a-IBS: AAAAAAGCTTATAATTATCCTTACTGACCGGCTAAGTGCGCCCAGATAGGGT A_243|244a-EBS1d: CAGATTGTACAAATGTGGTGATAACAGATAAGTCGGCTAACTTAACTTACCTTTCTTTGT A_243|244a-EBS2: TGAACGCAAGTTTCTAATTTCGGTTGTCAGTCGATAGAGGAAAGTGTCT	
tagD	n.d.			
wsfC	522|523s	0.191	C_522|523s-IBS: AAAAAAGCTTATAATTATCCTTAGATTCCCTAGTAGTGCGCCCAGATAGGGTG C_522|523s-EBS1d: CAGATTGTACAAATGTGGTGATAACAGATAAGTCCTAGTAGATAACTTACCTTTCTTTGT C_522|523s-EBS2: TGAACGCAAGTTTCTAATTTCGATTGAATCTCGATAGAGGAAAGTGTCT	C_KO_for: GGCTGCAGATACTTATGAAT C_KO_rev: TATAACAATAGGCGTATCCG
wsfD	136|137a	0.103	D_136|137a-IBS: AAAAAAGCTTATAATTATCCTTAACATACGAGTAGGTGCGCCCAGATAGGGT D_136|137a-EBS1d: CAGATTGTACAAATGTGGTGATAACAGATAAGTCGAGTAGGTTAACTTACCTTTCTTTGT D_136|137a-EBS2: TGAACGCAAGTTTCTAATTTCGATTTATGTTCGATAGAGGAAAGTGTCT	D_TM_for: GGATAGCTGACAATGGTGAT D_TM_rev: GTTAATATGATGCGGCTTCT
wsfE	101|102s	0.244	E_101|102s-IBS: AAAAAAGCTTATAATTATCCTTACTTTGCTGTAGAGTGCGCCCAGATAGGGTG E_101|102s-EBS1d: CAGATTGTACAAATGTGGTGATAACAGATAAGTCTGTAGATTTAACTTACCTTTCTTTGT E_101|102s-EBS2: TGAACGCAAGTTTCTAATTTCGGTTCAAAGTCGATAGAGGAAAGTGTCT	E_KO_for: TTATGGCCGAAGAATTGACTAGGT E_KO_rev: GAAGTCGATCCAACTTCCAATTGT

rmlA	n.d.			
rmlC	n.d.			
rmlB	n.d.			
rmlD	n.d.			
wsfF	116|117a	0.284	F_116|117a-IBS: AAAAAAGCTTATAATTATCCTTATCGTCCCGTATCGTGCGCCCAGATAGGGTG F_116|117a-EBS1d: CAGATTGTACAAATGTGGTGATAACAGATAAGTCCGTATCGTTAACTTACCTTTCTTTGT F_116|117a-EBS2: TGAACGCAAGTTTCTAATTTCGGTTGACGATCGATAGAGGAAAGTGTCT	F_KO_for: GCGGAACAACTTAATAGTCTATTGG F_KO_rev: GTGTTCGGTAGATTCTGTTGTAGTA
wsfG	93|94a	0.054	G_93|94a-IBS: AAAAAAGCTTATAATTATCCTTAAACCTCGATTTGGTGCGCCCAGATAGGGTG G_93|94a-EBS1d: CAGATTGTACAAATGTGGTGATAACAGATAAGTCGATTTGATTAACTTACCTTTCTTTG G_93|94a-EBS2: TGAACGCAAGTTTCTAATTTCGGTTAGGTTCCGATAGAGGAAAGTGTCT	G_KO_for: ATTATAACACGAAGCAACTAACC G_KO_rev: TTAAGAAGAAGTGCCACAGACATT
wsfP	1176|1177s	0.202	P_1176|1177s-IBS: AAAAAAGCTTATAATTATCCTTAAGACCCGAACGGGTGCGCCCAGATAGGGTG P_1176|1177s-EBS1d: CAGATTGTACAAATGTGGTGATAACAGATAAGTCGAACGGCCTAACTTACCTTTCTTTGT P_1176|1177s-EBS2: TGAACGCAAGTTTCTAATTTCGATTGGTCTTCGATAGAGGAAAGTGTCT	P_KO_for: TCTTATCCTTGGTGCCGGTACACTTG P_KO_rev: AGCCTGTAATTCCAGGACGCACAT
wsfH	418|419a	0.058	H_418|419a-IBS: AAAAAAGCTTATAATTATCCTTAGGTGGCCAGTAGGTGCGCCCAGATAGGGT H_418|419a-EBS1d: CAGATTGTACAAATGTGGTGATAACAGATAAGTCCAGTAGCCTAACTTACCTTTCTTTGT H_418|419a-EBS2: TGAACGCAAGTTTCTAATTTCGATTCCACCTCGATAGAGGAAAGTGTCT	H_KO_for: CCAATGTACAATGAGGAAGAA H_KO_rev: GGTAACTGATTAAGTGCATCTA

Table 4. Predicted gene products encoded by the *slg* gene cluster of *P. alvei* CCM 2051^T together with database homologies.

ORF	Length / Mol. mass	Conserved motifs and region		Related proteins			
				Name / putative function	Organism	Identity / similarity (%)	Accession No.
wsfB	781 / 87.5	Wzy_C (PF04932) TPR2 (PF07719)	347-411 690-723	O-antigen polymerase O-antigen polymerase	*Geobacillus sp.* Y412MC10 *Caldicellulosiruptor saccharolyticus* DSM 8903	43 / 65 23 / 45	ZP_03037638 YP_001181337
galE	328 / 36.3	Epimerase (PF01370)	3-251	UDP-glucose 4-epimerase UDP-glucose 4-epimerase	*Paenibacillus larvae subsp. larvae* BRL-230010 *Geobacillus sp*. Y412MC10	65 / 80 64 / 79	ZP_02328485 ZP_03040569
galU	290 / 33.0	NTP_transferase (PF00483)	5-216	UTP-glucose-1-phosphate uridylyltransferase UTP-glucose-1-phosphate uridylyltransferase	*Geobacillus sp.* Y412MC10 *Bacillus weihenstephanensis* KBAB4	73 / 85 69 / 82	ZP_03037639 YP_001647518
wzm	232 / 27.4	ABC2_membrane (PF01061)	1-191	Wzm ABC transporter, permease protein	*Aneurinibacillus thermoaerophilus* DSM 10155/G$^+$ *Vibrio sp.* MED222	58 / 77 52 / 74	AAS49124 ZP_01065570
wzt	434 / 48.7	ABC_tran (PF00005)	53-222	Wzt Wzt	*Aneurinibacillus thermoaerophilus* DSM 10155/G$^+$ *Geobacillus tepidamans* GS5-97	42 / 62 57 / 79	AAS49125 ABM68319
wsfA	520 / 60.5	GATase_2 (PF00310) Asn_synthase (PF00733)	2-157 224-475	asparagine synthase (glutamine-hydrolyzing) asparagine synthase (glutamine-hydrolyzing)	*Francisella philomiragia subsp. philomiragia* ATCC 25017 *Thioalkalivibrio sp.* HL-EbGR7	52 / 69 29 / 50	YP_001678197 YP_002514426
tagD	139 / 16.3	CTP_transf_2 (PF01467)	5-131	putative glycerol-3-phosphate cytidyltransferase glycerol-3-phosphate cytidyltransferase	*Francisella philomiragia subsp. philomiragia* ATCC 25017 *Clostridium perfringens* str. 13	64 / 86 58 / 78	YP_001678198 NP_561399
wsfC	1260 / 147.3	Glyphos_transf (PF04464) Glycos_transf_2 (PF00535) Glycos_transf_2 (PF00535)	99-279 558-729 947-1124	glycosyltransferase family 2 glycosyltransferase	*Anaeromyxobacter sp.* Fw109-5 *Rickettsia felis* URRWXCal2	38 / 54 39 / 58	YP_001378599 YP_246702
wsfD	457 /	PMT (PF02366)	15-264	transmembrane protein	*Bacillus cereus* G9241	30 / 51	ZP_00240365

	53.3			transmembrane protein	*Paenibacillus larvae subsp. larvae* BRL-230010	25 / 48	ZP_02327204
wsfE	364 / 42.3	Glycos_transf_1 (PF00534)	187-342	glycosyltransferase glycosyltransferase, group 2 family protein	*Clostridium acetobutylicum* ATCC 824 *Pseudomonas fluorescens* Pf-5	29 / 48 23 / 42	NP_349666 YP_259145
rmlA	247 / 54.6	NTP_transferase (PF00483)	2-236	glucose-1-phosphate thymidyltransferase glucose-1-phosphate thymidylyltransferase	*Geobacillus stearothermophilus* ATCC 12980 *Bacillus anthracis* str. Ames	76 / 84 70 / 84	AAQ23685 NP_843700
rmlC	183 / 20.8	dTDP_sugar_isom (PF00908)	3-177	dTDP-4-dehydrorhamnose 3,5-epimerase dTDP-dehydrorhamnose 3,5-epimerase	*Geobacillus sp.* Y412MC10 *Aneurinibacillus thermoaerophilus* DSM 10155/G^{+}	70 / 85 72 / 83	ZP_03037628 AAL18012
rmlB	341 / 38.5	Epimerase (PF01370)	3-241	dTDP-glucose 4,6-dehydratase dTDP-glucose 4,6-dehydratase	*Geobacillus sp.* Y412MC10 *Geobacillus stearothermophilus* NRS 2004/3a	76 / 88 72 / 85	ZP_03037627 AAR99612
rmlD	286 / 32.2	RmlD_sub_bind (PF04321)	1-281	dTDP-4-dehydrorhamnose reductase RmlD	*Geobacillus stearothermophilus* NRS 2004/3a *Geobacillus tepidamans* GS5-97	60 / 78 59 / 76	AAR99613 ABM68332
wsfF	314 / 36.8	Glycos_transf_2 (PF00535)	9-119	WsdG rhamnosyltransferase	*Aneurinibacillus thermoaerophilus* DSM 10155/G^{+} *Oenococcus oeni* ATCC BAA-1163	49 / 66 35 / 56	AAL18015 ZP_01544702
wsfG	299 / 34.1	Glycos_transf_2 (PF00535)	5-191	glycosyltransferase family 2 WsaD	*Geobacillus sp.* Y412MC10 *Geobacillus stearothermophilus* NRS 2004/3a	64 / 79 56 / 75	ZP_03037625 AAR99614
wsfP	468 / 54.6	Bac_transf (PF02397)	281-468	WsaP WsbP	*Geobacillus stearothermophilus* NRS 2004/3a *Geobacillus tepidamans* GS5-97	60 / 75 60 / 75	AAR99615 ABM68334
wsfH	336 / 38.4	Glycos_transf_2 (PF00535)	12-176	glycosyltransferase glycosyltransferase family 2	*Microcystis aeruginosa* NIES-843 *Arthrospira maxima* CS-328	58 / 81 61 / 79	YP_001655327 ZP_03275272

FIGURE LEGENDS

Figure 1. Genetic organization of the *slg* gene cluster of *P. alvei* CCM 2051^T. Predicted open reading frames are indicated by horizontal arrows with the respective gene designations indicated above the arrow. Genes encoding similar functions in S-layer glycan biosynthesis have a similar grey scaling code. Genes flanking the *slg* gene cluster are indicated in black and genes coding for proteins with unknown function are indicated in white. Genes indicated in light grey encode putative glycosyltransferases. *Wzm* and *wzt* (dark gray) encode the two components of the ABC transporter. Squares indicate genes encoding proteins involved in the biosynthesis of nucleotide-activated sugar precursors. The lightest grey indicates the *wsfB* gene encoding the OTase. Putative promoters and terminators are represented as flags and hairpins, respectively. The location of the Ll.LtrB insertion is indicated by vertical black arrows and by the numbers below, whereas positions are given relative to the initial ATG codon. Besides the predicted transcription units (1), the experimentally identified transcription units are depicted (2). The reverse transcription analysis and subsequent cDNA amplification are shown, whereas primer positions are depicted as vertical black arrows (3). The percentage G+C base composition is given below the cluster map.

Figure 2. RT-PCR analysis of total RNA of *P. alvei* CCM 2051^T. Reverse transcription was performed with the specific primer 4r targeted to *wsfA* (lane 2-7) or 8r annealing to *wsfF* (lane 9-20). Subsequent cDNA amplification was carried out with primer pairs 2f/2r annealing to *galU*/*wzm* (lane 2-4), with 3f/3r targeted to *wzt*/*wsfA* (lane 5-7), with 4f/4r annealing to *wsfA* (lane 9-11), with 5f/5r targeted to *wsfC*/*wsfD* (lane 12-14), with 6f/6r annealing to *wsfD*/*wsfE* (lane 15-17), and with 8f/8r amplifying *rmlB*/*wsfF* (lane 18-20). Lanes (a) show the specific PCR amplification products, using reverse transcribed single-strand cDNA as template; lanes (b) show control reactions, using DNase I-treated RNA without the cDNA-generating step as PCR template; lanes (c) show positive controls using genomic DNA as template. The 1 kb DNA Plus marker (Invitrogen) was used as DNA size marker (lanes 1, 8, and 20).

Figure 3. Predicted transmembrane topology of the OTase WsfB of *P. alvei* CCM 2051^T. The twelve transmembrane helices are shown as light grey boxes; the Wzy_C region (PF04932) and the TPR2 region (PF07719) are highlighted in grey and in black, respectively.

Figure 4. Alignment of the Wzt homologues from *E. coli* O9a (Ec) and *P. alvei* CCM 2051^T (Pa). Identical residues are highlighted in grey. Conserved sequence motifs common to all Wzt proteins are indicated. Residues involved in substrate binding in *E. coli* are highlighted with an asterisk.

Figure 5. Effect of insertional inactivation of genes from the *P. alvei* CCM 2051^T *slg* gene cluster. An aliquot (0.5 mg wet weight) of biomass from various *P. alvei* CCM 2051^T mutant strains was analyzed by SDS-PAGE followed by Coomassie Brilliant Blue G250 staining. The mutant strains carry the Ll.LtrB insertion as indicated. Tri-banded appearance corresponds to non-glycosylated (N), monoglycosylated (M), and diglycosylated (D) chimeric SpaA.

Figure 6. Mass spectrometry of glycopeptides. Glycopeptides derived from pronase digestion of SpaA protein produced in *P. alvei* CCM 2051^T wild-type (A) and *wsfD*::Ll.LtrB mutant strain (B) were analyzed by on-line MSMS, showing the sequences of the repeating unit and the core structure. The spectrum of *wsfH*::Ll.LtrB mutant strain is identical to that observed for the *wsfD*::Ll.LtrB mutant strain and hence not shown.

Figure 7. Working model of S-layer glycan biosynthesis in *P. alvei* CCM 2051^T. The initial transfer of a Gal residue from UDP-α-D-Gal to a lipid carrier is catalyzed by WsfP (A). The adaptor saccharide is formed by the α1,3-linkage of an L-Rha residue from dTDP-β-L-Rha to the linkage sugar D-Gal performed by WsfG, followed by the transfer of two additional α1,3-linked L-Rha residues by the action of WsfF (B). The glycan chain would be elongated by the activity of the aminosugar transferase WsfE and the tripartite transferase WsfC. WsfE may form the β1,4-linkage of a ManNAc residue from UDP-ManNAc to the third rhamnose residue. WsfC putatively adds a single glycerol phosphate from CDP-glycerol to the ManNAc residue of the adaptor oligosaccharide and may form the β1,3-linkage of a ManNAc residue to the third

rhamnose residue as well as the β1,4-linkage of a Gal to the ManNAc residues of the repeating units. The glycan chain is recognized by the carboxyterminal part of Wzt and exported by the ABC transporter system through the cytoplasmic membrane (C). The transfer of cytoplasmic Glc to the lipid carrier is carried out by WsfH and, which, after reorientation, is used at the external face of the cytoplasmic membrane by WsfD for α1,6-linkage of the Glc residues to ManNAc residues of the repeating units (D). The final transfer of the completed S-layer glycan to certain tyrosine residues of the S-layer protein is catalyzed by the oligosaccharyltransferase WsfB (E).

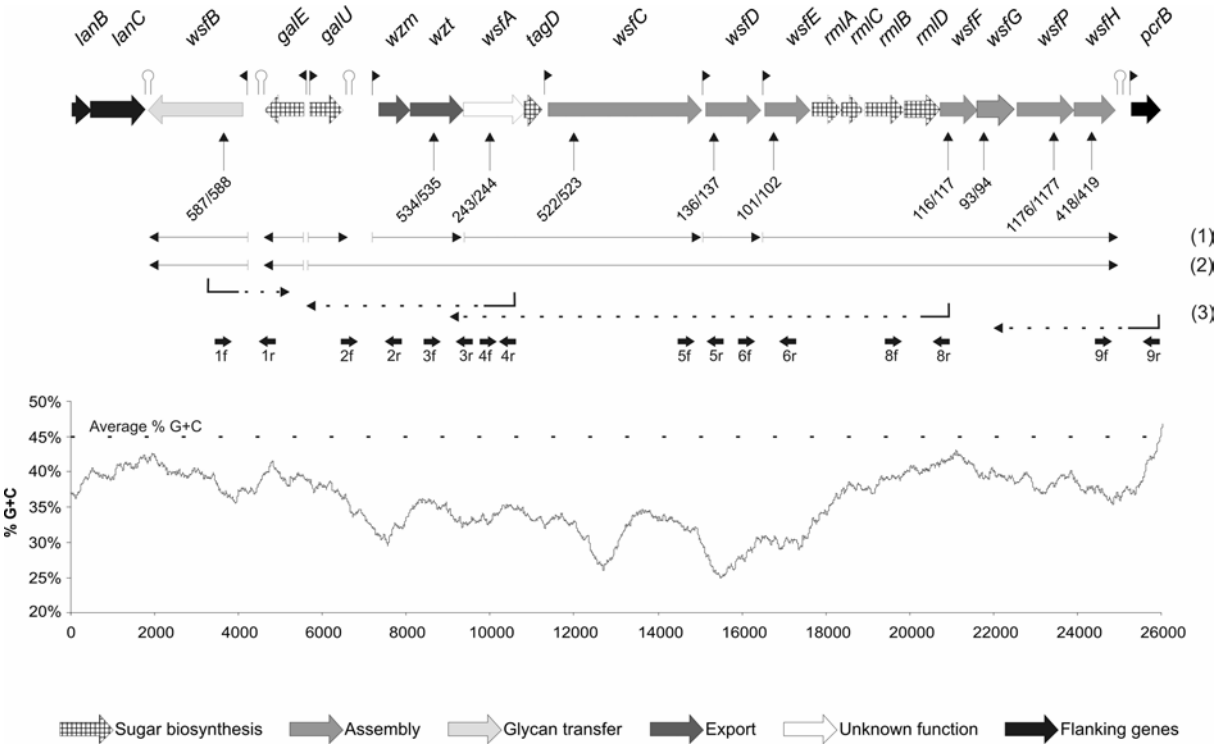

lanB
lanC
wsfB
galE
galU
wzm
wzt
wsfA
tagD
wsfC
wsfD
wsfE
rmlA
rmlC
rmlB
rmlD
wsfF
wsfG
wsfP
wsfH
pcrB
587/588
534/535
243/244
522/523
136/137
101/102
116/117
93/94
1176/1177
418/419
(1)
(2)
(3)
1f
1r
2f
2r
3f
3r
4f
4r
5f
5r
6f
6r
8f
8r
9f
9r
Average % G+C
% G+C
50%
45%
40%
35%
30%
25%
20%
0
2000
4000
6000
8000
10000
12000
14000
16000
18000
20000
22000
24000
26000
Sugar biosynthesis
Assembly
Glycan transfer
Export
Unknown function
Flanking genes

Figure 1

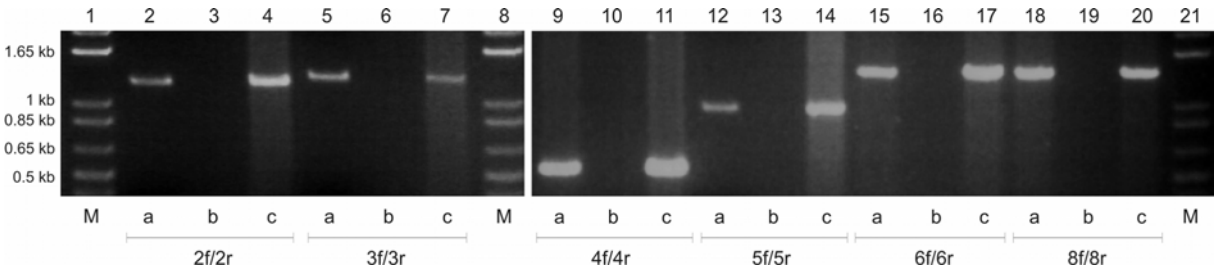

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21
1.65 kb
1 kb
0.85 kb
0.65 kb
0.5 kb
M a b c a b c M a b c a b c a b c a b c M
2f/2r
3f/3r
4f/4r
5f/5r
6f/6r
8f/8r

Figure 2

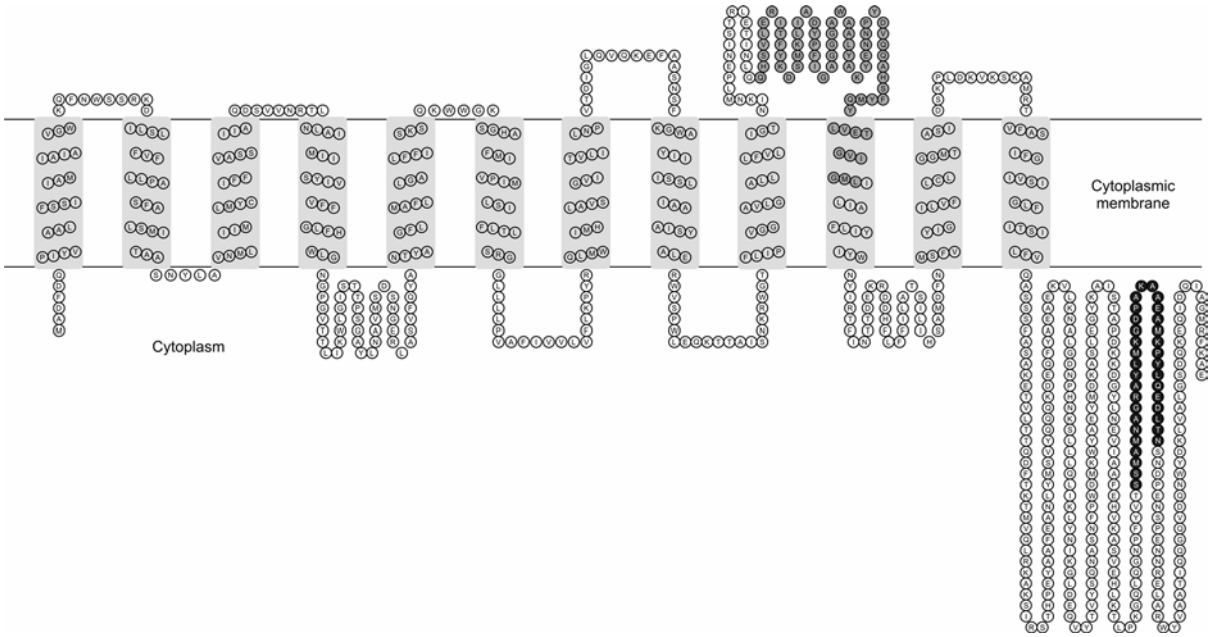


Figure 3

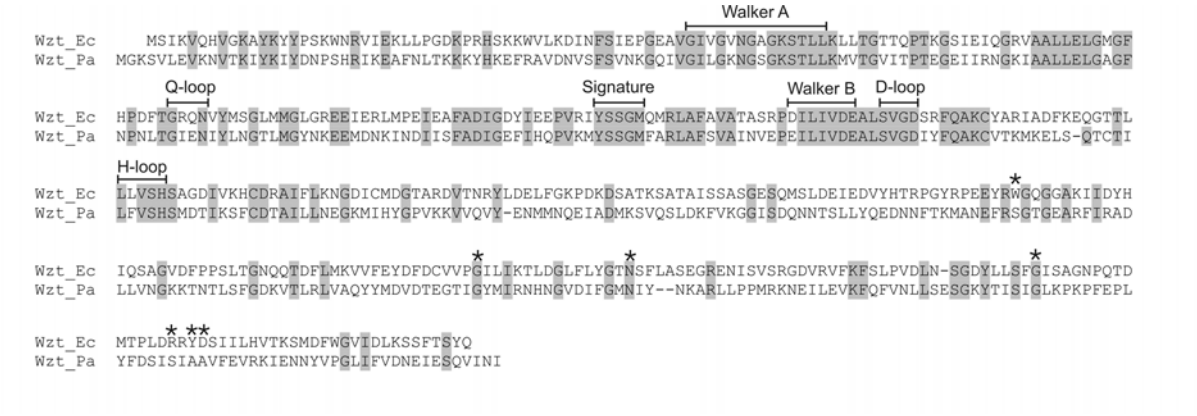

Walker A
Wzt_Ec MSIKVQHVGKAYKYYPSKWNRVIEKLLPGDKPRHSKKWVLKDINFSIEPGEAVGIVGVNGAGKSTLLKLLTGTTQPTKGSIEIQGRVAALLELGMGF
Wzt_Pa MGKSVLEVKNVTKIYKIYDNPSHRIKEAFNLTKKKYHKEFRAVDNVSFSVNKGQIVGILGKNGSGKSTLLKMVTGVITPTEGEIIRNGKIAALLELGAGF
Q-loop
Signature
Walker B
D-loop
Wzt_Ec HPDFTGRQNVYMSGLMMGLGREEIERLMPEIEAFADIGDYIEEPVRIYSSGMQMRLAFAVATASRPDILIVDEALSVGDSRFQAKCYARIADFKEQGTTL
Wzt_Pa NPNLTGIENIYLNGTLMGYNKEEMDNKINDIISFADIGEFIHQPVKMYSSGMFARLAFSVAINVEPEILIVDEALSVGDIYFQAKCVTKMKELS-QTCTI
H-loop
Wzt_Ec LLVSHSAGDIVKHCDRAIFLKNGDICMDGTARDVTNRYLDELFGKPDKDSATKSATAISSASGESQMSLDEIEDVYHTRPGYRPEEYRWGQGGAKIIDYH
Wzt_Pa LFVSHSMDTIKSFCDTAILLNEGKMIHYGPVKKVVQVY-ENMMNQEIADMKSVQSLDKFVKGGISDQNNTSLLYQEDNNFTKMANEFRSGTGEARFIRAD
Wzt_Ec IQSAGVDFPPSLTGNQQTDFLMKVVFEYDFDCVVPGILIKTLDGLFLYGTNSFLASEGRENISVSRGDVRVFKFSLPVDLN-SGDYLLSFGISAGNPQTD
Wzt_Pa LLVNGKKTNTLSFGDKVTLRLVAQYYMDVDTEGTIGYMIRNHNGVDIFGMNIY--NKARLLPPMRKNEILEVKFQFVNLLSESGKYTISIGLKPKPFEPL
Wzt_Ec MTPLDRRYDSIILHVTKSMDFWGVIDLKSSFTSYQ
Wzt_Pa YFDSISIAAVFEVRKIENNYVPGLIFVDNEIESQVINI

Figure 4

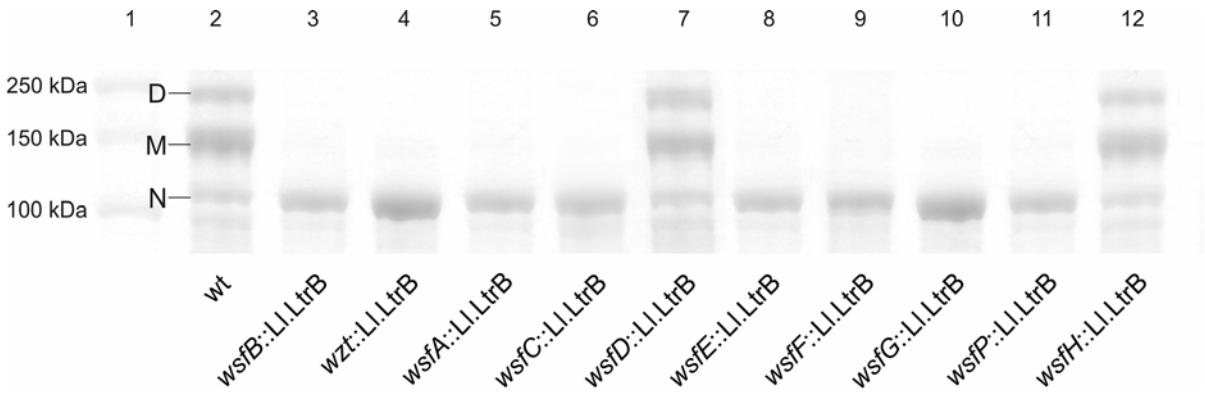

1
2
3
4
5
6
7
8
9
10
11
12
250 kDa
150 kDa
100 kDa
D
M
N
wt
wsfB::LI.LtrB
wzt::LI.LtrB
wsfA::LI.LtrB
wsfC::LI.LtrB
wsfD::LI.LtrB
wsfE::LI.LtrB
wsfF::LI.LtrB
wsfG::LI.LtrB
wsfP::LI.LtrB
wsfH::LI.LtrB

Figure 5

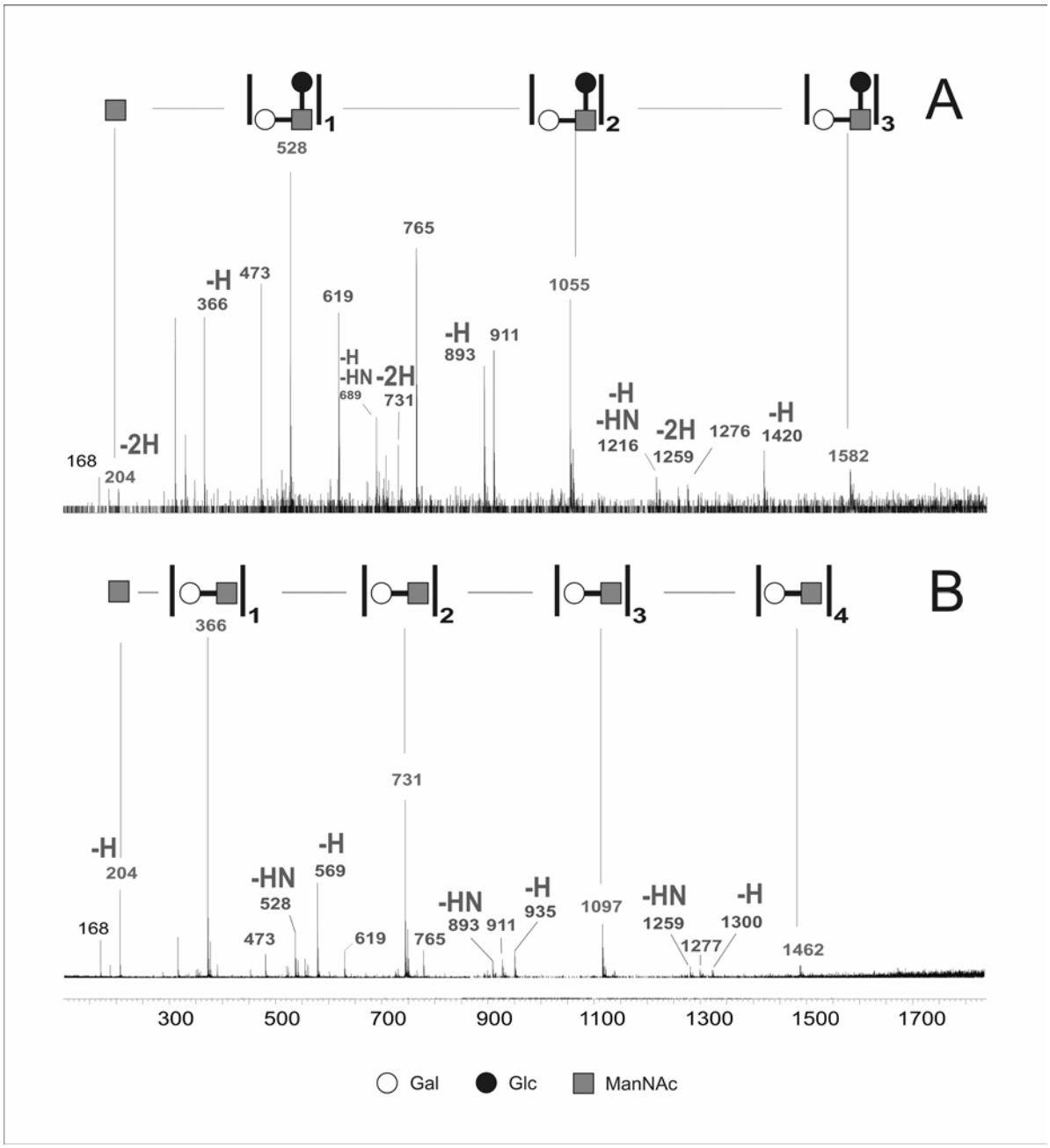

A
528
765
-H
366
473
619
1055
-H
893
911
-H
-HN
689
-2H
731
-H
-HN
1216
-2H
1259
1276
-H
1420
168
-2H
204
1582
B
366
731
-H
204
-H
569
-HN
528
-HN
893
911
-H
935
1097
-HN
1259
-H
1300
168
473
619
765
1277
1462
300
500
700
900
1100
1300
1500
1700
Gal
Glc
ManNAc

Figure 6

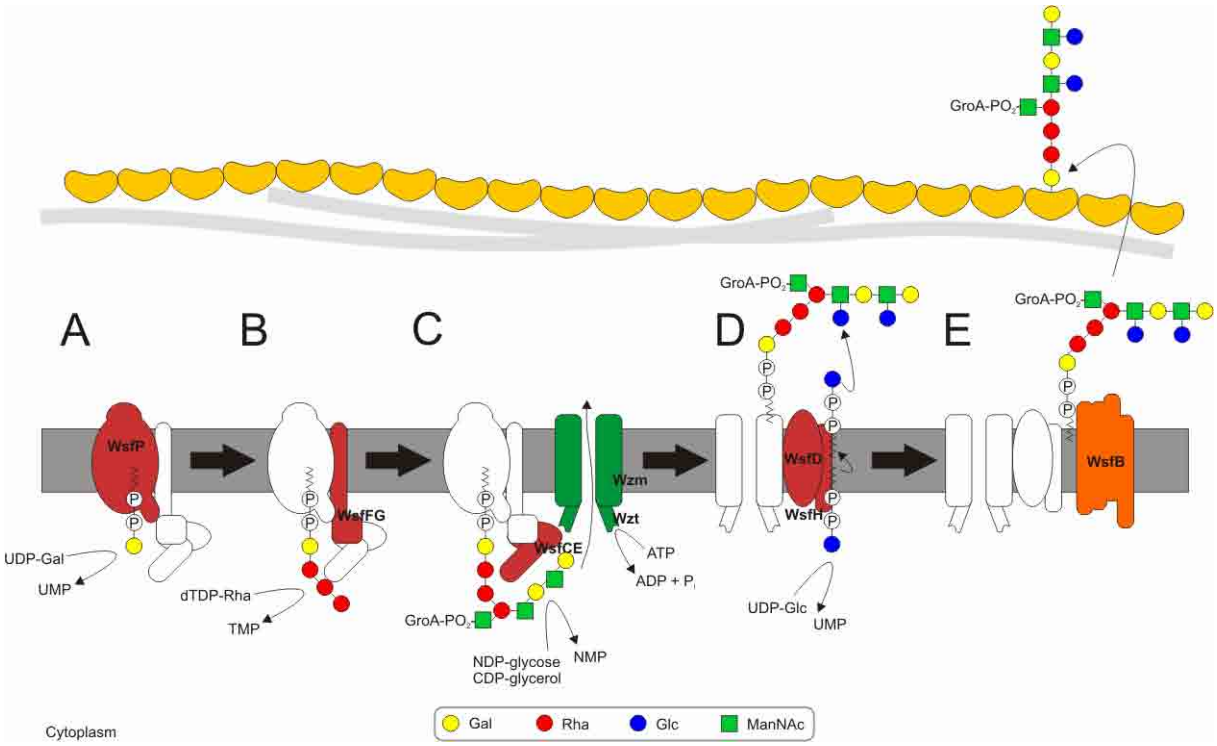

GroA-PO4
A
B
C
D
E
WsfP
WsfFG
WsfCE
Wzm
Wzt
WsfD
WsfH
WsfB
UDP-Gal
UMP
dTDP-Rha
TMP
GroA-PO4
NDP-glycose
CDP-glycerol
NMP
ATP
ADP + Pi
UDP-Glc
UMP
Cytoplasm
Gal
Rha
Glc
ManNAc

Figure 7

Manuscript in preparation

Surface display of chimeric glycoproteins by the surface (S-) layer system of *Paenibacillus alvei* CCM 2051^T

Kristof Zarschler, Birgit Kainz, Bettina Janesch, Paul Messner*, and Christina Schäffer*

Department of NanoBiotechnology, University of Natural Resources and Applied Life Sciences (BOKU), 1190 Vienna, Austria.

Running title: Surface display in *Paenibacillus alvei* CCM 2051^T

* Corresponding authors.

Mailing address: Department of NanoBiotechnology, University of Natural Resources and Applied Life Sciences (BOKU), Muthgasse 11, 1190 Vienna, Austria

Phone: +43-1-47654 ext. 2200

Fax: +43-1-4789112

E-mail: paul.messner@boku.ac.at

christina.schaeffer@boku.ac.at

ABSTRACT

The Gram-positive, mesophilic bacterium *Paenibacillus alvei* CCM 2051^T possesses a porous crystalline protein surface layer (S-layer) with oblique lattice symmetry composed of a single *O*-glycoprotein species. Here, we describe the isolation and the nucleotide sequence of the corresponding structural gene *spaA*. The open reading frame codes for a protein of 983 amino acids, including a signal peptide of 24 amino acids. The mature S-layer protein has a theoretical molecular mass of 105.95 kDa, a calculated pI of 5.83 and contains three S-layer homology domains being involved in anchoring of the glycoprotein subunits via a pyruvylated secondary cell wall polymer to the rigid peptidoglycan layer of the cell wall. Based on this S-layer glycoprotein, we established a system for displaying heterologous proteins and peptide epitopes on the bacterial cell surface by translational fusion to the C-terminus of SpaA, while maintaining their surface attachment. Immunoblot analysis, immunofluorescence staining and fluorescence microscopy revealed that the recombinant SpaA proteins with a carboxyterminal hexahistidine (SpaA_6HIS) or eGFP tag (SpaA_eGFP) are efficiently expressed and glycosylated as well as successfully displayed on the surfaces of the *P. alvei* CCM 2051^T cells. Exclusively non-glycosylated chimeric SpaA proteins were obtained, when the recently generated S-layer glycosylation-deficient *wsfP* mutant was used for surface display.

Additionally, we gained insights into the biosynthesis pathway of the pyruvylated secondary cell wall polymer, required for anchoring of the S-layer glycoprotein to the *P. alvei* cell wall, by identifying several putatively involved enzymes encoded upstream of the *spaA* gene.

INTRODUCTION

The creation and presentation of heterologous proteins or peptide epitopes on various cell surfaces by genetic engineering has become an intensely used strategy in a wide range of biotechnological applications, including live vaccine development and delivery (1), antibody production (2), peptide library screening (3), whole-cell biocatalysis (4), and bioremediation (5, 6). Numerous microbial display systems have been extensively reviewed for both, bacteria and yeast (7-12).

Outer membrane proteins, lipoproteins, autotransporters, S-layer proteins, ice nucleation proteins, flagellae and fimbriae are the most common anchoring motifs of

Gram-negative surface display systems, whereas for Gram-positive bacteria, the staphylococcal protein A and different S-layer proteins have been utilized to display various polypeptides and proteins (9).

For anchoring heterologous target proteins on the cell surface of bacteria, the S-layer proteins RsaA of *Caulobacter crescentus* and EA1 / Sap of *Bacillus anthracis*, respectively, have been used (13-19). For the former, the insertion of a protein G IgG-binding domain into certain sites of full length RsaA resulted in functional, immunoreactive surface display at very high density (20). For *Bacillus anthracis*, targeting of active levansucrase of *Bacillus subtilis* and immunogenic tetanus toxin fragment C of *Clostridium tetani* to the cell surface was achieved by translational fusion of the target proteins to the three S-layer homology (SLH) domains of EA1 and Sap, respectively (18, 19).

Amino- or carboxyterminal SLH domains of up to three approximately 55 amino acid-long sequences have been identified in a number of S-layer proteins, many cell wall bound exoenzymes and outer membrane proteins (21). Although the overall sequence similarity of SLH domains is rather low, a highly conserved TRAE motif has been identified to play a key role for the binding function of SLH domains (22). For S-layer proteins, it has been shown that these anchoring modules do not directly bind to peptidoglycan, but to secondary cell wall polymers (SCWPs) carrying negative charges from pyruvate residues. In *Bacillus anthracis* and *Thermus thermophilus*, pyruvylation of SCWPs was shown to be dependent on the presence and activity of the polysaccharide pyruvyltransferase CsaB (23-25). Further, in different *Bacillus* strains, a perfect correlation between the occurrence of CsaB homologues and the presence of SLH domains does exist, leading to the conclusion that the interaction between pyruvylated SCWP and SLH domains is widespread in bacteria and has been conserved during evolution (23).

The mesophilic, Gram-positive organism *Paenibacillus alvei* CCM 2051^T possesses a naturally *O*-glycosylated S-layer protein and a pyruvate-containing SCWP with the following structure: [(Pyr4,6)-β-D-Man*p*NAc-(1→4)-β-D-Glc*p*NAc-(1→3)]$_{n\sim11}$-(Pyr4,6)-β-D-Man*p*NAc-(1→4)-α-D-Glc*p*NAc-(1→. These pyruvate-containing and, hence, overall anionic glycan chains are linked via phosphate-containing groups to muramic acid residues of the peptidoglycan layer (26).

In this report, we describe the identification of the *spaA* gene coding for the S-layer protein of *P. alvei* CCM 2051^T and its utilization as a surface display system. To proof

the functional presentation of heterologous proteins and peptide epitopes on the cell surface of *P. alvei* CCM 2051^T we expressed full-length SpaA with a carboxyterminal hexahistidine or eGFP tag, both of which were shown to be surface located and glycosylated. To our knowledge, this is the first report on *in vivo* bacterial cell surface display of a functionalized glycoprotein, opening up new opportunities for controlled high-density display of bioactive glycoproteins with defined orientation. When the recently generated S-layer glycosylation-deficient *wsfP* mutant was used for surface display (27), exclusively non-glycosylated chimeric SpaA proteins were obtained.

To identify the genetic locus encoding the biosynthesis of the native anchoring structure of the S-layer glycoprotein to the cell wall of *P. alvei*, we sequenced the upstream region of *spaA* and detected several open reading frames coding for enzymes putatively being involved in the biosynthesis of the pyruvate-containing SCWP.

MATERIALS AND METHODS

Bacterial strains and growth conditions. *P. alvei* CCM 2051^T (Table 1) was obtained from the Czech Collection of Microorganisms (CCM, Brno, Czech Republic) and was grown at 37°C and 200 rpm in Luria-Bertani (LB) broth or on LB agar plates supplemented with 10 µg/ml chloramphenicol (Cm), when appropriate. *Escherichia coli* DH5α (Invitrogen, Lofer, Austria) was grown in LB broth at 37°C supplemented with 30 µg/ml Cm, when appropriate.

Analytical and general methods. Genomic DNA of *P. alvei* CCM 2051^T was isolated as described recently (27). Restriction and cloning enzymes were purchased from Invitrogen. The MinElute gel extraction kit (Qiagen, Vienna, Austria) was used to purify DNA fragments from agarose gels, and the MinElute reaction cleanup kit (Qiagen) was used to purify digested oligonucleotides and plasmids. Plasmid DNA from transformed cells was isolated with the Plasmid Miniprep kit (Qiagen). Agarose gel electrophoresis was performed as described elsewhere (28). Transformation of *E. coli* DH5α was done according to the manufacturer's protocol (Invitrogen). Transformands were screened by *in situ* PCR using RedTaq ReadyMix PCR mix (Sigma-Aldrich, Vienna, Austria); recombinant clones were analyzed by restriction

mapping and confirmed by sequencing (Agowa, Berlin, Germany). Transformation of *P. alvei* CCM 2051^T was performed as described recently (27). Sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) was carried out according to a standard protocol (29) using a Protean II electrophoresis apparatus (Bio-Rad, Vienna, Austria). Protein bands were visualized with Coomassie Brilliant Blue G250 staining reagent. Western blotting using a Mini Trans-Blot Cell (Bio-Rad) was performed to transfer the proteins to a polyvinylidene difluoride membrane (Bio-Rad). Anti-GFP antibody (Roche, Vienna, Austria) and anti-His-tag antibody (Novagen, Darmstadt, Germany) were used to detect eGFP and hexahistidine epitopes, respectively. Western blot development using goat anti-mouse IgG AP conjugate (Novagen) in conjunction with the INT/BCIP system (Roche) was accomplished according to the manufacturer's instructions. Isolation and purification of S-layer glycoprotein essentially followed published methods (30). Qiagen's Ni-NTA Spin Kit was used to purify hexahistidine tagged S-layer glycoprotein under denaturing conditions. Deglycosylation of S-layer glycoprotein and N-terminal sequencing (Edman degradation) was performed as published previously (31).

PCR and DNA sequencing. PCR (My Cycler^{TM}, Bio-Rad) was performed using the Herculase ® II Fusion DNA Polymerase (Stratagene, La Jolla, CA, USA). For each primer pair (Table 2), PCR conditions were optimized, and amplification products were purified using the MinElute PCR purification kit (Qiagen). Primers for PCR and DNA sequencing were purchased from Invitrogen. For sequence determination of the *spaA* gene including upstream and downstream regions, chromosome walking was applied as previously described (32, 33).

Sequence analysis. Nucleotide and protein sequences were analyzed using the BLASTN and BLASTP sequence homology analysis tools (National Center for Biotechnology Information, Bethesda, MD, USA). Open reading frames in the DNA sequence were identified by using the Clone Manager Professional Suite (SECentral, Cary, NC, USA) and the ORF Finder analysis tool (National Center for Biotechnology Information). For the identification of putative protein transmembrane-spanning domains and the presence and location of signal peptide cleavage sites, the TMHMM Server v. 2.0 transmembrane prediction program and the SignalP 3.0 Server (Center for Biological Sequence Analysis, Technical University of Denmark, Lyngby,

Denmark) were used, respectively. The GC content of a certain DNA sequence was determined using the GC Content and GC Skew program (Nano+Bio-Center, University of Kaiserslautern, Germany). For *in silico* reverse translation, the Sequence Manipulation Suite was used (34). Bacterial promoters, transcriptional terminators, operons and genes were predicted by the BProm and FindTerm modules of the FGenesB gene prediction program in Molquest software (SoftBerry Inc., Mount Kisco, NY, USA). The presence of conserved motifs in a given protein sequence was analyzed by the Pfam protein families database (35), the SUPERFAMILY database (36), and the Conserved Domain Finder of the National Center for Biotechnology Information (37-39). Physical and chemical parameters for a given protein were calculated using the ProtParam tool (40).

Construction of SpaA surface display constructs. The *P. alvei* expression vector pEXALV was used for construction of all SpaA surface display constructs (Figure 1).
A carboxyterminal hexahistidine tag was fused to the coding sequence of *spaA* by PCR using primers SP_SpaA_*Sph*I_for and SpaA_6HIS_STOP_*Kpn*I_rev, using genomic DNA of *P. alvei* CCM 2051^T as template. The ~3,000-bp PCR product was digested with *Sph*I and *Kpn*I and ligated into *Sph*I/*Kpn*I-linearized and dephosphorylated plasmid pEXALV. This construct was named pEXALV_SP_SpaA_6HIS.
For the creation of an S-layer-eGFP fusion protein, the DNA fragment encoding the *spaA* gene lacking the TAA stop codon was amplified by PCR using primers SP_SpaA_*Sph*I_for and SpaA_noSTOP_*Pst*I_rev, with genomic DNA of *P. alvei* CCM 2051^T as template. The ~3,000-bp PCR product was digested with *Sph*I and *Pst*I and ligated into *Sph*I/*Pst*I-linearized and dephosphorylated plasmid pEXALV. This construct was named pEXALV_SP_SpaA. The 742-bp *egfp* fragment was amplified by PCR from plasmid pEGFP-N1 using the primers eGFP_for_*Pst*I and eGFP_rev_*Kpn*I, digested with *Pst*I and *Kpn*I and cloned *in frame* into *Pst*I/*Kpn*I-linearized and dephosphorylated plasmid pEXALV_SP_SpaA. This construct was named pEXALV_SP_SpaA_eGFP.

Analysis of the cell surface expression and immunofluorescence staining of SpaA constructs. The surface accessibility of the displayed hexahistidine-tagged S-layer protein SpaA_6HIS on whole *P. alvei* cells was analyzed by direct

immunofluorescence staining in conjunction with fluorescence microscopy. Briefly, the cells transformed with pEXALV_SP_SpaA_6HIS were harvested after expression at an OD_{600} = ~0.6, resuspended and washed three times in phosphate-buffered saline (PBS). After resuspension in 200 µl of PBS, 10 µl of Penta-His Alexa Fluor 555 Conjugate (Qiagen) were added and incubated for two hours at room temperature on a horizontal shaker. After washing three times, the cells were resuspended in 500 µl of PBS and analyzed by fluorescence microscopy.

To directly assess the functional surface expression of the chimeric S-layer fusion protein SpaA_eGFP, whole *P. alvei* cells transformed with pEXALV_SP_SpaA_eGFP were analyzed by fluorescence microscopy. Therefore, cells expressing SpaA_eGFP were harvested at an OD_{600} = ~0.6, washed three times in PBS and finally resuspended in 500 µl of PBS.

Fluorescence microscope imaging of whole *P. alvei* cells was carried out with a Nikon Eclipse TE2000-S inverted fluorescence microscope with a Hg vapor lamp, a Nikon digital sight DS-Qi1Mc camera and the NIS-Elements imaging software using the TRITC (540/25 nm for excitation light, 605/55 nm for emission light) filter block for Alexa Fluor 555 and the GFP LP (480/40 nm for excitation light, long pass at 510 nm for emission light) filter block for eGFP.

RESULTS

General description of the organism. The cell surface of the type strain *P. alvei* CCM 2051^T (= ATCC 6344; DSM 29) is completely covered with an oblique S-layer lattice composed of identical glycoprotein species. Various aspects of its crystalline S-layer including ultrastructural characterization (41), glycosylation analysis (42, 43), and glycan biosynthesis (44) have been investigated in the past. In SDS-PAGE, the S-layer glycoprotein is separated into three bands with apparent molecular masses of approximately 105, 160, and 240 kDa, respectively. The two high molecular mass bands give a positive PAS staining reaction (27), which leads to the interpretation, that in the 160-kDa glycoprotein species only one glycosylation site is occupied and in the 240-kDa glycoprotein species two glycosylation sites are occupied, whereas the 105-kDa band represents non-glycosylated S-layer protein. Similar situations

were observed for the S-layer glycoproteins of *Geobacillus stearothermophilus* NRS 2004/3a and *G. tepidamans* GS5-97^T (45, 46).

Isolation and molecular characterization of *spaA*. After chemical deglycosylation of the S-layer glycoprotein, Edman degradation revealed the N-terminus of the mature protein to have the amino acid sequence ADAAKTTQEK. Based on this information, the degenerate oligonucleotide primer proof_wSpa_for was designed by *in silico* reverse translation. Thus, identification of the entire *spaA* gene was achieved using gene walking approaches, starting with the primer specific for the N-terminus of *spaA*. The sequence has been deposited at GenBank under the accession number FJ751775. The *spaA* gene revealed one ORF extending 2952 nt, encoding a putative protein of 983 amino acids (aa) with a calculated molecular mass of 108.55 kDa. The ORF starts with an ATG at nucleotide position 1, preceded by a typical ribosomal binding site (Shine-Dalgarno sequence) of 19 nt upstream of the start codon and a putative promoter region comprising a -10 sequence (TTGTATAAT) located 232 nt upstream of the translation start and a putative -35 sequence (TTTACG) starting 252 nt in front of the start codon. 7 nt downstream of the TAA stop codon, a putative ρ-independent transcriptional termination signal was identified. The terminator consists of a palindromic stem loop sequence of 39 nt with a perfect stem of 13 nt. The average G+C content of the whole *spaA* gene was calculated to be 38.6%.

Description of SpaA. The amino acid sequence ADAAKTTQEK gained by N-terminal sequencing of chemical deglycosylated, mature SpaA was identified at position 25 to 34, indicating that the first 24 aa constitute a signal sequence, which is cleaved at the Gly-Ala motif during biosynthetic protein processing. The overall amino acid composition of mature SpaA is within the typical data reported for S-layer proteins of Gram-positive bacteria (47), exhibiting a high content of hydrophobic and acidic amino acids and lacking cysteine residues. Mature SpaA has a calculated theoretical molecular mass of 105.95 kDa and a pI of 5.83. Analysis of the presence of conserved motifs in the S-layer protein revealed the presence of three SLH domains (PF00395) located in the aminoterminal part, spanning the regions from aa 24 to aa 65, aa 82 to aa 129 and aa 140 to aa 181 (Table 3).

Comparison of SpaA with S-layer proteins. A homology search using the BLASTP program showed that the amino acid sequence of SpaA shows high identities with other surface exposed proteins of different *Bacillaceae*, including the S-layer domain proteins of *Geobacillus sp.* Y412MC10 (ZP_03037712; 43% identity) and *Paenibacillus sp.* oral taxon 786 strain D14 (ZP_04851624; 43% identity) and the S-layer protein SlpC of *Lysinibacillus sphaericus* (ABQ00416; 35% identity). Less identities are given with several other proteins, such as the S-layer domain protein of *Paenibacillus sp.* JDR-2 (YP_003011599), the S-layer protein Sl1 of *Lysinibacillus fusiformis* (CAH03723) and the S-layer domain protein of *Anaerocellum thermophilum* DSM 6725 (YP_002574146).

Description of the SCWP biosynthesis locus. Upstream and downstream of the *spaA* gene, several genes indicative of an SCWP biosynthesis locus were identified (Figure 2). By *in silico* analysis using the *E. coli* σ^{70} promoter consensus sequence, putative promoter sites located upstream of *orf1*, *tagO*, *slhA*, *spaA*, and *orf7* were recognized. In addition to the ρ-independent bacterial terminator downstream of *spaA*, a single terminator was identified downstream of the *slhA* gene. The prediction of transcription units for the complete ~14 kb DNA fragment resulted in a polycistronic RNA containing *orf1*, *csaB* and *tagA* and four separate monocistronic RNAs for *tagO*, *slhA*, *spaA* and *orf7*.

The characteristics of these genes will be discussed in the order of their appearance, starting with *orf1*.

orf1. The gene product of *orf1* contains eight potential transmembrane domains in its carboxyterminal part and several domains of unknown function (DUFs) also found in putative bacterial cell surface and hypothetical transmembrane proteins. The signal peptide of 23 aa is followed by a 340 amino acid loop potentially facing the cell exterior. The Conserved Domain Finder (National Center for Biotechnology Information) found a weak similarity to the sublancin ABC transporter SunT of *Bacillus subtilis* 168, suggesting that this protein might act as an exporter (48).

csaB. Throughout the whole gene, *csaB* is highly homologous to genes coding for pyruvyltransferases (CsaB) in various *Bacillus* strains and the translation product

contains a polysaccharide pyruvyltransferase domain. In *Bacillus anthracis* and *Thermus thermophilus*, CsaB is involved in the addition of a pyruvyl group to the SCWP, a necessary modification for anchoring of cell wall associated proteins containing SLH domains (23, 25). The presence of a pyruvate-containing SCWP in *P. alvei* CCM 2051^T suggests, that CsaB is responsible for SCWP-pyruvylation in this organism (26).

tagA and *tagO*. The deduced 252 and 377 amino acid proteins encoded by *tagA* and *tagO* reveal high similarity to the glycosyltransferases TagA and TagO of different *Bacillaceae*, respectively. In *Bacillus subilis*, both enzymes are involved the biosynthesis of teichoic acids, where TagO couples *N*-acetylglucosamine (GlcNAc) to the membrane-embedded lipid undecaprenyl-pyrophosphate and TagA catalyzes the addition of *N*-acetylmannosamine (ManNAc) to produce the lipid-linked GlcNAc-ManNAc disaccharide (49, 50). Since a ManNAc-GlcNAc backbone disaccharide motif is found in the SCWP of *P. alvei* CCM 2051^T, both enzymes may be involved in the biosynthesis of this cell wall polysaccharide (26, 51).

slhA. The *slhA* gene product contains three carboxyterminal SLH domains and is similar to S-layer domain-containing proteins of various Gram-positive bacteria. In the central part of the putative protein, a bacterial Ig-like domain found in a variety of bacterial surface proteins was identified (52). A galactose-binding domain (CBM6) typical for proteins binding to specific ligands, such as cell-surface-attached carbohydrate substrates, was detected in the aminoterminal part. This observation leads to the suggestion that SlhA might be a cell surface-anchored exoenzyme or a receptor.

orf7. No homology was found for this incomplete gene or its putative protein product.

Expression and display of SpaA constructs on the cell surface. The expression of the hexahistidine tagged S-layer protein SpaA_6HIS and the chimeric fusion protein SpaA_eGFP in *P. alvei* CCM 2051^T wild-type and *wsfP*::Ll.LtrB mutant cells was confirmed by immunoblot analysis of whole-cell lysates using anti-His-tag and anti-GFP antibody, respectively (Figure 3A and B). In both cases, three bands corresponding to the non-glycosylated, monoglycosylated and diglycosylated forms

of SpaA_6HIS and SpaA_eGFP were detected from *P. alvei* CCM 2051^T wild-type cells carrying either pEXALV_SP_SpaA_6HIS or pEXALV_SP_SpaA_eGFP, respectively, while no such proteins were detected with *P. alvei* CCM 2051^T carrying the parental plasmid pEXALV (data not shown). Since the transfer of the completed glycan chain to the S-layer protein occurs on the external face of the cytoplasmic membrane, these results indicate that SpaA_6HIS and SpaA_eGFP are correctly expressed and targeted to the cell surface of *P. alvei* CCM 2051^T. For *P. alvei* CCM 2051^T *wsfP*::Ll.LtrB mutant cells, only a single band corresponding to non-glycosylated SpaA_6HIS and SpaA_eGFP was detected (Figure 3A and B).

To confirm the surface localization of the recombinant proteins, immunofluorescence microscopy was used. For SpaA_6HIS, *P. alvei* CCM 2051^T wild-type and *wsfP*::Ll.LtrB mutant cells were probed and fluorescently stained with the Penta-His Alexa Fluor 555 Conjugate. As shown in Figure 4A, cells harboring pEXALV_SP_SpaA_6HIS were brightly fluorescent, indicating that SpaA_6HIS was successfully displayed on the surface. The surface display of SpaA_eGFP was confirmed by direct fluorescence microscopy (Figure 4B). *P. alvei* CCM 2051^T wild-type and *wsfP*::Ll.LtrB mutant cells showed identical results in immunofluorescence microscopy, whereas cells carrying the parental plasmid pEXALV were not stained at all.

DISCUSSION

Frequently monomolecular isoporous S-layers cover prokaryotic cell surfaces during all stages of cell growth and cell division. An intact closed S-layer on an average-sized, rod-shaped prokaryotic cell consists of ~500,000 monomers making it ideally suited for highly efficient immobilization of a significant number of functional epitopes in a defined and precise orientation on the bacterial cell surface (53, 54).

In this report, we describe identification of the structural gene coding for the S-layer protein SpaA of *P. alvei* CCM 2051^T and development of a surface display system using this protein as a cell wall anchor to present a heterologous peptide epitope and a functional protein. This system is based on the continuous expression of plasmid-encoded S-layer chimera using a constitutive promoter system and its export, glycosylation as well as surface anchoring in competition to the wild-type S-layer glycoprotein.

To investigate the applicability of this system, we created the hexahistidine tagged S-layer protein SpaA_6HIS as well as the chimeric fusion protein SpaA_eGFP and demonstrated their cell surface localization by immunofluorescence staining and fluorescence microscopy. The *P. alvei* CCM 2051^T surface display vector pEXALV_SP_SpaA (Figure 5A) can be used for the translational fusion of various functional epitopes to the C-terminus of SpaA, their subsequent export and surface presentation (Figure 5B).

By immunoblot analysis, we detected for both surface displayed constructs, SpaA_6HIS and SpaA_eGFP, a non-glycosylated, monoglycosylated and diglycosylated form resulting from the native S-layer protein *O*-glycosylation system of wild-type *P. alvei* CCM 2051^T and reflecting the native SpaA *O*-glycosylation pattern. This indicates the suitability of this system for controlled surface display of heterologous glycoproteins. Since the ongoing investigation of the S-layer glycosylation machinery of *P. alvei* CCM 2051^T revealed some promising results (Zarschler *et al*., manuscript in preparation), this mesophilic organism became a favorite candidate for the design and presentation of S-layer *neo*glycoproteins by means of genetic carbohydrate engineering (55). Thereby, the S-layer glycosylation deficient mutant strain *wsfP*::Ll.LtrB, carrying an insertion in the *wsfP* gene, coding for the initiation enzyme of S-layer glycan biosynthesis, provides the possibility to switch on glycosylation of any SpaA chimera, when the activity of WsfP is reconstituted by plasmid-based expression of *wsfP* (27).

Besides the attachment of the target glycoprotein to the cell envelope, also its secretion into the medium could be a desired option. Therefore, we investigated the anchoring mechanism of the S-layer protein starting with the identification of three SLH domains located at the N-terminus of SpaA. Since these conserved motifs are known to play a key role in mediating cell wall binding of various proteins by interacting with SCWPs carrying pyruvate residues, it was not surprising to identify a pyruvyltransferase gene upstream of *spaA*. Furthermore, with *tagA* and *tagO*, both encoding glycosyltransferases, we identified two additional enzymes involved in SCWP-biosynthesis of *P. alvei* CCM 2051^T. Due to the similarity of TagO to an undecaprenyl-phosphate *N*-acetylglucosaminyl 1-phosphate transferase, we assume that TagO couples GlcNAc to the membrane-embedded lipid carrier undecaprenyl-pyrophosphate at the cytoplasmic side of the membrane. The next step in the biosynthesis pathway would be the addition of ManNAc catalyzed by the *N*-

acetylglucosaminyldiphosphoundecaprenol *N*-acetyl-β-D-mannosaminyltransferase TagA, to produce the lipid-linked GlcNAc-ManNAc disaccharide. The transfer of the pyruvate modification to ManNAc residue is accomplished by the pyruvyltransferase CsaB. Additional enzymes required for export and formation of the covalent pyrophosphate linkage between SCWP and peptidoglycan have not been identified, so far. The prediction of transcription units for the SCWP biosynthesis locus showed, that *orf1*, *csaB*, and *tagA* are located on a polycistronic mRNA. This finding is supported by the closed spacing and the absence of a promoter consensus sequence between these three genes. Due to the lack of reliable sequence similarities, the role of the gene product of *orf1* as part of this polycistronic mRNA remains speculative and needs to be further investigated. However, *tagO*, *slhA* and *spaA* are predicted to be transcribed monocistronically. The presence of three SLH domains located at the C-terminus of SlhA argues for its surface exposure and turns this protein into an eligible candidate for co-display of SpaA- and SlhA-fusion proteins interacting with each other or binding a target molecule in two different manners.

In conclusion, we have demonstrated cell surface display of a heterologous peptide epitope and a functional protein fused to the N-terminus of the S-layer glycoprotein SpaA of *P. alvei* CCM 2051^T. This concept is the starting point for future presentation of numerous peptides and proteins combined with functional glycans, which may be valuable in the fields of receptor mimics, vaccine development, or drug delivery.

ACKNOWLEDGEMENTS

We thank Andrea Scheberl and Sonja Zayni for excellent technical assistance. Financial support came from the Austrian Science Fund, project P20745-B11 (to P.M.) and projects P19047-B12 and P20605-12 (to C.S.), and the Hochschul-jubiläumsstiftung der Stadt Wien, project H-02229-2007 (to K.Z.).

REFERENCES

1. **Georgiou, G., C. Stathopoulos, P. S. Daugherty, A. R. Nayak, B. L. Iverson, and R. Curtiss, 3rd.** 1997. Display of heterologous proteins on the surface of microorganisms: from the screening of combinatorial libraries to live recombinant vaccines. Nature Biotechnology **15:**29-34.

2. **Miller, K. D., N. B. Pefaur, and C. L. Baird.** 2008. Construction and screening of antigen targeted immune yeast surface display antibody libraries. Current Protocols in Cytometry **45:**1-30.
3. **Daugherty, P. S.** 2007. Protein engineering with bacterial display. Current Opinion in Structural Biology **17:**474-480.
4. **Ishige, T., K. Honda, and S. Shimizu.** 2005. Whole organism biocatalysis. Current Opinion in Chemical Biology **9:**174-180.
5. **Wu, C. H., A. Mulchandani, and W. Chen.** 2008. Versatile microbial surface-display for environmental remediation and biofuels production. Trends in Microbiology **16:**181-188.
6. **Saleem, M., H. Brim, S. Hussain, M. Arshad, M. B. Leigh, and H. Zia ul.** 2008. Perspectives on microbial cell surface display in bioremediation. Biotechnology Advances **26:**151-161.
7. **Wernerus, H., and S. Ståhl.** 2004. Biotechnological applications for surface-engineered bacteria. Biotechnology and Applied Biochemistry **40:**209-228.
8. **Samuelson, P., E. Gunneriusson, P. A. Nygren, and S. Ståhl.** 2002. Display of proteins on bacteria. Journal of Biotechnology **96:**129-154.
9. **Lee, S. Y., J. H. Choi, and Z. Xu.** 2003. Microbial cell-surface display. Trends in Biotechnology **21:**45-52.
10. **Desvaux, M., E. Dumas, I. Chafsey, and M. Hebraud.** 2006. Protein cell surface display in Gram-positive bacteria: from single protein to macromolecular protein structure. FEMS Microbiology Letters **256:**1-15.
11. **Pepper, L. R., Y. K. Cho, E. T. Boder, and E. V. Shusta.** 2008. A decade of yeast surface display technology: where are we now? Combinatorial Chemistry and High Throughput Screening **11:**127-134.
12. **Gai, S. A., and K. D. Wittrup.** 2007. Yeast surface display for protein engineering and characterization. Current Opinion in Structural Biology **17:**467-473.
13. **Smit, J.** 2008. Heads up S-layer display: the power of many. Structure **16:**1151-1153.
14. **Bingle, W. H., J. F. Nomellini, and J. Smit.** 1997. Cell-surface display of a *Pseudomonas aeruginosa* strain K pilin peptide within the paracrystalline S-layer of *Caulobacter crescentus*. Molecular Microbiology **26:**277-288.

15. **Bingle, W. H., J. F. Nomellini, and J. Smit.** 1997. Linker mutagenesis of the *Caulobacter crescentus* S-layer protein: toward a definition of an N-terminal anchoring region and a C-terminal secretion signal and the potential for heterologous protein secretion. Journal of Bacteriology **179:**601-611.
16. **Bingle, W. H., J. F. Nomellini, and J. Smit.** 2000. Secretion of the *Caulobacter crescentus* S-layer protein: further localization of the C-terminal secretion signal and its use for secretion of recombinant proteins. Journal of Bacteriology **182:**3298-3301.
17. **Mesnage, S., E. Tosi-Couture, M. Mock, and A. Fouet.** 1999. The S-layer homology domain as a means for anchoring heterologous proteins on the cell surface of *Bacillus anthracis*. Journal of Applied Microbiology **87:**256-260.
18. **Mesnage, S., M. Weber-Levy, M. Haustant, M. Mock, and A. Fouet.** 1999. Cell surface-exposed tetanus toxin fragment C produced by recombinant *Bacillus anthracis* protects against tetanus toxin. Infection and Immunity **67:**4847-4850.
19. **Mesnage, S., E. Tosi-Couture, and A. Fouet.** 1999. Production and cell surface anchoring of functional fusions between the SLH motifs of the *Bacillus anthracis* S-layer proteins and the *Bacillus subtilis* levansucrase. Molecular Microbiology **31:**927-936.
20. **Nomellini, J. F., G. Duncan, I. R. Dorocicz, and J. Smit.** 2007. S-layer-mediated display of the immunoglobulin G-binding domain of streptococcal protein G on the surface of *Caulobacter crescentus*: development of an immunoactive reagent. Applied and Environmental Microbiology **73:**3245-3253.
21. **Engelhardt, H., and J. Peters.** 1998. Structural research on surface layers: a focus on stability, surface layer homology domains, and surface layer-cell wall interactions. Journal of Structural Biology **124:**276-302.
22. **May, A., T. Pusztahelyi, N. Hoffmann, R. J. Fischer, and H. Bahl.** 2006. Mutagenesis of conserved charged amino acids in SLH domains of *Thermoanaerobacterium thermosulfurigenes* EM1 affects attachment to cell wall sacculi. Archives of Microbiology **185:**263-269.
23. **Mesnage, S., T. Fontaine, T. Mignot, M. Delepierre, M. Mock, and A. Fouet.** 2000. Bacterial SLH domain proteins are non-covalently anchored to

the cell surface via a conserved mechanism involving wall polysaccharide pyruvylation. EMBO Journal **19:**4473-4484.

24. **Sára, M.** 2001. Conserved anchoring mechanisms between crystalline cell surface S-layer proteins and secondary cell wall polymers in Gram-positive bacteria? Trends in Microbiology **9:**47-49; discussion 49-50.
25. **Cava, F., M. A. de Pedro, H. Schwarz, A. Henne, and J. Berenguer.** 2004. Binding to pyruvylated compounds as an ancestral mechanism to anchor the outer envelope in primitive bacteria. Molecular Microbiology **52:**677-690.
26. **Schäffer, C., N. Müller, P. K. Mandal, R. Christian, S. Zayni, and P. Messner.** 2000. A pyrophosphate bridge links the pyruvate-containing secondary cell wall polymer of *Paenibacillus alvei* CCM 2051 to muramic acid. Glycoconjugate Journal **17:**681-690.
27. **Zarschler, K., B. Janesch, S. Zayni, C. Schäffer, and P. Messner.** 2009. Construction of a gene knockout system for application in *Paenibacillus alvei* CCM 2051^{T}, exemplified by the S-layer glycan biosynthesis initiation enzyme WsfP. Applied and Environmental Microbiology **75:**3077-3085.
28. **Sambrook, J., E. F. Fritsch, and T. Maniatis.** 1989. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
29. **Laemmli, U. K.** 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature **227:**680-685.
30. **Messner, P., and U. B. Sleytr.** 1988. Separation and purification of S-layers from Gram-positive and Gram-negative bacteria, p. 97-104. *In* I. C. Hancock and I. R. Poxton (ed.), Bacterial cell surface techniques. John Wiley & Sons, Chichester.
31. **Schäffer, C., T. Wugeditsch, H. Kählig, A. Scheberl, S. Zayni, and P. Messner.** 2002. The surface layer (S-layer) glycoprotein of *Geobacillus stearothermophilus* NRS 2004/3a. Analysis of its glycosylation. Journal of Biological Chemistry **277:**6230-6239.
32. **Kneidinger, B., M. Graninger, and P. Messner.** 2001. Chromosome walking by cloning of distinct PCR fragments. Biotechniques **30:**248-249.
33. **Pilhofer, M., A. P. Bauer, M. Schrallhammer, L. Richter, W. Ludwig, K. H. Schleifer, and G. Petroni.** 2007. Characterization of bacterial operons

consisting of two tubulins and a kinesin-like gene by the novel two-step gene walking method. Nucleic Acids Research **35:**E135.

34. **Stothard, P.** 2000. The sequence manipulation suite: JavaScript programs for analyzing and formatting protein and DNA sequences. Biotechniques **28:**1102-1104.
35. **Finn, R. D., J. Tate, J. Mistry, P. C. Coggill, S. J. Sammut, H. R. Hotz, G. Ceric, K. Forslund, S. R. Eddy, E. L. Sonnhammer, and A. Bateman.** 2008. The Pfam protein families database. Nucleic Acids Research **36:**D281-D288.
36. **Gough, J., K. Karplus, R. Hughey, and C. Chothia.** 2001. Assignment of homology to genome sequences using a library of hidden Markov models that represent all proteins of known structure. Journal of Molecular Biology **313:**903-919.
37. **Marchler-Bauer, A., J. B. Anderson, P. F. Cherukuri, C. DeWeese-Scott, L. Y. Geer, M. Gwadz, S. He, D. I. Hurwitz, J. D. Jackson, Z. Ke, C. J. Lanczycki, C. A. Liebert, C. Liu, F. Lu, G. H. Marchler, M. Mullokandov, B. A. Shoemaker, V. Simonyan, J. S. Song, P. A. Thiessen, R. A. Yamashita, J. J. Yin, D. Zhang, and S. H. Bryant.** 2005. CDD: a conserved domain database for protein classification. Nucleic Acids Research **33:**D192-D196.
38. **Marchler-Bauer, A., J. B. Anderson, M. K. Derbyshire, C. DeWeese-Scott, N. R. Gonzales, M. Gwadz, L. Hao, S. He, D. I. Hurwitz, J. D. Jackson, Z. Ke, D. Krylov, C. J. Lanczycki, C. A. Liebert, C. Liu, F. Lu, S. Lu, G. H. Marchler, M. Mullokandov, J. S. Song, N. Thanki, R. A. Yamashita, J. J. Yin, D. Zhang, and S. H. Bryant.** 2007. CDD: a conserved domain database for interactive domain family analysis. Nucleic Acids Research **35:**D237-D240.
39. **Marchler-Bauer, A., and S. H. Bryant.** 2004. CD-Search: protein domain annotations on the fly. Nucleic Acids Research **32:**W327-W331.
40. **Gasteiger, E., C. Hoogland, A. Gattiker, S. Duvaud, M. R. Wilkins, R. D. Appel, and A. Bairoch.** 2005. Protein identification and analysis tools on the ExPASy server, p. 571-607. *In* J. M. Walker (ed.), The Proteomics Protocols Handbook. Humana Press, Totowa, NJ.
41. **Sára, M., K. Moser-Thier, U. Kainz, and U. B. Sleytr.** 1990. Characterization of S-layers from mesophilic *Bacillaceae* and studies on their protective role towards muramidases. Archives of Microbiology **153:**209-214.

42. **Altman, E., J. R. Brisson, P. Messner, and U. B. Sleytr.** 1991. Structure of the glycan chain from the surface layer glycoprotein of *Bacillus alvei* CCM 2051. Biochemistry and Cell Biology **69:**72-78.

43. **Messner, P., R. Christian, C. Neuninger, and G. Schulz.** 1995. Similarity of "core" structures in two different glycans of tyrosine-linked eubacterial S-layer glycoproteins. Journal of Bacteriology **177:**2188-2193.

44. **Hartmann, E., P. Messner, G. Allmeier, and H. König.** 1993. Proposed pathway for biosynthesis of the S-layer glycoprotein of *Bacillus alvei*. Journal of Bacteriology **175:**4515-4519.

45. **Steiner, K., G. Pohlentz, K. Dreisewerd, S. Berkenkamp, P. Messner, J. Peter-Katalinic, and C. Schäffer.** 2006. New insights into the glycosylation of the surface layer protein SgsE from *Geobacillus stearothermophilus* NRS 2004/3a. Journal of Bacteriology **188:**7914-7921.

46. **Zayni, S., K. Steiner, A. Pföstl, A. Hofinger, P. Kosma, C. Schäffer, and P. Messner.** 2007. The dTDP-4-dehydro-6-deoxyglucose reductase encoding *fcd* gene is part of the surface layer glycoprotein glycosylation gene cluster of *Geobacillus tepidamans* GS5-97^{T}. Glycobiology**:**433-443.

47. **Sleytr, U. B., and T. J. Beveridge.** 1999. Bacterial S-layers. Trends in Microbiology **7:**253-260.

48. **Paik, S. H., A. Chakicherla, and J. N. Hansen.** 1998. Identification and characterization of the structural and transporter genes for, and the chemical and biological properties of, sublancin 168, a novel lantibiotic produced by *Bacillus subtilis* 168. Journal of Biological Chemistry **273:**23134-23142.

49. **Bhavsar, A. P., R. Truant, and E. D. Brown.** 2005. The TagB protein in *Bacillus subtilis* 168 is an intracellular peripheral membrane protein that can incorporate glycerol phosphate onto a membrane-bound acceptor *in vitro*. Journal of Biological Chemistry **280:**36691-36700.

50. **Formstone, A., R. Carballido-Lopez, P. Noirot, J. Errington, and D. J. Scheffers.** 2008. Localization and interactions of teichoic acid synthetic enzymes in *Bacillus subtilis*. Journal of Bacteriology **190:**1812-1821.

51. **Schäffer, C., and P. Messner.** 2005. The structure of secondary cell wall polymers: how Gram-positive bacteria stick their cell walls together. Microbiology **151:**643-651.

52. **Pavkov, T., E. M. Egelseer, M. Tesarz, D. I. Svergun, U. B. Sleytr, and W. Keller.** 2008. The structure and binding behavior of the bacterial cell surface layer protein SbsC. Structure **16:**1226-1237.

53. **Sleytr, U. B., E. M. Egelseer, N. Ilk, D. Pum, and B. Schuster.** 2007. S-Layers as a basic building block in a molecular construction kit. Febs Journal **274:**323-334.

54. **Sleytr, U. B.** 1978. Regular arrays of macromolecules on bacterial cell walls: structure, chemistry, assembly, and function. International Review of Cytology **53:**1-62.

55. **Schäffer, C., and P. Messner.** 2001. Glycobiology of surface layer proteins. Biochimie **83:**591-599.

Table 1. Bacterial strains and plasmids used in this study.

Strain or plasmid	Genotype and/or relevant characteristics	Source or reference
Strains		
P. alvei CCM 2051^T	wild-type isolate, Km^r	Czech Collection of Micro-organisms
P. alvei CCM 2051^T *wsfP*::Ll.LtrB	S-layer glycosylation deficient mutant carrying a targetron insertion at the *wsfP* gene; Km^r	(27)
Escherichia coli DH5α	F$^-$ ϕ80d*lacZ*Δ*M15* Δ(*lacZYA-argF*)*U169 deoR recA1 endA1 hsdR17* (rK$^-$ mK$^-$) *phoA supE44 thi-1 gyrA96 relA1* λ^-	Invitrogen
Plasmids		
pEXALV	*P. alvei* expression vector	(27)
pEGFP-N1	Expression vector for mammalian cells encoding a red-shifted variant of wild-type GFP	BD Biosciences

pEXALV_SP_SpaA_6HIS	pEXALV carrying the his-tagged *spaA* gene of *P. alvei* CCM 2051^T	This study
pEXALV_SP_SpaA	pEXALV carrying the *spaA* gene of *P. alvei* CCM 2051^T lacking the TAA stop codon	This study
pEXALV_SP_SpaA_eGFP	pEXALV carrying a *spaA-egfp* fusion construct	This study

Table 2. Oligonucleotide primers used for PCR amplification reactions.

Oligonucleotide	**Sequence (5' → 3')[a]**
proof_wSpa_for	GCIGAYGCIGCIAARACIACICARG
SP_SpaA_*Sph*I_for	aatcaGCATGCAGAAAAGATTGGCCCTTCTGCTTTCCG
SpaA_6HIS_STOP_*Kpn*I_rev	aatcaGGTACCttaatggtgatggtgatggtgCTTACCGGAGTATGTTCCAGGAAGG
SpaA_noSTOP_*Pst*I_rev	aatcaCTGCAGCTTACCGGAGTATGTTCCAGGAAGG
eGFP_for_*Pst*I	aatcaCTGCAGATGGTGAGCAAGGGCGAGGAGC
eGFP_rev_*Kpn*I	aatcaGGTACCTTACTTGTACAGCTCGTCCATGCC

[a]Artificial restriction sites are underlined

Table 3. Predicted gene products encoded by the SCWP biosynthesis locus of *P. alvei* CCM 2051^T together with database homologies.

ORF	Length / Mol. mass	Conserved motifs and regions	Related proteins			
			Name / putative function	Organism	Identity / similarity (%)	Accession No.
orf1	657 / 73.1	Signal peptide 1-23 DUF916 (PF06030) 1-30 DUF204 (PF02659) 341-395 DUF808 (PF05661) 460-641 DUF981 (PF06168) 521-558 DUF2061 (PF09834) 608-650	hypothetical protein GYMC10DRAFT_1551 hypothetical protein Pjdr2DRAFT_3486	*Geobacillus sp.* Y412MC10 *Paenibacillus sp.* JDR-2	52 / 72 48 / 66	ZP_03037718 ZP_02848379
csaB	396 / 43.5	PS_pyruv_trans (PF04230) 7-357	polysaccharide pyruvyltransferase CsaB protein	*Geobacillus sp.* Y412MC10 *Bacillus anthracis* str. Ames	49 / 64 37 / 56	ZP_03037717 NP_843396
tagA	252 / 28.6	Glyco_tran_WecB (PF03808) 60-231	glycosyltransferase, WecB/TagA/CpsF family *N*-acetylglucosaminyldiphosphoundecaprenol *N*-acetyl-β-D-mannosaminyltransferase	*Paenibacillus sp.* JDR-2 *Bacillus cereus* ATCC 14579	62 / 79 45 / 65	ZP_02848377 NP_835080
tagO	377 / 40.4	Signal peptide 1-33 Glycos_transf_4 (PF00953) 84-247	TagO protein undecaprenyl-phosphate *N*-acetylglucosaminyl 1-phosphate transferase	*Bacillus megaterium* DSM319 *Lysinibacillus sphaericus* C3-41	53 / 72 52 / 72	CAL44583 YP_001696879
slhA	1335 / 148.4	Signal peptide 1-31 Galactose-binding domain-like (CBM6) 91-200 Big_3 (PF07523) 493-513 SLH (PF00395) 1125-1169 SLH (PF00395) 1198-1242 SLH (PF00395) 1267-1319	S-layer domain protein hypothetical protein BBR47_54190	*Paenibacillus sp.* JDR-2 *Brevibacillus brevis* NBRC 100599	49 / 63 46 / 61	ZP_02848374 YP_002774900
orf7'	468 / 52.5	Signal peptide 1-20	hypothetical protein GYMC10DRAFT_0729 hypothetical protein Pjdr2DRAFT_3479	*Geobacillus sp.* Y412MC10 *Paenibacillus sp.* JDR-2	24 / 44 24 / 44	ZP_03036896 ZP_02848309

FIGURE LEGENDS

Figure 1. Schematic drawing of the SpaA_6HIS and SpaA_eGFP fusion constructs cloned into the expression vector pEXALV. Squares indicate the aminoterminal signal peptide (SP), whereas the three S-layer homology (SLH) domains are highlighted in grey. The hexahistidine and eGFP tags are shown as circles or black box, respectively.

Figure 2. Genetic organization of the proposed SCWP biosynthesis locus of *P. alvei* CCM 2051^T. Predicted open reading frames are indicated by horizontal arrows with the respective gene designations indicated above the arrow. Genes encoding similar functions in SCWP biosynthesis have a similar grey scaling code. Genes coding for proteins with unknown function are indicated in black. Genes highlighted in grey encode putative glycosyltransferases. The white arrow indicates the *csaB* gene encoding the pyruvyltransferase. The structural gene *spaA* is highlighted in lightest grey. Putative promoters and terminators are represented as flags and hairpins, respectively. Monocistronic and polycistronic mRNAs are depicted as vertical black arrows. The graphic representation of the G+C percentages is given below the locus map.

Figure 3. Western immunoblot detection of (A) SpaA_6HIS using an anti-His-tag and (B) of SpaA_eGFP using an antibody raised against GFP. The S-layer constructs were analyzed by SDS–PAGE (10% gel) and transfer of the samples to a PVDF membrane followed by immunoblot detection. Tri-banded appearance corresponds to non-glycosylated (N), monoglycosylated (M), and diglycosylated (D) chimeric SpaA produced by *P. alvei* CCM 2051^T wild-type cells (wt). For *P. alvei* CCM 2051^T *wsfP*::Ll.LtrB mutant cells, single bands corresponding to non-glycosylated SpaA chimera were detected (*wsfP*::Ll.LtrB).

Figure 4. Representative immunofluorescence images of control *P. alvei* CCM 2051^T wild-type cells harboring pEXALV (A, B) and recombinant *P. alvei* CCM 2051^T wild-type cells harboring pEXALV_SP_SpaA_6HIS (C) and pEXALV_SP_SpaA_eGFP (D), respectively. For immunofluorescence staining of surface-located SpaA_6HIS, cells were probed with the Penta-His Alexa Fluor 555 Conjugate (Qiagen). The

TRITC and the GFP LP filter blocks were used for detection of Alexa Fluor 555 (A, C) and eGFP (B, D), respectively. Corresponding brightfield images of the same cells are shown on the left. The immunofluorescence images of *wsfP*::Ll.LtrB mutant strain are identical to those observed for wild-type cells and, hence, not shown.

Figure 5. (A) Schematic representation of the *P. alvei* CCM 2051^T surface display vector pEXALV_SP_SpaA. P_{SgsE} indicates the *sgsE* S-layer gene promoter of *Geobacillus stearothermophilus* NRS 2004/3a. SP_SpaA highlights the open reading frame coding for the S-layer protein SpaA including the corresponding signal peptide. (B) Schematic picture showing the presentation of functional epitopes in a defined and precise orientation on the cell wall of *P. alvei* CCM 2051^T.


S-layer protein
Spacer
Tag
SP SLH SLH SLH
1 24
SpaA
983
HHHHHH
hexahistidine tag
SP SLH SLH SLH
1 24
SpaA
983
LQ
eGFP

Figure 1

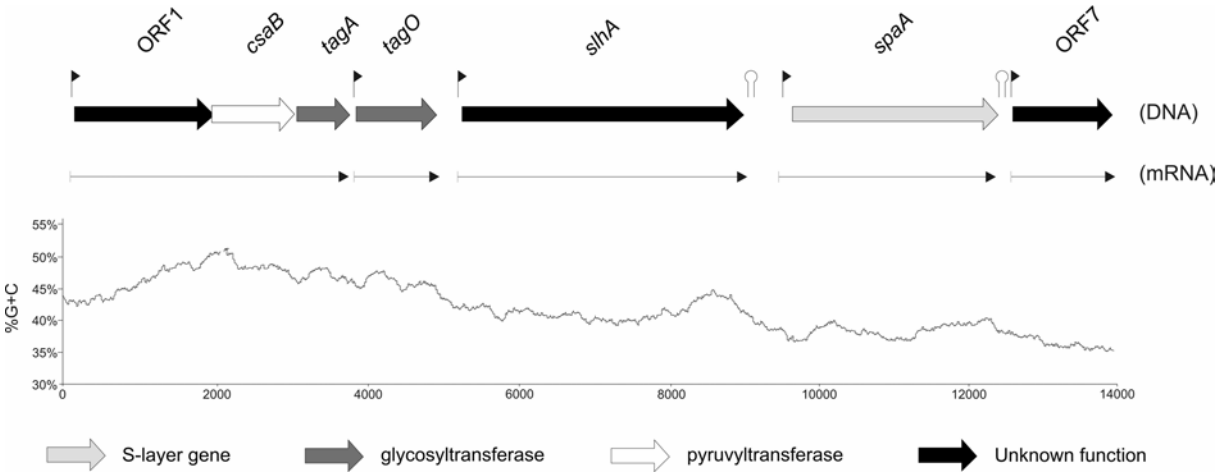

ORF1
csaB
tagA
tagO
slhA
spaA
ORF7
(DNA)
(mRNA)
%G+C
55%
50%
45%
40%
35%
30%
0
2000
4000
6000
8000
10000
12000
14000
S-layer gene
glycosyltransferase
pyruvyltransferase
Unknown function

Figure 2

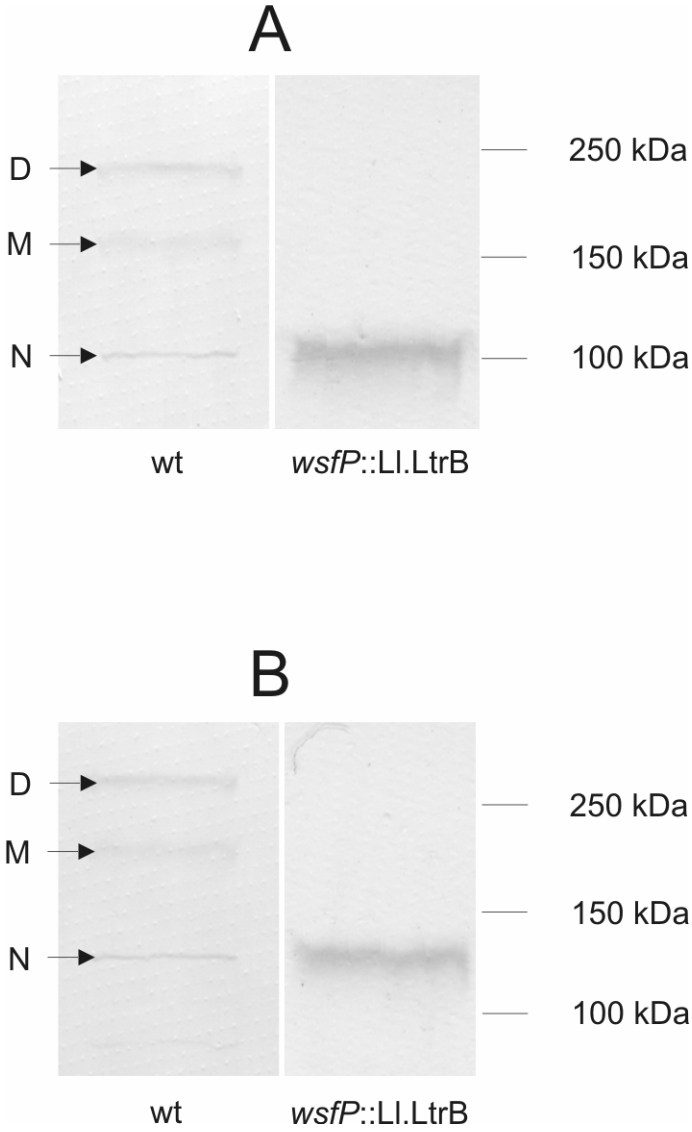

A
D
M
N
250 kDa
150 kDa
100 kDa
wt
wsfP::Ll.LtrB
B
D
M
N
250 kDa
150 kDa
100 kDa
wt
wsfP::Ll.LtrB

Figure 3

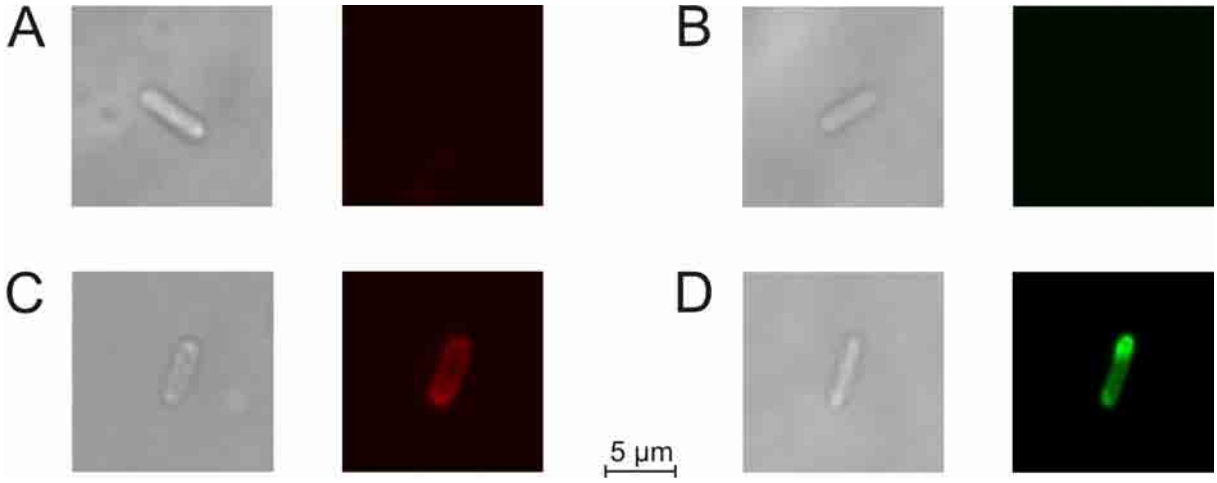


Figure 4

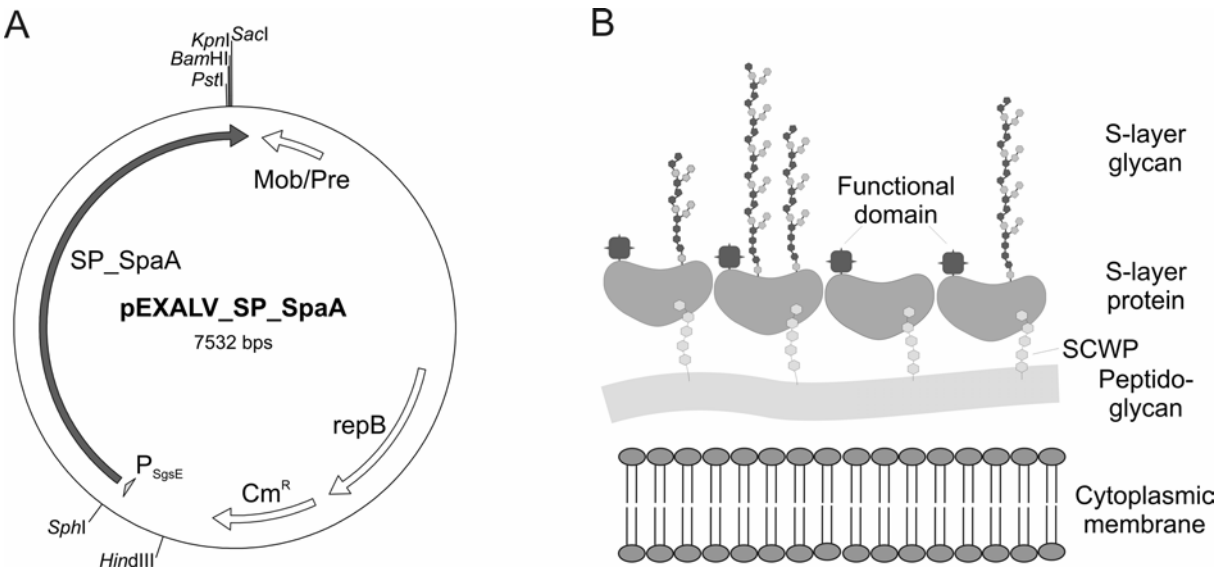

A
KpnI
SacI
BamHI
PstI
Mob/Pre
SP_SpaA
pEXALV_SP_SpaA
7532 bps
repB
P SgsE
Cm R
SphI
HindIII
B
Functional domain
S-layer glycan
S-layer protein
SCWP
Peptido-glycan
Cytoplasmic membrane

Figure 5

Applied and Environmental Microbiology, May 2009, p. 3077–3085 Vol. 75, No. 10
0099-2240/09/$08.00+0 doi:10.1128/AEM.00087-09

Construction of a Gene Knockout System for Application in *Paenibacillus alvei* CCM 2051^T, Exemplified by the S-Layer Glycan Biosynthesis Initiation Enzyme WsfP$^\nabla$

Kristof Zarschler, Bettina Janesch, Sonja Zayni, Christina Schäffer,* and Paul Messner*

Department für NanoBiotechnologie, Universität für Bodenkultur Wien, A-1180 Vienna, Austria

Received 14 January 2009/Accepted 17 March 2009

The gram-positive bacterium *Paenibacillus alvei* CCM 2051^T is covered by an oblique surface layer (S-layer) composed of glycoprotein subunits. The S-layer O-glycan is a polymer of [→3)-β-D-Gal*p*-(1[α-D-Glc*p*-(1→6)]→4)-β-D-Man*p*NAc-(1→] repeating units that is linked by an adaptor of -[GroA-2→OPO_2→4-β-D-Man*p*NAc-(1→4)]→3)-α-L-Rha*p*-(1→3)-α-L-Rha*p*-(1→3)-α-L-Rha*p*-(1→3)-β-D-Gal*p*-(1→ to specific tyrosine residues of the S-layer protein. For elucidation of the mechanism governing S-layer glycan biosynthesis, a gene knockout system using bacterial mobile group II intron-mediated gene disruption was developed. The system is further based on the *sgsE* S-layer gene promoter of *Geobacillus stearothermophilus* NRS 2004/3a and on the *Geobacillus-Bacillus-Escherichia coli* shuttle vector pNW33N. As a target gene, *wsfP*, encoding a putative UDP-Gal:phosphoryl-polyprenol Gal-1-phosphate transferase, representing the predicted initiation enzyme of S-layer glycan biosynthesis, was disrupted. S-layer protein glycosylation was completely abolished in the insertional *P. alvei* CCM 2051^T *wsfP* mutant, according to sodium dodecyl sulfate-polyacrylamide gel electrophoresis evidence and carbohydrate analysis. Glycosylation was fully restored by plasmid-based expression of *wsfP* in the glycan-deficient *P. alvei* mutant, confirming that WsfP initiates S-layer protein glycosylation. This is the first report on the successful genetic manipulation of bacterial S-layer protein glycosylation in vivo, including transformation of and heterologous gene expression and gene disruption in the model organism *P. alvei* CCM 2051^T.

Bacterial cell surface layer (S-layer) glycoproteins provide a unique self-assembly matrix that has been optimized by nature for regular and periodic display of glycans with nanometer scale accuracy (21, 31). Exploitation of this self-assembly system for surface display of functional, tailor-made glycans is an attractive alternative to the use of common cell surface anchors (7), with potential areas of application relating to any biological phenomenon that is based on carbohydrate recognition, such as receptor-substrate interaction, signaling, or cell-cell communication. A prerequisite for this endeavor is the availability of an S-layer glycoprotein-covered bacterium that is amenable to genetic manipulation. Despite the high application potential offered by the S-layer glycan display system, there are so far only two reports in the literature dealing with the genetic manipulation of S-layer glycoprotein-carrying bacteria. Both reports concern the gram-negative periodontal pathogen *Tannerella forsythia* ATCC 43037, but neither of them affects S-layer protein glycosylation (12, 24). In archaea, in contrast, molecular studies of S-layer protein glycosylation are quite advanced (1), but with the archaeal system, S-layer glycoprotein self-assembly, which is a prerequisite for the desired glycan display, has not been manageable in vitro so far.

Our model organisms and, hence, candidates for S-layer-mediated glycan display enabled by carbohydrate engineering techniques are members of the *Bacillaceae* family. Currently, the S-layer glycosylation system of the thermophilic bacterium *Geobacillus stearothermophilus* NRS 2004/3a is best understood (20, 23, 29, 33, 34). However, a drawback of this organism is its resistance to take up foreign DNA. Although described in the literature (13, 14, 37), transformation of thermophilic bacilli seems to be a strain-specific trait. Based on successful transformation experiments in our laboratory, the mesophilic bacterium *Paenibacillus alvei* CCM 2051^T (ATCC 6344; DSM 29) (formerly *Bacillus alvei* [4]) was chosen to set up a system for genetic manipulation. The bacterium is completely covered with an oblique S-layer lattice composed of glycoprotein species. Various aspects of its S-layer, including ultrastructural characterization (27), glycosylation analysis (2, 18), and glycan biosynthesis (11), have been investigated so far. The S-layer O-glycans are polymers of [→3)-β-D-Gal*p*-(1[α-D-Glc*p*-(1→6)]→4)-β-D-Man*p*NAc-(1→] repeating units that are linked via the adaptor -[GroA-2→OPO_2→4-β-D-Man*p*NAc-(1→4)]→3)-α-L-Rha*p*-(1→3)-α-L-Rha*p*-(1→3)-α-L-Rha*p*-(1→3)-β-D-Gal*p*-(1→ to specific tyrosine residues (2, 18) of the S-layer protein SpaA (GenBank accession number FJ751775).

Due to the presence of an identical adaptor saccharide backbone in the S-layer glycan of *G. stearothermophilus* NRS 2004/3a (29), where its biosynthesis is initiated by the UDP-Gal:phosphoryl-polyprenol Gal-1-phosphate transferase WsaP (33), it was conceivable that a homologous enzyme would initiate S-layer glycosylation in *P. alvei* CCM 2051^T. Considering that the S-layer protein glycosylation machinery has been found to be encoded by S-layer glycosylation (*slg*) gene clusters

* Corresponding author. Mailing address: Department für NanoBiotechnologie, Universität für Bodenkultur Wien, Gregor-Mendel-Strasse 33, A-1180 Vienna, Austria. Phone: 43-1-47654, ext. 2202. Fax: 43-1-4789112. E-mail for Paul Messner: paul.messner@boku.ac.at. E-mail for Christina Schäffer: christina.schaeffer@boku.ac.at.

$^\nabla$ Published ahead of print on 20 March 2009.

TABLE 1. Oligonucleotide primers used for PCR amplification reactions

Oligonucleotide	Sequence (5′ → 3′)[a]
P(SgsE)_*Hin*dIII_for	AATCAAAGCTTTGTTTTTGCACAAAATGTTTGCC
P(SgsE)_*Sph*I_rev	AATCAGCATGCAGCCTAAAATCCCCCTTCG
P(SgsE)_*Sph*I_for	AATCAGCATGCTGTTTTTGCACAAAATGTTTGCC
P(SgsE)_*Hin*dIII_rev	AATCAAAGCTTAAAGCCTAAAATCCCCCTTCG
Targe_*Sph*I_for	AATCAGCATGCGCTGGCGTAATAGCGAAGA
Targe_*Sph*I_rev	AATCAGCATGCTACCGCACAGATGCGTAAG
KO_*wsfP*_control_for_1	TCTTATCCTTGGTGCCGGTACACTTG
KO_*wsfP*_control_rev_1	AGCCTGTAATTCCAGGACGCACA
*wsfP*_for_*Sph*I	AATCAGCATGCTTCGCAAAAATCAAAGGTTTTTGTCGAAG
*wsfP*_rev_*Kpn*I	AATCAGGTACCTTAATATGCATTTTTATTTATAAACCCATTCC
*wsaP*_for_*Sph*I	AATCAGCATGCTGGTTAAGGTGATTAGAGGAAGAGAGCGG
*wsaP*_rev_*Kpn*I	AATCAGGTACCTTAATATGCATTTTTATTTACCAAACCATTGG
P_555 556s-IBS	AAAAAAGCTTATAATTATCCTTAGCTGACTGGAAAGTGCGCCCAGATAGGGTG
P_555 556s-EBS1d	CAGATTGTACAAATGTGGTGATAACAGATAAGTCTGGAAAACTAACTTACCTTTCTTTGT
P_555 556s-EBS2	TGAACGCAAGTTTCTAATTTCGATTTCAGCTCGATAGAGGAAAGTGTCT
P_654 655s-IBS	AAAAAAGCTTATAATTATCCTTAATTCTCGCACTAGTGCGCCCAGATAGGGTG
P_654 655s-EBS1d	CAGATTGTACAAATGTGGTGATAACAGATAAGTCGCACTACCTAACTTACCTTTCTTTGT
P_654 655s-EBS2	TGAACGCAAGTTTCTAATTTCGGTTAGAATCCGATAGAGGAAAGTGTCT
P_1176 1177s-IBS	AAAAAAGCTTATAATTATCCTTAAGACCCGAACGGGTGCGCCCAGATAGGGTG
P_1176 1177s-EBS1d	CAGATTGTACAAATGTGGTGATAACAGATAAGTCGAACGGCCTAACTTACCTTTCTTTGT
P_1176 1177s-EBS2	TGAACGCAAGTTTCTAATTTCGATTGGTCTTCGATAGAGGAAAGTGTCT
Cm_*Kpn*I_for	AATCAGGTACCAAGCCGATGAAGATGGA
Cm_*Kpn*I_rev	AATCAGGTACCACAGTCGGCATTATCTC

[a] Artificial restriction sites are underlined.

(21), degenerate primers for the *rml* genes catalyzing the dTDP-L-Rha biosynthesis required for building up the adaptor saccharide of the *P. alvei* CCM 2051^T S-layer glycan were used to define a point of entry into the glycosylation locus (K. Zarschler, B. Janesch, P. Messner, and C. Schäffer, unpublished data). Chromosome walking revealed the existence of an *slg* gene cluster of about 24 kb, including an open reading frame (ORF) predicted to encode the initiation enzyme of S-layer protein glycosylation. The corresponding gene, named *wsfP*, served as a first target for the gene knockout system developed in the course of the present study. This target was chosen because loss of function would be easily screenable, resulting in an S-layer glycosylation-deficient mutant. The gene knockout system constructed for insertional inactivation of the chromosomal *wsfP* gene of *P. alvei* CCM 2051^T is based on the commercially available bacterial mobile group II intron Ll.LtrB of *Lactococcus lactis*, in combination with further components available in our laboratory, including the broad-host-range S-layer gene promoter of *sgsE* from *G. stearothermophilus* NRS 2004/3a (22) and the *Geobacillus-Bacillus-Escherichia coli* shuttle vector pNW33N. Bacterial mobile group II introns are retroelements inserted into specific DNA target sites at high frequency by use of the retrohoming mechanism, by which the excised intron lariat RNA is inserted directly into a DNA target site and is then reverse transcribed by the associated intron-encoded enzyme protein (6, 8, 17). Since the DNA target site is recognized primarily by base pairing of intron RNA, which can be modified, and a few intron-encoded-enzyme-protein recognition positions, these introns can be inserted efficiently into any specific DNA target (9, 15, 35, 40).

The aim of this study is the development of a genetic tool for manipulation of S-layer protein glycosylation in *P. alvei* CCM 2051^T. For proof of concept, we specifically deal with (i) the construction of a broad-host-range gene knockout system based on the *L. lactis* Ll.LtrB intron; (ii) its modification for specific disruption of the *wsfP* gene on the *P. alvei* CCM 2051^T chromosome, encoding the putative initiation enzyme of S-layer glycan biosynthesis; and (iii) the reconstitution of enzyme activity by plasmid-based expression of *wsfP* and its predicted functional homologue *wsaP* from *G. stearothermophilus* NRS 2004/3a.

MATERIALS AND METHODS

Bacterial strains and growth conditions. *P. alvei* CCM 2051^T (Table 1) was obtained from the Czech Collection of Microorganisms (CCM; Brno, Czech Republic) and was grown at 37°C and 200 rpm in Luria-Bertani (LB) broth or on LB agar plates supplemented with 10 μg/ml chloramphenicol (Cm), when appropriate. *G. stearothermophilus* NRS 2004/3a (Table 1) was obtained from F. Hollaus (19) and grown on modified S-VIII medium at 55°C (29). *Escherichia coli* DH5α (Invitrogen, Lofer, Austria) was grown in LB broth at 37°C supplemented with 30 μg/ml Cm, when appropriate.

General methods. Genomic DNA of *G. stearothermophilus* NRS 2004/3a and of *P. alvei* CCM 2051^T was isolated by using a Genomic Tip 100 kit (Qiagen, Vienna, Austria) according to the manufacturer's instructions, except that for the latter organism, cells were broken by repeated freezing and thawing cycles (10 times), because of its resistance toward lysozyme (27). Restriction enzymes, calf intestinal alkaline phosphatase, and T4 DNA ligase were purchased from Invitrogen. A MinElute gel extraction kit (Qiagen) was used to purify DNA fragments from agarose gels, and a MinElute reaction cleanup kit (Qiagen) was used to purify digested oligonucleotides and plasmids. Plasmid DNA from transformed cells was isolated with a QIAprep Spin Miniprep kit (Qiagen). Agarose gel electrophoresis was performed as described elsewhere (26). PCR (My Cycler; Bio-Rad, Hercules, CA) was performed using an Expand Long Range dNTPack (Roche, Vienna, Austria). PCR conditions were optimized for each primer pair (Table 2), and amplification products were purified using a MinElute PCR purification kit (Qiagen). Transformation of *E. coli* DH5α was done according to the manufacturer's protocol (Invitrogen). Transformants were screened by in situ PCR using RedTaq ReadyMix PCR mix (Sigma-Aldrich, Vienna, Austria); recombinant clones were analyzed by restriction mapping and confirmed by sequencing (Agowa, Berlin, Germany). Sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) was carried out according to a standard protocol (16), using a Protean II electrophoresis apparatus (Bio-Rad). Protein bands were visualized with Coomassie brilliant blue G250 staining reagent. Periodic acid-Schiff (PAS) staining for carbohydrates was performed according to the method of Hart and coworkers (10).

TABLE 2. Bacterial strains and plasmids used in this study

Strain or plasmid	Genotype and/or relevant characteristic(s)	Source
Strains		
P. alvei CCM 2051^T	Wild-type isolate; Km^r	CCM
P. alvei CCM 2051 *wsfP*::Ll.LtrB	*P. alvei* CCM 2051^T carrying a targetron insertion at the *wsfP* locus; Km^r	This study
Escherichia coli DH5α	F$^-$ ϕ80d*lacZ*ΔM15 Δ(*lacZYA-argF*)*U169 deoR recA1 endA1 hsdR17* (r_K^- m_K^-) *phoA supE44 thi-1 gyrA96 relA1* λ$^-$	Invitrogen
Plasmids		
pNW33N	*Geobacillus-Bacillus-E. coli* shuttle vector; Cm^r	*Bacillus* Genetic Stock Center
pEXALV	pNW33N carrying the *sgsE* S-layer gene promoter of *G. stearothermophilus* NRS 2004/3a; Cm^r	This study
pNW33NΔ*Hin*dIII	pNW33N without its unique HindIII restriction site; Cm^r	This study
pJIR750ai	*Clostridium perfringens-E. coli* shuttle vector carrying alpha-toxin gene (*plc*) targetron; Cm^r	Sigma-Aldrich
pJIR750ai_P(SgsE)	pJIR750ai carrying the *sgsE* S-layer gene promoter of *G. stearothermophilus* NRS 2004/3a; Cm^r	This study
pTT_*plc*	pNW33N without HindIII carrying the *sgsE* S-layer gene promoter of *G. stearothermophilus* NRS 2004/3a in front of the *plc* intron cassette of pJIR750ai; Cm^r	This study
pTT_*wsfP*555	pTT_*plc* targeted for insertion between positions 555 and 556 from the initial ATG codon of *wsfP*; Cm^r	This study
pTT_*wsfP*654	pTT_*plc* targeted for insertion between positions 654 and 655 from the initial ATG codon of *wsfP*; Cm^r	This study
pTT_*wsfP*1176	pTT_*plc* targeted for insertion between positions 1176 and 1177 from the initial ATG codon of *wsfP*; Cm^r	This study
pEXALV_*wsfP*	pEXALV carrying the *wsfP* gene of *P. alvei* CCM 2051^T; Cm^r	This study
pEXALV_*wsaP*	pEXALV carrying the *wsaP* gene of *G. stearothermophilus* NRS 2004/3a; Cm^r	This study

Transformation of *P. alvei* CCM 2051^T. Transformation of *P. alvei* CCM 2051^T followed the protocol of Turgeon et al. (36), with some modifications. Briefly, the organism was grown to an optical density at 600 nm (OD_{600}) of 0.2 to 0.3. Subsequently, the culture was washed five times with ice-cold electroporation buffer (250 mM sucrose-1 mM HEPES-1 mM $MgCl_2$-10% glycerol, pH 7.0), resuspended in 1/500 of a culture volume, and stored in 50-μl aliquots at −70°C. In transformation studies, plasmid DNA originating from the *E. coli-G. stearothermophilus-Bacillus subtilis* shuttle vector pNW33N (*Bacillus* Genetic Stock Center, Columbus, OH) (Table 1) was used, and electroporation was done at a capacitance of 25 μF, using a Gene Pulser II apparatus connected to a pulse controller (Bio-Rad). For optimization of electroporation conditions, voltage was varied between 5.0 and 20 kV/cm, and resistance was set to 100 Ω, 200 Ω, and 400 Ω. Five hundred nanograms of plasmid DNA was added to an aliquot of electrocompetent cells, and the mixture was transferred into a prechilled 1-mm electroporation cuvette (Bio-Rad). Immediately after application of the pulse, the cell suspension was diluted with 4 ml of prewarmed casein-peptone soymeal-peptone broth (Sigma-Aldrich), containing 250 mM sucrose, 5 mM $MgCl_2$, and 5 mM $MgSO_4$, and incubated for 2 h at 37°C, allowing expression of the antibiotic resistance marker. Finally, cells were spread on LB agar supplemented with Cm and incubated overnight at 37°C.

Construction of an expression vector for *P. alvei* CCM 2051^T. A ~400-bp DNA fragment containing the *sgsE* surface layer gene promoter of *G. stearothermophilus* NRS 2004/3a (22) was amplified from *G. stearothermophilus* NRS 2004/3a genomic DNA with primers P(SgsE)_*Hin*dIII_for and P(SgsE)_*Sph*I_rev, digested with HindIII and SphI, and ligated into HindIII/SphI-linearized and dephosphorylated pNW33N plasmid. The resulting plasmid was named pEXALV.

Construction of a *wsfP* gene knockout mutant. A schematic diagram for construction of a shuttle plasmid containing the *wsfP* targetron used to construct the *P. alvei* CCM 2051^T *wsfP* mutant is given in Fig. 1. Plasmids are listed in Table 2.

(i) Deletion of a HindIII restriction site from pNW33N. Plasmid pNW33N was digested with HindIII, and the 5′ overhangs were filled in to form blunt ends by a large (Klenow) fragment of DNA polymerase I. The modified plasmid was self-ligated and transformed into *E. coli* DH5α, and the loss of the unique HindIII restriction site was verified. The resulting plasmid was named pNW33NΔ*Hin*dIII (Fig. 1A).

(ii) Insertion of P(*sgsE*) in front of the intron cassette of pJIR750i. The *sgsE* promoter was amplified from genomic DNA of *G. stearothermophilus* NRS 2004/3a by PCR using primers P(SgsE)_*Sph*I_for and P(SgsE)_*Hin*dIII_rev. The resulting fragment was digested with SphI and HindIII, cloned into SphI/HindIII-linearized and dephosphorylated pJIR750ai plasmid, and transformed into *E. coli* DH5α. Thereby, the promoter region P(*cpb2*) of the β-2 toxin gene (*cpb2*) from *Clostridium perfringens* in front of the α-toxin gene (*plc*) targetron was replaced by the *sgsE* surface layer gene promoter P(SgsE) of *G. stearothermophilus* NRS 2004/3a. The resulting plasmid was named pJIR750ai_P(SgsE) (Fig. 1B).

(iii) Transfer of the promoter-intron cassette construct into pNW33NΔ*Hin*dIII. Purified plasmid DNA of pJIR750ai_P(SgsE) was used as a template for PCR with primers Targe_*Sph*I_for and Targe_*Sph*I_rev. The resulting ~3,900-bp fragment containing P(SgsE), the Ll.LtrA ORF, and the *plc* targetron was digested with SphI, cloned into SphI-linearized and dephosphorylated pNW33NΔ*Hin*dIII plasmid, and transformed into *E. coli* DH5α. The plasmid was named pTT_*plc* (Fig. 1C).

(iv) Modification of the intron cassette for targeting to the putative *wsfP* gene of *P. alvei* CCM 2051^T. The Ll.LtrB targetron was retargeted to be inserted into the putative *wsfP* gene of *P. alvei* CCM 2051^T by using a computer algorithm that identifies potential insertion sites and directly designs PCR primers for modifying the intron RNA to base pair with these sites (TargeTron; Sigma-Aldrich). For gene interruption and stable insertion, the insertion sites with the lowest E-values and, for this reason, with high intron insertion efficiency were used. There are three short sequence elements involved in the base pairing interaction between the DNA target site (IBS1, IBS2, and δ′) and intron RNA (EBS1, EBS2, and δ). Modifications of intron RNA sequences (EBS1, EBS2, and δ) to base pair with the *wsfP* target site sequences were introduced via PCR by primer-mediated mutation with the primer sets comprising P_555|556s-IBS, P_555|556s-EBS1d, P_555|556s-EBS2, and P_654|655s-IBS; P_654|655s-EBS1d, P_654|655s-EBS2, and P_1176|1177s-IBS; and P_1176|1177s-EBS1d and P_1176|1177s-EBS2. The amplified 353-bp fragment was subsequently digested with HindIII and BsrGI and ligated into pTT_*plc* vector digested with the same restriction enzymes (Fig. 1D). The three resulting vectors were named pTT_*wsfP*555, pTT_*wsfP*654, and pTT_*wsfP*1176.

(v) Creation of a *wsfP* gene knockout with the *wsfP* targetron. pTT_*wsfP*555, pTT_*wsfP*654, and pTT_*wsfP*1176 were electroporated into *P. alvei* CCM 2051^T, and the cell suspension was plated on LB supplemented with Cm. Integration of the intron was assayed by colony PCR, using primers KO_*wsfP*_control_for_1 and KO_*wsfP*_control_rev_1, which hybridize to flanking sequences of the insertion sites.

(vi) Confirmation of *wsfP* gene insertion. For proof of insertion of the intron at the correct position, the PCR product obtained from genomic DNA of *P. alvei* CCM 2051^T *wsfP*::Ll.LtrB upon use of the primer pair comprising KO_*wsfP*_control_for_1 and KO_*wsfP*_control_rev_1 was sequenced.

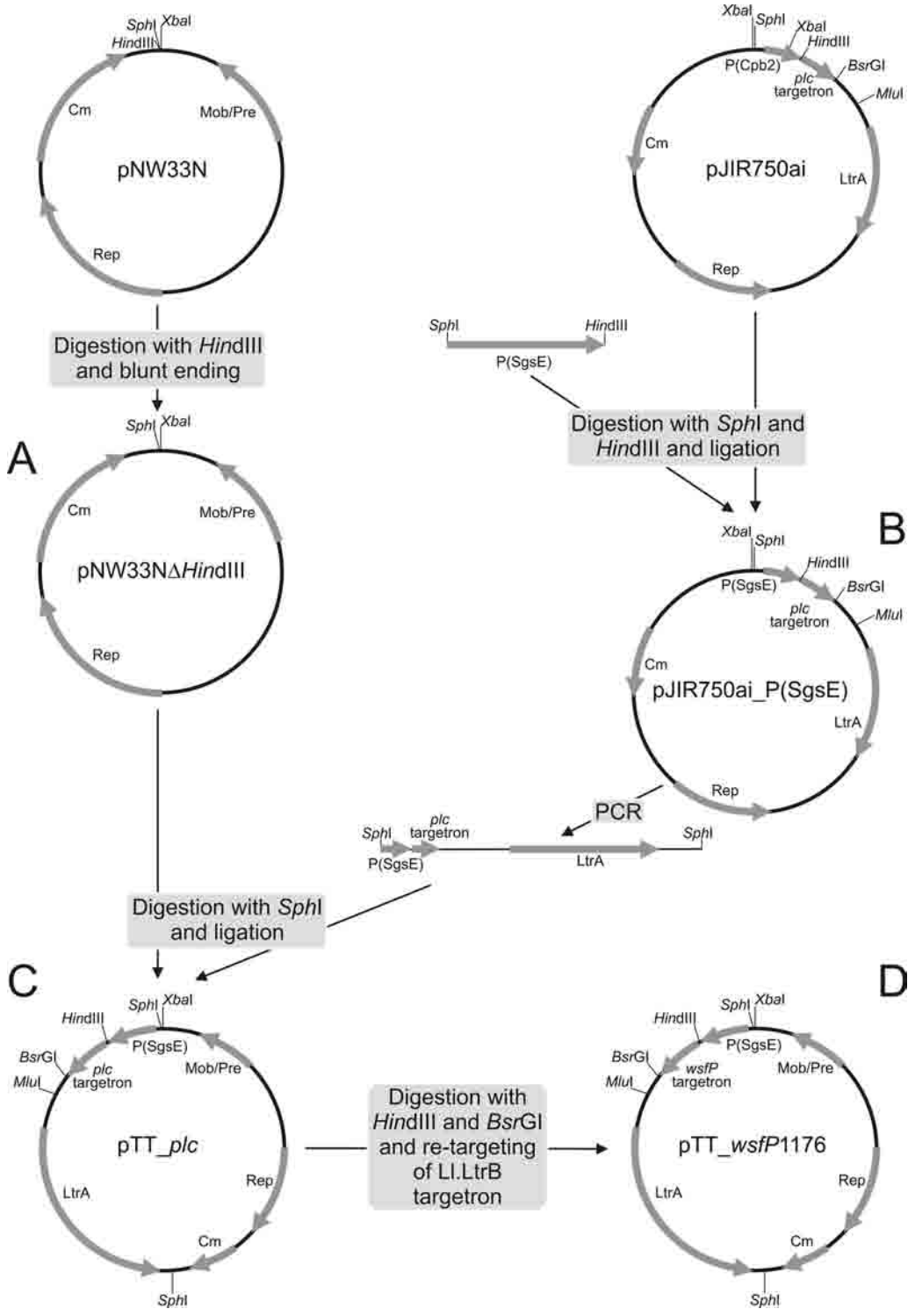


FIG. 1. Schematic drawing of the construction of the shuttle plasmid pTT_*wsfP*1176, containing the *wsfP* targetron.

Analysis of S-layer glycosylation in *P. alvei* CCM 2051^T wild-type cells and in *P. alvei* CCM 2051^T *wsfP*::LI.LtrB. The presence or absence of S-layer protein glycosylation on intact bacterial cells was monitored by SDS-PAGE followed by PAS staining (3) and by high-performance anion-exchange chromatography–pulsed electrochemical detection with a CarboPAc PA-1 column (Dionex, Sunnyvale, CA) after hydrolysis of crude S-layer preparations with trifluoroacetic acid (2, 30).

Reconstitution of enzyme activity in *P. alvei* CCM 2051^T *wsfP*::LI.LtrB by plasmid-based enzyme expression. The coding sequence of *wsfP* was amplified from genomic DNA of *P. alvei* CCM 2051^T by using primers *wsfP*_for_*Sph*I and *wsfP*_rev_*Kpn*I. The ~1,400-bp PCR product was digested with SphI and KpnI and ligated into SphI/KpnI-linearized and dephosphorylated pEXALV plasmid. This construct was named pEXALV_*wsfP*. Similarly, the coding sequence of *wsaP* from *G. stearothermophilus* NRS 2004/3a was cloned into

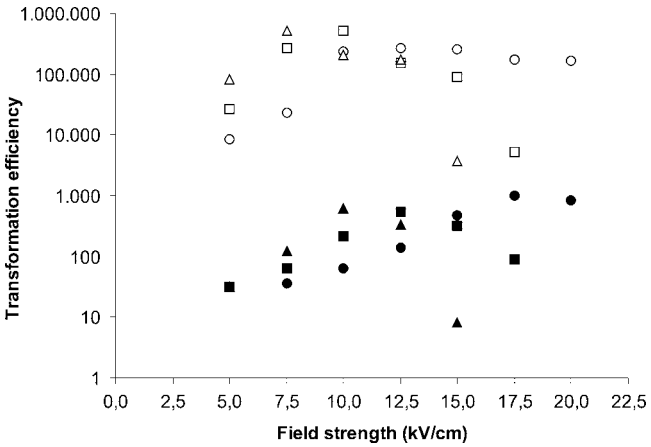


FIG. 2. Determination of optimal electroporation parameters for wild-type cells (●, ■, and ▲) and *wsfP* mutant cells (○, □, and △) of *P. alvei* CCM2051^T. The relationship between the numbers of transformants obtained per μg of DNA (pNW33N) and per 10^6 competent cells and the applied voltage is shown. Electroporation experiments were performed with cultures from the early growth phase (OD_{600}, ~0.2 to 0.3) at voltages ranging from 5 to 20 kV/cm and at resistance levels of 100 Ω (●/○), 200 Ω (■/□), or 400 Ω (▲/△).

pEXALV, using the primer pair comprising *wsaP*_for_*Sph*I and *wsaP*_rev_*Kpn*I and genomic DNA of *G. stearothermophilus* NRS 2004/3a as a template. The resulting construct was named pEXALV_*wsaP*. Each construct was transformed into *P. alvei* CCM 2051^T *wsfP*::Ll.LtrB, and reconstitution of UDP-Gal:phosphoryl-polyprenol Gal-1-phosphate transferase activity was analyzed. As a negative control, *P. alvei* CCM 2051^T transformants harboring pEXALV without *wsfP* were used.

RESULTS

Determination of optimal electroporation conditions for wild-type *P. alvei* CCM 2051^T cells. For transformation studies, *P. alvei* CCM 2051^T cells from the early logarithmic growth phase (OD_{600} of ~0.2 to 0.3) were used. From the different electroporation settings applied, an electric field at 100 Ω/25 μF/17.5 kV · cm^{-1} gave the best result; a transformation efficiency of 1×10^3 transformants per μg of plasmid DNA (pNW33N) and per 10^6 competent cells was obtained (Fig. 2).

Description of the putative initiation enzyme WsfP of S-layer glycan biosynthesis in *P. alvei* CCM 2051^T. On the basis of the structural identity of adaptor saccharide backbones in the S-layer glycans of *P. alvei* CCM 2051^T (18) and *G. stearothermophilus* NRS 2004/3a (29), where its biosynthesis is initiated by the UDP-Gal:phosphoryl-polyprenol Gal-1-phosphate transferase WsaP (33), it was conceivable that a homologous enzyme would initiate S-layer protein glycosylation in *P. alvei* CCM 2051^T. The putative initiation enzyme of SpaA glycosylation was chosen as a first target for the gene knockout system to be developed in the course of the present study, because the glycosylation-deficient phenotype resulting from its disruption would be easily screenable by SDS-PAGE and PAS staining. Disruption of the *wsfP* gene should result in the prevention of the initiation reaction and, thus, in the complete loss of S-layer glycans. For this purpose, the bacterial chromosome of *P. alvei* CCM 2051^T was searched for a putative S-layer glycosylation (*slg*) gene cluster as present in all other S-layer glycoprotein-carrying bacteria investigated so far (21). The chromosome walking strategy leading to the identification of the *slg* gene cluster of *P. alvei* CCM 2051^T will be published elsewhere (K. Zarschler, B. Janesch, P. Messner, and C. Schäffer, unpublished data). Specifically, an ORF of 1,407 bp encoding a putative UDP-Gal:phosphoryl-polyprenol Gal-1-phosphate transferase, named *wsfP* (for nomenclature, see reference 21), was identified. The ORF shows high similarity to WsaP (identity = 60%, similarity = 75%; GenBank accession number FJ751776). Typical of a member of the polyisoprenylphosphate hexose-1-phosphate transferase family, whose representatives transfer hexose-1-P residues from UDP-hexoses to a lipid carrier (33), the topological model of WsfP shows five transmembrane helices, a central loop facing the periplasmic space, and a highly conserved carboxy-terminal cytosolic tail containing the catalytic domain (25) (Fig. 3).

Implementation of the bacterial mobile group II intron Ll.LtrB for *wsfP* gene disruption in *P. alvei* CCM 2051^T. For *wsfP* gene disruption in *P. alvei* CCM 2051^T, a broad-host-range gene knockout system based on the *L. lactis* Ll.LtrB intron was constructed by following the strategy of Chen and coworkers (5). The *sgsE* S-layer gene promoter of *G. stearothermophilus* NRS 2004/3a, known to work also in *Bacillus subtilis* (22), was placed in front of the intron cassette of pJIR750ai, composed of the intron RNA and the ORF coding for the LtrA protein. This promoter-intron cassette construct was finally transferred into the *Geobacillus-Bacillus-E. coli* shuttle vector pNW33N to create a plasmid-borne Ll.LtrB mobile group II intron for gene disruption in *P. alvei* CCM 2051^T. Prior to the retargeting of the Ll.LtrB mobile group II intron for insertion into the putative *wsfP* gene on the *P. alvei* CCM2051^T chromosome, the gene was analyzed by a computer algorithm for identification of potential insertion sites. The algorithm predicted 11 intron insertion sites across the 1,407-bp *wsfP* gene. For gene interruption, the insertion sites with high intron insertion efficiency between positions 555 and 556 (E-value = 0.094), 654 and 655 (E-value = 0.010), and 1176 and 1177 (E-value = 0.202) were selected for intron modification (positions are given relative to the initial ATG codon; lower E-values correspond to higher predicted intron insertion efficiencies; target sites with E-values of <0.5 are predicted to be efficient introns). PCRs using primers designed by the algorithm for retargeting the intron by primer-mediated mutation were performed, and donor plasmids containing the *wsfP* targetrons were constructed. The plasmids containing the targetrons P555, P654, and P1176, named pTT_*wsfP*555, pTT_*wsfP*654, and pTT_*wsfP*1176, respectively, were transformed into *P. alvei* CCM 2051^T by electroporation. Analysis of 28 Cm-resistant *P. alvei* CCM 2051^T colonies for *wsfP* disruption showed that one colony transformed with pTT_*wsfP*1176 contained both wild-type (0.78-kb PCR product) and intron-inserted (1.68-kb PCR product) *wsfP* (Fig. 4). The rest of the colonies also contained the vector, which is the criterion for selection of colonies, but for unknown reasons, no intron insertion has occurred in the *wsfP* gene. The observation of getting both a wild-type gene and an intron-inserted gene by PCR screening of bacterial colonies was also described by Chen et al. (5) when screening for an α-toxin gene (*plc*) knockout in *Clostridium perfringens* ATCC 3624 by using a plasmid-borne Ll.LtrB mobile group II intron. Since intron RNA insertion occurs in some but not all of the progeny cells of a

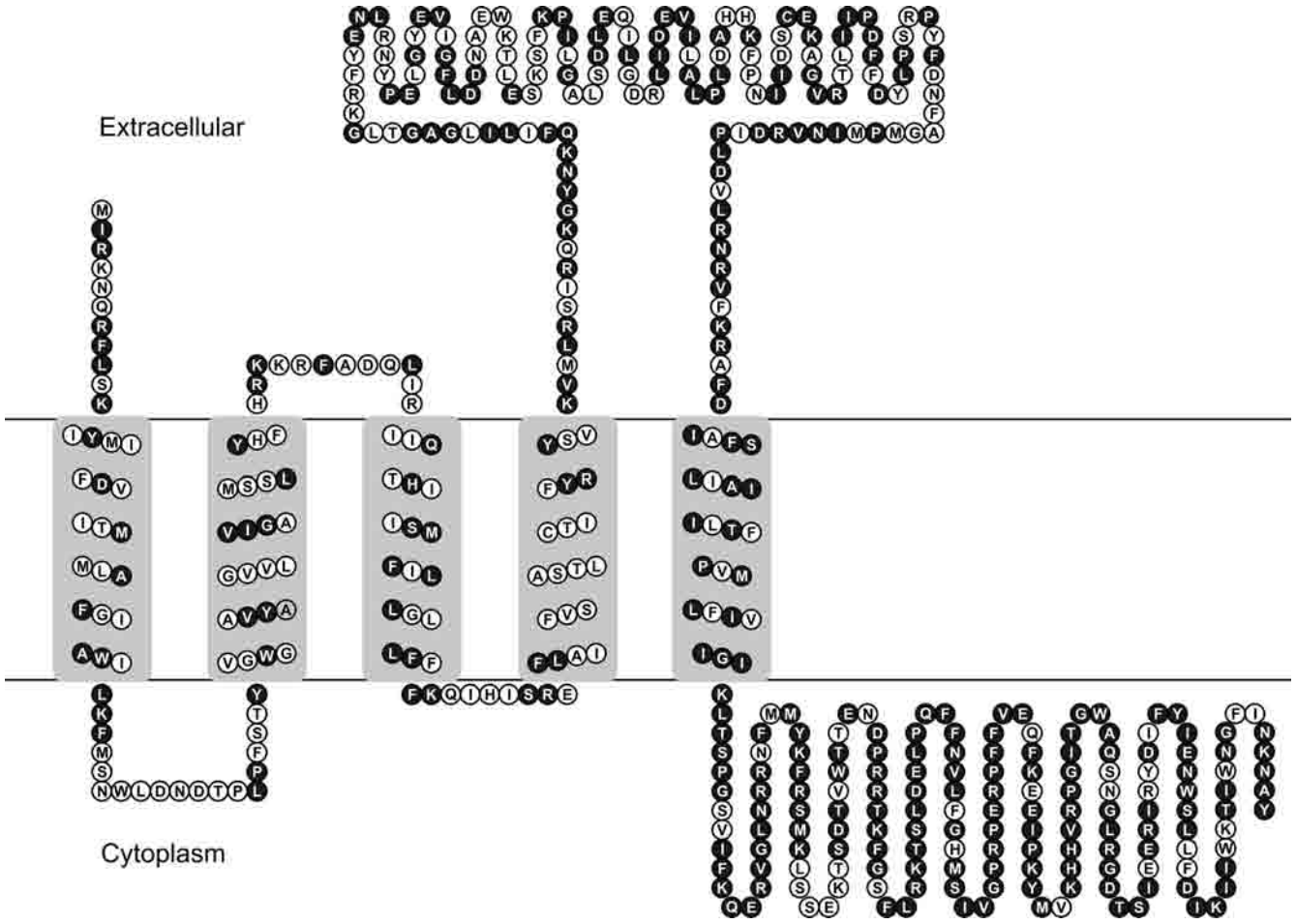


FIG. 3. Predicted topology of the WsfP protein of *P. alvei* CCM 2051^T. Shown are the five transmembrane helices (boxed), the central extracellular loop, and the carboxy-terminal cytosolic tail. Black amino acid residues are identical to corresponding amino acids in the functional WsaP homologue of *G. stearothermophilus* NRS 2004/3a.

single transformed bacterium, an isolated colony contains some cells with the intron-inserted gene and some cells with the wild-type gene. Therefore, bacteria from a colony containing both the wild-type gene and the intron-inserted gene were singularized on LB supplemented with Cm by streaking, and 74 colonies were screened again, using the same primer pair. This time, 36 colonies showed only the intron-inserted *wsfP* gene, 28 colonies only the wild-type *wsfP* gene, and 8 colonies both. A colony possessing only intron-inserted *wsfP*, named *P. alvei* CCM 2051^T *wsfP*::Ll.LtrB, was selected for further analysis.

Proof of functionality of the developed gene knockout system. To confirm the absence of the functional WsfP enzyme and, thus, the loss of S-layer glycoprotein glycan biosynthesis in *P. alvei* CCM 2051^T *wsfP*::Ll.LtrB, SDS-PAGE of intact cells accompanied by PAS staining and carbohydrate analysis of S-layer extract was performed. A clone showing exclusively the intron-inserted *wsfP* gene was cultivated in 5 ml LB medium supplemented with Cm, and an aliquot of biomass was loaded on an SDS-PAGE gel. As shown in Fig. 5, lanes 2 and 6, the *wsfP* mutant shows only a single S-layer protein band at ~105 kDa, representing nonglycosylated SpaA S-layer protein (the molecular mass estimated from the gel is in accordance with the molecular mass of 105.9 kDa, as calculated from the amino acid sequence), while the S-layer protein of *P. alvei* CCM 2051^T wild-type cells possessing an intact *wsfP* gene migrates in three distinct bands, with apparent molecular masses of ~240, ~160, and ~105 kDa, with the upper two bands representing different glycoforms of SpaA (Fig. 5, lanes 1 and 5). This experiment clearly demonstrated that WsfP is the initiation enzyme of SpaA S-layer protein glycosylation.

This finding is further supported by the comparative sugar composition analysis of the crude S-layer fraction derived from *P. alvei* CCM 2051^T wild-type cells and from the *wsfP* mutant (Fig. 6). The results have to be interpreted in the light of a secondary cell wall polymer (SCWP), composed of [(Pyr4,6)-β-D-Man*p*NAc-(1→4)-β-D-Glc*p*NAc-(1→3)]$_{n\sim11}$-(Pyr4,6)-β-D-Man*p*NAc-(1→4)-α-D-Glc*p*NAc-(1→*O*)-P→] repeats, which mediates attachment of the S-layer to the bacterial cell wall, being also contained in the samples (28). Consequently, the wild-type S-layer sample that possesses an overall degree of glycosylation of ~2.5% contains ManNAc-GlcNAc at an approximate molar ratio of 1:1, in addition to the S-layer glycan components Gal-ManNAc-Glc-Rha at an approximate molar ratio of 7:7:7:1. Quantification of sugars indicates an S-layer

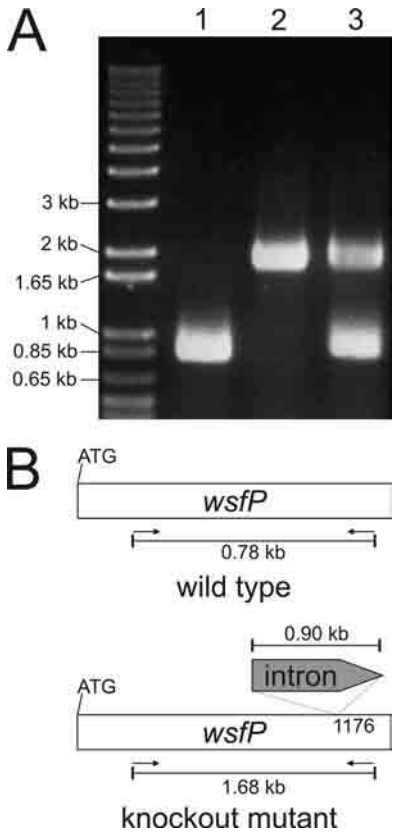


FIG. 4. Bacterial mobile group II intron-mediated gene disruption of *wsfP* in *P. alvei* CCM 2051^T. (A) Screening of Cm-resistant *P. alvei* CCM 2051^T colonies for intron insertion by in situ PCR using primers KO_*wsfP*_control_for_1 (→) and KO_*wsfP*_control_rev_1 (←). A PCR product obtained from a wild-type colony (lane 1), a PCR fragment obtained from a *wsfP* mutant (lane 2), and PCR products obtained from a colony containing both wild-type and intron-inserted *wsfP* (lane 3) are shown. (B) Schematic drawing of the *wsfP* gene with (bottom) and without (top) intron insertion, indicating the positions of primers KO_*wsfP*_control_for_1 (→) and KO_*wsfP*_control_rev_1 (←).

protein that carries, on average, two glycan chains of ~21 repeating units, with an SCWP of ~11 repeating being associated with the protein. The *wsfP* mutant, in contrast, is completely devoid of galactose and rhamnose, while the components of the SCWP can be clearly identified at the correct molar ratio. This comparative analysis clearly demonstrates that the developed gene knockout system is fully functional in abolishing S-layer protein SpaA glycosylation in *P. alvei* CCM 2051^{T}and serves as an additional proof of WsfP function. On the basis of the elucidated S-layer glycan structure (2, 18), it is conceivable that the nonstoichiometrically high glucose content detected in either analysis originates from an impurity present in the crude samples.

After subculturing of *P. alvei* CCM 2051^T *wsfP*::Ll.LtrB containing plasmid pTT_*wsfP*1176 without selective antibiotic for 10 days by replica plating, Cm-sensitive *wsfP* mutant clones lacking plasmid DNA but still showing *wsfP* gene disruption and the loss of S-layer glycans were isolated. The absence of the vector was confirmed by obtaining a negative PCR result using the primer pair comprising Cm_*Kpn*I_for and Cm_*Kpn*I_rev for amplifying the Cm resistance cassette (data not shown); *wsfP* disruption was verified by obtaining a PCR product of 1.68 kb, using the primer pair comprising KO_*wsfP*_control_for_1 and KO_*wsfP*_control_rev_1 (Fig. 4).

Reconstitution of S-layer glycan biosynthesis by plasmid-based expression of *wsfP* and *wsaP*. For the final proof of function of WsfP, reconstitution of S-layer glycosylation was analyzed. Transformation of pEXALV_*wsfP* into the Cm-sensitive *wsfP* mutant resulted in plasmid-based expression of the functional WsfP protein, as demonstrated by reconstitution of S-layer glycoprotein glycan biosynthesis (Fig. 5, lanes 3 and 7). Restoration of S-layer protein glycosylation was observed in *P. alvei* CCM2051^T *wsfP*::Ll.LtrB also after heterologous expression of WsaP from *G. stearothermophilus* NRS 2004/3a, expressed from pEXALV_*wsaP* (Fig. 5, lanes 4 and 8). However, in this experiment, glycosylation was obviously less efficient, with the nonglycosylated SpaA protein band appearing more intense and the glycoform band migrating at ~240 kDa appearing less intense on the gel than in the homologous expression approach. Nevertheless, these data confirm the initial assumption that WsfP and WsaP are functional homologues.

Electrocompetence of *P. alvei* CCM 2051^T *wsfP*::Ll.LtrB cells. Since there is speculation that glycosylation of surface proteins may affect the transformation efficiency of cells, *P. alvei* CCM 2051^T *wsfP*::Ll.LtrB cells were analyzed for their electrocompetence. Following the established procedure (see above), optimal electroporation conditions were determined for the *wsfP* mutant by using pNW33N plasmid DNA. A transformation efficiency up to 5×10^5 transformants per μg of

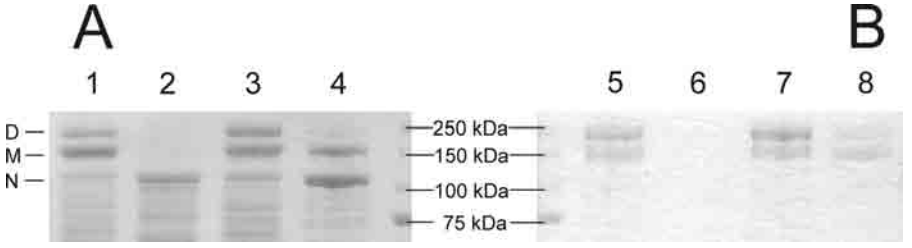


FIG. 5. SDS-PAGE gels showing the S-layer glycosylation profile of *P. alvei* CCM 2051^T wild-type cells (lanes 1 and 5), *wsfP* mutant cells (lanes 2 and 6), and *wsfP* mutant cells after reconstitution with WsfP (lanes 3 and 7) and WsaP (lanes 4 and 8) upon plasmid-based expression. Results are shown for Coomassie brilliant blue G250 staining (A) and PAS staining for carbohydrate (B). Nonglycosylated (N), monoglycosylated (M) and diglycosylated (D) S-layer SpaA proteins are indicated on the left. SDS-PAGE was performed using a 10% gel, and 10 μg and 20 μg of protein were loaded for Coomassie and PAS staining, respectively.

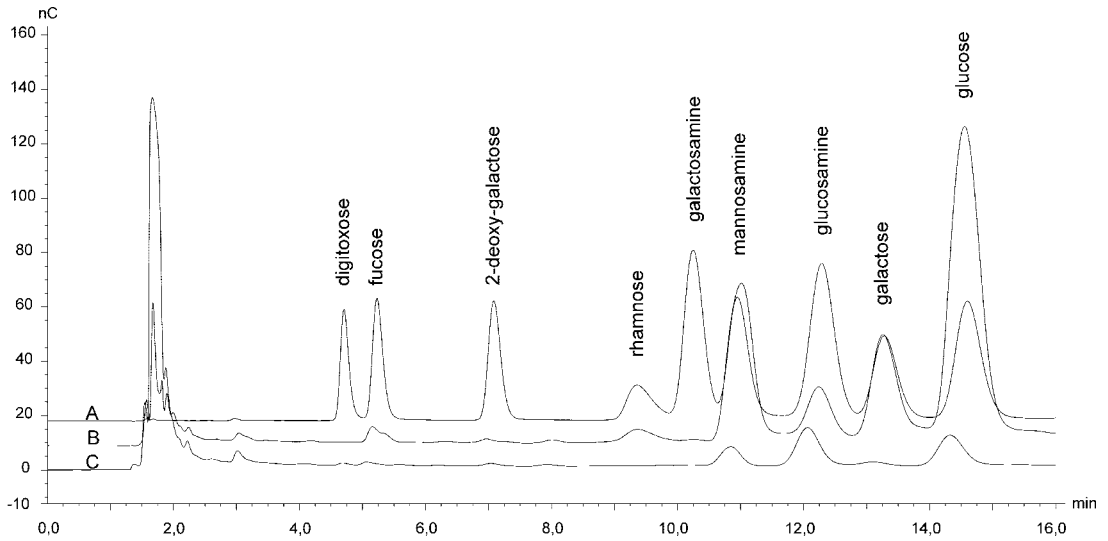


FIG. 6. Dionex carbohydrate analysis of S-layer extracts from *P. alvei* CCM 2051^T wild-type and *wsfP* mutant cells. (A) Standards (1 nmol each); (B) S-layer from wild-type cells (30 μg); (C) S-layer from *wsfP*::Ll.LtrB cells (175 μg).

plasmid DNA and per 10^6 competent cells was obtained by applying an electric field at 200 Ω/25 μF/10 kV · cm^{-1} or 400 Ω/25 μF/7.5 kV · cm^{-1} (Fig. 3). This corresponds to a factor of 500 for improvement of transformation efficiency for mutant cells versus wild-type cells of *P. alvei* CCM 2051^T.

DISCUSSION

Due to the lack of suitable tools for genetic manipulation of bacterial S-layer glycosylation pathways, progress in the elucidation of the glycan biosynthesis mechanism, which is a prerequisite for the desired production of functional S-layer *neo*-glycoproteins, was limited to in vitro testing of individual enzymes from these pathways (34) and to heterologous carbohydrate-engineering approaches in the past (32).

For the envisaged in vivo display of functional glycans via the S-layer anchor, in this work, a reliable and effective tool for the production of gene knockout mutants and for the expression of heterologous genes in the model organism *P. alvei* CCM 2051^T was developed. A targetron gene knockout system was constructed by cloning the Ll.LtrB group II intron, controlled by the *sgsE* surface layer gene promoter of *G. stearothermophilus* NRS 2004/3a, into the shuttle plasmid pNW33N; retargeting the intron; and producing insertional mutants after transformation of the plasmid into *P. alvei* CCM 2051^T. During the past 10 years, bacterial mobile group II introns have become a versatile instrument for site-specific chromosomal insertion in various prokaryotic species (5, 15, 38, 39). The requirements for adapting targetrons to specific needs are their sufficient expression via an inducible or constitutive promoter from a plasmid replicating in the host organism and the possibility of transferring this DNA into the desired host. From the adaptation of a targetron-based gene disruption system to *P. alvei* CCM 2051^T, an essential tool for elucidating molecular details about S-layer protein glycosylation has evolved. In this system, plasmid pNW33N and the *sgsE* S-layer gene promoter of *G. stearothermophilus* NRS 2004/3a are integral components. Since this plasmid replicates in thermophilic and mesophilic *Bacillaceae*, and the *sgsE* S-layer gene promoter drives gene expression in several thermophilic and mesophilic bacterial species (22), the system is likely to be applicable to different organisms within the radiation of *Bacillus* and related taxa.

For proof of functionality of targetron-mediated gene disruption in *P. alvei* CCM 2051^T, the *wsfP* gene was chosen. This gene codes for a putative UDP-Gal:phosphoryl-polyprenol Gal-1-phosphate transferase and shows high similarity to the gene coding for WsaP, the initiation enzyme of S-layer glycan biosynthesis in *G. stearothermophilus* NRS 2004/3a (33). Mutant cells of *P. alvei* CCM 2051^T carrying the Ll.ltrB intron in the chromosomal *wsfP* gene inserted between positions 1176 and 1177 from the initial ATG codon lost the ability to glycosylate their cognate S-layer protein SpaA. This effect was completely restored by the expression of plasmid-encoded WsfP. Heterologously expressed WsaP from *G. stearothermophilus* NRS 2004/3a also reconstituted the S-layer glycosylation process, albeit less efficiently, which might be due to the thermophilic origin of this initiation enzyme. By applying the constructed tool to the *wsfP* target, the first enzyme from the otherwise largely unknown S-layer glycan biosynthesis pathway of *P. alvei* CCM 2051^T (11) could be functionally characterized as initiating UDP-Gal:phosphoryl-polyprenol Gal-1-phosphate transferase.

In summary, in the course of the present study, an effective tool for gene disruption and heterologous gene expression in *P. alvei* CCM 2051^T was established, with *P. alvei* CCM 2051^T being the first gram-positive S-layer glycoprotein-carrying organism amenable to this kind of genetic engineering. The observation that *P. alvei* CCM 2051^T *wsfP*::Ll.LtrB cells show clearly improved transformation efficiency in comparison to wild-type cells, which may be due to spatial hindrance or charge repulsion effects between the DNA molecules and the S-layer glycan on the wild type, hampering the passage through the cell envelope, may have implications for the envisaged utilization of *P. alvei* CCM 2051^T as a means for surface display of functional, recombinant glycans. Thus, this work is opening

up new possibilities for the future design of functional glycans on S-layer proteins for in vivo and in vitro applications.

ACKNOWLEDGMENTS

We thank Andrea Scheberl for excellent technical assistance.

Financial support came from the Austrian Science Fund, project P20745-B11 (to P.M.) and projects P19047-B12 and P20605-12 (to C.S.), and the Hochschuljubiläumsstiftung der Stadt Wien, project H-02229-2007 (to K.Z.).

REFERENCES

1. **Abu-Qarn, M., J. Eichler, and N. Sharon.** 2008. Not just for Eukarya anymore: protein glycosylation in Bacteria and Archaea. Curr. Opin. Struct. Biol. **18:**544–550.
2. **Altman, E., J. R. Brisson, P. Messner, and U. B. Sleytr.** 1991. Structure of the glycan chain from the surface layer glycoprotein of *Bacillus alvei* CCM 2051. Biochem. Cell Biol. **69:**72–78.
3. **Altman, E., C. Schäffer, J.-R. Brisson, and P. Messner.** 1995. Characterization of the glycan structure of a major glycopeptide from the surface layer glycoprotein of *Clostridium thermosaccharolyticum* E207-71. Eur. J. Biochem. **229:**308–315.
4. **Ash, C., F. G. Priest, and M. D. Collins.** 1993. Molecular identification of rRNA group 3 bacilli (Ash, Farrow, Wallbanks and Collins) using a PCR probe test. Proposal for the creation of a new genus *Paenibacillus.* Antonie van Leeuwenhoek **64:**253–260.
5. **Chen, Y., B. A. McClane, D. J. Fisher, J. I. Rood, and P. Gupta.** 2005. Construction of an α toxin gene knockout mutant of *Clostridium perfringens* type A by use of a mobile group II intron. Appl. Environ. Microbiol. **71:** 7542–7547.
6. **Cousineau, B., D. Smith, S. Lawrence-Cavanagh, J. E. Mueller, J. Yang, D. Mills, D. Manias, G. Dunny, A. M. Lambowitz, and M. Belfort.** 1998. Retrohoming of a bacterial group II intron: mobility via complete reverse splicing, independent of homologous DNA recombination. Cell **94:**451–462.
7. **Desvaux, M., E. Dumas, I. Chafsey, and M. Hébraud.** 2006. Protein cell surface display in Gram-positive bacteria: from single protein to macromolecular protein structure. FEMS Microbiol. Lett. **256:**1–15.
8. **Eskes, R., J. Yang, A. M. Lambowitz, and P. S. Perlman.** 1997. Mobility of yeast mitochondrial group II introns: engineering a new site specificity and retrohoming via full reverse splicing. Cell **88:**865–874.
9. **Guo, H., M. Karberg, M. Long, J. P. Jones III, B. Sullenger, and A. M. Lambowitz.** 2000. Group II introns designed to insert into therapeutically relevant DNA target sites in human cells. Science **289:**452–457.
10. **Hart, C., B. Schulenberg, T. H. Steinberg, W. Y. Leung, and W. F. Patton.** 2003. Detection of glycoproteins in polyacrylamide gels and on electroblots using Pro-Q Emerald 488 dye, a fluorescent periodate Schiff-base stain. Electrophoresis **24:**588–598.
11. **Hartmann, E., P. Messner, G. Allmeier, and H. König.** 1993. Proposed pathway for biosynthesis of the S-layer glycoprotein of *Bacillus alvei*. J. Bacteriol. **175:**4515–4519.
12. **Honma, K., H. K. Kuramitsu, R. J. Genco, and A. Sharma.** 2001. Development of a gene inactivation system for *Bacteroides forsythus*: construction and characterization of a BspA mutant. Infect. Immun. **69:**4686–4690.
13. **Imanaka, T., M. Fujii, I. Aramori, and S. Aiba.** 1982. Transformation of *Bacillus stearothermophilus* with plasmid DNA and characterization of shuttle vector plasmids between *Bacillus stearothermophilus* and *Bacillus subtilis*. J. Bacteriol. **149:**824–830.
14. **Isono, K.** 1970. Transformation of amylase producing ability in *Bacillus stearothermophilus*. Jpn. J. Genet. **45:**285–291.
15. **Karberg, M., H. Guo, J. Zhong, R. Coon, J. Perutka, and A. M. Lambowitz.** 2001. Group II introns as controllable gene targeting vectors for genetic manipulation of bacteria. Nat. Biotechnol. **19:**1162–1167.
16. **Laemmli, U. K.** 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature **227:**680–685.
17. **Lambowitz, A. M., and S. Zimmerly.** 2004. Mobile group II introns. Annu. Rev. Genet. **38:**1–35.
18. **Messner, P., R. Christian, C. Neuninger, and G. Schulz.** 1995. Similarity of "core" structures in two different glycans of tyrosine-linked eubacterial S-layer glycoproteins. J. Bacteriol. **177:**2188–2193.
19. **Messner, P., F. Hollaus, and U. B. Sleytr.** 1984. Paracrystalline cell wall surface layers of different *Bacillus stearothermophilus* strains. Int. J. Syst. Bacteriol. **34:**202–210.
20. **Messner, P., D. Pum, and U. B. Sleytr.** 1986. Characterization of the ultrastructure and the self-assembly of the surface layer of *Bacillus stearothermophilus* strain NRS 2004/3a. J. Ultrastruct. Mol. Struct. Res. **97:**73–88.
21. **Messner, P., K. Steiner, K. Zarschler, and C. Schäffer.** 2008. S-layer nanoglycobiology of bacteria. Carbohydr. Res. **343:**1934–1951.
22. **Novotny, R., H. Berger, T. Schinko, P. Messner, C. Schäffer, and J. Strauss.** 2008. A temperature-sensitive expression system based on the *Geobacillus stearothermophilus* NRS 2004/3a *sgsE* surface-layer gene promoter. Biotechnol. Appl. Biochem. **49:**35–40.
23. **Novotny, R., C. Schäffer, J. Strauss, and P. Messner.** 2004. S-layer glycan-specific loci on the chromosome of *Geobacillus stearothermophilus* NRS 2004/3a and dTDP-L-rhamnose biosynthesis potential of *G. stearothermophilus* strains. Microbiology **150:**953–965.
24. **Sakakibara, J., K. Nagano, Y. Murakami, N. Higuchi, H. Nakamura, K. Shimozato, and F. Yoshimura.** 2007. Loss of adherence ability to human gingival epithelial cells in S-layer protein-deficient mutants of *Tannerella forsythensis*. Microbiology **153:**866–876.
25. **Saldias, M. S., K. Patel, C. L. Marolda, M. Bittner, I. Contreras, and M. A. Valvano.** 2008. Distinct functional domains of the *Salmonella enterica* WbaP transferase that is involved in the initiation reaction for synthesis of the O antigen subunit. Microbiology **154:**440–453.
26. **Sambrook, J., E. F. Fritsch, and T. Maniatis.** 1989. Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
27. **Sára, M., K. Moser-Thier, U. Kainz, and U. B. Sleytr.** 1990. Characterization of S-layers from mesophilic *Bacillaceae* and studies on their protective role towards muramidases. Arch. Microbiol. **153:**209–214.
28. **Schäffer, C., N. Müller, P. K. Mandal, R. Christian, S. Zayni, and P. Messner.** 2000. A pyrophosphate bridge links the pyruvate-containing secondary cell wall polymer of *Paenibacillus alvei* CCM 2051 to muramic acid. Glycoconj. J. **17:**681–690.
29. **Schäffer, C., T. Wugeditsch, H. Kählig, A. Scheberl, S. Zayni, and P. Messner.** 2002. The surface layer (S-layer) glycoprotein of *Geobacillus stearothermophilus* NRS 2004/3a. Analysis of its glycosylation. J. Biol. Chem. **277:** 6230–6239.
30. **Schäffer, C., T. Wugeditsch, C. Neuninger, and P. Messner.** 1996. Are S-layer glycoproteins and lipopolysaccharides related? Microb. Drug Resist. **2:**17–23.
31. **Sleytr, U. B., C. Huber, N. Ilk, D. Pum, B. Schuster, and E. M. Egelseer.** 2007. S-layers as a tool kit for nanobiotechnological applications. FEMS Microbiol. Lett. **267:**131–144.
32. **Steiner, K., A. Hanreich, B. Kainz, P. G. Hitchen, A. Dell, P. Messner, and C. Schäffer.** 2008. Recombinant glycans on an S-layer self-assembly protein: a new dimension for nanopatterned biomaterials. Small **4:**1728–1740.
33. **Steiner, K., R. Novotny, K. Patel, E. Vinogradov, C. Whitfield, M. A. Valvano, P. Messner, and C. Schäffer.** 2007. Functional characterization of the initiation enzyme of S-layer glycoprotein glycan biosynthesis in *Geobacillus stearothermophilus* NRS 2004/3a. J. Bacteriol. **189:**2590–2598.
34. **Steiner, K., R. Novotny, D. B. Werz, K. Zarschler, P. H. Seeberger, A. Hofinger, P. Kosma, C. Schäffer, and P. Messner.** 2008. Molecular basis of S-layer glycoprotein glycan biosynthesis in *Geobacillus stearothermophilus*. J. Biol. Chem. **283:**21120–21133.
35. **Toro, N., J. I. Jimenez-Zurdo, and F. M. Garcia-Rodriguez.** 2007. Bacterial group II introns: not just splicing. FEMS Microbiol. Rev. **31:**342–358.
36. **Turgeon, N., C. Laflamme, J. Ho, and C. Duchaine.** 2006. Elaboration of an electroporation protocol for *Bacillus cereus* ATCC 14579. J. Microbiol. Methods **67:**543–548.
37. **Welker, N. E.** 1993. The genetic map of *Bacillus stearothermophilus* NUB36, p. 483–487. *In* A. L. Sonenshein, J. A. Hoch, and R. Losick (ed.), *Bacillus subtilis* and other gram-positive bacteria: biochemistry, physiology, and molecular genetics. American Society for Microbiology, Washington, DC.
38. **Yao, J., and A. M. Lambowitz.** 2007. Gene targeting in gram-negative bacteria by use of a mobile group II intron ("targetron") expressed from a broad-host-range vector. Appl. Environ. Microbiol. **73:**2735–2743.
39. **Yao, J., J. Zhong, Y. Fang, E. Geisinger, R. P. Novick, and A. M. Lambowitz.** 2006. Use of targetrons to disrupt essential and nonessential genes in *Staphylococcus aureus* reveals temperature sensitivity of Ll.LtrB group II intron splicing. RNA **12:**1271–1281.
40. **Zhong, J., M. Karberg, and A. M. Lambowitz.** 2003. Targeted and random bacterial gene disruption using a group II intron (targetron) vector containing a retrotransposition-activated selectable marker. Nucleic Acids Res. **31:** 1656–1664.

Available online at www.sciencedirect.com
ScienceDirect
Carbohydrate Research 343 (2008) 1934–1951
Carbohydrate RESEARCH

Review

S-layer nanoglycobiology of bacteria

Paul Messner,* Kerstin Steiner, Kristof Zarschler and Christina Schäffer*

Universität für Bodenkultur Wien, Zentrum für NanoBiotechnologie A-1180 Wien, Gregor-Mendel-Strasse 33, Austria

Received 2 October 2007; received in revised form 5 December 2007; accepted 20 December 2007
Available online 16 January 2008

Presented at Eurocarb 14th Lübeck, Germany, September 2007

Abstract—Cell surface layers (S-layers) are common structures of the bacterial cell envelope with a lattice-like appearance that are formed by a self-assembly process. Frequently, the constituting S-layer proteins are modified with covalently linked glycan chains facing the extracellular environment. S-layer glycoproteins from organisms of the *Bacillaceae* family possess long, O-glycosidically linked glycans that are composed of a great variety of sugar constituents. The observed variations already exceed the display found in eukaryotic glycoproteins. Recent investigations of the S-layer protein glycosylation process at the molecular level, which has lagged behind the structural studies due to the lack of suitable molecular tools, indicated that the S-layer glycoprotein glycan biosynthesis pathway utilizes different modules of the well-known biosynthesis routes of lipopolysaccharide O-antigens. The genetic information for S-layer glycan biosynthesis is usually present in S-layer glycosylation (*slg*) gene clusters acting in concert with housekeeping genes. To account for the nanometer-scale cell surface display feature of bacterial S-layer glycosylation, we have coined the neologism 'nanoglycobiology'. It includes structural and biochemical aspects of S-layer glycans as well as molecular data on the machinery underlying the glycosylation event. A key aspect for the full potency of S-layer nanoglycobiology is the unique self-assembly feature of the S-layer protein matrix. Being aware that in many cases the glycan structures associated with a protein are the key to protein function, S-layer protein glycosylation will add a new and valuable component to an 'S-layer based molecular construction kit'. In our long-term research strategy, S-layer nanoglycobiology shall converge with other functional glycosylation systems to produce 'functional' S-layer *neo*glycoproteins for diverse applications in the fields of nanobiotechnology and vaccine technology. Recent advances in the field of S-layer nanoglycobiology have made our overall strategy a tangible aim of the near future.

Keywords: S-layer glycoprotein; Glycosylation enzymes; Glycosylation gene cluster; Nanobiotechnology; Carbohydrate engineering; Self-assembly

Contents

* Corresponding authors. Tel.: +43 1 47654 2202; fax: +43 1 478 9112 (P.M.); e-mail addresses: paul.messner@boku.ac.at; christina.schaeffer@boku.ac.at

0008-6215/$ - see front matter © 2008 Elsevier Ltd. All rights reserved.
doi:10.1016/j.carres.2007.12.025

1. Introduction

1.1. The S-layer protein self-assembly system

Prokaryotic surface layers (S-layers) have been described for the first time time in 1952. We now approach the sixth decade of S-layer research.[1,2] S-layers, in general, are two-dimensional crystalline arrays formed of individual subunits by an entropy-driven self-assembly event, which completely cover bacterial cells.[3] Over time, we faced a considerable change in the focus of S-layer research, accounting for novel areas and demands opening up in conjunction with applied research. At the very beginning of S-layer research, identification and ultrastructural characterization of S-layers from prokaryotes of different habitats and sources were the main tasks (for reviews see Refs. 3–6), since the mid 1970s chemical characterization of archaeal as well as of bacterial S-layer proteins became the central research focus (for reviews, see Refs. 7–15). The mid-1970s were also the time when glycosylation of S-layers had been discovered.[16,17] Almost 30 years later, it is now established knowledge that glycosylation represents the major modification of S-layer proteins, both in archaea and in bacteria.[18] (Please note that the term archaebacteria was replaced by archaea and the term eubacteria was replaced by bacteria[19]).

In 1986, the first S-layer gene sequences encoding the two S-layer proteins of *Brevibacillus brevis* 47 have been published.[20] Now more than 50 sequences of S-layer genes are available in public data bases, with ten of them encoding S-layer proteins with confirmed glycosylation. These are the S-layer proteins of the archaea *Haloarcula japonica* TR-1 (GenBank D87290), *Halobacterium salinarum* R_1M_1 (GenBank J02767), *Haloferax volcanii* DS2 (GenBank M62816), *Methanothermus fervidus* DSM 2088 (GenBank X58297), *Methanothermus sociabilis* DSM 3496 (GenBank X58296), and of the bacteria *Geobacillus stearothermophilus* NRS 2004/3a (GenBank AF328862), *Aneurinibacillus thermoaerophilus* DSM 10155/G^+ (GenBank AY395579), *A. thermoaerophilus* L420-91^T (GenBank AY395578), *Geobacillus tepidamans* GS5-97^T(GenBank AY883421), and *Thermoanaerobacter kivui* DSM 2030 (GenBank M31069). With the demonstration of the feasibility of recombinant S-layer protein production in heterologous expression systems, new avenues for S-layer protein research have opened up, putting forward the use of S-layer protein self-assembly systems for a wide spectrum of applications (for reviews, see Refs. 21–24). Regardless of the source of the S-layer, either after isolation from the bacterial cell wall by treatment with chaotropic agents or after heterologous expression in a suitable host, S-layer subunits characteristically self-assemble into monomolecular, two-dimensional arrays with oblique (p1, p2), square (p4) or hexagonal (p3, p6) symmetry, which are identical with those observed on the respective native bacterial cell.[25] These so-called self-assembly structures can have flat, cylindrical or even vesicular appearance, be either mono- or multilayered, and have variable dimensions (up to several μm^2). Usually, there is an anisotropical charge distribution and considerable differences in the surface topography of the self-assembly structures. These are formed either in suspension, on solid supports (e.g., gold chips, silicon wafers, plastic materials), on air-liquid interfaces, on liposomes or on lipid layers.[24,26] Due to the spacing between the subunits within the two-dimensional S-layer lattices ranging between 3 and 35 nm, S-layers are regarded unique patterning elements for basic and applied research in the field of nanobiotechnology, including both life and non-life sciences.[23,24,27]

1.2. Prokaryotic glycosylation

The increasing evidence that prokaryotes can glycosylate proteins, especially the finding that several human pathogens contain glycoproteins in their surface appendages,[28] has put an end to the doctrine that restricted glycoproteins to eukaryotes. About a decade ago, Sandercock and colleagues[29] tried to sort the accumulated literature on prokaryotic glycoproteins and they discriminated between non-S-layer glycoproteins and S-layer glycoproteins. Among non-S-layer glycoproteins are glycosylated enzymes, such as cellulases and xylanases, membrane-associated glycoproteins, surface-associated glycoproteins, and glycosylated antigens that are shed into the surrounding environment by the organisms.[18] The best investigated glycoproteins from this group are the flagella and pili of archaea and bacteria,[30] and the N-glycoprotein species of *Campylobacter jejuni*.[31,32] A common feature of this group of

glycoproteins is that the glycans are usually relatively short (about 1–20 glycoses) in comparison to the long-chain glycans from bacterial S-layer glycoproteins (with up to approximately 150 glycoses).[33] In contrast, glycans from archaeal S-layer glycoproteins are usually also relatively short.[34]

With S-layer glycoproteins constituting a major class of prokaryotic glycoproteins, much of the current knowledge about prokaryotic glycosylation has been derived from S-layer glycoprotein research. Initial analyses originate from the S-layer glycoprotein of the halophilic archaeon *H. salinarum*.[16] This work was expanded in great detail by Sumper, Wieland, and co-workers in the 1980s. The authors reported not only on the structures but also on the biosynthesis of halobacterial S-layer glycoproteins; furthermore, they provided the first sequence data of the gene encoding a glycosylated archaeal S-layer protein.[12,35] Other archaea that were investigated in greater detail were the haloarchaeon *H. volcanii*,[15,34,36] and the methanogens *Methanothermus fervidus*[13] and *Methanococcus voltae*.[37] Recently, *M. voltae* flagellins have been shown to contain a novel N-linked trisaccharide. Analysis of trypsin-generated peptides derived from the *M. voltae* S-layer glycoprotein revealed a modification by the same trisaccharide, suggesting a common glycosylation process for the two proteins.[38]

As mentioned before, at about the same time when glycosylation on haloarchaea was reported for the first time, Sleytr and Thorne[17] have discovered glycosylation of S-layer proteins from the bacteria *Thermoanaerobacter thermohydrosulfuricus* and *Thermoanaerobacterium thermosaccharolyticum*. Since then, S-layer glycoproteins from several other bacteria have been extensively studied,[39] leading to the awareness of the wide distribution of S-layer glycoproteins among bacteria. Based on a considerable body of S-layer glycan structures from organisms of the *Bacillaceae* family, investigation of the S-layer protein glycosylation process at the molecular level has been initiated about 10 years ago. This endeavor has been lagging behind the structural work due to the lack of suitable molecular tools. When eventually in 2002 the first S-layer gene sequence of the S-layer glycoprotein carrying bacterium *G. stearothermophilus* NRS 2004/3a became available,[40] and only two years later it became evident that S-layer protein glycosylation of the investigated bacterium is encoded by an S-layer glycosylation (*slg*) gene cluster,[41] a novel research direction has emerged, for which we have coined the neologism S-layer 'nanoglycobiology'.[39]

1.3. S-layer nanoglycobiology

S-layer nanoglycobiology accounts for the nanometer-scale cell surface display feature of bacterial S-layer glycosylation and includes structural and biochemical aspects of S-layer glycans as well as molecular data on the machinery underlying the glycosylation event. A key aspect for the full potency of S-layer nanoglycobiology is the unique self-assembly feature of the S-layer protein matrix.[39] Being aware that in many cases the glycan structures associated with a protein are the key to protein function,[42,43] S-layer protein glycosylation will add a new and very valuable component to an 'S-layer based molecular construction kit'.[23] Methods for organizing functional materials at the nanometer level are essential for the development of novel fabrication techniques.[23,24] In particular, molecular self-assembly systems that exploit the molecular scale manufactory precision of biological systems are the prime candidates in nanobiotechnology.

Considering that, at the current state of knowledge, S-layer glycoprotein glycans represent 'nonsense structures' (no discrete function could be attributed to any bacterial S-layer glycan so far), the recent demonstration of the functional transfer of protein glycosylation pathways into the experimental model organism *Escherichia coli*[44,45] opens new avenues for functionalization of S-layer proteins by glycoengineering. It is conceivable that this technology, as was the case for DNA and protein engineering, will become an important tool both in basic and applied S-layer glycoprotein research to analyze the S-layer protein glycosylation process and to artificially equip S-layer proteins with 'functional' glycosylation motifs for (nano)biotechnological or biomedical purposes.

2. Perspectives: S-layer *neo*glycoprotein production

Engineering of tailor-made self-assembly S-layer glycoproteins, the so-called S-layer *neo*glycoproteins, will decisively change our capabilities in influencing and controlling complex biological systems and to conceptuate novel self-assembly nanomaterials.

In our long-term research strategy, the detailed knowledge of the S-layer polypeptide and of the glycosylation feature shall converge (Fig. 1). In principle, alteration of the native S-layer glycan or assembly of completely new glycans on permissive sites of the S-layer protein portion can be envisaged. In any case, the detailed and molecular understanding of the native S-layer protein glycosylation process is a prerequisite. Due to the complexity of S-layer glycan biosynthesis, involving a large number of enzymes for nucleotide sugar biosynthesis, glycosyl transfer reactions, polymerization, membrane transfer of the oligosaccharide chain, and its ligation to distinct sites on the target protein, S-layer glycoproteins have escaped (nano)biotechnological applications so far.

To utilize an S-layer protein as a target for engineered glycosylation, based on the knowledge of the amino acid

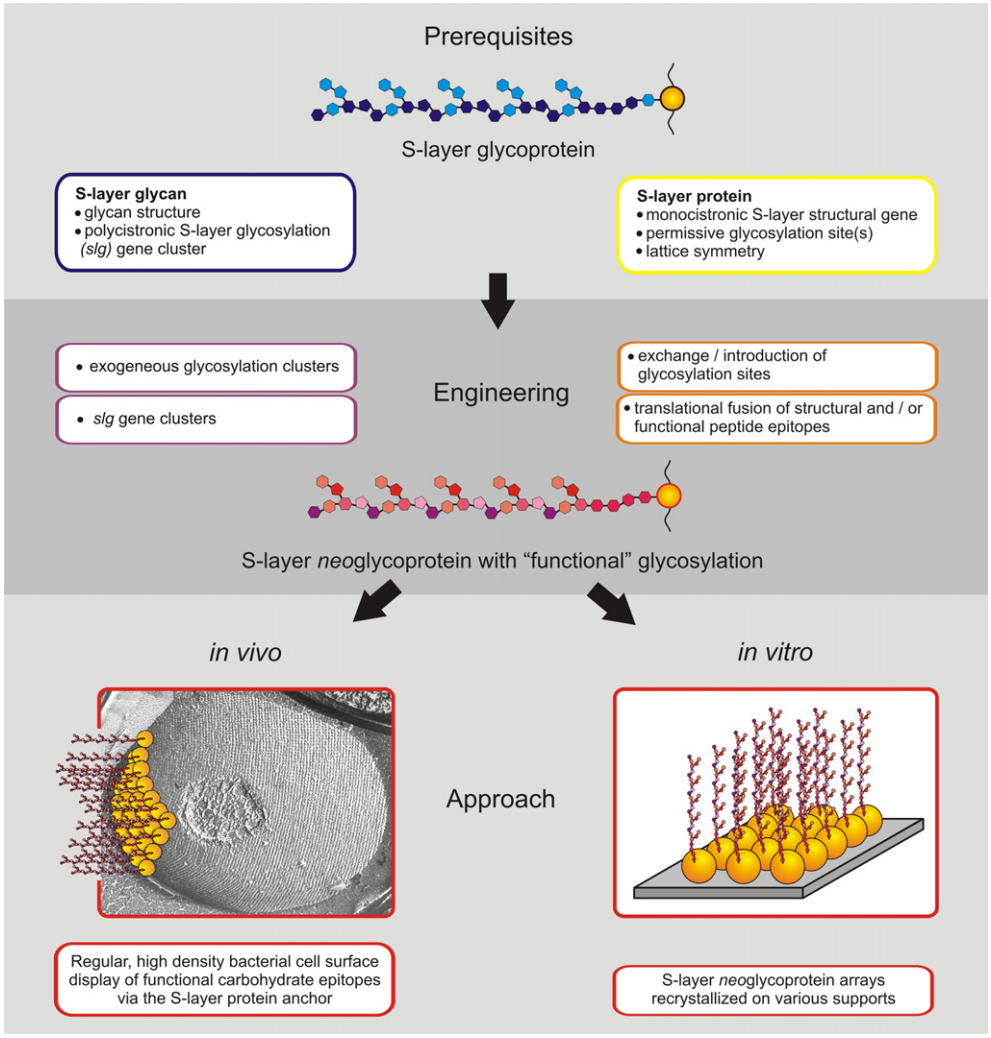


Figure 1. Illustration of prerequisites and strategies for the production of S-layer *neo*glycoproteins.

sequence, the native glycosylation sites, and potentially additional permissive sites for glycosylation have to be determined that allow introduction of exogenous glycosylation sequences into the S-layer protein. In addition, incorporation of structural or functional domains into the S-layer protein by protein engineering techniques will allow tuning S-layer *neo*glycoprotein properties for specific purposes.

Aiming at the nanobiotechnological utilization of the S-layer protein self-assembly feature, in combination with 'functional' glycosylation, it will be important to investigate for any given glycan structure, whether a change in the self-assembly behavior of the S-layer protein takes place upon recombinant glycosylation. It was important to learn from comparative self-assembly studies of native S-layer glycoproteins and the corresponding recombinant, and, consequently, non-glycosylated proteins, that glycosylation does not affect S-layer lattice formation (A. Scheberl, C. Schäffer, P. Messner, unpublished observation).

Our strategy follows two principal lines of development. The first one is the in vivo display of 'functional' glycoproteins on the surface of bacteria enabled by means of recombinant DNA technology. This has become an increasingly used strategy in various applications in microbiology, nanobiotechnology, and vaccinology.[46] Besides outer membrane proteins, lipoproteins, autotransporters, or subunits of surface appendages, which are being evaluated for that kind of applications, the use of the S-layer (glyco)protein cell surface anchor is a very attractive and promising alternative. An impressing example related to this line of development was stated by Paton and co-workers,[47] who demonstrated that a recombinant *E. coli* that displayed a Shiga toxin receptor mimic on its cell surface was capable of adsorbing and neutralizing Shiga toxins with very high efficiency. The in vitro line of development utilizes the recrystallization capability of the S-layer portion on a broad spectrum of supports. In either line, the S-layer 'anchor' offers the unique

advantage of providing a crystalline, regular 'immobilization matrix' that should eventually allow the controlled and periodic surface display of 'functional' glycosylation motifs (Fig. 1).

The conceptuation of S-layer *neo*glycoproteins clearly benefits from the data accumulated in the course of the transfer of the *C. jejuni* N-glycosylation machinery (*pgl* system) together with the *C. jejuni* target protein AcrA into *E. coli*, enabling heterologous expression of a recombinant N-glycoprotein.[44,45] These studies furthermore revealed differences between the prokaryotic and eukaryotic N-glycosylation systems. Bacteria obviously require a negatively charged amino acid at the −2 position to the glycosylated Asn in the sequon for N-glycosylation to occur, resulting in the stringent acceptor sequence D/E-YN-X-S/T (Y, XaP).[48] Concerning the PglB oligosaccharyl::protein transferase of *C. jejuni*, which is responsible for the transfer of the glycan to the AcrA protein as the final step of the glycosylation event, relaxed oligosaccharide substrate specificity has been reported, allowing the transfer of different glycans from the lipid carrier undecaprenyl pyrophosphate to an acceptor protein.[45,49] The corresponding enzyme from the S-layer protein glycosylation pathway has been identified as one of the key modules of S-layer *neo*glycoprotein production. Elucidation of the mechanism of action and the substrate specificity of this enzyme (named WsaB) from *G. stearothermophilus* NRS 2004/3a is currently being investigated in our laboratory. The obvious differences in the protein glycosylation process between bacteria and eukaryotes will clearly have to be accounted for, when aiming at the 'humanization' of S-layer proteins through 'functional' glycosylation.

As S-layer *neo*glycoprotein production represents a fresh area of research, the benefits of S-layer *neo*glycoproteins for potential nanobiotechnology applications may currently be only deduced form the successful cell surface display of foreign peptide epitopes. Among many examples in the literature for peptide epitope display via the S-layer protein anchor (for summary, see Ref. 24), connecting to the field of nanoglycobiology, we have recently constructed a chimeric S-layer protein displaying the fully active RmlA enzyme that is involved in the biosynthesis of nucleotide-activated L-rhamnose.[50] Recent advances in the field of S-layer nanoglycobiology that will be discussed in this article have made our overall strategy a tangible aim of the near future.

3. Diversity of bacterial S-layer protein glycosylation

3.1. Organisms possessing a glycosylated S-layer

Since the early days of S-layer glycoprotein research, it was evident that these cell surface components occur on archaea as well as on bacteria. S-layer glycoproteins have been known for their occurrence among the major lineages of archaea,[15,51] with most of the data concerning *H. salinarum*,[12] *M. fervidus*,[52] and *H. volcanii*.[36] Among bacteria, for a long time only Gram-positive members of the *Bacillaceae* family have been known to possess S-layer glycoproteins, including the species *G. stearothermophilus*, *G. tepidamans*, *Paenibacillus alvei*, *A. thermoaerophilus*, *Thermoanaerobacterium thermosaccharolyticum, Thermoanaerobacter thermohydrosulfuricus*, *T. kivui*, and *Desulfotomaculum nigrificans*.[18,39] Only very recently there were the first reports on the occurrence of glycosylated S-layer proteins in the Gram-negative species *Tannerella forsythia*[53] and *Bacteroides distasonis*.[54] Evidence was obtained from biochemical analyses and so far nothing is known about either glycan structure or linkage of the glycans to the S-layer protein portion. However, in contrast to the known S-layer glycoproteins from *Bacillaceae* investigated by our group in the past,[18,39] these glycosylated S-layer proteins originate from potential pathogens and, therefore, might be of medical relevance.[55] The relation of S-layer protein glycosylation to pathogenicity might implicate a first, discrete function of S-layer glycans, and, therefore, this research direction is currently being established in our laboratory. Concerning an overall function of S-layer glycoproteins in non-pathogenic bacteria, it is conceivable that, by representing the outermost cell surface structure of a bacterium, they participate in diverse cell surface phenomena and, simultaneously, contribute to a high diversification potential of the bacterial cell surface, which may be advantageous in the competitive natural habitat.

3.2. Biochemical and structural insights into S-layer protein glycosylation

The glycosylation degree of S-layer proteins generally varies between 1% and 10% (w/w), but is also subject to change, depending on the laboratory cultivation conditions of the bacteria. Presently, about 15 different S-layer glycoprotein glycan structures have been fully or at least partially elucidated and there are currently more than 25 further indications for glycan modifications of S-layer proteins according to biochemical evidence.[39,56] Figure 2 gives a comprehensive overview on S-layer glycoprotein glycan structures from bacteria. A common feature of almost all bacterial S-layer glycoproteins is the presence of long glycans made of repeats. The observed structures and glycosidic linkage types exceed by far the display found in eukaryotic glycoproteins. Commonly, bacterial S-layer glycan chains are long linear or branched homo- or heterosaccharides with 50–150 glycoses that constitute about 15–50 repeating units.[18,39] The monosaccharide constituents of bacterial S-layer glycan chains include a wide range

of neutral hexoses, 6-deoxyhexoses, and amino sugars. In addition, this spectrum is further extended by rare sugars, such as Qui*p*3NAc, Fuc*p*3NAc, D-Rha*p*, D-Fuc*p* or D-*glycero*-D-*manno*-heptose, which are otherwise

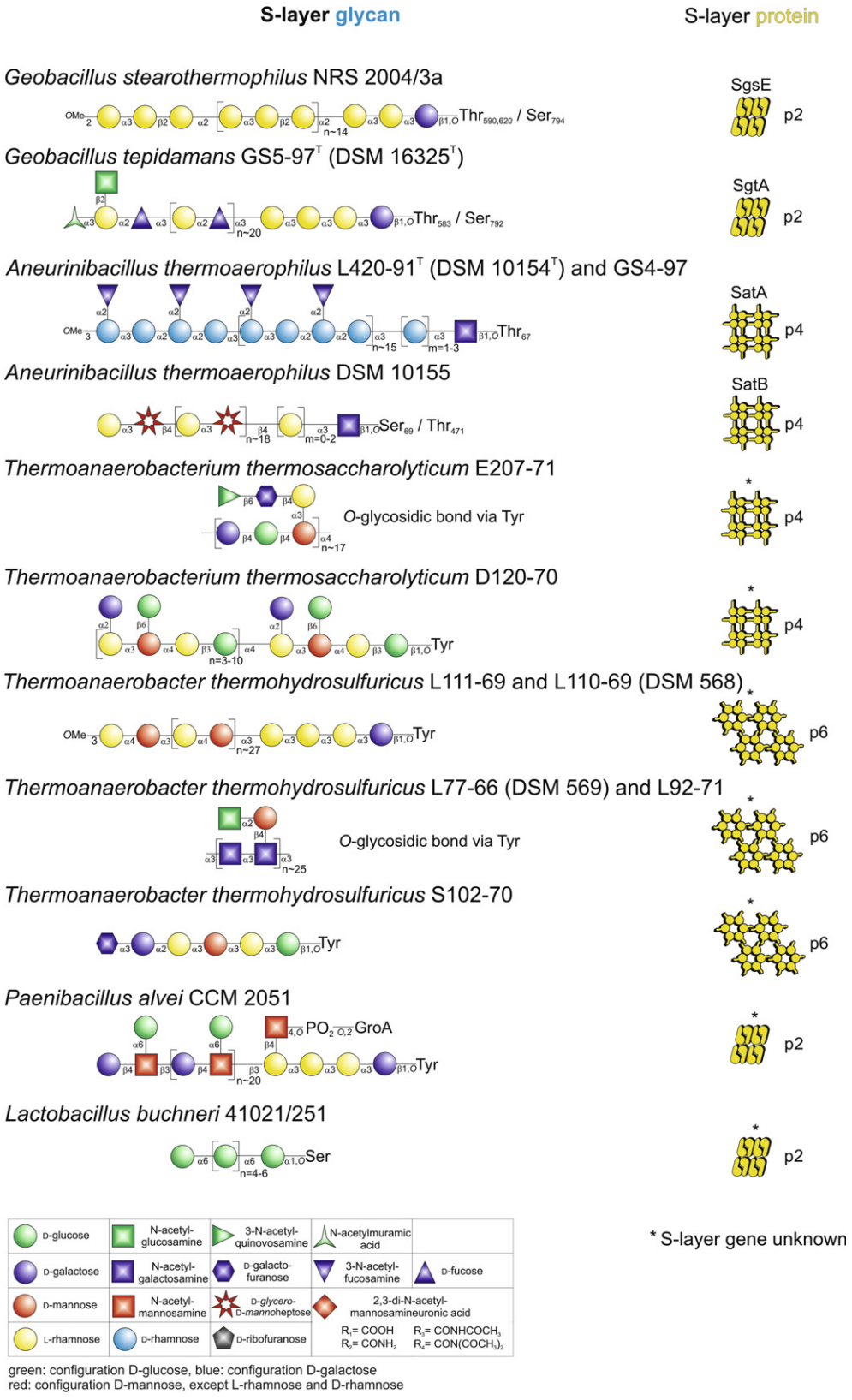


Figure 2. Bacterial S-layer glycan structures and attachment sites of S-layer proteins.

typical constituents of lipopolysaccharide (LPS) O-antigens of Gram-negative bacteria.[57] The typical linkages of these S-layer glycans to the protein portion are O-glycosidic linkages to serine, threonine, and tyrosine; so far, *N*-glycans have been found only in archaea.[39]

Bacterial S-layer glycoproteins possess a tripartite structure, which compares to lipopolysaccharide (LPS) O-antigens. This scheme comprises a glycan chain built of a variable number of repeating units that are linked via a variable core oligosaccharide to the S-layer protein backbone.[39,58,59] In the case where the oligosaccharide core is missing, the carbohydrate chain is directly bound to the glycosylated amino acid via the first repeating unit; the anomeric configuration of the linkage sugar may remain unchanged (no core) or be inverted (pseudocore). In some S-layer glycoproteins, capping of the terminal sugar residue at the non-reducing end of the glycan chain with non-carbohydrate constituents such as *O*-methyl-groups (2-*O*-Me, 3-*O*-Me) is present.[40,60] In this regard, a very interesting modification has been observed just recently. The S-layer glycoprotein glycan of *G. tepidamans* GS5-97^T is capped by α-(*R*)-*N*-acetylmuramic acid, the key component of bacterial peptidoglycan, at carbon 3, and by β-*N*-acetylglucosamine at carbon 2 of the terminal rhamnose residue.[61] A possible explanation for the presence or absence of a capping motif at the non-reducing end of the glycan chains might be derived, when relating S-layer glycoprotein glycan biosynthesis to LPS O-antigen biosynthesis.[57] There, depending on the route of biosynthesis, methylation was reported to function as termination signal for chain elongation.[57,62]

Detailed structural and genetic analyses have shown that bacterial S-layer glycoproteins, for example, from *G. stearothermophilus* NRS 2004/3a, *G. tepidamans* GS5-97^T, and *A. thermoaerophilus* strains, possess only a small number of potential glycosylation sites.[40,59] For instance, on SgsE, the S-layer protein of *G. stearothermophilus* NRS 2004/3a, the amino acids Thr-590, Thr-620, and Ser-794 have been identified as potential glycosylation sites. For the S-layer protein SgtA of *G. tepidamans*, also three distinct glycosylation sites are obvious, in *A. thermoaerophilus* DSM 10155/G$^+$ the amino acids Ser-69 and Thr-471 of SatB represent putative glycosylation sites, and for *A. thermoaerophilus* L420-91^T position Thr-67 of SatA has so far been identified as a glycosylation site. However, it is not known to which extent each of these sites is occupied in the mature glycoprotein. Consequently, an S-layer glycoprotein array is a mixture of variably glycosylated S-layer protein species. In summary, the glycosylation event adds an enormous variation potential to the respective S-layer protein. All features of bacterial S-layer glycoproteins known to date, including potential modifications and glycosylation sites, are summarized in Figure 2.

3.3. The S-layer protein portion

One of the prerequisites for the production of self-assembly S-layer *neo*glycoproteins is the detailed knowledge of the S-layer protein matrix.

Independent of the presence of glycosylation, S-layer proteins *per se* are water-insoluble proteins. Results of amino acid analysis indicated large amounts of glutamic acid and aspartic acid (~15 mol %), a high lysine content (~10 mol %), and large amounts of hydrophobic amino acids (~40–60 mol %). Hydrophilic and hydrophobic amino acids do not form extended clusters. Most S-layer subunits are weakly acidic proteins with isoelectric points in the range 4–6, with the exception of the S-layer proteins of lactobacilli.[24] Secondary structure predictions derived from comparison of protein sequence data and circular dichroism measurements revealed that S-layer proteins have an average α-helix content of ~20% and a β-sheet content of ~40%; aperiodic folding and the β-turn content may vary between 5% and 45%.[23]

As no tertiary structure of a full-size S-layer protein is known so far, our approach for 'functional' S-layer protein glycosylation is the utilization of the native glycosylation sites on the protein, because these sites can be expected to be located in surface-exposed loops within the bulk of the S-layer protein, allowing the display of attached glycans. The identification of the native S-layer glycosylation sites is based on the availability of the primary sequence of the protein, with which the sequences of proteolytically derived S-layer glycopeptides are aligned. The amino acid sequence of an S-layer protein can in principle be obtained either by a genomic approach or a proteomic approach, with the former being more frequently applied. From a recent combined approach, the primary sequence of the S-layer protein SgtA of *G. tepidamans* GS5-97^T was obtained. The purified S-layer protein was applied to proteolytic digests and the obtained peptides were analyzed by nano-ESI-QTOF tandem mass spectrometry.[63] Alignment of the sequenced peptides of SgtA showed high similarity to the N-terminus of the S-layer protein SgsE of *G. stearothermophilus* NRS 2004/3a. The identified conserved amino acid sequences were used for the design of degenerate primers for PCR amplification reactions to enable further sequencing, which finally resulted in the complete sequence of the S-layer structural gene *sgtA*.[64]

Studies on S-layer proteins from different bacteria revealed that S-layer proteins are multidomain proteins, comprising a cell-wall targeting region, which is involved in anchoring of the protein to the peptidoglycan of the bacterial cell wall by interaction with a species specific secondary cell wall polymer,[65,66] and a self-assembly domain, which makes up the larger part of the protein. Based on amino acid sequence alignment and on experimental data, for S-layer glycoproteins

from the *Bacillaceae*, which are serving as a base for S-layer *neo*glycoprotein production, two structural organization principles exist.[23,24] For the species *G. stearothermophilus*, the self-assembly domain is located at the C-terminal region of the S-layer protein,[67] while S-layer proteins of the species *A. thermoaerophilus* possess an N-terminal self-assembly domain. The finding that all glycosylation sites that have been identified on S-layer proteins so far are located within the proposed self-assembly region of the S-layer protein is of high relevance for the conceptuation of S-layer *neo*glycoproteins. Aiming at a highly efficient S-layer *neo*glycoprotein production, deletion of the cell wall binding domain of the S-layer protein without affecting the protein's self-assembly capability or its glycosylation potential should be possible in principle. Indeed, it was shown that deleting 130 or 330 amino acids from the N-terminus of SgsE of *G. stearothermophilus* NRS 2004/3a apparently does not affect S-layer self-assembly.[50]

3.4. Understanding S-layer protein glycosylation patterns

S-layer glycoprotein arrays are usually composed of individual, high-molecular mass glycoprotein subunits.[18] A first indication for the presence of a glycosylated S-layer protein in a given organism can be easily inferred from a positive carbohydrate staining reaction of a highly abundant protein band on an SDS-PA gel of intact bacterial cells, appearing in an estimated molecular mass range between 45 and 200 kDa.[24] Figure 3 gives an overview of the diversity of S-layer protein glycosylation patterns according to the migration behavior of glycoproteins on SDS-PA gels, visualized by Coomassie Blue and periodic acid-Schiff staining (PAS) reaction. It is evident from SDS-PAGE analyses that S-layer glycoproteins migrate as single or multiple bands, or even have a ladder-like appearance, an effect known from LPS O-antigens and referred to as nanoheterogeneity in the context of S-layer nanoglycobiology. In any case, also the corresponding, non-glycosylated S-layer protein is present, which is in accordance with the finding that glycosylation is lagging behind S-layer protein synthesis (R. Novotny, A. Scheberl, and C. Schäffer, unpublished data).

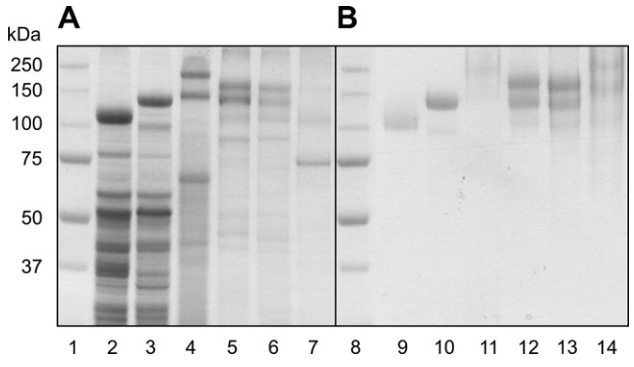


Figure 3. SDS-PAGE analysis of S-layer protein glycosylation patterns on intact bacterial cells by (A) Coomassie Blue staining and by (B) periodic acid-Schiff (PAS) staining reaction. Lanes 1, 8, molecular mass standard; lanes 2, 9, *A. thermoaerophilus* L420-91^T; lanes 3, 10, *A. thermoaerophilus* DSM 10155/G$^+$; lanes 4, 11, *P. alvei* CCM 2051^T; lanes 5, 12, *G. stearothermophilus* NRS 2004/3a; lanes 6, 13, *G. tepidamans* GS5-97^T, lanes 7, 14, *T. thermosaccharolyticum* E207-71. Amounts of 5 μg (A) and 10 μg (B) were loaded onto the gel.

SDS-PAGE analysis of the S-layer glycoprotein of *T. thermosaccharolyticum* E207-71 reveals an apparent molecular mass of 75 kDa for the non-glycosylated protein and nanoheterogeneity of the glycoprotein migrating as 14 different bands in the mass range of 103–213 kDa (Fig. 3, lanes 7, 14).[68] The glycosylated subunits of *A. thermoaerophilus* DSM 10155/G$^+$ possess an apparent molecular mass of 143 kDa, whereas the non-glycosylated band is downshifted to 76 kDa (Fig. 3, lanes 3, 10), which is in accordance with the calculated molecular mass derived from the amino acid sequence of the mature structural protein SatB (78.3 kDa). *A. thermoaerophilus* L420-91^T shows only one glycosylated band at an apparent molecular mass of 109 kDa on SDS-PAGE and a non-glycosylated band at 76 kDa (Fig. 3, lanes 2, 9), which corresponds to the calculated molecular mass of SatA (81.4 kDa). *P. alvei* CCM 2051^T displays three bands on SDS-PAGE with apparent molecular masses of 105, 155, and 240 kDa, of which the two high-mass bands are glycosylated (Fig. 3, lanes 4, 11). On an SDS-PA gel, the mature S-layer glycoproteins of *G. stearothermophilus* NRS 2004/3a (Fig. 3, lanes 5, 12) and of *G. tepidamans* GS5-97^T (Fig. 3, lanes 6, 13) are separated into four bands with apparent molecular masses of 93, 119, 147, and 170 kDa, and 93, 119, 140, and 166 kDa, respectively.[59] In either case, the three high-molecular-mass bands give a positive PAS staining reaction for carbohydrates. The 93-kDa bands are non-glycosylated, and the estimated molecular mass concurs with the calculated masses for the respective S-layer protein after cleavage of the 30-amino acid signal peptide from the precursor protein, that is, 93.7 kDa for SgsE and 92.3 kDa for SgtA.

As the detailed knowledge of the glycosylation pattern is another prerequisite for S-layer *neo*glycoprotein production, the interpretation of SDS-PAGE profiles of S-layer glycosylation was recently refined by mass spectrometry approaches. This was necessary, because it is known that glycoproteins have a retarded electrophoretic mobility due to the attached glycan portion, resulting in too high molecular mass estimates, and that the resolution of individual glycoprotein bands on a PA gel is not good enough to display glycan chain length distribution in detail. To address these questions, mass spectrometry (MS) is a more reliable alternative.[69–71]

Already in 1995, it was shown for the S-layer glycoprotein of *T. thermosaccharolyticum* E207-71 that the

mass of 75.62 kDa obtained by matrix-assisted laser desorption/ionization (MALDI)-MS (mass accuracy, 0.2%) was in accordance with the value of ~75 kDa derived from SDS-PAGE analysis.[68,72] A second broad, only partially resolved signal centered around 87.62 kDa corresponded presumably to the glycoprotein with carbohydrate-related heterogeneity. With the S-layer glycoprotein SgsE of *G. stearothermophilus* NRS 2004/3a, it has been demonstrated, how MS approaches can elucidate S-layer glycosylation patterns based on the known S-layer glycan structure.[59] As described above, on SDS-PA gels, the S-layer glycoprotein preparation of this bacterium is separated into three glycoprotein species. Previous NMR analyses have revealed that the S-layer glycans consist of trisaccharide repeats with the structure →[2)-α-L-Rha*p*-(1→3)-β-L-Rha*p*-(1→2)-α-L-Rha*p*-(1→]$_n$, with the terminal rhamnose residue at the non-reducing end modified on carbon-2 by O-methylation, and of a short core saccharide consisting of α-(1→3)-linked L-Rha*p* residues, attached to carbon 3 of a β-D-galactose residue that serves as the linkage sugar to distinct sites on the S-layer polypeptide backbone.[40] As these data could not fully explain the SDS-PAGE banding pattern, infrared MALDI orthogonal time-of-flight MS (IR-MALDI-oTOFMS), nanoelectrospray ionization quadrupol time-of-flight mass spectrometry (nano-ESI-QTOFMS), and Fourier transform ion cyclotron resonance infrared multiphoton dissociation mass spectrometry (FTICR-IRMPDMS) were adapted for analysis of this high-molecular-mass S-layer glycoprotein and glycopeptides thereof.[59,73] Optimization and adaptation of IR-MALDI-oTOF MS to the water-insoluble S-layer glycoprotein allowed the determination of the average masses of the three inherently heterogenic glycoprotein species of SgsE to be 101.66 kDa, 108.68 kDa, and 115.73 kDa, corresponding to SgsE with up to three attached glycan chains, because the average mass differences between two neighboring peaks of the singly charged ions were 7.09 kDa, which is the average mass of a glycan chain with 15 repeating units. However, nanoheterogeneity of the glycan chain cannot be detected by this method due to peak broadening resulting from reduced resolution in the high-mass range and adduct formation of the matrix. Analysis of isolated glycopeptides by nanoESI-QTOFMS and FTICR-IRMPDMS enabled clear assignment of nanoheterogeneity to each glycan chain, with each of them revealing the most prevalent variation between 12 and 18 trisaccharide repeating units and the possibility of extension of the already known di-L-rhamnose core region by one additional L-Rha residue, and unambiguous identification of a third glycosylation site on the SgsE S-layer protein, namely, at position Thr-590, in addition to the known sites of Thr-620 and Ser-794. These data have led to the current interpretation that in the 101.66-kDa glycoprotein species only one glycosylation site is occupied, in the 108.68-kDa glycoprotein species two glycosylation sites are occupied, and in the 115.73-kDa glycoprotein species three glycosylation sites are occupied, whereas the 94.46-kDa band represents non-glycosylated S-layer protein. A similar situation was observed for the S-layer glycoprotein SgtA of *G. tepidamans* GS5-97^T, where four peaks corresponding to the four bands in SDS-PAGE could be determined to be 94.1, 101.3, 108.5, and 115.7 kDa.[64]

4. The molecular machinery behind S-layer protein glycosylation

4.1. Model organisms

Currently, *G. stearothermophilus* NRS 2004/3a is the best investigated model organism for addressing questions relevant for S-layer nanoglycobiology. However, to obtain a more general insight into the S-layer glycosylation process there are five more organisms under investigation in our laboratory; these are *G. tepidamans* GS5-97^T, *A. thermoaerophilus* DSM 10155/G$^+$ and L420-91^T, *T. thermosaccharolyticum* E207-71, and *P. alvei* CCM 2051^T. The selection of these organisms is based on the availability of the S-layer glycan structure; the current status of research on these organisms is quite different. It is important to note that *P. alvei* CCM 2051^T is so far the only one of the selected organisms that can uptake foreign DNA, making it a prime candidate for in vivo display of S-layer *neo*glycoproteins (K. Zarschler, C. Schäffer, P. Messner, unpublished data).

All model organisms are Gram-positive, moderately thermophilic or mesophilic, spore-forming bacteria that originate either from soil samples (*G. stearothermophilus* NRS 2004/3a), from extraction plants of Austrian beet sugar factories (*T. thermosaccharolyticum* E207-71, *G. tepidamans* GS5-97^T, *A. thermoaerophilus* DSM 10155/G$^+$ and L420-91^T), or from foul brood of bees (*P. alvei* CCM 2051^T).

Both *Geobacillus* strains, the well characterized *G. stearothermophilus* NRS 2004/3a and the just recently classified *G. tepidamans* GS5-97^T, possess S-layer glycoproteins assembling into arrays with oblique symmetry.[74,75] While the S-layer glycan of *G. stearothermophilus* NRS 2004/3a is a poly-L-rhamnan chain consisting of trisaccharide repeats with the structure →[2)-α-L-Rha*p*-(1→3)-β-L-Rha*p*-(1→2)-α-L-Rha*p*-(1→]$_n$, the more complex glycan of *G. tepidamans* GS5-97^T is composed of disaccharide repeats with the structure →[3)-α-L-Rha*p*-(1→2)-α-D-Fuc(1→]$_n$.[40,61] The S-layer glycoprotein SatA of *A. thermoaerophilus* L420-91^T constitutes a square S-layer lattice and its glycan chain is composed of identical hexasaccharide repeats containing D-rhamnose and 3-acetamido-3,6-dideoxy-D-galact-

ose (D-Fuc*p*3NAc) in the molar ratio of 2:1.[60,76] Another member of the species *Aneurinibacillus*, namely *A. thermoaerophilus* DSM 10155/G$^+$, possesses the only heptose-containing S-layer glycan known so far. It is composed of disaccharide repeats of D-rhamnose and D-*glycero*-D-*manno*-heptose residues.[58,77] *T. thermosaccharolyticum* E207-71 possesses the most complex repeating unit structure of our model organisms with a branched S-layer hexasaccharide repeat composed of D-Gal*f*, L-Rha, D-Gal, D-Glc, D-Man, and the rare glycose 3-*N*-acetylquinovosamine (D-Qui*p*3NAc).[68]

Due to the ease of its transformability, we are recently focusing on the mesophilic bacterium *P. alvei* CCM 2051^T. More than 40 years ago, a granular, macromolecular pattern in the cell wall of this bacterium was observed and several years later, the structure of the S-layer and its glycan chain have been investigated.[78–81] Cells of *P. alvei* CCM 2051^T are covered with an oblique S-layer glycoprotein lattice. The S-layer glycan is composed of branched trisaccharide repeats with the constituents D-Gal, D-Glc, and D-ManNAc; in addition, a 2-phosphoglyceric acid residue is present in the core of this S-layer glycan.[81]

4.2. S-layer glycosylation gene clusters

An important milestone toward understanding S-layer protein glycosylation was the identification and sequencing of several S-layer glycosylation (*slg*) gene clusters. Analysis of S-layer glycosylation on the molecular level clearly benefited from the detailed molecular knowledge of the biosynthesis routes of LPS O-antigens and capsular polysaccharides.[82] Currently, most data are available from the organisms *G. stearothermophilus* NRS 2004/3a (GenBank AF328862),[83] *G. tepidamans* GS5-97^T (GenBank AY883421)[64] and *A. thermoaerophilus* strains L420-91^T (GenBank AY442352),[41] and DSM 10155/G$^+$ (GenBank AF324836).[41] In addition, a partial *slg* gene cluster sequence is available from *T. thermosaccharolyticum* E207-71 (GenBank AY422724)[41] (Table 1, Fig. 4).

Based on the common principle that sugars are incorporated into growing glycan chains from the respective nucleotide-activated precursor, we surveyed the literature for the genes involved in the biosynthesis of nucleotide-activated L-rhamnose (dTDP-β-L-rhamnose),[84] D-fucose (dTDP-α-D-fucose),[64] D-rhamnose (GDP-α-D-rhamnose),[85] 3-*N*-acetylfucosamine (dTDP-α-D-Fuc*p*3NAc),[86] D-*glycero*-α-D-*manno*-heptose (GDP-D-*glycero*-α-D-*manno*-heptose),[87,88] and 3-*N*-acetylquinovosamine (dTDP-α-D-Qui*p*3NAc),[89] all of which are constituents of the S-layer glycans of our model organisms. For entry into the *slg* gene clusters, degenerate primers were designed based on these gene sequences. Further up- and downstream sequencing eventually revealed the presence of whole gene clusters. The current picture of the *slg* gene clusters is the following. Depending on the complexity of the encoded S-layer glycan, the clusters are ~16 to ~25 kb in size and transcribed as polycistronic units.[41] They include nucleotide sugar pathway genes that are arranged consecutively, glycosyl transferase genes, glycan processing genes, and transporter genes. The presence of insertion sequences and the decrease of the G + C content at the *slg* locus in comparison to the respective bacterial genome suggest that the investigated organisms have acquired their specific S-layer glycosylation potential by lateral gene transfer. From the assigned genes, it is evident that none of the *slg* gene clusters encodes the biosynthesis of the nucleotide-activated linkage sugar of the S-layer glycan (UDP-Gal and UDP-GalNAc, respectively). Thus, S-layer protein glycosylation additionally requires the participation of housekeeping genes that map outside the cluster. The gene encoding the respective S-layer target protein is transcribed monocistronically and independently of the *slg* cluster genes (compare with Fig. 4). Its chromosomal location is not necessarily in close vicinity to the *slg* gene cluster.

For deducing a common organization principle of *slg* gene clusters, as known from the clusters encoding the biosynthesis of other bacterial polysaccharides, such as the LPS O-antigens of Gram-negative bacteria,[57,90] or the exopolysaccharides of lactic acid bacteria,[91] the number of *slg* clusters sequenced so far is too low. Furthermore, the current sequence information does not allow the identification of specific genes on the chromosome of the *Bacillaceae*, such as the *galF* and *gnd* genes in *E. coli* and *Salmonella enterica*,[90] or the *hemH* and *gsk* genes in *Yersinia enterocolitica*,[92] between which the *slg* locus is preferentially located. Nevertheless, recently, the comparison of the *slg* gene clusters of *G. stearothermophilus* NRS 2004/3a and *G. tepidamans* GS5-97^T revealed that the clusters are organized in a similar way.[64] The first three ORFs downstream of the S-layer gene include highly homologous genes coding for a putative ligase and a putative rhamnosyltransferase and are followed by the ABC-2 transporter encoding genes *wzm* and *wzt*. The subsequent sequence segments are different. In *G. tepidamans* GS5-97^T, there are five ORFs including four putative glycosyltransferase genes and the dTDP-4-dehydro-6-deoxyglucose reductase gene *fcd*. This segment is terminated by three transposases and one small ORF with unknown function. In *G. stearothermophilus* NRS 2004/3a, only two ORFs in that region of the gene cluster are found, a putative methyltransferase gene and a putative glycosyltransferase gene. Downstream of these variable parts of the *slg* gene clusters, the *rmlACBD* genes, a putative rhamnosyltransferase gene and the putative UDP-galactose lipid carrier transferase gene, are located. These parts of both gene clusters have the highest homology (up to 96% similarity). The

Table 1. Data base homologies of the genes contained in the *slg* gene clusters and S-layer structural genes

Nomenclature[a]		aa	MW (kDa)	G + C%	Assigned protein function (database link)
New	Old				
(a) Description of the slg gene cluster of Geobacillus stearothermophilus NRS 2004/3a (GenBank accession number AF328862)					
wsaA	ORFG101	169	19.9	43	TPR-repeat containing protein (ABM68315)
wsaB	ORFG102	526	59.9	37.4	Lipid A core::surface polymer ligase (ABM68316)
wsaC	ORFG103	324	38.2	33.0	Rhamnosyltransferase (ABM68317)
wzm	*wzm*	268	30.4	33.0	ABC transporter integral membrane protein (ZP_01501790)
wzt	*wzt*	409	46.2	33.7	ABC transporter nucleotide-binding protein (NP_643907)
wsaE	ORFG106	1127	132.6	31.8	Methyltransferase (YP_984868)
wsaF	ORFG107	413	48.3	30.0	Rhamnosyltransferase (ZP_01037878)
rmlA	*rmlA*	299	33.2	39.6	Glucose-1-phosphate thymidylyltransferase (ABM68329)
rmlC	*rmlC*	183	21.3	38.4	dTDP-dehydrorhamnose 3,5-epimerase (ABM68330)
rmlB	*rmlB*	342	38.8	38.4	dTDP-D-glucose 4,6-dehydratase (ABM68331)
rmlD	*rmlD*	282	31.9	38.5	dTDP-dehydrorhamnose reductase (ABM68332)
wsaD	ORFG112	289	33.4	37.0	Rhamnosyltransferase (ABM68333)
wsaP	ORFG113	471	54.5	37.9	UDP-galactose-lipid carrier transferase (ABM68334)
transposase	*istA*	184	21.7	42.2	Transposase (NP_634718)
(b) Description of the slg gene cluster of Geobacillus tepidamans GS5-97^T (GenBank accession number AY883421)					
wsbA		160	19.2	41.7	TPR-domain containing protein (AAR99603)
wsbB		526	59.7	37.1	Lipid A core::O-antigen ligase (AAR99604)
wsbC		325	38.5	32.1	Rhamnosyltransferase (AAR99605)
wzm		269	32.0	27.1	ABC transporter integral membrane protein (AAS49124)
wzt		395	45.5	29.1	ABC transporter nucleotide-binding protein (BAA82537)
wsbG		283	33.9	32.3	Glycosyltransferase (EDN15657)
wsbH		434	51.2	27.9	Glycosyltransferase (YP_323850)
wsbI		332	39.3	29.8	Rhamnosyltransferase (BAA19639)
wsbJ		414	48.6	31.2	Glycosyltransferase (YP_061583)
fcd		308	35.5	26.4	dTDP-4-dehydro-6-deoxy-glucose reductase (ZP_01501805)
transposase		181	21.4	41.3	Transposase (CAA79750)
transposase		122	14.5	38.8	Transposase (BAD18133)
transposase		59	7.1	46.2	Transposase (NP_243386)
wsbL		62	7.1	43.0	Unknown function
rmlA		299	33.3	41.6	Glucose-1-phosphate thymidylyltransferase (AAR99610)
rmlC		183	21.3	37.2	dTDP-4-dehydrorhamnose 3,5-epimerase (AAR99611)
rmlB		339	38.7	37.7	dTDP-D-glucose 4,6-dehydratase (AAR99612)
rmlD		282	32.0	37.8	dTDP-4-dehydrorhamnose reductase (AAR99613)
wsbD		250	33.3	37.1	Rhamnosyltransferase (AAR99614)
wsbP (incomplete ORF)		441	54.0	39.2	UDP-galactose-lipid carrier transferase (AAR99615)
(c) Description of the slg gene cluster of Aneurinibacillus thermoaerophilus L420-91^T (GenBank accession number AY442352)					
wscA	ORFA101	883	97.0	48.4	S-layer associated protein (YP_076994)
wscC	ORFA102	188	21.2	44.0	Transcriptional regulator (YP_753396)
gmd	*gmd*	341	39.0	37.0	GDP-mannose-4,6-dehydratase (YP_685009)
rmd	*rmd*	309	34.5	35.9	GDP-4-dehydro-6-deoxy-D-mannose-reductase (ZP_00768807)
wzm	*wzm*	261	30.5	33.2	ABC transporter integral membrane protein (YP_727348)
wzt	*wzt*	408	45.7	35.4	ABC transporter nucleotide-binding protein (BAA28325)
wscD	ORFA107	467	55.4	32.0	Methyltransferase (NP_616119)
wscE	ORFA108	353	41.2	31.0	Glycosyltransferase (YP_303938)
wscF	ORFA109	393	46.4	32.2	Hexosyltransferase (YP_303939)
transposase	ORFA110	103	11.94	41.1	Transposase (YP_001124563)
wscG	ORFA111	746	86.2	29.4	Integral membrane protein (AAK53474)
fdtA	*fdtA*	139	16.0	34.8	dTDP-4-keto-6-deoxy-D-glucose-3,4-ketoisomerase (EDN36731)
fdtC	*fdtC*	192	21.0	35.1	dTDP-D-Fuc*p*3N acetylase (ZP_01774133)
fdtB	*fdtB*	363	41.0	35.4	dTDP-3-keto-6-deoxy-D-galactose aminase (ABF93267)
wscH	ORFA115	310	35.8	34.4	Glycosyltransferase (YP_001420868)
wscI	ORFA116	119	13.7	25.8	Integral membrane (GtrA-like) protein (YP_365446)
rmlA	*rmlA*	305	33.8	36.0	Glucose-1-phosphate thymidylyltransferase (ABM68329)
rmlB	*rmlB*	343	39.7	33.7	dTDP-D-glucose 4,6-dehydratase (ZP_01872234)
transposase	ORFA119	132	15.7	47.0	Transposase (NP_243386)
transposase	ORFA120	172	19.7	47.5	Transposase (BAD18254)
manC	*manC*	464	52.8	30.1	Mannose-1-phosphate guanylyltransferase (YP_149157)

Table 1 (*continued*)

Nomenclature[a]		aa	MW (kDa)	G + C%	Assigned protein function (database link)
New	Old				
wscJ	ORFA122	389	44.9	30.4	Mannosyltransferase (CAJ74004)
wscK	ORFA123	377	43.1	27.1	Glycosyltransferase (ZP_01425292)
wscL	ORFA124 (fragment)	111	12.6	48.3	Mannosyltransferase (NP_285363)
(d) Description of the slg gene cluster of Aneurinibacillus thermoaerophilus DSM 10155/G$^+$ (GenBank accession number AF324836)					
wsdA	ORFA201	883	97.0	48.2	S-layer associated protein (YP_076994)
wsdC	ORFA202	188	21.1	44.0	Transcription regulator (YP_753396)
wzm	*wzm*	267	31.4	32.3	ABC transporter integral membrane protein (YP_001232541)
wzt	*wzt*	435	49.4	33.6	ABC transporter nucleotide-binding protein (YP_001311813)
wsdD	ORFA205	303	34.7	33.8	Methyltransferase (YP_001311812)
wsdE	ORFA206	618	72.4	32.4	Glycosyltransferase (ZP_01624791)
wsdF	ORFA207	438	50.9	34.8	Glycosyltransferase (YP_294944)
hddA	*hddA*	341	38.1	36.5	D-*Glycero*-D-*manno*-heptose 7-phosphate kinase (ZP_00592568)
gmhA	*gmhA*	198	21.5	36.0	Sedoheptulose 7-phosphate isomerase (YP_179590)
hddC	*hddC*	230	26.6	35.1	D-*Glycero*-D-*manno*-heptose 1-phosphate guanylyltransferase (NP_809389)
gmhB	*gmhB*	179	20.8	31.1	D-*Glycero*-D-*manno*-heptose 1,7-bisphosphate phosphatase (YP_001060251)
transposase	ORFA212	114	13.7	35.7	Transposase fragment (YP_001127525)
transposase	ORFA213	78	8.9	38.9	Transposase fragment (YP_001127525)
rmlA	*rmlA*	296	32.9	43.0	Glucose-1-phosphate thymidylyltransferase (ABM68329)
rmlC	*rmlC*	182	20.7	44.0	dTDP-6-deoxy-D-glucose-3,5-epimerase (AAQ23680)
rmlB	*rmlB*	341	39.1	41.4	dTDP-D-glucose 4,6-dehydratase (ABM68331)
rmlD	*rmlD*	282	31.7	45.2	dTDP-4-dehydrorhamnose reductase (AAQ23682)
wsdG	ORFA218 (partial)	295	23.2	41.0	Sugar transferase (YP_385128)
(e) Description of the incomplete slg gene cluster of Thermoanaerobacterium thermosaccharolyticum E207-71 (*GenBank accession number* AY422724)					
glf	*glf*	372	44.6	27.1	UDP-galactopyranose mutase (YP_149169)
wse1	ORFT102	321	37.5	26.6	Rhamnosyltransferase (NP_275486)
wse2	ORFT103	278	33.6	20.7	Glycosyltransferase (ZP_01372717)
qdtC	*qdtC*	265	29.7	29.2	dTDP-D-Qui*p*3N acetylase (ZP_01065559)
qdtA	*qdtA*	136	16.0	26.5	dTDP-4-keto-6-deoxy-D-glucose-3,4-ketoisomerase (AAS55720)
qdtB	*qdtB*	365	41.5	29.4	dTDP-3-keto-6-deoxy-D-glucose aminase (YP_001309709)
wzx	*wzx*	491	55.8	25.3	Polysaccharide transporter (YP_001392372)
rmlB	*rmlB*	351	40.7	32.2	dTDP-D-glucose 4,6-dehydratase (ZP_00778370)
rmlA	*rmlA*	302	33.6	35.0	Glucose-1-phosphate thymidylyltransferase (ZP_00778369)
rmlD	*rmlD*	294	33.5	32.8	dTDP-dehydrorhamnose reductase (ZP_00778368)
rmlC	*rmlC* (partial)	(83)	(9.7)	(32.0)	dTDP-4-dehydrorhamnose 3,5-epimerase (ZP_00778367)

(f) Description of S-layer structural genes encoding target proteins for glycosylation

Species	Strain	Gene	G + C%	Precursor/signal peptide (aa)	MW (kDa)	pI	Glycosylation sites on protein precursor	Protein sequence accession number
Geobacillus stearothermophilus	NRS 2004/3a	*sgsE*	47.6	903/30	93.7	6.01	Thr-590, Thr-620, Ser-794	AAL46630
Geobacillus tepidamans	GS5-97^T	*sgtA*	38.0	901/30	92.3	5.38	Ser-792, Thr-583	AAX46285
Aneurinibacillus thermoaerophilus	L420-91^T	*satA*	43.6	759/30	78.5	5.72	Thr-67	AAS44591
Aneurinibacillus thermoaerophilus	DSM 10155/G$^+$	*satB*	44.2	738/30	75.4	5.15	Ser-69, Thr-794	AAS44592

[a] The identical color code for the components involved in S-layer glycoprotein glycan biosynthesis is used in Figures 4, 6 and in Table 1: light blue, monosaccharide biosynthesis; red, glycan assembly; orange, glycan transfer; green, export; gray, transposase; black, unknown function; dark blue, elongated glycan chain; pink, linkage glycose; yellow, S-layer protein.

glycoproteins of both organisms possess an extended tripartite structure.[39] It seems that the variable part in the center of the *slg* gene clusters is responsible for the biosynthesis of the individual repeating units and terminating elements, whereas the region with higher homology codes for proteins involved in assembling the core region, transport of the glycan to the cell surface, and its ligation to the S-layer protein. This resembles the organization of

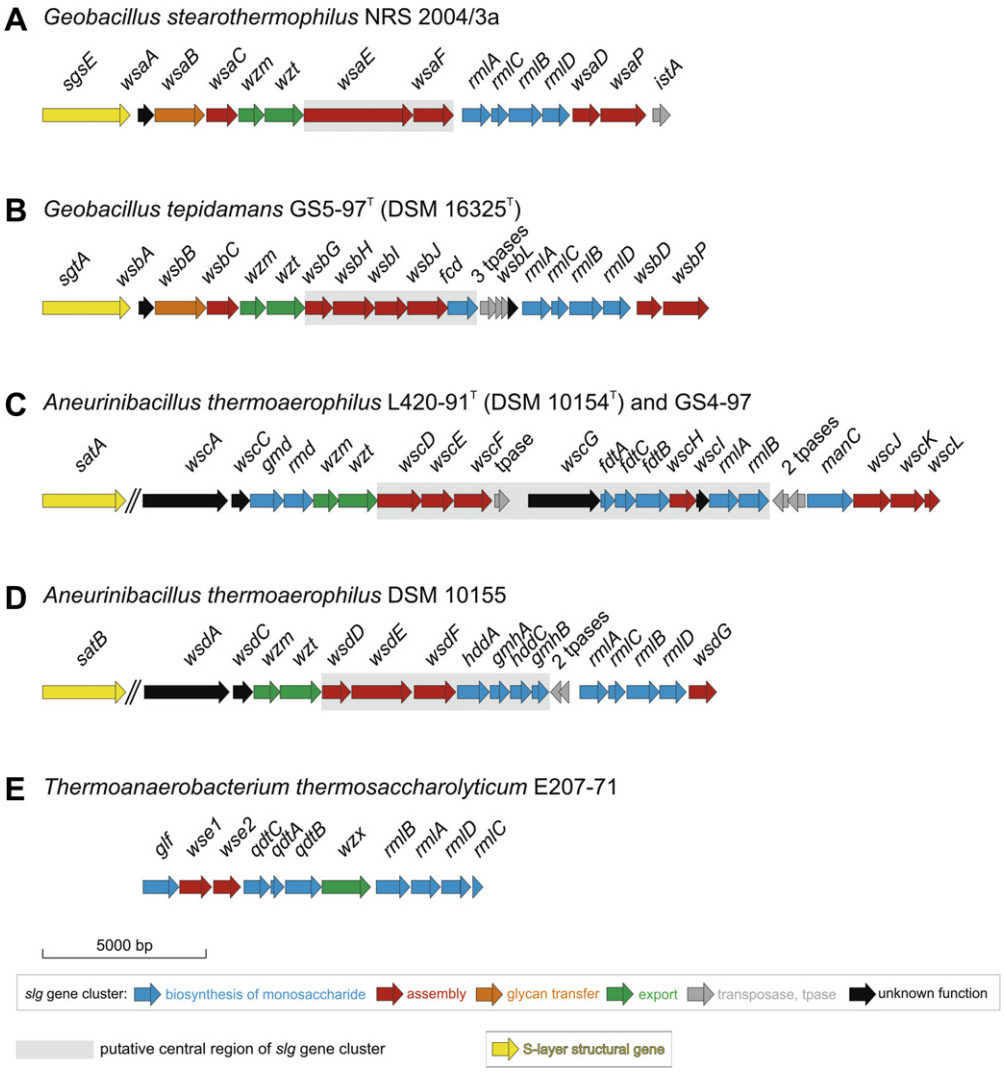


Figure 4. Genetic organization of *slg* gene clusters of (A) *G. stearothermophilus* NRS 2004/3a; (B) *G. tepidamans* GS5-97^T; (C) *A. thermoaerophilus* L420-91^T; (D) *A. thermoaerophilus* DSM 10155/G$^+$; and (E) *T. thermosaccharolyticum* E207-71. The identical color code for the components involved in S-layer glycoprotein glycan biosynthesis is used in Figures 4, 6 and in Table 1: light blue, monosaccharide biosynthesis; red, glycan assembly; orange, glycan transfer; green, export; gray, transposase; black, unknown function; dark blue, elongated glycan chain; pink, linkage glycose; yellow, S-layer protein.

O-antigen gene clusters in Gram-negative organisms, where the variability of O-antigens is considered as a result of recombination events in the central region of the O-antigen gene clusters.[93,94]

To conform the designation to the bacterial polysaccharide gene nomenclature[95] (BPGD database, http://www.microbio.usyd.edu.au/BPGD), we named the genes of the *slg* gene clusters *wsx***Y**, where *w* stands for glycan biosynthesis, *s* for S-layer-associated, *x**, starting from *a*, indicates the bacterial organism, in which a given *slg* gene cluster occurs, and *Y**, starting from *A*, reflects the consecutive order of the genes within a cluster. In case of an incomplete gene cluster, the capital letters are replaced by consecutive numbers, starting from 1. There are two exceptions, *B*, independent on the position within the cluster, is always reserved for the gene encoding the oligosaccharyl::protein transferase and *P* is reserved for the initiation enzyme.

4.3. Enzymes from S-layer protein glycosylation pathways

Most of the protein functions encoded by the different genes contained in the *slg* gene clusters have been preliminary assigned according to data base alignments. Functional characterization has so far been performed only with different proteins that are involved in the biosynthesis of nucleotide sugars and for the initiation enzyme of S-layer glycan biosynthesis WsaP.

4.3.1. Nucleotide sugar biosynthesis enzymes. Based on the identification of the S-layer glycan specific nucleotide sugar genes in the *slg* gene clusters, the encoded

proteins were cloned and overexpressed in *E. coli*. This, together with the establishment of recombinant enzymes functional assays led to the characterization of the biosynthesis pathways for dTDP-β-L-Rha*p*,[84] dTDP-α-D-Fuc*p*3NAc,[86] dTDP-α-D-Qui*p*3NAc,[89] GDP-D-*glycero*-α-D-*manno*-heptose,[87] GDP-α-D-Rha*p*,[85] and dTDP-α-D-Fuc*p*[64] in Gram-positive organisms.

Comparable to lipopolysaccharide O-antigen biosynthesis in Gram-negative bacteria, dTDP-β-L-rhamnose is synthesized by *A. thermoaerophilus* DSM 10155/G$^+$ in a four-step reaction sequence from dTTP and glucose-1-phosphate by the enzymes glucose-1-phosphate thymidylyltransferase (RmlA), dTDP-D-glucose-4,6-dehydratase (RmlB), dTDP-4-dehydrorhamnose-3,5-epimerase (RmlC), and dTDP-4-dehydrorhamnose reductase (RmlD).[84] RmlA and RmlB are also involved in the biosynthesis of dTDP-α-D-Fuc*p*3NAc in *A. thermoaerophilus* L420-91^T, in the production of dTDP-α-D-Qui*p*3NAc in *T. thermosaccharolyticum* E207-71, and in the biosynthesis of dTDP-α-D-fucose in *G. tepidamans* GS5-97^T, producing the important key intermediate dTDP-4-dehydro-6-deoxyglucose[96] (Fig. 5). In the first case, the production of dTDP-α-D-Fuc*p*3NAc is completed by the sequential action of a dTDP-4-keto-6-deoxy-D-glucose-3,4-ketoisomerase (FdtA), a dTDP-3-keto-6-deoxy-D-galactose aminase (FdtB), and a dTDP-α-D-Fuc*p*3N acetylase (FdtC). FdtA was the first isomerase described that is capable of synthesizing dTDP-3-keto-(dehydro)-6-deoxy-galactose from dTDP-4-dehydro-6-deoxyglucose.[86] In the second case, dTDP-4-dehydro-6-deoxyglucose is processed by the enzymes dTDP-4-keto-6-deoxy-D-glucose-3,4-ketoisomerase (QdtA), dTDP-3-keto-6-deoxy-D-glucose aminase (QdtB), and dTDP-D-Qui*p*3N acetylase (QdtC) to form dTDP-α-D-Qui*p*3NAc.[89] In the last case, the biosynthesis of dTDP-α-D-fucose is completed by the dTDP-4-dehydro-6-deoxyglucose reductase (Fcd), using NADH as a cofactor.[64]

The two enzymes responsible for the biosynthesis of the nucleotide activated form of the D-enantiomer of rhamnose, GDP-α-D-rhamnose, are the GDP-D-mannose dehydratase (Gmd) converting GDP-D-mannose to GDP-4-dehydro-6-deoxy-D-mannose with NADPH as cofactor, and the reductase Rmd catalyzing the reduction of the keto-intermediate to the final product using both NADH and NADPH as hydride donor. Gmd was identified as a novel bifunctional enzyme exhibiting both dehydratase and reductase activities.[85]

For biosynthesis of the nucleotide-activated form of D-*glycero*-D-*manno*-heptose in *A. thermoaerophilus* DSM 10155/G$^+$, the four enzymes GmhA, HddA, GmhB, and HddC are required. The isomerase GmhA catalyzes the conversion of D-sedoheptulose-7-phosphate to D-*glycero*-D-*manno*-heptose-7-phosphate and the phosphokinase HddA adds a phosphate group to form D-*glycero*-D-*manno*-heptose-1,7-bisphosphate. The phosphatase GmhB removes the phosphate at the C-7 position, and the intermediate D-*glycero*-α-D-*manno*-heptose-1-phosphate is finally activated with GTP by the pyrophosphorylase HddC to yield the final product GDP-D-*glycero*-α-D-*manno*-heptose. It should be noted that the heptose residue is synthesized as GDP-D-*glycero*-α-D-*manno*-heptose, whereas ADP-L-*glycero*-β-D-*manno*-heptose is the precursor of the inner core

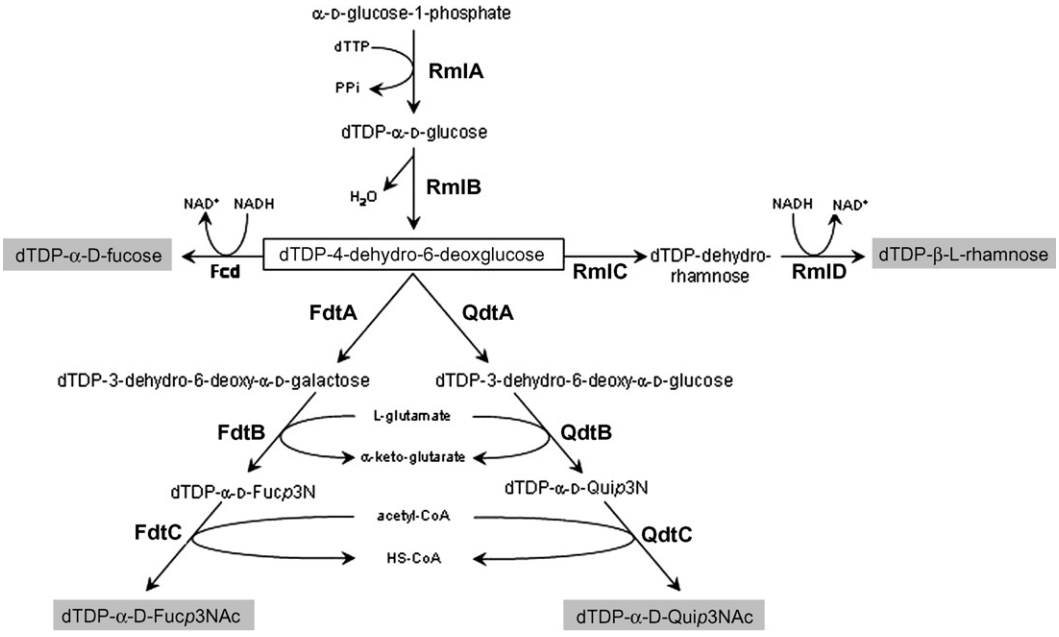


Figure 5. Nucleotide sugar biosynthetic pathways of activated sugars required for S-layer glycoprotein glycan biosyntheses utilizing dTDP-4-dehydro-6-deoxyglucose as key intermediate.

lipopolysaccharide biosynthesis of organisms like *E. coli* or *Salmonella enterica*.[87,88]

From these data, it is obvious that the different S-layer protein glycosylation pathways provide a spectrum of rare enzymes that may be used for glycoengineering purposes in heterologous hosts. Furthermore, some of these enzymes, for example, most Rml enzymes from *A. thermoaerophilus* DSM 10155/G^{+}, exhibit significantly higher stability at 37 °C than the enzymes from the mesophilic strain *S. enterica*, due to their origin from thermophilic organisms. This advantage could lead to the development of improved high-throughput screening systems for specific sugars.[97]

4.3.2. Initiation enzyme. Following the identification of the initiation enzyme as one of the key modules of S-layer protein glycosylation, recently, the initiation enzyme of S-layer glycoprotein biosynthesis has been characterized in the model organism *G. stearothermophilus* NRS 2004/3a.[98] The 471-aa membrane protein WsaP of *G. stearothermophilus* NRS 2004/3a shows high homology to glycosyltransferases, for example, WbaP, which catalyzes the first step in polysaccharide biosynthesis by transferring a hexose-1-phosphate residue from UDP-hexoses (galactose and glucose) to a phosphoryl-ated lipid carrier.[99] To assess the functional domain of the enzyme, different truncated forms of the protein were designed and heterologously expressed in *E. coli*.[98] WsaP is capable of reconstituting K30 antigen and O-antigen biosynthesis in the WbaP-deficient strains *E. coli* CWG 466 and *S. enterica* MSS2, respectively. In vitro assays of isolated membranes of *E. coli* harboring WsaP confirmed the galactosyltransferase activity and the catalytic site is located at the C-terminal half of WsaP, including one transmembrane domain. Undecaprenol phosphate is recognized as an acceptor molecule for WsaP in the *E. coli* background, which is a necessary prerequisite for combining the S-layer protein O-glycosylation system with other polysaccharide biosynthesis pathways for S-layer *neo*glycoprotein production.

5. Proposed S-layer protein glycosylation pathway

It is obvious that the biosynthesis of an S-layer glycoprotein is a very complex process in which the glycosylation event has to be coordinated with the amount of the synthesized S-layer protein, its translocation through the cell wall, and its incorporation into the existing S-layer lattice. Considering that up to 20% of the total cellular

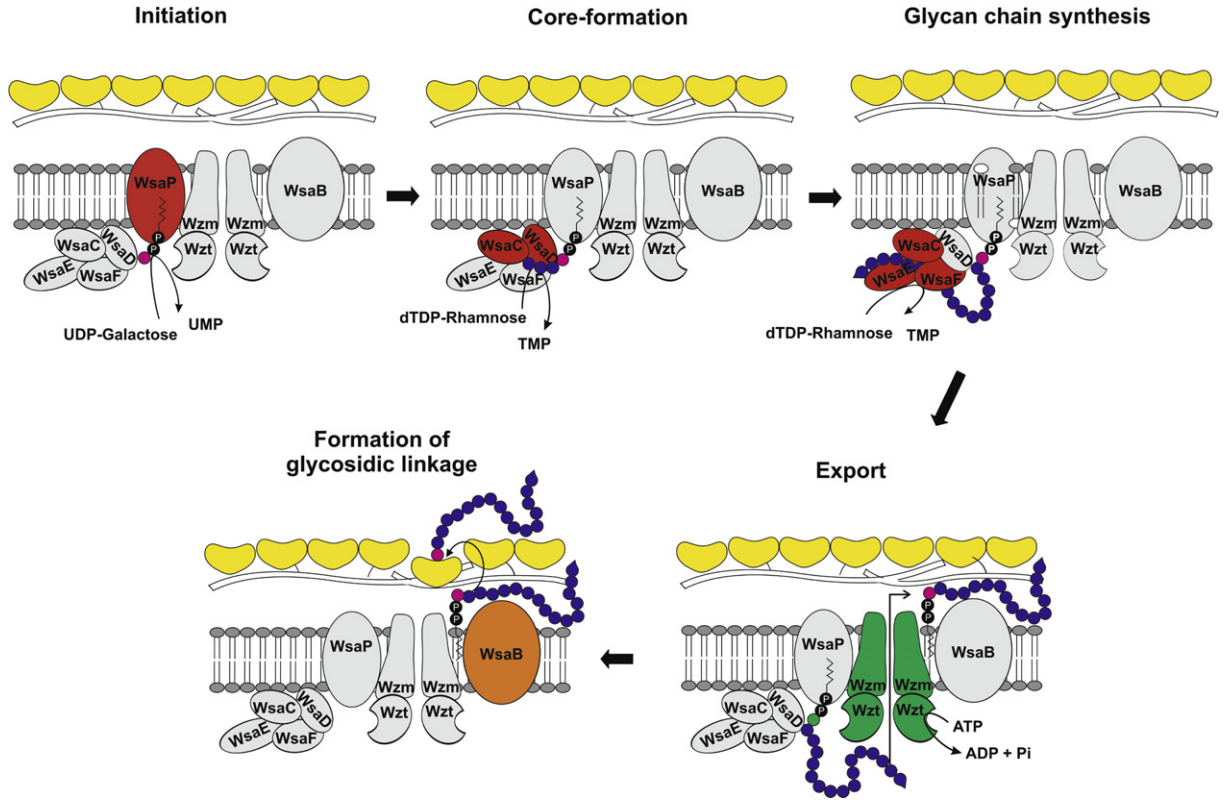


Figure 6. Schematic representation of the proposed biosynthesis route of S-layer glycoproteins, exemplified with *G. stearothermophilus* NRS 2004/3a. The identical color code for the components involved in S-layer glycoprotein glycan biosynthesis is used in Figures 4, 6 and in Table 1: light blue, monosaccharide biosynthesis; red, glycan assembly; orange, glycan transfer; green, export; gray, transposase; black, unknown function; dark blue, elongated glycan chain; pink, linkage glycose; yellow, S-layer protein.

protein synthesis effort of a bacterium may be devoted to S-layer protein production, it is conceivable that the S-layer protein glycosylation machinery has to be very effective. However, preliminary data indicate that the glycosylation event is lagging behind protein production (R. Novotny, A. Scheberl, and C. Schäffer, unpublished data). It is currently neither known whether S-layer glycosylation occurs co- or posttranslationally, nor whether the S-layer target protein is still unfolded or already prefolded when the oligosaccharyl::protein transferase exerts its activity.

Accumulated data on the S-layer nanoglycobiology of the model organism *G. stearothermophilus* NRS 2004/3a have led to the proposal of a pathway for S-layer protein O-glycosylation. The proposed pathway clearly benefits from the knowledge of the two principal routes of LPS O-antigen biosynthesis and of the *C. jejuni* protein N-glycosylation system.[44,49,100] O-Glycosylation of this S-layer glycoprotein generally seems to follow a similar pathway as is described for N-glycosylation of bacterial proteins.[101,102] The presence of a predicted ABC-2-type transporter system and the absence of a putative polymerase in the *slg* gene cluster of *G. stearothermophilus* NRS 2004/3a[83] indicate that the S-layer glycan chains are most probably synthesized in a process comparable to the ABC transporter-dependent pathway of LPS O-polysaccharide biosynthesis, although being initiated by a WbaP homologue instead of a WecA homologue, which usually serves as the initiation enzyme in that pathway.[57] Our current model implicates WsaP in the first step of synthesis whereby galactose is transferred from its nucleotide-activated form (UDP-Gal) to a membrane-associated lipid carrier at the cytoplasmic face of the plasma membrane[98] (Fig. 6). Chain extension presumably would continue in the cytoplasm by processive addition of rhamnose residues from dTDP-β-L-rhamnose to the non-reducing terminus of the lipid-linked glycan chain. Chain growth is predicted to be terminated by 2-O-methylation of the terminal repeating unit, catalyzed by an O-methyltransferase. A similar modification was recently described as chain length termination signal in the biosynthesis of O8 and O9 antigens.[62] The complete glycan chain would then be transported across the membrane by a process involving an ABC transporter and eventually transferred to the S-layer protein by the oligosaccharyl::protein transferase WsaB in a reaction comparable to that described recently for protein N-glycosylation of *C. jejuni*.[44,49,100]

As several other *slg* gene clusters contain an ABC-transporter (Fig. 4),[83] and numerous glycan chains are modified at the non-reducing end (Fig. 2),[60,61,103] the described model for S-layer protein glycosylation might be widely valid. However, the presence of a putative flippase Wzx in the *slg* gene cluster of *T. thermosaccharolyticum* E207-71 might be taken as an indication that S-layer glycoprotein glycans may also be synthesized via a route that is more related to the *wzy*-dependent O-antigen biosynthesis pathway.[57,83]

6. Conclusions

This review article covers recent advances made in the field of bacterial S-layer nanoglycobiology. From the accumulated data, important conclusions can be drawn for the future design of 'functional' glycans on S-layer proteins. The knowledge of the enzyme apparatus involved in S-layer glycoprotein glycan biosynthesis and the understanding of the underlying mechanisms should eventually allow the alteration or the rational design of S-layer protein glycosylation patterns to obtain bioactive S-layer *neo*glycoproteins. In principal, such compounds may be either used for nanobiotechnological bottom-up strategies to build nanoarrays or will be presented by the engineered bacteria in vivo. Controlled surface display of heterologous (glyco)proteins has become an increasingly used strategy in various applications in microbiology, nanobiotechnology, and vaccinology. The common trend of glycoengineering is reflected by several recent review articles on that topic.[104–109] Nanobiotechnology applications of tailored S-layer *neo*glycoproteins may include the fields of receptor mimics, vaccine design, or drug delivery using carbohydrate recognition.

Acknowledgments

We want to acknowledge our technicians Sonja Zayni and Andrea Scheberl, whose contributions to the advancement of S-layer nanoglycobiology by far exceeded routine work. Financial support came from the Austrian Science Fund, Projects P18013-B10 (to P.M.) and P19047-B12 (to C.S.), and the Hochschuljubiläumsstiftung der Stadt Wien, Project H-1809/2006 (to C.S.).

References

1. Houwink, A. L.; Le Poole, J. B. *Physikalische Verhandlungen* **1952**, *3*, 98.
2. Houwink, A. L. *Biochim. Biophys. Acta* **1953**, *10*, 360–366.
3. Sleytr, U. B. *Int. Rev. Cytol.* **1978**, *53*, 1–64.
4. Murray, R. G. E. In Guze, L. B., Ed.; Microbial Protoplasts, Spheroplasts and L-Forms; Williams & Wilkins: Baltimore, 1968; pp 1–18.
5. Glauert, A. M.; Thornley, M. J. *Annu. Rev. Microbiol.* **1969**, *23*, 159–198.
6. Kawata, T. *Electron Microsc.* **1981**, *15*, 115–122.
7. Wieland, F.; Lechner, J.; Bernhardt, G.; Sumper, M. *FEBS Lett.* **1981**, *132*, 319–323.

8. Sleytr, U. B.; Messner, P. *Annu. Rev. Microbiol.* **1983**, *37*, 311–339.
9. Kandler, O.; König, H. In *The Bacteria*; Woese, C. R., Wolfe, R. S., Eds.; Archaebacteria; Academic Press: New York, 1985; Vol. VIII, pp 413–457.
10. Koval, S. F.; Murray, R. G. E. *Microbiol. Sci.* **1986**, *3*, 357–361.
11. Smit, J. In *Bacterial Outer Membranes as Model Systems*; Inouye, M., Ed.; John Wiley & Sons: New York, 1987; pp 343–376.
12. Lechner, J.; Wieland, F. *Annu. Rev. Biochem.* **1989**, *58*, 173–194.
13. Kandler, O.; König, H. In *The Biochemistry of Archaea (Archaebacteria)*; Kates, M., Kushner, D. J., Matheson, A. T., Eds.; Elsevier: Amsterdam, 1993; pp 223–259.
14. Beveridge, T. J. *Curr. Opin. Struct. Biol.* **1994**, *4*, 204–212.
15. Sumper, M.; Wieland, F. T. In *Glycoproteins*; Montreuil, J., Vliegenthart, J. F. G., Schachter, H., Eds.; Elsevier: Amsterdam, 1995; pp 455–473.
16. Mescher, M. F.; Strominger, J. L. *J. Biol. Chem.* **1976**, *251*, 2005–2014.
17. Sleytr, U. B.; Thorne, K. J. I. *J. Bacteriol.* **1976**, *126*, 377–383.
18. Messner, P.; Schäffer, C. In *Progress in the Chemistry of Organic Natural Products*; Herz, W., Falk, H., Kirby, G. W., Eds.; Springer: Wien, 2003; Vol. 85, pp 51–124.
19. Woese, C. R.; Kandler, O.; Wheelis, M. L. *Proc. Natl. Acad. Sci. U.S.A.* **1990**, *87*, 4576–4579.
20. Tsuboi, A.; Uchihi, R.; Tabata, R.; Takahashi, Y.; Hasiba, H.; Sasaki, T.; Yamagata, H.; Tsukagoshi, N.; Udaka, S. *J. Bacteriol.* **1986**, *168*, 365–373.
21. Udaka, S.; Tsukagoshi, N.; Yamagata, H. *Biotechnol. Gen. Eng. Rev.* **1989**, *7*, 113–146.
22. Sleytr, U. B.; Messner, P.; Pum, D.; Sára, M. *Angew. Chem., Int. Ed.* **1999**, *38*, 1034–1054.
23. Sleytr, U. B.; Egelseer, E. M.; Ilk, N.; Pum, D.; Schuster, B. *FEBS J.* **2007**, *274*, 323–334.
24. Sleytr, U. B.; Huber, C.; Ilk, N.; Pum, D.; Schuster, B.; Egelseer, E. M. *FEMS Microbiol. Lett.* **2007**, *267*, 131–144.
25. Sleytr, U. B.; Sára, M.; Pum, D.; Schuster, B.; Messner, P.; Schäffer, C. In *Biopolymers*; Steinbüchel, A., Fahnestock, S. R., Eds.; Polyamides and Complex Proteinaceous Matrices I; Wiley-VCH: Weinheim, 2002; Vol. 7, pp 285–338.
26. Sleytr, U. B.; Sára, M.; Pum, D.; Schuster, B. In *Supramolecular Polymers*; Ciferri, A., Ed., 2nd ed.; Taylor & Francis: Boca Raton, FL, 2005; pp 583–616.
27. Sleytr, U. B.; Sára, M. *Trends Biotechnol.* **1997**, *15*, 20–26.
28. Benz, I.; Schmidt, M. A. *Mol. Microbiol.* **2002**, *45*, 267–276.
29. Sandercock, L. E.; MacLeod, A. M.; Ong, E.; Warren, R. A. J. *FEMS Microbiol. Lett.* **1994**, *118*, 1–8.
30. Logan, S. M. *Microbiology* **2006**, *152*, 1249–1262.
31. Szymanski, C. M.; Logan, S. M.; Linton, D.; Wren, B. W. *Trends Microbiol.* **2003**, *11*, 233–238.
32. Szymanski, C. M.; Wren, B. W. *Nature Rev. Microbiol.* **2005**, *3*, 225–237.
33. Schäffer, C.; Messner, P. *Biochimie* **2001**, *83*, 591–599.
34. Eichler, J.; Adams, M. W. W. *Microbiol. Mol. Biol. Rev.* **2005**, *69*, 393–425.
35. Lechner, J.; Sumper, M. *J. Biol. Chem.* **1987**, *262*, 9724–9729.
36. Sumper, M.; Berg, E.; Mengele, R.; Strobl, I. *J. Bacteriol.* **1990**, *172*, 7111–7118.
37. Kalmokoff, M. L.; Jarrell, K. F. *J. Bacteriol.* **1991**, *173*, 7113–7125.
38. Voisin, S.; Houliston, R. S.; Kelly, J.; Brisson, J.-R.; Watson, D.; Bardy, S. L.; Jarrell, K. F.; Logan, S. M. *J. Biol. Chem.* **2005**, *280*, 16586–16593.
39. Schäffer, C.; Messner, P. *Glycobiology* **2004**, *14*, 31R–42R.
40. Schäffer, C.; Wugeditsch, T.; Kählig, H.; Scheberl, A.; Zayni, S.; Messner, P. *J. Biol. Chem.* **2002**, *277*, 6230–6239.
41. Novotny, R.; Pfoestl, A.; Messner, P.; Schäffer, C. *Glyconjugate J.* **2004**, *20*, 435–447.
42. Apweiler, R.; Hermjakob, H.; Sharon, N. *Biochim. Biophys. Acta* **1999**, *1473*, 4–8.
43. Spiro, R. G. *Glycobiology* **2002**, *12*, 43R–56R.
44. Wacker, M.; Linton, D.; Hitchen, P. G.; Nita-Lazar, M.; Haslam, S. M.; North, S. J.; Panico, M.; Morris, H. R.; Dell, A.; Wren, B. W.; Aebi, M. *Science* **2002**, *298*, 1790–1793.
45. Feldman, M. F.; Wacker, M.; Hernandez, M.; Hitchen, P. G.; Marolda, C. L.; Kowarik, M.; Morris, H. R.; Dell, A.; Valvano, M. A.; Aebi, M. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 3016–3021.
46. Samuelson, P.; Gunneriusson, E.; Nygren, P.-Å.; Ståhl, S. *J. Biotechnol.* **2002**, *96*, 129–154.
47. Paton, A. W.; Morona, R.; Paton, J. C. *Nat. Med.* **2000**, *6*, 265–270.
48. Kowarik, M.; Young, N. M.; Numao, S.; Schulz, B. L.; Hug, I.; Callewaert, N.; Mills, D. C.; Watson, D. C.; Hernandez, M.; Kelly, J. F.; Wacker, M.; Aebi, M. *EMBO J.* **2006**, *25*, 1957–1966.
49. Wacker, M.; Feldman, M. F.; Callewaert, N.; Kowarik, M.; Clarke, B. R.; Pohl, N. L.; Hernandez, M.; Vines, E. D.; Valvano, M. A.; Whitfield, C.; Aebi, M. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 7088–7093.
50. Schäffer, C.; Novotny, R.; Küpcü, S.; Zayni, S.; Scheberl, A.; Friedmann, J.; Sleytr, U. B.; Messner, P. *Small* **2007**, *3*, 1549–1559.
51. Kandler, O. *Progr. Botany* **1993**, *54*, 1–24.
52. Kärcher, U.; Schröder, H.; Haslinger, E.; Allmaier, G.; Schreiner, R.; Wieland, F.; Haselbeck, A.; König, H. *J. Biol. Chem.* **1993**, *268*, 26821–26826.
53. Lee, S.-W.; Sabet, M.; Um, H.-S.; Yang, J.; Kim, H. C.; Zhu, W. *Gene* **2006**, *371*, 102–111.
54. Fletcher, C. M.; Coyne, M. J.; Bentley, D. L.; Villa, O. F.; Comstock, L. E. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 2413–2418.
55. Messner, P. *J. Bacteriol.* **2004**, *186*, 2517–2519.
56. Schäffer, C.; Graninger, M.; Messner, P. *Proteomics* **2001**, *1*, 248–261.
57. Raetz, C. R. H.; Whitfield, C. *Annu. Rev. Biochem.* **2002**, *71*, 635–700.
58. Wugeditsch, T.; Zachara, N. E.; Puchberger, M.; Kosma, P.; Gooley, A. A.; Messner, P. *Glycobiology* **1999**, *9*, 787–795.
59. Steiner, K.; Pohlentz, G.; Dreisewerd, K.; Berkenkamp, S.; Messner, P.; Peter-Katalinić, J.; Schäffer, C. *J. Bacteriol.* **2006**, *188*, 7914–7921.
60. Schäffer, C.; Müller, N.; Christian, R.; Graninger, M.; Wugeditsch, T.; Scheberl, A.; Messner, P. *Glycobiology* **1999**, *9*, 407–414.
61. Kählig, H.; Kolarich, D.; Zayni, S.; Scheberl, A.; Kosma, P.; Schäffer, C.; Messner, P. *J. Biol. Chem.* **2005**, *280*, 20292–20299.
62. Clarke, B. R.; Cuthbertson, L.; Whitfield, C. *J. Biol. Chem.* **2004**, *279*, 35709–35718.

63. Steiner, K. *Unravelling Details of the S-Layer Glycoprotein of Geobacillus stearothermophilus NRS 2004/3a*; Doctoral Thesis, University of Natural Resources and Applied Life Sciences: Vienna, Austria, 2006; 146 p.
64. Zayni, S.; Steiner, K.; Pfoestl, A.; Hofinger, A.; Kosma, P.; Schäffer, C.; Messner, P. *Glycobiology* **2007**, *17*, 433–443.
65. Schäffer, C.; Messner, P. *Microbiology* **2005**, *151*, 643–651.
66. Ferner-Ortner, J.; Mader, C.; Ilk, N.; Sleytr, U. B.; Egelseer, E. M. *J. Bacteriol.* **2007**, *189*, 7154–7158.
67. Rünzler, D.; Huber, C.; Moll, D.; Köhler, G.; Sára, M. *J. Biol. Chem.* **2004**, *279*, 5207–5215.
68. Altman, E.; Schäffer, C.; Brisson, J.-R.; Messner, P. *Eur. J. Biochem.* **1995**, *229*, 308–315.
69. Peter-Katalinić, J. *Methods Enzymol.* **2005**, *405*, 139–171.
70. Harvey, D. J. *Exp. Rev. Proteomics* **2005**, *2*, 87–101.
71. Hitchen, P. G.; Dell, A. *Microbiology* **2006**, *152*, 1575–1580.
72. Allmaier, G.; Schäffer, C.; Messner, P.; Rapp, U.; Mayer-Posner, F. J. *J. Bacteriol.* **1995**, *177*, 1402–1404.
73. Bindila, L.; Steiner, K.; Schäffer, C.; Messner, P.; Mormann, M.; Peter-Katalinić, J. *Anal. Chem.* **2007**, *79*, 3271–3279.
74. Küpcü, Z.; März, L.; Messner, P.; Sleytr, U. B. *FEBS Lett.* **1984**, *173*, 185–190.
75. Schäffer, C.; Franck, W. L.; Scheberl, A.; Kosma, P.; McDermott, T. R.; Messner, P. *Int. J. Syst. Evol. Microbiol.* **2004**, *54*, 2361–2368.
76. Kosma, P.; Neuninger, C.; Christian, R.; Schulz, G.; Messner, P. *Glycoconjugate J.* **1995**, *12*, 99–107.
77. Kosma, P.; Wugeditsch, T.; Christian, R.; Zayni, S.; Messner, P. *Glycobiology* **1995**, *5*, 791–796, Erratum: Glycobiology, **1996**, *6*, 5.
78. Yoshii, Z. *Proc. 6th Int. Congr. Electron Microsc., Kyoto* **1966**, *2*, 277.
79. Sára, M.; Moser-Thier, K.; Kainz, U.; Sleytr, U. B. *Arch. Microbiol.* **1990**, *153*, 209–214.
80. Altman, E.; Brisson, J.-R.; Messner, P.; Sleytr, U. B. *Biochem. Cell Biol.* **1991**, *69*, 72–78.
81. Messner, P.; Christian, R.; Neuninger, C.; Schulz, G. *J. Bacteriol.* **1995**, *177*, 2188–2193.
82. Whitfield, C. *Annu. Rev. Biochem.* **2006**, *75*, 39–68.
83. Novotny, R.; Schäffer, C.; Strauss, J.; Messner, P. *Microbiology* **2004**, *150*, 953–965.
84. Graninger, M.; Kneidinger, B.; Bruno, K.; Scheberl, A.; Messner, P. *Appl. Environ. Microbiol.* **2002**, *68*, 3708–3715.
85. Kneidinger, B.; Graninger, M.; Adam, G.; Puchberger, M.; Kosma, P.; Zayni, S.; Messner, P. *J. Biol. Chem.* **2001**, *276*, 5577–5583.
86. Pfoestl, A.; Hofinger, A.; Kosma, P.; Messner, P. *J. Biol. Chem.* **2003**, *278*, 26410–26417.
87. Kneidinger, B.; Graninger, M.; Puchberger, M.; Kosma, P.; Messner, P. *J. Biol. Chem.* **2001**, *276*, 20935–20944.
88. Kneidinger, B.; Marolda, C.; Graninger, M.; Zamyatina, A.; McArthur, F.; Kosma, P.; Valvano, M. A.; Messner, P. *J. Bacteriol.* **2002**, *184*, 363–369.
89. Pföstl, A.; Zayni, S.; Hofinger, A.; Kosma, P.; Schäffer, C.; Messner, P. *Biochemical J.* **2008**, *410*, 187–194.
90. Reeves, P. R.; Wang, L. *Curr. Top. Microbiol. Immunol.* **2002**, *264*, 109–135.
91. Jolly, L.; Stingele, F. *Int. Dairy J.* **2001**, *11*, 733–745.
92. Skurnik, M. *Adv. Exp. Med. Biol.* **2003**, *529*, 187–197.
93. Wang, L.; Romana, L. K.; Reeves, P. R. *Genetics* **1992**, *130*, 429–443.
94. Kaplan, J. B.; Perry, M. B.; MacLean, L. L.; Furgang, D.; Wilson, M. E.; Fine, D. H. *Infect. Immun.* **2001**, *69*, 5375–5384.
95. Reeves, P. R.; Hobbs, M.; Valvano, M. A.; Skurnik, M.; Whitfield, C.; Coplin, D.; Kido, N.; Klena, J.; Maskell, D.; Raetz, C. R. H.; Rick, P. D. *Trends Microbiol.* **1996**, *4*, 495–503.
96. Stein, A.; Kula, M.-R.; Elling, R.; Verseck, S.; Klaffke, W. *Angew. Chem., Int. Ed. Engl.* **1995**, *34*, 1748–1749.
97. Ma, Y.; Stern, R. J.; Scherman, M. S.; Vissa, V. D.; Yan, W.; Jones, V. C.; Zhang, F.; Franzblau, S. G.; Lewis, W. H.; McNeil, M. R. *Antimicrob. Agents Chemother.* **2001**, *45*, 1407–1416.
98. Steiner, K.; Novotny, R.; Patel, K.; Vinogradov, E.; Whitfield, C.; Valvano, M. A.; Messner, P.; Schäffer, C. *J. Bacteriol.* **2007**, *189*, 2590–2598.
99. Valvano, M. A. *Front. Biosci.* **2003**, *8*, 452–471.
100. Kowarik, M.; Numao, S.; Feldman, M. F.; Schulz, B. L.; Callewaert, N.; Kiermaier, E.; Catrein, I.; Aebi, M. *Science* **2006**, *314*, 1148–1150.
101. Linton, D.; Dorrell, N.; Hitchen, P. G.; Amber, S.; Karlyshev, A. V.; Morris, H. R.; Dell, A.; Valvano, M. A.; Aebi, M.; Wren, B. W. *Mol. Microbiol.* **2005**, *55*, 1695–1703.
102. Glover, K. J.; Weerapana, E.; Imperiali, B. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 14255–14259.
103. Bock, K.; Schuster-Kolbe, J.; Altman, E.; Allmaier, G.; Stahl, B.; Christian, R.; Sleytr, U. B.; Messner, P. *J. Biol. Chem.* **1994**, *269*, 7137–7144.
104. Endo, T.; Koizumi, S. *Curr. Opin. Struct. Biol.* **2000**, *5*, 536–541.
105. Mendez, C.; Salas, J. A. *Trends Biotechnol.* **2001**, *11*, 449–456.
106. Saxon, E.; Bertozzi, C. R. *Annu. Rev. Cell Dev. Biol.* **2001**, *17*, 1–23.
107. Desvaux, M.; Dumas, E.; Chafsey, I.; Hébraud, M. *FEMS Microbiol. Lett.* **2006**, *256*, 1–15.
108. Campbell, C. T.; Sampathkumar, S.-G.; Yarema, K. J. *Mol. BioSyst.* **2007**, *3*, 187–194.
109. Thibodeaux, C. J.; Melançon, C. E.; Liu, H.-w. *Nature* **2007**, *446*, 1008–1016.

THE JOURNAL OF BIOLOGICAL CHEMISTRY VOL. 283, NO. 30, pp. 21120–21133, July 25, 2008

Molecular Basis of S-layer Glycoprotein Glycan Biosynthesis in *Geobacillus stearothermophilus*[*] [S]

Received for publication, March 6, 2008, and in revised form, May 23, 2008 Published, JBC Papers in Press, May 30, 2008, DOI 10.1074/jbc.M801833200

Kerstin Steiner[‡1], René Novotny[‡2], Daniel B. Werz[§3], Kristof Zarschler[‡], Peter H. Seeberger[§], Andreas Hofinger[¶], Paul Kosma[¶], Christina Schäffer[‡4], and Paul Messner[‡5]

From the [‡]Center for NanoBiotechnology and the [¶]Department of Chemistry, University of Natural Resources and Applied Life Sciences, Wien, Austria, and the [§]Laboratory for Organic Chemistry, Swiss Federal Institute of Technology, Zürich, Switzerland

The Gram-positive bacterium *Geobacillus stearothermophilus* NRS 2004/3a possesses a cell wall containing an oblique surface layer (S-layer) composed of glycoprotein subunits. *O*-Glycans with the structure [→2)-α-L-Rha*p*-(1→3)-β-L-Rha*p*-(1→2)-α-L-Rha*p*-(1→]$_{n = 13–18}$, a 2-*O*-methyl group capping the terminal repeating unit at the nonreducing end and a →2)-α-L-Rha*p*- [(1→3)-α-L-Rha*p*]$_{n = 1–2}$(1→3)- adaptor are linked via a β-D-Gal*p* residue to distinct sites of the S-layer protein SgsE. S-layer glycan biosynthesis is encoded by a polycistronic *slg* (surface layer glycosylation) gene cluster. Four assigned glycosyltransferases named WsaC–WsaF, were investigated by a combined biochemical and NMR approach, starting from synthetic octyl-linked saccharide precursors. We demonstrate that three of the enzymes are rhamnosyltransferases that are responsible for the transfer of L-rhamnose from a dTDP-β-L-Rha precursor to the nascent S-layer glycan, catalyzing the formation of the α1,3- (WsaC and WsaD) and β1,2-linkages (WsaF) present in the adaptor saccharide and in the repeating units of the mature S-layer glycan, respectively. These enzymes work in concert with a multifunctional methylrhamnosyltransferase (WsaE). The N-terminal portion of WsaE is responsible for the *S*-adenosylmethionine-dependent methylation reaction of the terminal α1,3-linked L-rhamnose residue, and the central and C-terminal portions are involved in the transfer of L-rhamnose from dTDP-β-L-rhamnose to the adaptor saccharide to form the α1,2- and α1,3-linkages during S-layer glycan chain elongation, with the methylation and the glycosylation reactions occurring independently. Characterization of these enzymes thus reveals the complete molecular basis for S-layer glycan biosynthesis.

Glycosylation is the most common posttranslational modification of proteins (1–3), playing important roles in all living organisms (4–7). Although glycosylation has popularly been considered to be restricted to eukaryotes, prokaryotic glycosylation is an emerging field that has opened new avenues in basic and applied research. Among the best investigated prokaryotic glycoproteins are surface layer (S-layer)[6] glycoproteins of *Bacillaceae* (11, 12). Due to the ability of native and recombinant S-layer (glyco)proteins to self-assemble into two-dimensional arrays in solution and on various supports (*e.g.* silica, polymers, liposomes, lipid films, or membranous structures), they are considered promising candidates for *in vivo* and *in vitro* glycan display approaches with nanometer scale periodicity (13–15). Understanding the basic principles of prokaryotic protein glycosylation is a prerequisite for engineering of tailor-made ("functional") glycan motifs on S-layer proteins for applications in microbiology, nanobiotechnology, and vaccinology (8–10).

The S-layer glycoprotein of the Gram-positive organism *Geobacillus stearothermophilus* NRS 2004/3a serves as a model system for elucidating the basic principles of prokaryotic protein glycosylation. The S-layer glycan with the structure, 2-*O*Me-α-L-Rha*p*-(1→3)-β-L-Rha*p*-(1→2)-α-L-Rha*p*-(1→[2)-α-L-Rha*p*-(1→3)-β-L-Rha*p*-(1→2)-α-L-Rha*p*-(1→]$_{n = 13–18}$ 2)-α-L-Rha*p*-(1→[3)-α-L-Rha*p*]$_{n = 1–2}$-(1→3)-β-D-Gal*p*-(1→ is *O*-glycosidically linked to threonine 590, threonine 620, and serine 794 of the S-layer protein SgsE (16, 17).

Based on sequence comparisons, we hypothesize that the genetic information for the enzymes involved in S-layer glycan biosynthesis of *G. stearothermophilus* NRS 2004/3a is organized in a polycistronic *slg* (S-layer glycosylation) gene cluster (GenBank™ accession number AF328862) (18). Besides four previously uncharacterized, putative glycosyltransferases, which are the focus of the present study, the *slg* gene cluster encodes enzymes catalyzing the biosynthesis of dTDP-β-L-rh-

* This work was supported by Austrian Science Fund Projects P19047-B12 (to C. S.) and P18013-B10 (to P. M.), Hochschuljubiläumsstiftung der Stadt Wien Project H-1809/2006 (to C. S.), ETH Zürich, and the Swiss National Science Foundation (to P. H. S.). The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

[S] The on-line version of this article (available at http://www.jbc.org) contains supplemental Tables S1–S3 and Figs. S1–S7.

[1] Present address: Centre of Biomolecular Sciences, University of St. Andrews, St. Andrews, UK.

[2] Present address: Institute for Applied Genetics and Cell Biology, University of Natural Resources and Applied Life Sciences (Wien, Austria).

[3] Recipient of a Deutsche Forschungsgemeinschaft Emmy Noether Fellowship. Present address: Institute for Organic and Biomolecular Chemistry, Georg-August-Universität (Göttingen, Germany).

[4] To whom correspondence may be addressed. Tel.: 43-1-47654-2203; Fax: 43-1-4789112; E-mail: christina.schaeffer@boku.ac.at.

[5] To whom correspondence may be addressed. Tel.: 43-1-47654-2202; Fax: 43-1-4789112; E-mail: paul.messner@boku.ac.at.

[6] The abbreviations used are: S-layer, bacterial cell surface layer; ABC, ATP-binding cassette; CHAPS, 3-[(3-cholamidopropyl)dimethylammonio]-1-propanesulfonate; COSY, correlation spectroscopy; dTDP-β-L-Rha, dTDP-β-L-rhamnose; ESI-QTOF, electrospray ionization quadrupole time-of-fight; MS, mass spectrometry; Gal*p*, galactopyranose; Gal, galactose; LDAO, lauryldimethylamine *N*-oxide; LPS, lipopolysaccharide; *O*-PS, *O*-polysaccharide; MS^2, MS/MS; Rha*p*, rhamnopyranose; SAM, S-adenosylmethionine; und-P, undecaprenyl phosphate; und-PP, undecaprenylpyrophosphate.

amnose, ABC transporter components, an initiating glycosyltransferase (WsaP), and an oligosaccharyltransferase (WsaB). Additionally, UDP-Gal that is required for S-layer glycan formation is derived from the general metabolism of the bacterium (19). Generally, *slg* gene clusters show high homology to gene clusters involved in *O*-polysaccharide (*O*-PS) biosynthesis of Gram-negative bacteria (12, 20). The presence of a homopolymeric S-layer glycan chain (poly-L-rhamnan) in *G. stearothermophilus* NRS 2004/3a and the occurrence of the *wzm*/*wzt* genes encoding ABC transporter components in the *slg* gene cluster indicate that the biosynthesis of this S-layer glycan may follow a route similar to the ABC transporter-dependent pathway of *O*-PS chains (21). On the other hand, with the functional characterization of the initiation enzyme WsaP (a WbaP homologue) of S-layer glycan biosynthesis as a Gal-1-phosphate:lipid carrier transferase, it is evident that also a biosynthesis module from the ABC transporter-independent *O*-PS biosynthesis route is utilized (22). More precisely, in *G. stearothermophilus* NRS 2004/3a, WsaP replaces the GlcNAc-1-phosphate:undecaprenyl phosphate (und-P) transferase activity of WecA, which is the characteristic initiation reaction of ABC transporter-dependent pathways of *O*-PS biosynthesis (23).

In this report, we demonstrate the function of the four assigned glycosyltransferases of the *slg* gene cluster of *G. stearothermophilus* NRS 2004/3a, named WsaC–WsaF, which should be responsible for the consecutive transfer of rhamnose residues from a nucleotide diphosphate-activated precursor to the growing S-layer glycan chain while catalyzing the formation of the α1,2-, α1,3-, and β1,2-linkages. In this context, we identify WsaE as a novel multifunctional enzyme, catalyzing not only the consecutive transfer of L-rhamnose residues from dTDP-β-L-rhamnose to the adaptor saccharide of the S-layer glycan chain but also the *S*-adenosylmethionine (SAM)-dependent methylation reaction of the terminal α-L-rhamnose residue of the glycan chain. It is shown that the glycosylation reaction and the methylation reaction occur independently. Further we show that the four glycosyltransferases working in concert are sufficient for S-layer poly-L-rhamnan biosynthesis in *G. stearothermophilus* NRS 2004/3a.

EXPERIMENTAL PROCEDURES

Bacterial Strains and Growth Conditions—Bacterial strains and plasmids are listed in Table S1. *G. stearothermophilus* NRS 2004/3a was grown in S-VIII medium at 55 °C (16) and *Escherichia coli* was grown in Luria-Bertani broth at 37 °C. Growth media were supplemented with kanamycin (50 μg/ml), when appropriate.

Sequence Analysis—Protein sequences were analyzed using the BLASTP on-line sequence homology analysis tools (National Center for Biotechnology Information, Bethesda, MD). Putative transmembrane helices were identified with the TMHMM2.0 program (Center for Biological Sequence Analysis, Lyngby, Denmark). Sequence alignments were carried out with the ClustalW program (available on the World Wide Web).

Construction of Plasmids—Various plasmids for enzyme overexpression in *E. coli* were constructed (Tables S1 and S2). To obtain proteins with an N-terminal His_6 tag, the entire coding sequences of WsaC, WsaD, WsaE, WsaF, RmlB, RmlC, and RmlD and/or truncated versions thereof were amplified by PCR with the primer pairs pET-WsaC_for/pET-WsaC_rev (WsaC), pET-WsaC_for/pET-WsaC_I_rev (WsaC_I), pET-WsaD_for/pET-WsaD_rev (WsaD), pET-WsaD_for/pET-WsaD_I_rev (WsaD_I), pET-WsaE_for/pET-WsaE_rev (WsaE), pET-WsaE_for/pET-WsaE_M_rev (WsaE_M), pET-WsaE_for/pET-WsaE_N_rev (WsaE_N), pET-WsaE_B_for/pET-WsaE_rev (WsaE_B), pET-WsaE_C_for/pET-WsaE_rev (WsaE_C), pET-WsaE_B_for/pET-WsaE_A_rev (WsaE_A), pET-WsaF_for/pET-WsaF_rev (WsaF), RmlB_for/RmlB_rev (RmlB), RmlC_for/RmlC_rev (RmlC), and RmlD_for/RmlD_rev (RmlD), respectively (Table S2). Amplification products were digested with NdeI/XhoI (WsaC, WsaD, RmlC, and RmlD), NheI/SstI (WsaF), or NheI/XhoI (WsaE and RmlB) and inserted into the dephosphorylated expression vector pET28a (Novagen, Madison, WI), which was linearized with the same restriction enzymes. The resulting plasmids were named pNGB220 (pET28a-WsaC) and pNGB221 (pET28a-WsaC_I), pNGB230 (pET28a-WsaD) and pNGB231 (pET28a-WsaD_I), pNGB240 (pET28a-WsaE), pNGB241 (pET28a-WsaE_M), pNGB242 (pET28a-WsaE_N), pNGB243 (pET28a-WsaE_B), pNGB244 (pET28a-WsaE_C), pNGB245 (pET28a-WsaE_A), pNGB250 (pET28a-WsaF), pNGB261 (pET28a-RmlB), pNGB262 (pET28a-RmlC), and pNGB263 (pET28a-RmlD) and were transformed into the *E. coli* expression host BL21 Star (DE3).

Protein Overexpression—Expression of recombinant proteins in *E. coli* BL21 Star (DE3) was initiated by the addition of 1 mM isopropyl-β-D-thiogalactopyranoside to $A_{600} \sim 0.8$ cultures, and cultivation was continued for an additional 4 h. Expression of recombinant proteins was monitored by SDS-PAGE (24). Protein bands were visualized with Coomassie Blue R-250 staining reagent. Semidry blotting of proteins to a polyvinylidene difluoride membrane (Bio-Rad) was performed as described previously (22). Development of the blot with anti-His tag monoclonal antibody (Novagen) was done according to the manufacturer's instructions.

Purification of Recombinant Proteins—Biomass from 400 ml of *E. coli* BL21Star (DE3) expression cultures was harvested by centrifugation (15 min, 4,500 × *g*, 4 °C). The pellet was washed with 40 ml of cold saline and resuspended in 10 ml of cold buffer A (50 mM ammonium/acetate, pH 7.5, 1 mM EDTA). Cells were lysed by ultrasonication (Branson sonifier 450; Branson, Danbury, CT; duty cycle 60%, output 6), applying eight cycles of 10 pulses with 1-min breaks, each. Unlysed cells were removed by centrifugation (4,500 × *g*, 15 min, 4 °C), and the membrane fraction was separated from the cell-free lysate by ultracentrifugation (200,000 × *g*, 60 min, 4 °C), followed by a washing step with buffer A. Soluble enzymes (WsaE, WsaF, RmlB, RmlC, and RmlD) were directly purified on 1-ml HisTrap™ HP prepacked columns (GE Healthcare), using an Amersham Biosciences FPLC™ system (flow rate 1 ml/min). The column was equilibrated in His A buffer (20 mM sodium phosphate, 0.5 M NaCl, 20 mM imidazole, pH 7.4). Proteins were eluted using a step gradient up to 0.5 M imidazole in His A buffer. Isolated membranes harboring WsaC or WsaD were extracted on ice for 3 h to overnight with 2% CHAPS or LDAO in His A buffer (Sigma). After

centrifugation at 20,800 × *g* for 1 h, the supernatants of WsaC and WsaD were purified by HisTrap™ HP chromatography as described above. The addition of 0.2% CHAPS or 0.1 mM LDAO to all buffers was necessary to keep the proteins in solution. Fractions were analyzed by SDS-PAGE, and the protein content of the pools was determined using the Bio-Rad Bradford reagent.

Biosynthesis of dTDP-β-L-Rhamnose— dTDP-β-L-rhamnose was synthesized from dTDP-glucose using purified RmlB, RmlC, and RmlD enzymes as described previously (25).

Chemical Synthesis of Octyl Saccharide Substrates—For the functional enzyme assays, the following synthetic octyl-linked saccharides were used (Fig. S1): β-D-Gal-(1→*O*)-octyl (M_r 292.37) (**I**), α-L-Rha-(1→3)-β-D-Gal-(1→*O*)-octyl (M_r = 438.51) (**II**), and α-L-Rha-(1→3)-α-L-Rha-(1→3)-β-D-Gal-(1→*O*)-octyl (M_r = 584.65) (**III**). The chemical synthesis of (**I–III**) was based on commonly used monosaccharide building blocks (26–29).

Rhamnosyltransferase Assay with Octyl Saccharides—The synthetic substrates were dissolved in 2-propanol (**I** and **III**) or methanol (**II**) at a concentration of 20 nmol/μl (0.29 mg/50 μl (**I**), 0.44 mg/50 μl (**II**), and 0.28 mg/25 μl (**III**)). Reaction products of rhamnosyltransferases used as substrates in subsequent assays were dissolved in water. For the assay, either lysed cells (WsaC and WsaD) or cell-free extract and purified protein (WsaE and WsaF), respectively, were used. The reaction mixture contained 1 μl (20 nmol) of substrate, 2 μl (40 nmol) of dTDP-β-L-Rha, the relevant transferases, and 10 μl of 100 mM $MnCl_2$ in a final volume of 100 μl of buffer A (50 mM ammonium acetate, pH 7.5, 1 mM EDTA). Incubation was performed at 37 °C for 1 h, and the reaction was stopped by the addition of 1.2 ml of chloroform/methanol (3:2) followed by extraction for 20 min. After removal of insoluble material, the supernatant was dried under a stream of nitrogen. Samples were resuspended in 10 μl of chloroform/methanol (3:2) and analyzed by TLC. Octyl-linked products were separated on silica Gel 60 aluminum TLC plates (20 × 20 cm; thickness, 0.25 mm; Merck) using the solvent system chloroform/methanol/water (65:25:4). Carbohydrates were detected with thymol reagent (30).

Methyltransferase Activity Assay—This assay was based on the rhamnosyltransferase activity assay. Different forms of WsaE (WsaE, WsaE_M and WsaE_N) were mixed with dTDP-β-L-Rha, octyl substrates (α-L-Rha-(1→3)-α-L-Rha-(1→3)-β-D-Gal-(1→*O*)-octyl (**III**), α-L-Rha-(1→2)-α-L-Rha-(1→3)-α-L-Rha-(1→3)-β-D-Gal-(1→*O*)-octyl (**VII**), or α-L-Rha-(1→3)-β-L-Rha-(1→2)-α-L-Rha-(1→3)-β-D-Gal-(1→*O*)-octyl (**IX**)), and $MnCl_2$, and 1 μl of labeled *S*-[*methyl*-^{3}H]adenosylmethionine (10 Ci/mmol; PerkinElmer Life Sciences) was added to a final volume of 100 μl. Incubation was performed at 37 °C for 1 h, and the reaction products were separated by TLC. For autoradiography, the TLC plates were exposed to an Eastman Kodak Co. BioMax MS film (Sigma) for 2 days at −70 °C.

ESI-QTOF MS—For MS analysis of the rhamnosyltransferase reaction products, the assays were scaled up 10-fold. When purified enzyme was used (WsaE and WsaF), the reaction products were directly applied to preparative TLC. When WsaC was used, octyl-linked reaction products were extracted and hydrolyzed under alkaline conditions to destroy phospholipids. Subsequently, the products were dissolved in chloroform/methanol (2:1), mixed with 8 parts of ethanol and 1 part of 1 M NaOH (final concentration 0.1 M NaOH, pH 12), and incubated for 30 min at 37 °C (31). After the addition of ethyl acetate, alkali-stable products were extracted with chloroform/methanol, as described for the enzyme assay, and analyzed by TLC. For removal of remaining substrates, the samples were applied to preparative TLC. One of the lanes was stained with thymol, and the products were scratched off of the unstained plate at the corresponding height and extracted with 1 ml of chloroform/methanol (3:2), chloroform/methanol (1:2), methanol, and methanol/water (1:1), respectively. After removal of the silica material by centrifugation, the supernatant was dried under a stream of nitrogen and analyzed by ESI-QTOF MS.

For MS analysis of the products from the methyltransferase reaction, the assay was scaled up 20-fold using HisTrap-purified methyltransferase and unlabeled SAM (Sigma). After completion of the reaction, the protein was removed using Microcon® centrifugal filter units (M_r cut-off 10,000; Millipore). The sample was desalted using RP-18 Sep-Pak cartridges (Waters, Milford, MA). After activating the cartridge with 50% acetonitrile, it was equilibrated with MilliQ water, and the sample dissolved in 100 μl of MilliQ water was applied. After washing with 3 ml of MilliQ water, the product was eluted with 1 ml of 25, 50, 75, and 100% acetonitrile, respectively. Fractions of interest were combined and dried under a stream of nitrogen.

Positive mode mass spectrometry was performed on a Waters Micromass Q-TOF Ultima Global apparatus (Waters Micromass, Manchester, UK). Samples were subjected to off-line infusion ESI-QTOF MS after dilution in 75% methanol containing 0.1% formic acid to a final concentration of about 50 pmol/μl. Spectra acquisition was performed using 2–3 kV capillary and 100 V cone voltage. Desolvation gas flow was set at 450 liters/h and cone gas at 50 liters/h. Samples were injected at a flow rate of 3–5 μl/min. The instrument was controlled by MassLynx 4.0 software (Waters Micromass).

Sample Preparation for NMR Studies—All large scale reaction mixtures contained 10 mM $MnCl_2$ and were incubated overnight. Lysed cells containing WsaC (8.5 ml) were incubated with 5 mg (8.5 μmol) of α-L-Rha-(1→3)-α-L-Rha-(1→3)-β-D-Gal-(1→*O*)-octyl (**III**) and 15.5 μmol of dTDP-β-L-Rha. Purified WsaF (6 ml) was either reacted with 3 mg (6.8 μmol) of α-L-Rha-(1→3)-β-D-Gal-(1→*O*)-octyl (**II**) and 10 μmol of dTDP-β-L-Rha, or WsaF (4.3 ml) was reacted with 0.7 mg of the α-L-Rha-(1→2)-α-L-Rha-(1→3)-α-L-Rha-(1→3)-β-D-Gal-(1→*O*)-octyl (**VII**) reaction product of WsaE and 2 μmol of dTDP-β-L-Rha. For investigation of the α1,2-rhamnosyltransferase activity of WsaE, 8.3 ml of cell-free supernatant of *E. coli* expressing WsaE was incubated with 5 mg (8.5 μmol) of α-L-Rha-(1→3)-α-L-Rha-(1→3)-β-D-Gal-(1→*O*)-octyl (**III**) dissolved in 200 μl of methanol together with 510 μl of dTDP-β-L-Rha (10.2 μmol). For identification of α1,3-rhamnosyltransferase activity, 4.3 ml of WsaE-containing lysate were reacted with 0.5 mg of the β-L-Rha-(1→2)-α-L-Rha-(1→3)-β-D-Gal-(1→*O*)-octyl (**IV**) reaction product of WsaF and 1.5 μmol of dTDP-β-L-Rha. Membranes and aggregates were removed by ultracentrifugation at 200,000 × *g* for 45 min, and the pellet was reextracted with methanol. This fraction was

combined with the supernatant and centrifuged using Microcon® centrifugal filter units (M_r cut-off 10,000) to remove the proteins. The filtrate was dried and analyzed by TLC, as described above. Dried reaction mixtures were resuspended in 0.5 ml of 20% methanol and applied to a Sephadex G10 column (1.5 × 117 cm) with 20% methanol as eluent. Elution was monitored by spotting 1 μl of each fraction on a TLC plate followed by staining with thymol reagent. Fractions positive for carbohydrate were analyzed by TLC, and the fractions containing the products were pooled, dried in a SpeedVac vacuum concentrator, and desalted using RP-18 Sep-Pak Cartridges (Waters) with a discontinuous gradient of acetonitrile (25, 50, 75, and 100% acetonitrile) for elution of the reaction products.

NMR Spectroscopy—All NMR spectra were recorded at 297 K on a Bruker DPX 400 instrument at 400.13 MHz for proton and 100.62 MHz for carbon using a 5-mm broad band probe with *z*-gradients. The sample contained freeze-dried material dissolved in 600 μl of D_2O in a 5-mm NMR tube. The ppm scales were calibrated to 2,2-dimethyl-2-silapentane-5-sulfonic acid for proton (0.00 ppm) and external 1,4-dioxane for carbon (67.40 ppm). The NMR spectra were recorded using Bruker standard homonuclear pulse programs such as COSY and mlevph. For heteronuclear experiments, Bruker standard pulse programs were used as heteronuclear single quantum coherence spectroscopy for one-bond proton-carbon correlations and heteronuclear multiple bond correlation spectroscopy for multiple-bond proton-carbon correlations. All homonuclear and heteronuclear spectra were recorded with 1K × 1K data points and zero-filled in both dimensions.

Rhamnosyltransferase Assay with E. coli Membranes Harboring WsaP—E. coli C43 (DE3) cells harboring either pET28_WsaP, coding for the Gal-1-phosphate:lipid carrier transferase WsaP of *G. stearothermophilus* NRS 2004/3a (22), or pET28a_WsaD were lysed separately. For the assay, 88 μl of each or both lysates were mixed with 100 mM $MnCl_2$ (final concentration 10 mM), 84 pmol (25 nCi) of radiolabeled UDP-D-[^{14}C]galactose (GE Healthcare), and, optionally, 20 nmol of dTDP-β-L-Rha. The reaction was performed for 20 h at 37 °C and terminated by the addition of 1.25 ml of chloroform/methanol (3:2) (32). After shaking for 20 min, the samples were centrifuged, and the organic phase was transferred to a new tube. The organic phase was washed with 150 μl of 40 mM $MgCl_2$. The upper phase obtained after centrifugation was removed, and the lower organic phase was washed with 400 μl of pure solvent upper phase chloroform/methanol/water/1 M $MgCl_2$ (18:294:282:1). The organic phase containing the lipid-linked reaction products was dried under a stream of nitrogen, resuspended in 10 μl of chloroform/methanol (3:2), and applied to TLC and autoradiography as described above.

RESULTS

Sequence Analysis of the Transferases WsaC, WsaD, WsaE, and WsaF—Considering that the S-layer glycan of *G. stearothermophilus* NRS 2004/3a contains α- and β-linked L-rhamnose residues and the putative donor required for rhamnosyltransferases would be dTDP-β-L-rhamnose, a data base search for retaining and inverting glycosyltransferases was performed. Furthermore, the presence of a methylated terminal rhamnose residue let us hypothesize that a SAM-dependent methyltransferase may be present in the *slg* gene cluster of this organism. The results of the various BLAST searches revealed the presence of GT-2 and GT-4 transferases in the *slg* gene cluster. During our previous annotations of this gene cluster, we named the glycosyltransferase-encoding genes *wsaC*, *wsaD*, *wsaE*, and *wsaF* (12). The predicted primary structures of the encoded WsaC, WsaD, and WsaF proteins reveal proteins of the range of 300–400 amino acids, which is an expected size for a glycosyltransferase catalyzing a single reaction; these sequences also contain the D*X*D signature motif, which is commonly found in both inverting (*i.e.* inverting the stereochemistry of the donor) and retaining (*i.e.* retaining the stereochemistry of the donor) glycosyltransferases.

WsaC (protein accession number AAR99605) is assigned to the inverting transferases of the GT-2 family (33). An aspartate residue, Asp^{39}, in the model protein SpsA (34), potentially corresponding to Asp^{40}-Asp^{41} in WsaC, may coordinate N-3 of the uracil base, providing specificity for U and T as the sugar donor. In the so-called D*X*DD motif (DQDD; aa 96–99), the middle Asp residue binds the hydroxyl groups on the ribose moiety, whereas the third Asp residue binds a divalent metal ion. An additional Asp residue at position 191 is thought to function as a catalytic base activating the acceptor for nucleophilic attack at C-1 by deprotonation (35). As the substrate for WsaC is dTDP-β-L-Rha, an inverting transferase would be an α-rhamnosyltransferase. In the case of WsaD (AAR99614), there is homology to the inverting α-rhamnosyltransferases of *Shigella flexneri*, specifically matching the N-terminal (V/I)*X*(V/I)*XDX*$_2$S signature motif (36). Members of GT-2 family have a highly conserved N-terminal domain and a less conserved C terminus, due to the differences in substrate specificity (Fig. S2). The topology model of both the WsaC and the WsaD protein predicts a single, transmembrane-spanning domain at the C-terminal region. The predicted cytoplasmic WsaF (AAR99609) protein contains EX_7E motifs typically found in retaining glycosyltransferases of the GT-4 family (37). Since WsaF is putatively a retaining glycosyltransferase, we hypothesize that it would be responsible for the formation of the β1,2-linkage in the repeating unit trisaccharide of the *G. stearothermophilus* NRS 2004/3a S-layer protein glycan.

Finally, the *wsaE* gene encodes a predicted cytoplasmic protein of 1,127 amino acids (AAR99608); data base analysis of its sequence suggested that it may contain three functional domains (Fig. 1). The N-terminal portion of WsaE (aa 70–150) revealed homology to SAM-dependent methyltransferases (Fig. S3), whereas the central and C-terminal portions contain a glycosyltransferase domain, each belonging to the inverting GT-2 family (aa 604–765 and 863–1,043). The first glycosyltransferase domain contains a conserved DD motif and the D*X*DD motif. The second glycosyltransferase domain contains ED motifs and a D*X*E motif (Fig. S4).

Expression, Localization, and Purification of the Transferases WsaC, WsaD, WsaE, and WsaF—N-terminally His_6-tagged forms of WsaC, WsaD, and WsaE and truncated versions thereof as well as WsaF were produced in *E. coli* BL21 Star (DE3). According to the Western immunoblot evidence (Fig. S5), both WsaC and WsaD were smaller than expected, corre-

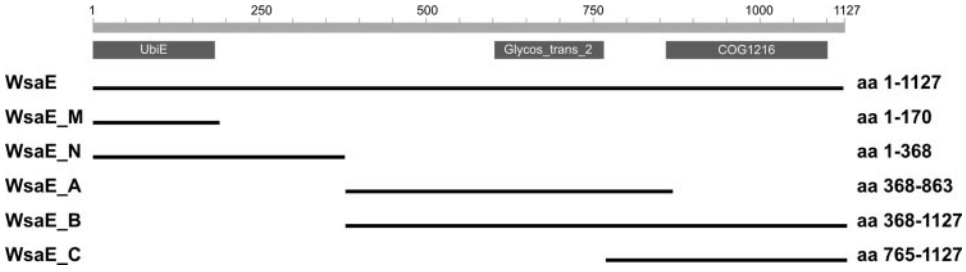


FIGURE 1. **Predicted topology of the WsaE protein of *G. stearothermophilus* NRS 2004/3a as a basis for designing different forms of WsaE to assess functional domains of the enzyme.** *WsaE*, aa 1-1,127, full size; *WsaE_M*, aa 1–170, including the C terminus with the UbiE motive; *WsaE_N*, aa 1–368, including the C terminus with the UbiE motive and additional 200 aa; *WsaE_A*, aa 368–863, including the first rhamnosyltransferase domain; *WsaE_B*, aa 368–1127, including both rhamnosyltransferase domains; *WsaE_C*, aa 765–1127, including the second rhamnosyltransferase domain.

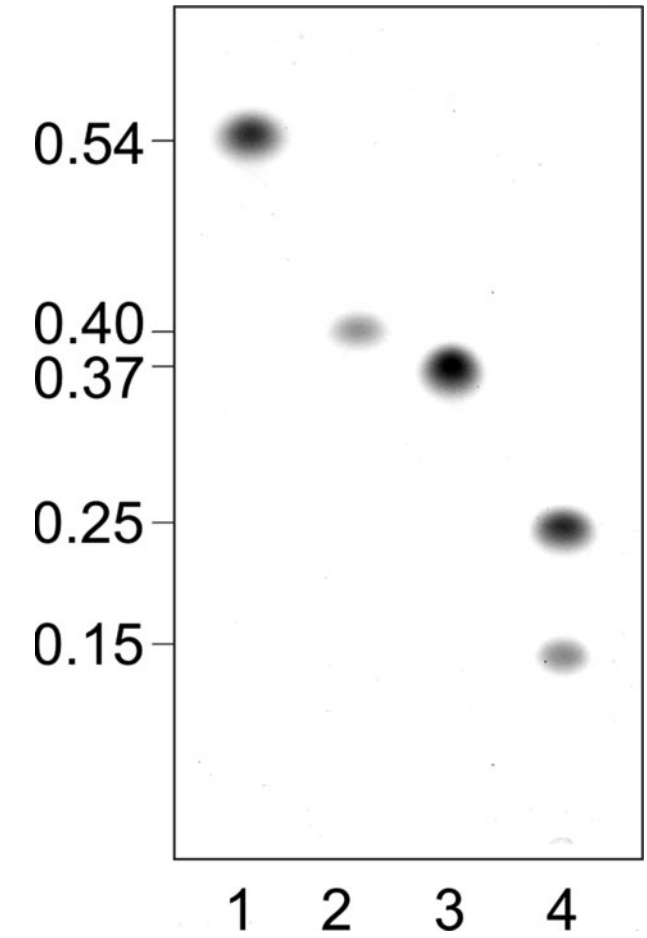


FIGURE 2. **TLC pattern of purified products from WsaC and WsaF activity assays using octyl-linked oligosaccharides.** Products were separated on silica TLC plates with chloroform/methanol/water (65:25:4) as solvent. *Lane 1*, substrate (**II**) α-L-Rha-(1→3)-β-D-Gal-(1→*O*)-octyl; *lane 2*, product of WsaF (**IV**); *lane 3*, substrate (**III**) α-L-Rha-(1→3)-α-L-Rha-(1→3)-β-D-Gal-(1→*O*)-octyl; *lane 4*, products of WsaC (**V** and **VI**). Staining was performed with thymol reagent.

sponding to proteins of ~37 and 30 kDa, respectively (Fig. S5*A*, *lanes 2* and *4*), as opposed to the calculated molecular masses of 40 kDa (WsaC) and 35.6 kDa (WsaD). A similar migration behavior has already been reported for the O-antigen biosynthesis enzyme WbbE from *Salmonella enterica*, for which also one transmembrane domain is predicted (38). The influence of the protein transmembrane domain resulting in too low molecular mass estimates was confirmed through analysis of C-terminally truncated forms of WsaC (WsaC_I; Fig. S5A, *lane 3*) and WsaD (WsaD_I; Fig. S5*A*, *lane 5*), both of which were devoid of the transmembrane region and could be detected at the calculated size on the Western immunoblot. WsaF (Fig. S5*A*, *lane 7*) as well as WsaE and its truncated forms (for details about the truncations, see Fig. 1 and Fig. S5*B*, *lanes 2–6*) and WsaF migrated at the correct size. This supports the topology models predicting WsaC and WsaD to be transmembrane proteins and WsaE and WsaF to be soluble proteins.

The cytoplasmic proteins (WsaC_I, WsaD_I, WsaE, and WsaF) could be directly purified by nickel affinity chromatography, whereas full-size WsaC and WsaD were shown to be accumulated in the membrane fraction obtained from lysed cells of *E. coli* BL21 Star (DE3) expression cultures (data not shown). After numerous attempts with different detergents (Triton X-100, octyl glucopyranoside, CHAPS, and LDAO) at various concentrations, WsaC and WsaD could be purified by using 2% CHAPS or LDAO for membrane extraction and the addition of detergent to all buffers used for nickel affinity chromatography. Detergent was necessary to keep the purified enzymes in solution.

Substrate Specificity of the Rhamnosyltransferases WsaC, WsaD, WsaF, and WsaE—Due to our hypothesis that the four predicted proteins WsaC-F would possess rhamnosyltransferase activity, we tested the activity of the recombinant forms of these proteins with dTDP-β-L-Rha (as donor) and various octyl-linked saccharides (as acceptors). The following conclusions regarding substrate specificity of the individual enzymes have to be interpreted in light of the *in vitro* system under study and of the lack of kinetic characterization of the enzyme reactions due to limited amounts of enzymes and substrates. Enzyme preparations (either cell-free lysates for WsaC and WsaD or supernatants after ultracentrifugation in the case of WsaE and WsaF) or FPLC-purified enzymes were assayed for their ability to transfer dTDP-β-L-Rha to the synthetic octyl-linked saccharides β-D-Gal-(1→*O*)-octyl (**I**), α-L-Rha-(1→3)-β-D-Gal-(1→*O*)-octyl (**II**), and α-L-Rha-(1→3)-α-L-Rha-(1→3)-β-D-Gal-(1→*O*)-octyl (**III**). According to the TLC evidence using chloroform/methanol/water (65:25:4) as solvent system, none of the enzymes used β-D-Gal-(1→*O*)-octyl (**I**) as substrate (Fig. S6, *lanes 1*, *12*, *24*, and *28*). This result suggested another mechanism for the generation of α-L-Rha-(1→3)-β-D-Gal-(1→*O*)-octyl, which is discussed below.

WsaF was able to transfer one Rha residue to the α-L-Rha-(1→3)-β-D-Gal-(1→*O*)-octyl substrate (**II**), resulting in a new product (**IV**) with a slightly higher R_f value (R_f ~ 0.40) than that of α-L-Rha-(1→3)-α-L-Rha-(1→3)-β-D-Gal-(1→*O*)-octyl (**III**) (R_f ~ 0.37) (Fig. 2, compare *lanes 2* and *3*, and Fig. S6, *lanes 28–30*), indicating that the newly formed linkage is not α1,3. At this stage of experimentation, this finding would be consistent with the formation of either an α1,2- or a β1,2-linkage, accord-

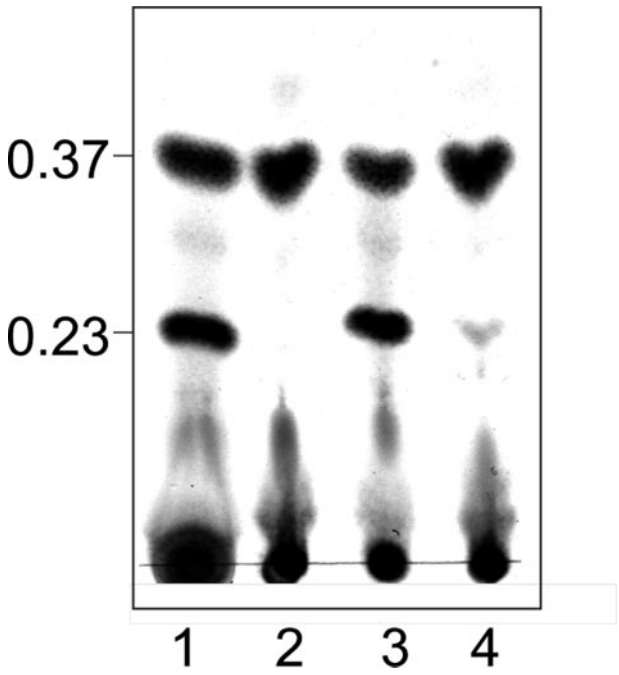


FIGURE 3. **TLC pattern of rhamnosyltransferase activity assays of the different truncated forms of WsaE using α-L-Rha-(1→3)-α-L-Rha-(1→3)-β-D-Gal-(1→*O*)-octyl (III) as substrate.** The products were separated on silica TLC plates with chloroform/methanol/water 65:25:4 as solvent; the representative part of the TLC plate is shown. *Lane 1*, WsaE; *lane 2*, WsaE_A; *lane 3*, WsaE_B; *lane 4*, WsaE_C. Staining was performed with thymol.

ing to the S-layer glycan structure of *G. stearothermophilus* NRS 2004/3a (16).

WsaC generated a new product (**V**) when incubated with α-L-Rha-(1→3)-α-L-Rha-(1→3)-β-D-Gal-(1→*O*)-octyl (**III**) and dTDP-β-L-Rha (Fig. S6, *lanes 12–14*). The new product (**V**) had a lower R_f value ($R_f \sim 0.25$) than the substrate (Fig. 2, *lane 4*). When incubated overnight, a second, minor product (**VI**) with an even lower R_f value of ~0.15 was synthesized in lower yield (Fig. 2, *lane 4*). Interestingly, the truncated form WsaC_I did not react (Fig. S6, *lanes 20–22*). Apparently, as is also the case for WsaP, the transmembrane domain is essential for anchoring of the enzyme in the membrane to maintain its activity (22).

WsaE was shown to react with α-L-Rha-(1→3)-α-L-Rha-(1→3)-β-D-Gal-(1→*O*)-octyl (**III**), yielding one product (**VII**) migrating with an R_f value of ~0.23 (Fig. 3, *lane 1*, and Fig. S6, *lanes 24–26*). To identify the enzymatic activity of the two rhamnosyltransferase domains, the three truncated forms WsaE_A (N-terminal domain), WsaE_B (both domains), and WsaE_C (C-terminal domain) were assayed. WsaE_B showed with substrate (**III**) activity comparable with that of full-length WsaE (Fig. 3, *lane 3*); product formation with WsaE_C was less pronounced (Fig. 3, *lane 4*), whereas WsaE_A did not react at all (Fig. 3, *lane 2*). These results indicate that both glycosyltransferase domains are necessary for full activity of the enzyme, with WsaE_C being the critical domain for formation of the α1,2-linkage. Obviously, the N-terminal methyltransferase domain is not crucial for the transfer of rhamnose from dTDP-β-L-Rha, since WsaE_B, lacking the methyltransferase domain, is fully active. Since none of the synthetic substrates contained a β1,2-linked rhamnose residue, which would be the substrate for the α1,3-rhamnosyltransferase required to form the repeats of the S-layer glycan, the reaction product of WsaF, which is β-L-Rha-(1→2)-α-L-Rha-(1→3)-β-D-Gal-(1→*O*)-octyl (**IV**), was used as a substrate for WsaC and WsaE. Unexpectedly, no product was formed by WsaC, whereas two products were formed by WsaE according to the TLC evidence (data not shown; see NMR data below).

WsaD did not react with any of the substrates (Fig. S6, *lanes 1–3*). None of the transferases gave a new product when incubated with dTDP-β-L-Rha alone or with the donor substrates alone (Fig. S6, *lanes 4–7*, *15–18*, *27*, and *31*) as negative controls.

WsaE and WsaF retained their activity after purification by affinity chromatography. In contrast, according to an *in vitro* assay of purified WsaC containing LDAO or CHAPS, the enzyme activity of these preparations was lost, suggesting that a lipid environment is necessary for enzymatic activity of WsaC or that the enzyme loses critical folding properties during purification. This was confirmed by the inactivity of WsaC_I lacking the transmembrane domain (see above). Concerning usage of CHAPS, conflicting data exist in the literature, ranging from promotion (39) to loss of enzyme activity (40).

Analysis of the Rhamnosyltransferase Products by ESI-QTOF MS—To further support the composition of the different octyl-linked reaction products, purified products obtained from large scale *in vitro* assays were analyzed by ESI-QTOF MS. The masses obtained for the different products were in accordance with the glycan chain length predicted from the migration of the samples on TLC plates (Table S3). To confirm the identity of the sugar residues, the product ions were fragmented by MS^2. For instance, the MS^2 spectrum of the singly charged ion at *m/z* 899.40 corresponds to the product (**VI**) of the *in vitro* assay of WsaC (Fig. S7). The singly charged ions at *m/z* 753.34, 607.29, 461.23, and 315.18 correspond to the subsequent loss of four rhamnoses.

Analysis of the Rhamnosyltransferase Products by NMR—For the characterization of the linkage specificity of WsaC, WsaE and WsaF, respectively, the newly synthesized octyl-linked saccharides were analyzed by NMR spectroscopy. To obtain sufficient amounts of material, the large scale reaction mixtures were applied to chromatography on Sephadex G-10 with 20% methanol as solvent. Although separation of remaining substrate and products was possible for all enzymes, the two products (**V** and **VI**) of WsaC could not be separated. The yields of individual reactions given below do not reflect absolute reaction yields, because some of the enzyme products were sacrificed in order to achieve purity, and due to limited amounts of substrates the reaction conditions were not optimized. For comparison with the enzyme products, NMR data of the substrate octyl trisaccharide (**III**) were fully assigned and compared favorably with those of the corresponding glycosyl units in the S-layer core glycan (41) (Table 1).

For WsaC, the conversion of 5 mg of α-L-Rha-(1→3)-α-L-Rha-(1→3)-β-D-Gal-(1→*O*)-octyl (**III**) yielded 2.9 mg of purified product. Within the detection limit of the NMR spectrometer, only signals of product **V** were observed in the NMR spectra (Fig. 4*a*). Product **V** was shown to be a tetrasaccharide fully amenable for NMR analysis. The ^1H NMR spectrum

TABLE 1

400.13 MHz ^{1}H and 100.62 MHz ^{13}C NMR chemical shift data (in ppm) and *J* coupling constants of the rhamnosyl oligosaccharides (III, V, VII, and IV)

Chemical shift data were measured at 297 K for solutions in D_2O. ^{13}C NMR data were obtained from a proton-decoupled one-dimensional spectrum for compounds **III**, **IV**, and **V** and from HSQC spectra of compound **VII**.

	Carbohydrate signals						Octyl		
	1	2	3	4	5	6	1	2	CH_3
Compound III									
A →3)-β-D-Gal*p*-(1→									
^{1}H	4.42	3.61	3.68	4.00	3.70	3.77/3.73	3.89/3.67	1.62	0.85
$J_{H,H}$ (Hz)	(7.8)	(9.9)	81.36	(3.1, 1.0)	75.83	61.63	71.46	29.51	14.19
^{13}C	103.32	70.89[a]		69.23					
B →3)-α-L-Rhap-(1→									
^{1}H	5.00	4.15	3.89	3.54	3.80	1.28			
$J_{H,H}$ (Hz)	(1.8)	(3.1)	(9.7)	(9.7)	70.12	(6.2)			
^{13}C	103.05	70.67	79.15	72.08		17.46[b]			
C α-L-Rhap-(1→									
^{1}H	5.04	4.06	3.84	3.45	3.80	1.29			
$J_{H,H}$ (Hz)	(1.7)	(3.3)	(9.7)	(9.6)	69.93	(6.3)			
^{13}C	103.22	70.96[a]	70.96[a]	72.81		17.38[b]			
Compound V									
A →3)-β-D-Gal*p*-(1→									
^{1}H	4.43	3.61	3.68	4.00	3.69	3.78-3.72	3.94/3.68	1.65-1.58	0.85
$J_{H,H}$ (Hz)	(7.6)	(9.6)	81.38	(3.2)	75.84	61.63	71.47	29.57	14.19
^{13}C	103.32	70.89[a]		69.24					
B →3)-α-L-Rhap-(1→									
^{1}H	5.02[b]	4.14[c]	3.90	3.56[d]	3.88-3.82	1.29[e]			
$J_{H,H}$ (Hz)	(2.0)	(3.6)	(9.6)	(9.6)	70.12	(6.2)			
^{13}C	103.05[f]	70.76[a]	79.17	72.08		17.44			
C →3)-α-L-Rhap-(1→									
^{1}H	5.01[b]	4.17[c]	3.90	3.54[d]	3.92-3.88	1.28[e]			
$J_{H,H}$ (Hz)	(2.0)	(3.2)	(9.6)	(9.6)	70.12	17.44[g]			
^{13}C	103.26[f]	70.96[a]	79.17	72.08					
D α-L-Rhap-(1→									
^{1}H	5.04	4.07	3.84	3.46	ND[h]	1.29[e]			
$J_{H,H}$ (Hz)	(1.6)	(3.2)	(9.6)	(9.6)	69.93	17.37[g]			
^{13}C	100.82	70.96[a]	70.65[a]	72.79					
Compound VII									
A →3)-β-D-Gal*p*-(1→									
^{1}H	4.40	3.55	3.66	3.97	~3.66	~3.70	3.83/3.64	1.59	0.82
$J_{H,H}$ (Hz)	(7.6)	(9.6)	81.5	(2.9, 1.0)	75.9	61.7	71.5	29.5	14.8
^{13}C	103.4	70.7		69.3					
B →3)-α-L-Rhap-(1→									
^{1}H	4.97	4.10	3.76	3.52	3.80-3.89	1.26[c]			
$J_{H,H}$ (Hz)	(1.6)	(3.3)	(10.1)	(10.1)	ND	(6.0)			
^{13}C	103.1	70.7	79.1	72.5		17.4			
C →2)-α-L-Rhap-(1→									
^{1}H	5.17	4.03	3.91	3.45	3.80-3.89	1.25[c]			
$J_{H,H}$ (Hz)	(1.6)	(3.2)	(9.8)	(9.9)	ND	(6.1)			
^{13}C	101.7	78.5	70.0	73.0		17.4			
D α-L-Rhap-(1→									
^{1}H	4.92	4.03	3.86	3.40	3.63	1.22			
$J_{H,H}$ (Hz)	(1.6)	(3.2)	(9.6)	(9.6)	70.6	(6.4)			
^{13}C	103.1	71.0	71.0	73.0		17.4			
Compound IV									
A →3)-β-D-Gal*p*-(1→									
^{1}H	4.43	3.59	3.68	4.01	3.70	3.75/3.65	3.92/3.67	1.59	0.82
$J_{H,H}$ (Hz)	(7.8)	(9.9)	81.45	(3.0, 1.0)	75.74	61.51	71.37[b]	29.41	14.11
^{13}C	103.20	70.86[a]		69.12					
$J_{C,H}$ (Hz)	(163.8)								
B →2)-α-L-Rhap-(1→									
^{1}H	5.14	4.25	3.84	3.48	3.84	1.28			
$J_{H,H}$ (Hz)	(1.5)	(3.5)	(9.8)	(9.7)	70.04[c]	(6.3)			
^{13}C	100.74	79.14	70.16[c]	73.01[d]		17.28			
$J_{C,H}$ (Hz)	(172.5)								
C β-L-Rhap-(1→									
^{1}H	4.70	4.01	3.58	3.38	3.38	1.31			
$J_{H,H}$ (Hz)	(1.5)	(3.3)	71.61	(9.5)	72.91[d]	(5.7)			
^{13}C	99.13	70.69[a]		72.61[d]		17.28			
$J_{C,H}$ (Hz)	(160.5)								

[a,b,c] Assignments are interchangeable.
[d] ND, not determined.

recorded at 400 MHz revealed three closely spaced signals in the anomeric region attributable to α-rhamnopyranosyl units. Heteronuclear single quantum coherence spectroscopy and COSY experiments allowed a complete assignment of the spin-spin correlations and provided evidence for the substitution at carbon-3 due to the glycosylation shifts observed for units **B** and **C**, respectively, whereas the ^{13}C NMR shifts for residue **D** were consistent with a terminal rhamnose moiety (Table 1 and

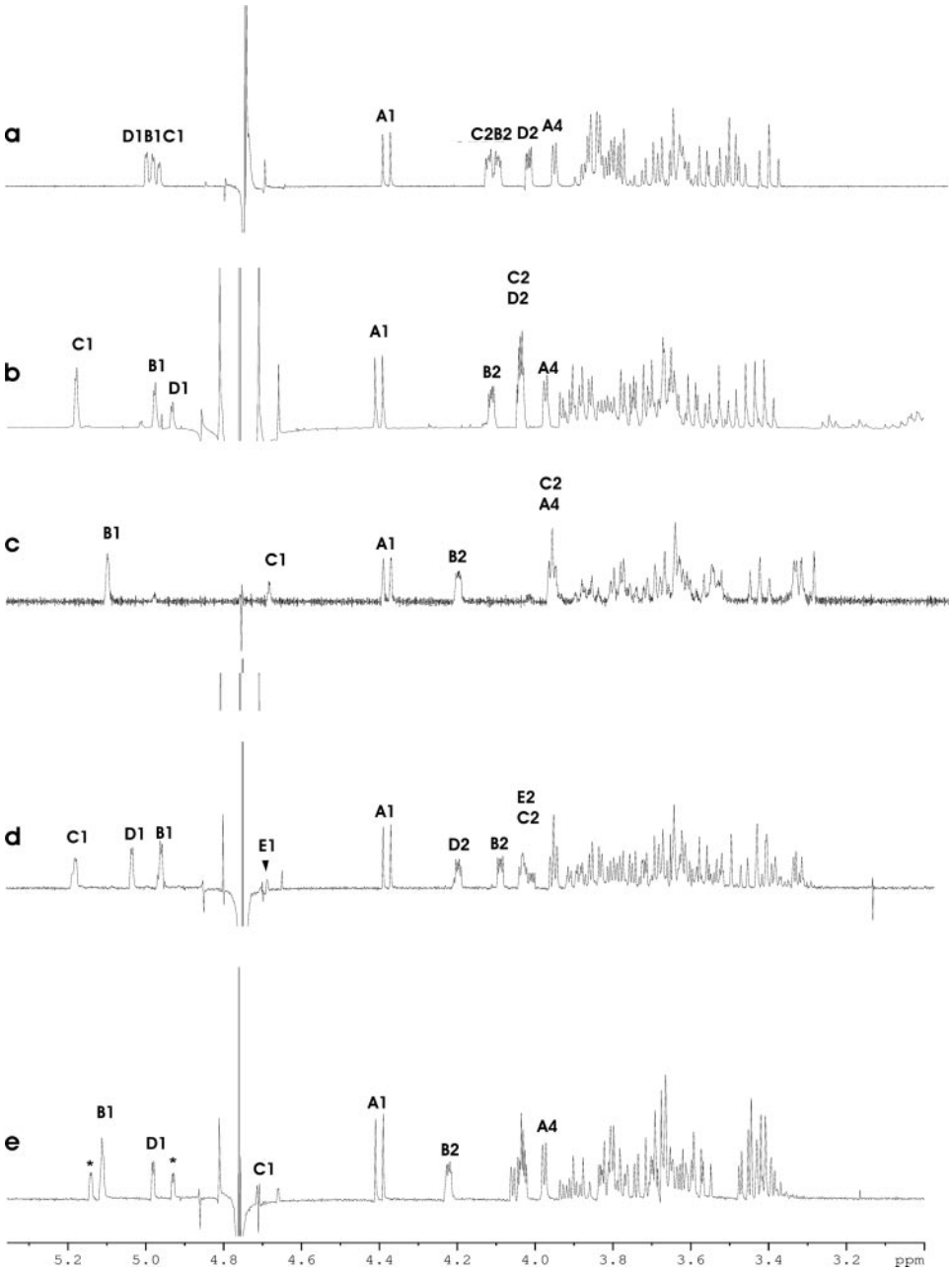


FIGURE 4. **400 MHz ^{1}H NMR spectra of the anomeric and bulk region of the products V (*a*), VII (*b*), IV (*c*), VIII (*d*), and IX (*e*), measured in D_2O at 297 K.** The *Arabic numerals* refer to protons in sugar residues denoted by *letters*, as shown in Fig. 5. *, the anomeric proton signals of the terminal disaccharide part in the tentative pentasaccharide component (**X**).

Incubation of WsaE with 5 mg of starting material (**III**) and dTDP-β-L-Rha yielded 0.96 mg of purified reaction product (**VII**). Analysis of the spin-spin connectivities by homo- and heteronuclear correlation experiments allowed the identification of unit **B** as a 3-substituted α-rhamnosyl moiety based on the downfield-shifted ^{13}C NMR signal of C-3 (Table 1 and Fig. 4*b*), whereas the anomeric proton of unit **C** was connected to a C-2 carbon atom displaying a downfield shift to 78.5 ppm. The remaining anomeric proton of residue **D** could be assigned to a terminal α-rhamnosyl unit (Fig. 5). Since both H-2 protons of units **C** and **D** had identical chemical shifts, the assignment of H-1 of unit **C** was inferred from the upfield-shifted signal of the corresponding anomeric carbon signal in unit **C** at 101.7 ppm (Table 1). The NMR analysis thus clearly established the structure of the tetrasaccharide octyl glycoside (**VII**) and thereby identified WsaE as an enzyme with α1,2-rhamnosyltransferase activity.

WsaF converted 3 mg of α-L-Rha-(1→3)-β-D-Gal-(1→*O*)-octyl (**II**) to 0.59 mg of new product (**IV**). The ^{1}H NMR spectrum of the product indicated a new, upfield-shifted signal hidden under the solvent water peak (Fig. 4*c*). Measurement of the heteronuclear $J_{C\text{-}1,H\text{-}1}$ coupling constants clearly proved that residue **C** had the β-anomeric configuration and confirmed the α-configuration of unit **B** (Table 1 and Fig. 5). In addition, the downfield-shifted ^{13}C NMR signal of carbon-2 in unit **B** proved C-2 as the linkage site. Thus, WsaF is a retaining glycosyltransferase responsible for the β1,2-linkage in the polyrhamnan. However, α-L-Rha-(1→3)-β-D-Gal-(1→*O*)-octyl (**II**) is not the natural substrate for the β1,2-rhamnosyltransferase in the biosynthesis of the glycan chain. It would need an α1,2-linked rhamnose substrate according to the glycan structure of *G. stearothermophilus* NRS 2004/3a (16). Since WsaF could only use α-L-Rha-(1→3)-β-D-Gal-(1→*O*)-octyl (**II**) and not α-L-Rha-(1→3)-α-L-Rha-(1→3)-β-D-Gal-(1→*O*)-octyl (**III**) (Fig. S6, *lanes 29* and *30*), it does not generally recognize α1,3-linked rhamnoses. An explanation for the recognition could be that the combination of α1,3-linked rhamnose with galactose and octyl can bind to the catalytic domain. Due to the unavail-

Fig. 5). Moreover, due to the similar substitution pattern of both units **B** and **C**, the intensities of the C-2, C-3, C-4, and C-5 ^{13}C NMR signals corresponded to two atoms, respectively (data not shown). Thus, NMR analysis of the major WsaC product (**V**) clearly demonstrated that WsaC is an α1,3-rhamnosyltransferase, which is able to transfer at least one rhamnose residue to an α1,3-linked rhamnose. Considering the results from TLC and ESI-QTOF MS analysis of WsaC products (Fig. S7), it is very likely that WsaC is also capable of transferring one more α1,3-linked rhamnose to product **V**, yielding compound **VI**, which implies that this enzyme is responsible for the core variability present in the S-layer glycan (17).

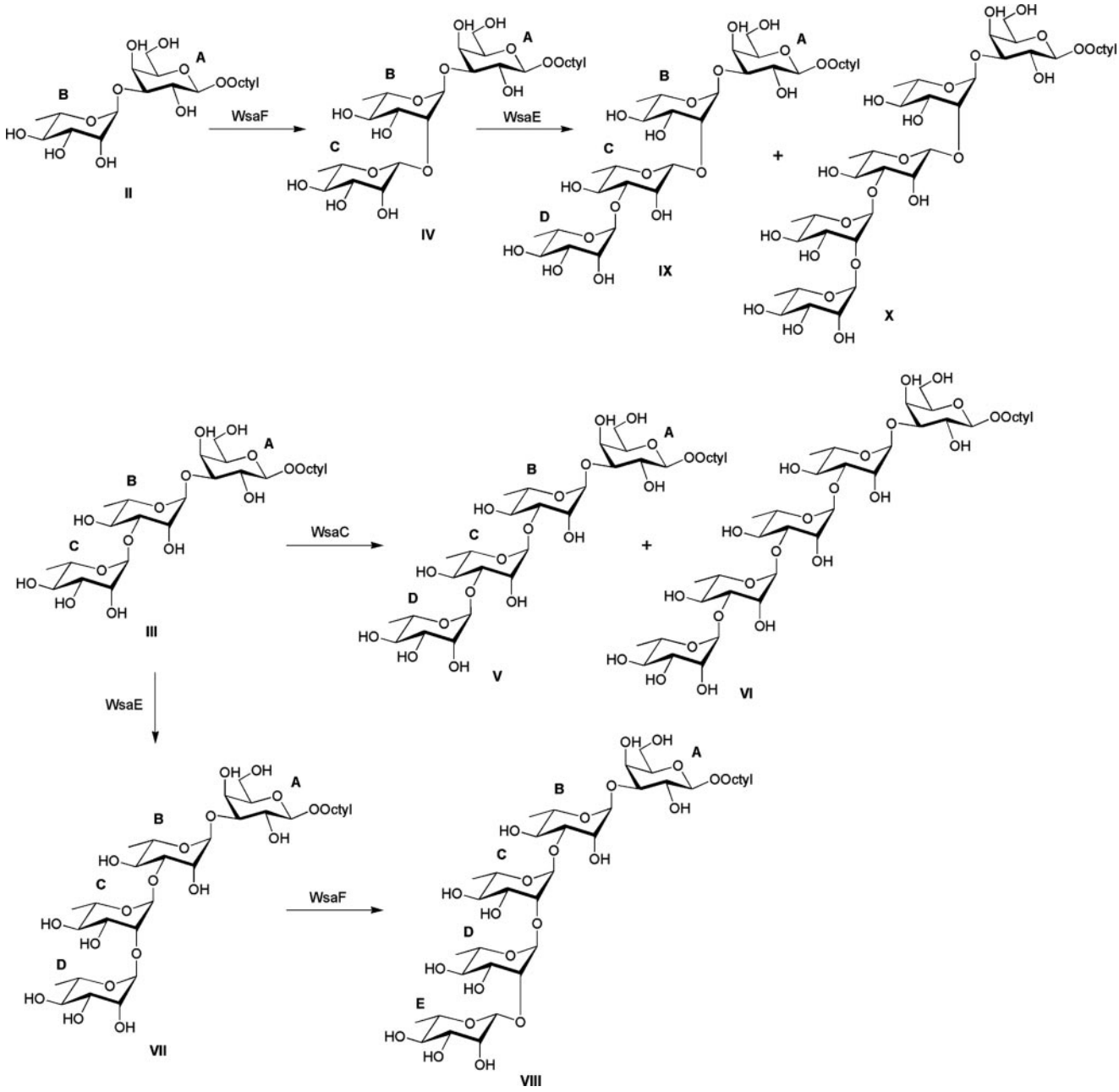


FIGURE 5. **Reaction scheme for the conversion of the synthetic substrates II and III to products IV, V, and VII and subsequently to VI, VIII, IX, and X, involving the enzymes WsaC, WsaE, and WsaF of *G. stearothermophilus* NRS 2004/3a.** For the assignment of glycoses **A**, **B**, **C**, **D**, and **E**, see Table 1.

ability of structural data on WsaF or a homologous protein, this explanation remains speculative.

The natural substrate for WsaF is the α-L-Rha-(1→2)-α-L-Rha-(1→3)-α-L-Rha-(1→3)-β-D-Gal-(1→*O*)-octyl product (**VII**) of WsaE (see above). After NMR analysis, this product (0.7 mg) was used for an *in vitro* activity assay of WsaF. The new product (**VIII**) (0.31 mg) was analyzed by TLC (data not shown), purified, and applied to NMR analysis, which confirmed that WsaF is indeed a β1,2-rhamnosyltransferase. Although a full assignment of the product mixture could not be accomplished, the spectral features of the distal disaccharide part structure in **VIII** were similar to that of compound **IV** (Fig. 4*d*). In particular, upfield-shifted anomeric ^{1}H and ^{13}C NMR signals of the terminal rhamnose were consistent with the β-anomeric configuration, whereas a downfield-shifted C-2 signal (77.8 ppm) was observed for unit **D** (data not shown).

To analyze the second reaction product of WsaE, a large scale reaction with 0.5 mg of β-L-Rha-(1→2)-α-L-Rha-(1→3)-β-D-Gal-(1→*O*)-octyl (**IV**; see above) yielded 0.25 mg of a purified product mixture. The ^{1}H NMR spectrum of the mixture revealed the presence of additional anomeric protons indicative of a mixture containing a tetrasaccharide (**IX**) and a pentasaccharide (**X**) (Fig. 4*e*). Although a full assignment of the tetrasaccharide (**IX**) could not be accomplished, the ^{13}C NMR signal of carbon-3 of the β-rhamnosyl residue **C** was observed at 81.1 ppm, indicative of α-substitution at position 3. The anomeric $^{1}H/^{13}C$ signals of the terminal α-linked rhamnose unit **D** were observed at 5.01/101.61 ppm, respectively (Fig. 5). Two minor

additional anomeric proton and carbon signals were observed with chemical shifts (5.16/100.88 ppm and 4.95/102.65 ppm, respectively) similar to the terminal disaccharide part of compound **VII** (Fig. 4*b*). These data thus suggest a further chain elongation of unit **D** by a rhamnose unit in α1,2-linkage. Thus, NMR analysis provided evidence that WsaE is obviously a multifunctional enzyme, which is also responsible for the transfer of an α1,3-linked rhamnose residue to the β1,2-linked rhamnose and, most probably, the subsequent addition of an α1,2-linked rhamnose.

Methyltransferase Activity of WsaE—To determine if the putative methyltransferase domain of WsaE is responsible for the methylation of the non-reducing end of the terminal repeating unit of the glycan chain of *G. stearothermophilus* NRS 2004/3a, a cell free lysate of *E. coli* BL21 Star (DE3) expressing WsaE was incubated with [^{3}H]SAM and α-L-Rha-(1→3)-α-L-Rha-(1→3)-β-D-Gal-(1→*O*)-octyl (**III**), either with or without addition of dTDP-β-L-Rha. According to the TLC evidence WsaE was able to methylate α-L-Rha-(1→3)-α-L-Rha-(1→3)-β-D-Gal-(1→*O*)-octyl (**III**) (Fig. 6, lanes 1, 2 and 11, 12). Only this product with an R_f value of 0.47 was formed, even if dTDP-β-L-Rha was added and one additional α1,2-linked rhamnose residue was linked to the substrate (compare with Fig. 3, lane 1), indicating that WsaE cannot methylate α1,2-linked rhamnose. This result was verified by MS analysis of a scaled up reaction mixture containing unlabeled SAM and dTDP-β-L-Rha (Table 2). Peaks corresponding to $[M + H]^+$, $[M + NH_4]^+$, and $[M + Na]^+$ of unmethylated substrate L-Rha-(1→3)-α-L-Rha-(1→3)-β-D-Gal-(1→*O*)-octyl (**III**) and product α-L-Rha-(1→2)-α-L-Rha-(1→3)-α-L-Rha-(1→3)-β-D-Gal-(1→*O*)-octyl (**VII**) as well as methylated substrate (**III**) could be observed, whereas no peak corresponded to the expected mass of methylated product (**VII**), showing that this reaction product of WsaE, which was proven by NMR to contain a terminal α1,2-rhamnose, could not be methylated.

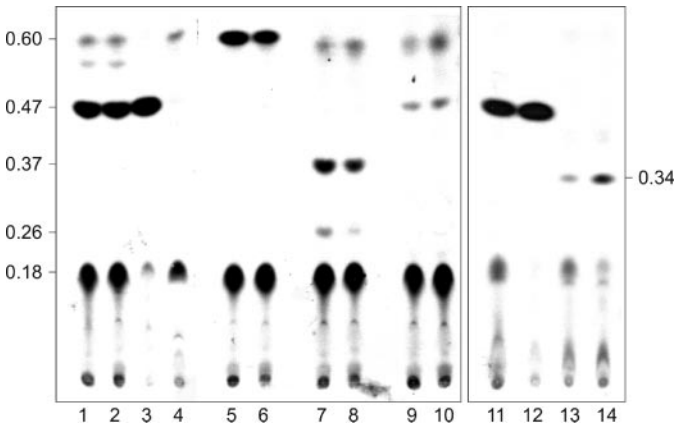


FIGURE 6. **Autoradiogram of methyltransferase activity assays of WsaE and the different truncated versions of WsaE, using α-L-Rha-(1→3)-α-L-Rha-(1→3)-β-D-Gal-(1→*O*)-octyl (III) (*lanes 1–4*, *11*, and *12*), α-L-Rha-(1→3)-β-D-Gal-(1→*O*)-octyl (II) (*lanes 5* and *6*), the reaction products of WsaC α-L-Rha-(1→3)-α-L-Rha-(1→3)-α-L-Rha-(1→3)-β-D-Gal-(1→*O*)-octyl (V and VI) (*lanes 7* and *8*), α-L-Rha-(1→2)-α-L-Rha-(1→3)-α-L-Rha-(1→3)-β-D-Gal-(1→*O*)-octyl (VII) (*lanes 9* and *10*), or α-L-Rha-(1→3)-β-L-Rha-(1→2)-α-L-Rha-(1→3)-β-D-Gal-(1→*O*)-octyl (IX) (*lanes 13* and *14*) as substrate.** In *lanes 1*, *6*, *8*, *10*, *12*, and *14*, dTDP-β-L-Rha was added; all *lanes* except for *lanes 3* and *4* contain WsaE. *Lane 3*, WsaE_N; *lane 4*, WsaE_M. The products were separated on silica TLC plates with chloroform/methanol/water (65:25:4) as solvent.

To confirm that the N terminus of WsaE is responsible for the methyltransferase activity of the enzyme, the two truncated forms WsaE_M (including the UbiE motive) and WsaE_N (including the UbiE motive and additional 200 aa) were incubated with [^{3}H]SAM. Compared with full-size WsaE, WsaE_N (Fig. 6, *lane 3*) showed an even higher activity, whereas the shorter form WsaE_M was inactive (Fig. 6, *lane 4*). These results confirmed that the N terminus is sufficient for methyltransferase activity. To further establish that WsaE methylates only α1,3-linked rhamnose residues, independent of the glycan chain length, α-L-Rha-(1→3)-β-D-Gal-(1→*O*)-octyl (**II**); the reaction products of WsaC, which are α-L-Rha-(1→3)-α-L-Rha-(1→3)-α-L-Rha-(1→3)-β-D-Gal-(1→*O*)-octyl (**V**) and (**VI**); and both reaction products of the rhamnosyltransferase domains of WsaE, which are α-L-Rha-(1→2)-α-L-Rha-(1→3)-α-L-Rha-(1→3)-β-D-Gal-(1→*O*)-octyl (**VII**) and α-L-Rha-(1→3)-β-L-Rha-(1→2)-α-L-Rha-(1→3)-β-D-Gal-(1→*O*)-octyl (**IX**), were used as substrates for WsaE. All reactions were done with and without dTDP-β-L-Rha but did not show any difference (compare neighboring *lanes* in Fig. 6). WsaE methylated all α1,3-linked rhamnoses, resulting in products with an R_f value of 0.60 for substrate (**II**) (Fig. 6, *lanes 5* and *6*), with R_f values of 0.37 and 0.26 for the products (**V**) and (**VI**) of WsaC (Fig. 6, *lanes 7* and *8*) and with an R_f value of 0.34 for the α1,3-linked product (**IX**) of WsaE (Fig. 6, *lanes 13* and *14*). Note that all R_f values of methylated products are increased as compared with the R_f values of the unmethylated substrates (compare with Figs. 2 and 4). WsaE was not able to methylate the α1,2-linked rhamnose of (**VII**) (Fig. 6, *lanes 9* and *10*). The faint band at the same R_f value ($R_f \sim 0.47$) as the reaction product of WsaE with α-L-Rha-(1→3)-α-L-Rha-(1→3)-β-D-Gal-(1→*O*)-octyl (**III**) most likely corresponds to methylation of residual substrate (**III**) being present as minor impurity in the tetrasaccharide sample α-L-Rha-(1→2)-α-L-Rha-(1→3)-α-L-Rha-(1→3)-β-D-Gal-(1→*O*)-octyl (**VII**). These results confirm the methylation of the α1,3-linked rhamnose of the terminal repeating unit of the glycan chain of *G. stearothermophilus* NRS 2004/3a (16) and furthermore show that WsaE indeed possesses a relaxed specificity for the glycan chain length.

Substrate Specificity of WsaD—Considering the overall composition of the S-layer glycan of *G. stearothermophilus* NRS

TABLE 2
ESI-QTOF MS analysis of octyl-linked products of an *in vitro* methylrhamnosyltransferase activity assay of WsaE with substrate III

Assignment	$[M + H]^+$		$[M + NH_4]^+$		$[M + Na]^+$	
	Experimental	Theoretical	Experimental	Theoretical	Experimental	Theoretical
Rha-Rha-Gal-octyl (**III**)	585.35	585.31		602.33	607.36	607.29
Methyl-Rha-Rha-Gal-octyl	599.36	599.33	616.39	616.35	621.34	621.31
Rha-Rha-Rha-Gal-octyl (**VII**)	731.40	731.37	748.44	748.39	753.39	753.35
Methyl-Rha-Rha-Rha-Gal-octyl		745.39		762.41		767.37

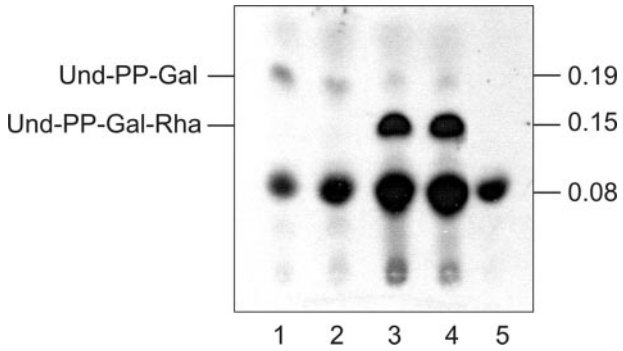


FIGURE 7. **TLC autoradiogram showing lipid carrier-linked intermediates synthesized in an *in vitro* assay using membranes isolated from *E. coli* C43 (DE3) harboring WsaP and/or WsaD and ^{14}C-labeled-UDP-Gal.** *Lane 1*, WsaP; *lane 2*, WsaP with dTDP-Rha; *lane 3*, WsaP and WsaD; *lane 4*, WsaP and WsaD with dTDP-Rha; *lane 5*, WsaD with dTDP-Rha. Products were separated on silica plates by TLC using the solvent system chloroform/methanol/water (65:25:4). The representative part of the TLC plate is shown.

2004/3a, the specificities that could be attributed to the transferases WsaC, WsaE, and WsaF (this study), and the lack of activity of WsaD in the aforementioned experiments, it was hypothesized that WsaD should correspond to the rhamnosyltransferase, which transfers the first rhamnose residue from dTDP-β-L-Rha to the lipid-linked galactose in the course of S-layer glycan biosynthesis. Consequently, WsaD should be able to recognize the reaction product of WsaP, which is lipid-PP-linked galactose, as substrate. For this experiment, equal amounts of lysed *E. coli* C43 (DE3) cells expressing either of the enzymes were mixed with ^{14}C-labeled UDP-Gal and unlabeled dTDP-β-L-Rha and incubated overnight. As negative controls, each enzyme was incubated alone. The reaction products were analyzed by TLC and autoradiography (Fig. 7, *lanes 1–5*). In all reaction mixtures, a spot at an R_f value of 0.08 was visible. However, the reaction product with an R_f value of 0.19 was formed only in the presence of WsaP (Fig. 7, *lanes 1–4*), indicating that it corresponds to und-PP-Gal as described before (22). The formation of this product was not dependent on the addition of dTDP-β-L-Rha (Fig. 7, compare *lanes 1* and *2*). A new product with a lower R_f value ($R_f \sim 0.15$) than the reaction product of WsaP ($R_f \sim 0.19$) was synthesized when WsaD was added to the reaction mixture (Fig. 7, *lanes 3* and *4*). Interestingly, this product was also formed without the addition of dTDP-β-L-Rha (Fig. 7, compare *lanes 3* and *4*). Apparently, dTDP-β-L-Rha present in *E. coli* C43 (DE3) cells can be used for the reaction (42). The negative control WsaD alone did not give any new product (Fig. 7, *lane 5*). These findings indicate that WsaD is responsible for the transfer of the first rhamnose to the lipid-linked und-PP-Gal intermediate.

DISCUSSION

In our study to determine the molecular basis for S-layer glycoprotein glycan biosynthesis in *G. stearothermophilus* NRS 2004/3a, we used recombinant enzymes as well as chemically synthesized substrates, linked to a hydrophobic octyl spacer, representing partial structural elements of the mature S-layer glycan. In *G. stearothermophilus* NRS 2004/3a, the glycan chain is linked to the protein via β-galactose (Gal) (16). Thus, β-Gal-octyl (**I**), was tested as a first substrate. It was unexpected to find that β-Gal-octyl (**I**) was not recognized by any of the transferases encoded by the *slg* gene cluster of the organism. A possible explanation might be that for the activity of the first rhamnosyltransferase, a pyrophosphate linkage of the galactose to the lipid carrier as is present in the native substrate would be essential. Another explanation for the lack of reactivity with β-Gal-octyl (**I**) might be that actually an α-linked Gal residue would be needed as a substrate for the first rhamnosyltransferases to act on, in the case that the possible oligosaccharyltransferase is inverting the linkage type during the transfer of the glycan onto the S-layer protein. Recently, the *pgl* (protein *N*-glycosylation) gene cluster of *Campylobacter jejuni* was described (43), where for the *in vitro* characterization of the glycosyltransferase PglA, synthetic und-PP-α-Bac was used as a substrate (44, 45). Although the glycan chain is linked via β-bacillosamine to Asn of the protein (46), the α-substrate is recognized by PglA, indicating that the glycan chain is α-linked to und-P and that the oligosaccharyltransferase PglB (47) is inverting the anomeric configuration of the linkage. Although nothing is known about the reaction mechanism of the respective oligosaccharyltransferase WsaB of *G. stearothermophilus* NRS 2004/3a, it might well be that protein glycosylation generally follows a common mechanism, involving the inversion of the stereochemistry at the reducing end of the glycan chain during its transfer from the lipid carrier onto the protein.

There have been previous models for LPS biosynthesis in Gram-negative bacteria (21) but none to date for a cell wall glycoconjugate of Gram-positive bacteria. Based on our data on a Gram-positive bacterium, we can propose a stepwise biosynthesis of the S-layer glycan of *G. stearothermophilus* NRS 2004/3a. The presence of a predicted ABC-2-type transporter and the absence of a putative polymerase in the *slg* gene cluster of *G. stearothermophilus* NRS 2004/3a (19) indicate that the S-layer glycan chains are most probably synthesized in a process comparable with the ABC transporter-dependent pathway of LPS *O*-PS biosynthesis (21). In our system, the previously characterized WsaP protein, which is a homologue of WbaP, serves as the initiation enzyme instead of a WecA, which usually acts as the initiation enzyme in that pathway (21, 22, 48). With the knowledge of the substrate specificities and the linkage types formed by the action of the different rhamnosyltransferases (this study), we can propose the following biosynthesis model for the S-layer glycan assembly (Fig. 8). WsaP is implicated in the first step of S-layer glycan synthesis, whereby galactose is transferred from its nucleotide-activated form (UDP-Gal) to a membrane-associated lipid carrier at the cytoplasmic face of the plasma membrane (22). In the next step, the first adaptor rhamnose is α1,3-linked from dTDP-β-L-rhamnose to the lipid-bound galactose by the action of WsaD, followed by the transfer of one or two additional α1,3-linked rhamnoses to the first α1,3-linked rhamnose by the action of WsaC to complete the adaptor. The different numbers of α1,3-linked rhamnoses transferred by WsaC would be reflected by the core variability of the S-layer glycan (17). Adaptor formation has also been described for other homopolymeric glycan chains, such as *E. coli* O8, O9a, and O9 antigen (49) or *Klebsiella pneumoniae* O3 and O5, where the adaptor is synthesized by the action of the mannosyltransferases WbdC and WbdB or homologues

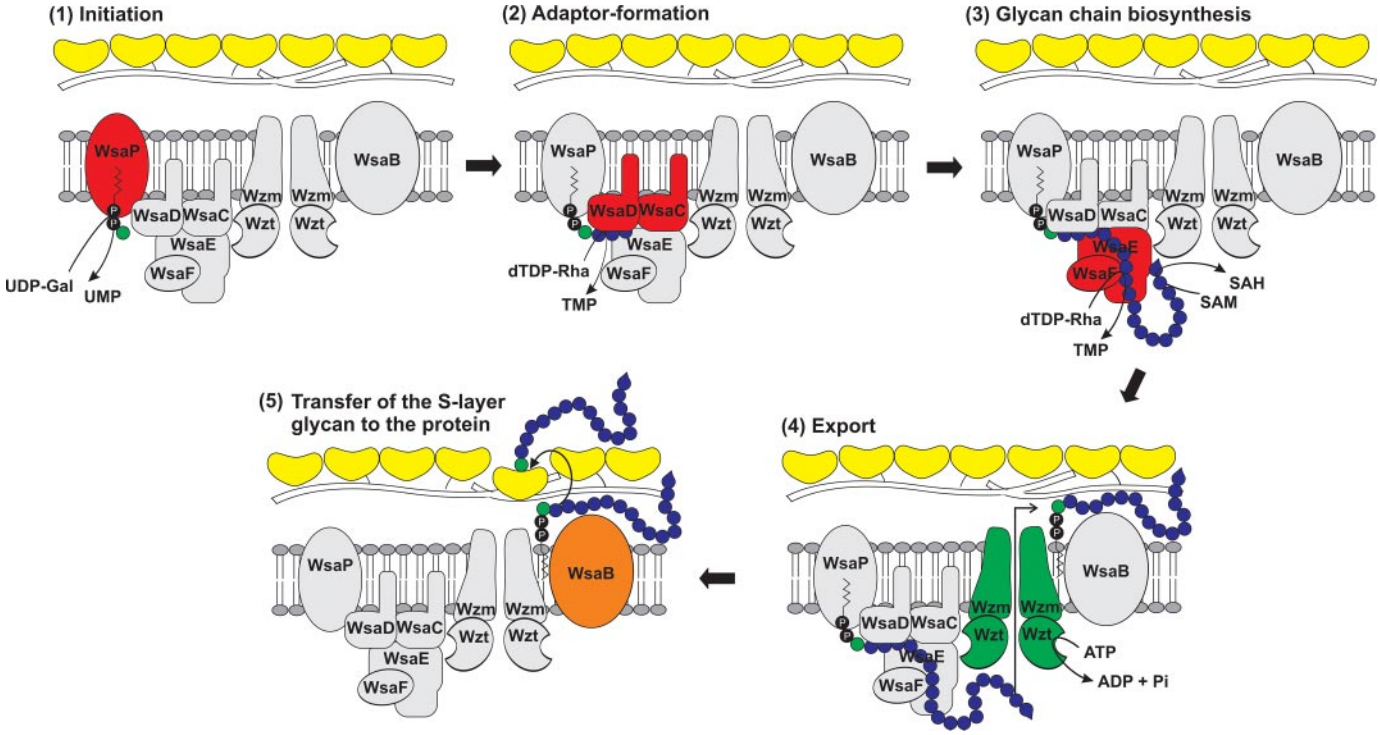


FIGURE 8. **Proposed model of S-layer glycoprotein glycan biosynthesis in *G. stearothermophilus* NRS 2004/3a.** *1*, initiation; transfer of a Gal residue from UDP-α-D-Gal to a lipid carrier catalyzed by WsaP. *2*, adaptor formation; α1,3-linkage of a rhamnose residue from dTDP-β-L-Rha to the primer by the action of WsaD, followed by the transfer of one or two additional α1,3-linked rhamnoses by the action of WsaC. *3*, glycan chain biosynthesis; formation of repeating unit-like structures by action of the rhamnosyltransferases WsaE and WsaF, whereby WsaE is forming the α1,2- and the α1,3-linkages, and WsaF is forming the β1,2-linkage. Chain growth is terminated by 2-*O*-methylation of the terminal repeating unit, catalyzed by the *O*-methyltransferase domain of WsaE. *4*, export; the Wzt component of the Wzm/Wzt ABC transporter system is predicted to be responsible for binding of the 2-*O*-methylated glycan chain and its subsequent export through the membrane. *5*, transfer of the S-layer glycan to the protein. The final transfer of the completed S-layer glycan to the S-layer protein would be catalyzed by the oligosaccharyltransferase WsaB.

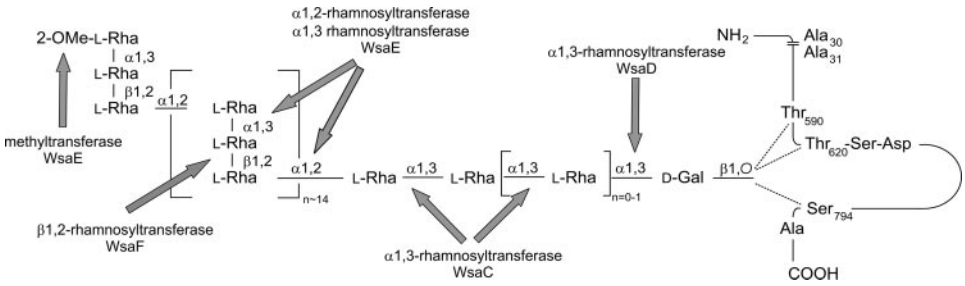


FIGURE 9. **S-layer glycoprotein glycan of *G. stearothermophilus* NRS 2004/3a (16).** The sites of action of the different rhamnosyltransferases and the methyltransferase are indicated with *arrows*. Note that the α1,2- and α1,3-rhamnosyltransferase activities of WsaE and the β1,2-rhamnosyltransferase activity of WsaF also apply for the terminal repeating unit; for simplification, this is not indicated in the figure.

thereof, to transfer the initial mannosyl residue to the priming und-PP-GlcNAc, followed by the addition of another α1,3-linked mannosyl residue. In *G. stearothermophilus* NRS 2004/3a, chain extension and formation of repeating unit-like structures presumably continues in the cytoplasm by the processive action of the two rhamnosyltransferases WsaE and WsaF, whereby WsaE is a multifunctional enzyme forming the α1,2- and the α1,3-linkage, and WsaF is a β1,2-rhamnosyltransferase. We can imagine a scenario in which an enzyme complex is formed that exerts multiple reaction cycles, with each transferase recognizing the reaction product of the foregoing enzyme. First, WsaE recognizes the α1,3-linked rhamnose product from the adaptor formed by WsaC and adds an α1,2-linked rhamnose to the saccharide intermediate, which is subsequently recognized by WsaF, which β1,2-links the next following rhamnose. The β1,2-linked rhamnose would then serve as acceptor for the transfer of rhamnose in α1,3-linkage and, subsequently, of another rhamnose in α1,2-linkage, both catalyzed by WsaE, with the latter transfer reaction being already the first step of the next cycle.

The variation in the final chain length as shown by analysis of the native glycan (17) is potentially governed by methylation of the terminal rhamnose residue; specifically, 2-*O*-methylation directly blocks the hydroxyl group that would be the acceptor for any subsequent rhamnosyltransferase reaction (16). This reaction is catalyzed by an *O*-methyltransferase, which is located at the N-terminal protein portion of WsaE. Nothing is known, so far, about how the methylation event and S-layer glycan chain length termination are coordinated and regulated in *G. stearothermophilus* NRS 2004/3a. The *in vitro* assays of this study have shown that WsaE seems to methylate terminal α1,3-linked rhamnoses, even of very short glycan chains. Since *G. stearothermophilus* NRS 2004/3a is not amenable to transformation, no knock-out mutants of this WsaE domain could

be designed so far to prove the influence of WsaE on the glycan chain length *in vivo* and to show effects on the organism. Nonreducing terminal modifications, such as methylation, have been identified on a number of glycans exported by ABC transporters in Gram-positive as well as in Gram-negative organisms; on the S-layer glycans of *Thermoanaerobacter thermohydrosulfuricus* L111-69 (41), *Aneurinibacillus thermoaerophilus* L420–91^T (50), and *G. tepidamans* GS5-97^T (51); on the O9a and O8 polysaccharides of *E. coli* (52, 53); and several other glycans (54). The gene clusters responsible for the biosynthesis of these polysaccharides encode a Wzt homologue with an extended C-terminal domain (12, 55), which was described to be responsible for the binding of the modified glycan chain and its subsequent export through the membrane in the biosynthesis of O8 and O9a *O*-PS (54).

The involvement of multifunctional enzymes in the biosynthesis of polysaccharides is not unusual. In *E. coli* O8 and O9a polysaccharides, for instance, chain extension is performed by a multidomain serotype-specific mannosyltransferase WbdA (56). In contrast to WsaE of *G. stearothermophilus* NRS 2004/3a, methylation or phosphomethylation of the nonreducing end occurs in separate steps by the action of WdbD. Although it is reported that methylation plays a critical role in chain length determination and that overexpression of WdbD reduces *O*-PS chain length in serotypes O8 and O9a (57), it is unclear how WdbD activity is regulated.

The described enzymatic reactions give a complete picture of how the structure of the S-layer glycan of *G. stearothermophilus* NRS 2004/3a, consisting of trisaccharide repeats with the structure →[2)-α-L-Rha*p*-(1→3)-β-L-Rha*p*-(1→2)-α-L-Rha*p*-(1→]$_n$, with the terminal rhamnose residue at the nonreducing end modified at the hydroxyl group of carbon-2 by *O*-methylation, a short adaptor saccharide of α1,3-linked L-Rhap residues, and a β-D-galactose residue serving as linker to threonine 590, threonine 620, and serine 794 of the S-layer polypeptide backbone (16, 17) (Fig. 9), might be assembled. The completed methylated glycan chain would then be recognized and transported across the membrane by a process involving an ABC transporter (55) and, eventually, be transferred to the S-layer protein by an oligosaccharyltransferase WsaB. A comparable transferase function was described recently for PglL and PilO, which are involved in protein *O*-glycosylation in *Neisseria meningitidis* and *Pseudomonas aeruginosa*, respectively (58), and for PglB involved in *N*-glycosylation in *C. jejuni* (47, 59, 60).

Considering that several other *slg* gene clusters contain an ABC transporter (19) and numerous S-layer glycan chains are modified at the nonreducing end (16, 41, 50, 51), the biosynthesis model proposed for *G. stearothermophilus* NRS 2004/3a might be more generally valid for S-layer protein glycosylation. In future studies, it will be interesting to show how these enzymes interact and work in concert in S-layer glycan biosynthesis of *G. stearothermophilus* NRS 2004/3a and how the modality of the glycan chain length is regulated. The understanding of S-layer glycan biosynthesis and the characterization of the involved enzymes are important steps toward the goal of engineering functional S-layer *neo*-glycoproteins (11, 12, 61) for applications in the fields of nanobiotechnology, biomimetics, and vaccine and drug design.

REFERENCES

1. Helenius, A., and Aebi, M. (2004) *Annu. Rev. Biochem.* **73,** 1019–1049
2. Szymanski, C. M., and Wren, B. W. (2005) *Nat. Rev. Microbiol.* **3,** 225–237
3. Eichler, J., and Adams, M. W. (2005) *Microbiol. Mol. Biol. Rev.* **69,** 393–425
4. Rudd, P. M., Elliott, T., Cresswell, P., Wilson, I. A., and Dwek, R. A. (2001) *Science* **291,** 2370–2376
5. Sharon, N., and Lis, H. (2004) *Glycobiology* **14,** 53R–62R
6. Ohtsubo, K., and Marth, J. D. (2006) *Cell* **126,** 855–867
7. Seeberger, P. H., and Werz, D. B. (2007) *Nature* **446,** 1046–1051
8. Elliott, S., Lorenzini, T., Asher, S., Aoki, K., Brankow, D., Buck, L., Busse, L., Chang, D., Fuller, J., Grant, J., Hernday, N., Hokum, M., Hu, S., Knudten, A., Levin, N., Komorowski, R., Martin, F., Navarro, R., Osslund, T., Rogers, G., Rogers, N., Trail, G., and Egrie, J. (2003) *Nat. Biotechnol.* **21,** 414–421
9. Stevens, J., Blixt, O., Tumpey, T. M., Taubenberger, J. K., Paulson, J. C., and Wilson, I. A. (2006) *Science* **312,** 404–410
10. Schäffer, C., Novotny, R., Küpcü, S., Zayni, S., Scheberl, A., Friedmann, J., Sleytr, U. B., and Messner, P. (2007) *Small* **3,** 1549–1559
11. Schäffer, C., and Messner, P. (2004) *Glycobiology* **14,** 31R–42R
12. Messner, P., Steiner, K., Zarschler, K., and Schäffer, C. (2008) *Carbohydr. Res.* **343,** 1934–1951
13. Schuster, B., Pum, D., Sára, M., and Sleytr, U. B. (2006) *Mini Rev. Med. Chem.* **6,** 909–920
14. Sleytr, U. B., Egelseer, E. M., Ilk, N., Pum, D., and Schuster, B. (2007) *FEBS J.* **274,** 323–334
15. Sleytr, U. B., Huber, C., Ilk, N., Pum, D., Schuster, B., and Egelseer, E. M. (2007) *FEMS Microbiol. Lett.* **267,** 131–144
16. Schäffer, C., Wugeditsch, T., Kählig, H., Scheberl, A., Zayni, S., and Messner, P. (2002) *J. Biol. Chem.* **277,** 6230–6239
17. Steiner, K., Pohlentz, G., Dreisewerd, K., Berkenkamp, S., Messner, P., Peter-Katalinić, J., and Schäffer, C. (2006) *J. Bacteriol.* **188,** 7914–7921
18. Novotny, R., Schäffer, C., Strauss, J., and Messner, P. (2004) *Microbiology* **150,** 953–965
19. Novotny, R., Pfoestl, A., Messner, P., and Schäffer, C. (2004) *Glycoconj. J.* **20,** 435–447
20. Power, P. M., and Jennings, M. P. (2003) *FEMS Microbiol. Lett.* **218,** 211–222
21. Raetz, C. R. H., and Whitfield, C. (2002) *Annu. Rev. Biochem.* **71,** 635–700
22. Steiner, K., Novotny, R., Patel, K., Vinogradov, E., Whitfield, C., Valvano, M. A., Messner, P., and Schäffer, C. (2007) *J. Bacteriol.* **189,** 2590–2598
23. Valvano, M. A. (2003) *Front. Biosci.* **8,** S452–S471
24. Laemmli, U. K. (1970) *Nature* **227,** 680–685
25. Graninger, M., Kneidinger, B., Bruno, K., Scheberl, A., and Messner, P. (2002) *Appl. Environ. Microbiol.* **68,** 3708–3715
26. Fürstner, A., and Müller, T. (1999) *J. Am. Chem. Soc.* **121,** 7814–7821
27. Love, K. R., and Seeberger, P. H. (2004) *Angew. Chem. Int. Ed.* **43,** 602–605
28. Werz, D. B., and Seeberger, P. H. (2005) *Angew. Chem. Int. Ed.* **44,** 6315–6318
29. Werz, D. B., Ranzinger, R., Herget, S., Adibekian, A., von der Lieth, C.-W., and Seeberger, P. H. (2007) *ACS Chem. Biol.* **2,** 685–691
30. Adachi, S. (1965) *J. Chromatogr.* **17,** 295–299
31. Wolucka, B. A., and de Hoffmann, E. (1998) *Glycobiology* **8,** 955–962
32. Folch, J., Lees, M., and Sloane-Stanley, G. H. (1957) *J. Biol. Chem.* **226,** 497–509
33. Saxena, I. M., Brown, R. M., Jr., Fevre, M., Geremia, R. A., and Henrissat, B. (1995) *J. Bacteriol.* **177,** 1419–1424
34. Charnock, S. J., and Davies, G. J. (1999) *Biochemistry* **38,** 6380–6385
35. Tarbouriech, N., Charnock, S. J., and Davies, G. J. (2001) *J. Mol. Biol.* **314,** 655–661
36. Morona, R., Macpherson, D. F., Van Den, B. L., Carlin, N. I., and Manning, P. A. (1995) *Mol. Microbiol.* **18,** 209–223
37. Geremia, R. A., Petroni, E. A., Ielpi, L., and Henrissat, B. (1996) *Biochem. J.* **318,** 133–138
38. Keenleyside, W. J., Clarke, A. J., and Whitfield, C. (2001) *J. Bacteriol.* **183,** 77–85
39. Rush, J. S., Rick, P. D., and Waechter, C. J. (1997) *Glycobiology* **7,** 315–322

40. Riley, J. G., Menggad, M., Montoya-Peleaz, P. J., Szarek, W. A., Marolda, C. L., Valvano, M. A., Schutzbach, J. S., and Brockhausen, I. (2005) *Glycobiology* **15,** 605–613
41. Bock, K., Schuster-Kolbe, J., Altman, E., Allmaier, G., Stahl, B., Christian, R., Sleytr, U. B., and Messner, P. (1994) *J. Biol. Chem.* **269,** 7137–7144
42. Wang, Q., Fang, X., Bai, B., Liang, X., Shuler, P. J., Goddard, W. A., III, and Tang, Y. (2007) *Biotechnol. Bioeng.* **98,** 842–853
43. Szymanski, C. M., Logan, S. M., Linton, D., and Wren, B. W. (2003) *Trends Microbiol.* **11,** 233–238
44. Glover, K. J., Weerapana, E., and Imperiali, B. (2005) *Proc. Natl. Acad. Sci. U. S. A.* **102,** 14255–14259
45. Weerapana, E., Glover, K. J., Chen, M. M., and Imperiali, B. (2005) *J. Am. Chem. Soc.* **127,** 13766–13767
46. Young, N. M., Brisson, J. R., Kelly, J., Watson, D. C., Tessier, L., Lanthier, P. H., Jarrell, H. C., Cadotte, N., St Michael, F., Aberg, E., and Szymanski, C. M. (2002) *J. Biol. Chem.* **277,** 42530–42539
47. Wacker, M., Feldman, M. F., Callewaert, N., Kowarik, M., Clarke, B. R., Pohl, N. L., Hernandez, M., Vines, E. D., Valvano, M. A., Whitfield, C., and Aebi, M. (2006) *Proc. Natl. Acad. Sci. U. S. A.* **103,** 7088–7093
48. Whitfield, C. (2006) *Annu. Rev. Biochem.* **75,** 39–68
49. Kido, N., Torgov, V. I., Sugiyama, T., Uchiya, K., Sugihara, H., Komatsu, T., Kato, N., and Jann, K. (1995) *J. Bacteriol.* **177,** 2178–2187
50. Schäffer, C., Müller, N., Christian, R., Graninger, M., Wugeditsch, T., Scheberl, A., and Messner, P. (1999) *Glycobiology* **9,** 407–414
51. Kählig, H., Kolarich, D., Zayni, S., Scheberl, A., Kosma, P., Schäffer, C., and Messner, P. (2005) *J. Biol. Chem.* **280,** 20292–20299
52. Sugiyama, T., Kido, N., Kato, Y., Koide, N., Yoshida, T., and Yokochi, T. (1998) *J. Bacteriol.* **180,** 2775–2778
53. Vinogradov, E., Frirdich, E., MacLean, L. L., Perry, M. B., Petersen, B. O., Duus, J. O., and Whitfield, C. (2002) *J. Biol. Chem.* **277,** 25070–25081
54. Cuthbertson, L., Kimber, M. S., and Whitfield, C. (2007) *Proc. Natl. Acad. Sci. U. S. A.* **104,** 19529–19534
55. Cuthbertson, L., Powers, J., and Whitfield, C. (2005) *J. Biol. Chem.* **280,** 30310–30319
56. Kido, N., Sugiyama, T., Yokochi, T., Kobayashi, H., and Okawa, Y. (1998) *Mol. Microbiol.* **27,** 1213–1221
57. Clarke, B. R., Cuthbertson, L., and Whitfield, C. (2004) *J. Biol. Chem.* **279,** 35709–35718
58. Faridmoayer, A., Fentabil, M. A., Mills, D. C., Klassen, J. S., and Feldman, M. F. (2007) *J. Bacteriol.* **189,** 8088–8098
59. Wacker, M., Linton, D., Hitchen, P. G., Nita-Lazar, M., Haslam, S. M., North, S. J., Panico, M., Morris, H. R., Dell, A., Wren, B. W., and Aebi, M. (2002) *Science* **298,** 1790–1793
60. Kowarik, M., Numao, S., Feldman, M. F., Schulz, B. L., Callewaert, N., Kiermaier, E., Catrein, I., and Aebi, M. (2006) *Science* **314,** 1148–1150
61. Steiner, K., Hanreich, A., Kainz, B., Hitchen, P. G., Dell, A., Messner, P., and Schäffer, C. (2007) *Small*, in press

MOLECULAR BASIS OF S-LAYER GLYCOPROTEIN GLYCAN BIOSYNTHESIS IN *GEOBACILLUS STEAROTHERMOPHILUS**

Kerstin Steiner[‡,1], René Novotny[‡,2], Daniel B. Werz[§,3], Kristof Zarschler[‡], Peter H. Seeberger[§], Andreas Hofinger[¶], Paul Kosma[¶], Christina Schäffer[‡,4], and Paul Messner[‡,4]

SUPPLEMENTARY METHODS

Materials-All chemicals used were reagent grade and used as supplied except where noted. Dichloromethane (CH_2Cl_2) was purchased from JT Baker and purified by a Cycle-Tainer Solvent Delivery System. Pyridine was refluxed over calcium hydride and distilled prior to use. Analytical thin-layer chromatography was performed on E. Merck silica gel 60 F_{254} plates (0.25 mm). Compounds were visualized by dipping the plates in a cerium sulfate ammonium molybdate solution or a sulfuric acid/methanol solution followed by heating. Liquid chromatography was performed using forced flow of the indicated solvent on Sigma H-type silica (10–40 mm). ^{1}H NMR spectra were obtained on a Varian VXR-300 (300 MHz), Bruker-600 (600 MHz) and are reported in parts per million (δ) relative to $CHCl_3$ (7.26 ppm) or in the case of CD_3OD as solvent relative to TMS (0.00 ppm). Coupling constants (J)[1] are reported in Hertz. ^{13}C NMR spectra were obtained on a *Varian* VXR-300 (75 MHz), Bruker-600 (150 MHz) and are reported in δ relative to $CDCl_3$ (77.0 ppm) as an internal reference or to TMS (0.00 ppm).

Synthesis of acceptor substrate-The synthesis of the β-D-Gal-(1→*O*)-octyl (**I**), α-L-Rha-(1→3)-β-D-Gal-(1→*O*)-octyl (**II**) and α-L-Rha-(1→3)-α-L-Rha-(1→3)-β-D-Gal-(1→*O*)-octyl (**III**) acceptors is shown in Fig. S1. Building block dibutyl 4,6-di-*O*-benzyl-3-*O*-fluorenylmethoxycarbonyl-2-*O*-pivaloyl-β-D-galactopyranosyl phosphate **a** (27) was reacted using common glycosylation conditions with 1-octanol to afford **b**. Fmoc deprotection furnished compound **c** (*n*-octyl 4,6-di-*O*-benzyl-2-*O*-pivaloyl-β-D-galactopyranoside), which served as a starting point for the synthesis of all three substrates. For the synthesis of the trisaccharide (**III**) an (2+1) approach was envisioned. For this approach the dirhamnose building block **f** was synthesized. The MP protected rhamnose unit **d** (28) was glycosylated at 0 °C with another rhamnose fragment to furnish **e**. A removal of the MP group using cerium ammonium nitrate (CAN) in aqueous solution and the subsequent reaction with trichloroacetonitrile yielded the dirhamnose building block **f** in good yield.

A removal of the pivaloyl group of compound **c** using strong basic conditions and the global deprotection of all the benzyl groups afforded β-D-Gal-(1→*O*)-octyl (**I**). For the synthesis of disaccharide (**II)**, compound **c** was glycosylated using a rhamnose building block with Ac in position 2 to ensure α-selectivity. Column chromatography to purify the reation products yielded an inseparable mixture of the desired compound together with (Overman)-rearranged trichloroacetimidate. Therefore, this mixture was subjected to basic deprotection removing acetyl (Ac) and pivaloyl (Piv) moieties. Using this approach disaccharide **g** was obtained as a pure compound. Global deprotection using $Pd(OH)_2$/C and H_2 in methanol/dichloromethane afforded disaccharide (**II**).

Glycosylating agent **f** as well as octyl galactoside **c** were utilized for the assembly of the trisaccharide **h**. A deprotection of all ester moieties followed by a hydrogenation in order to remove all benzyl groups afforded the completely deprotected trisaccharide (**III**). All the deprotected trisaccharide was purified by silica gel chromatography using mixtures of dichloromethane/methanol in order to remove all traces of palladium which could cause problems in biological experiments.

***n*-Octyl 4,6-Di-*O*-benzyl-2-*O*-pivaloyl-β-D-galactopyranoside (c)**-Compound **a** (960 mg, 1.14 mmol) (27) was azeotroped three times with toluene and dried *in vacuo*. Dichloromethane (25 ml) were added and the solution cooled to -40 °C. 1-Octanol (296 mg, 359 µl, 2.28 mmol) and afterwards TMSOTf (253 mg, 206 µl, 1.14 mmol) were added. The mixture was stirred for 1 h and quenched by addition of some drops of pyridine before the solvent was removed *in vacuo*. The resulting crude product was purified by flash chromatography (5:1 hexane/EtOAc) to afford 631 mg (71%) of **b** as a colorless oil. Compound **b** (585 mg, 0.75 mmol) was dissolved in 8 ml of DMF. Piperidine (2 ml) was added and the mixture stirred for 90 min at room temperature. The solvent was removed *in vacuo* and the residue purified by column chromatography (4:1, hexane/EtOAc) to afford 413 mg (99%) of **c** as a

colorless oil: $[\alpha]_D$: +17.0 (c = 0.41, $CHCl_3$). IR (thin film, $CHCl_3$): 3000, 1742, 1440, 1260, 1077 cm^{-1}. ^1H NMR ($CDCl_3$, 300 MHz) δ 0.90 (t, *J* = 6.4 Hz, 3H), 1.25 (s, 9H), 1.28 (b, 10H), 1.55 (b, 2H), 2.39 (d, *J* = 9.6 Hz, 2H), 3.45 (quart, *J* = 6.9 Hz, 1H), 3.65 (m, 4H), 3.86 (m, 2H), 4.38 (d, *J* = 7.8 Hz, 1H), 4.52 (m, 2H), 4.71 (s, 2H), 5.00 (dd, *J* = 10.2, 7.8 Hz, 1H), 7.25-7.39 (m, 10H). ^{13}C NMR ($CDCl_3$, 75 MHz) δ 14.2, 22.8, 26.1, 27.2, 29.3, 29.5, 29.7, 31.9, 38.9, 68.3, 69.8, 73.4, 73.5, 75.4, 76.5, 101.0, 127.8, 128.0, 128.1, 128.4, 128.4, 137.6, 138.0, 178.5. ESI-MS: *m/z* $[M + Na]^+$ calculated 579.3292, observed 579.3283.

***p*-Methoxyphenyl 2-*O*-Acetyl-3,4-di-*O*-benzyl-α-L-rhamnopyranosyl-(1→3)-2-*O*-acetyl-4-*O*-benzyl-α-L-rhamnopyranoside (e)**-Rhamnose building block **d** (237 mg, 0.589 mmol) (28) and rhamnosyl trichloroacetimidate (406 mg, 0.766 mmol) (26) were azeotroped three times with toluene and dried *in vacuo*. Dichloromethane (9 ml) were added and the solution cooled to 0 °C. TMSOTf (20 mg, 11 µl, 0.09 mmol) was added. The mixture was stirred for 90 min and quenched by addition of some drops of pyridine. Afterwards the solvent was removed *in vacuo*. The resulting crude product was purified by column chromatography (7:3, hexane/EtOAc) to afford 448 mg (98%) of **e** as a colorless oil: $[\alpha]_D$: -45.3 (c = 1.08, $CHCl_3$). IR (thin film, $CHCl_3$): 3005, 2933, 1739, 1585, 1503, 1456, 1369, 1226, 1097, 1046 cm^{-1}. ^1H NMR ($CDCl_3$, 300 MHz) δ 1.32 (pt, *J* = 7.2 Hz, 6H), 2.15 (s, 3H), 2.17 (s, 3H), 3.47 (t, *J* = 9.6 Hz, 1H), 3.54 (t, *J* = 9.6 Hz, 1H), 3.77 (s, 3H), 3.84 (m, 1H), 3.92 (m, 2H), 4.34 (dd, *J* = 9.3 Hz, 3.3 Hz, 1H), 4.48 (d, *J* = 11.4 Hz, 1H), 4.62-4.68 (m, 3H), 4.85 (d, *J* = 10.8 Hz, 1H), 4.93 (d, *J* = 11.1 Hz, 1H), 5.11 (ps, 1H), 5.32 (m, 1H), 5.37 (ps, 1H), 5.49 (m, 1H), 6.81 (pd, *J* = 9.3 Hz, 2H), 6.97 (pd, *J* = 9.3 Hz, 2H), 7.25-7.39 (m, 15H). ^{13}C NMR ($CDCl_3$, 75 MHz) δ 18.0, 18.1, 21.4, 21.2, 55.7, 68.4, 68.8, 69.1, 71.7, 72.1, 75.3, 75.5, 77.6, 79.8, 80.1, 95.9, 99.8, 114.5, 117.6, 127.6, 127.8, 127.8, 127.9, 127.9, 128.3, 128.4, 137.8, 138.3, 149.9, 154.9, 170.0, 170.1. MALDI-MS: *m/z* $[M + Na]^+$ calculated 793.3195, observed 793.3182.

2-*O*-Acetyl-3,4-di-*O*-benzyl-α-L-rhamnopyranosyl-(1→3)-2-*O*-acetyl-4-*O*-benzyl-α-L-rhamnopyranosyl trichloroacetimidate (f)-Disaccharide **e** (420 mg, 0.545 mmol) was suspended in a mixture of acetonitrile and water (20 ml, 1:1). The mixture was stirred for 2 h at room temperature until TLC control experiments showed that no starting material was left. The mixture was poured onto brine and extracted 2x with EtOAc. The combined organic phases were washed twice with water, dried over Na_2SO_4 and concentrated. The resulting crude product was purified by column chromatography (2:1, hexane/EtOAc) to afford 334 mg (92%) of the hemiacteal as a colorless oil. This hemiacetal (285 mg, 0.429 mmol) was dissolved in dichloromethane (5 ml) and trichloroacetonitrile (5 ml). NaH (5 mg) was added and the mixture was stirred for 90 min at room temperature. The solvent was removed *in vacuo*, column chromatography (3:1, hexane/EtOAc) furnished 305 mg (88%) of **f** as a slightly yellow oil: $[\alpha]_D$: -29.0 (c = 0.70, $CHCl_3$). IR (thin film, $CHCl_3$): 2995, 1744, 1672, 1615, 1451, 1369, 1092 cm^{-1}. ^1H NMR ($CDCl_3$, 300 MHz) δ 1.29 (d, *J* = 6.3 Hz, 3H), 1.36 (d, *J* = 6.0 Hz, 3H), 2.14 (s, 3H), 2.17 (s, 3H), 3.46 (t, *J* = 9.3 Hz, 1H), 3.59 (t, *J* = 9.6 Hz, 1H), 3.82 (m, 1H), 3.88-3.99 (m, 2H), 4.25 (dd, *J* = 9.6 Hz, 3.3 Hz, 1H), 4.51 (d, *J* = 11.1 Hz, 1H), 4.62-4.70 (m, 3H), 4.84 (d, *J* = 11.1 Hz, 1H), 4.92 (d, *J* = 11.1 Hz, 1H), 5.09 (ps, 1H), 5.31 (m, 1H), 5.50 (m, 1H), 6.21 (ps, 1H), 7.26-7.39 (m, 15H), 8.71 (s, 1H). ^{13}C NMR ($CDCl_3$, 75 MHz) δ 18.0, 18.1, 21.1, 21.1, 68.9, 69.0, 70.7, 70.7, 71.7, 75.2, 75.7, 76.1, 79.7, 79.8, 90.8, 94.3, 99.7, 127.6, 127.6, 127.9, 128.0, 128.3, 128.3, 128.4, 137.5, 137.8, 138.3, 159.9, 169.7, 169.9. MALDI-MS: *m/z* $[M + Na]^+$ calculated 830.1872, observed 830.1858.

***n*-Octyl β-D-galactopyranoside (I)**-Monosaccharide **c** (255 mg, 0.459 mmol) was dissolved in THF (20 ml), LiOH solution (1 M, 25 ml) and hydrogen peroxide solution (15 ml) were added at -5 °C. The reaction mixture was stirred for 2 days while warming to room temperature. Then methanol (15 ml) and potassium hydroxide (3 M, 30 ml) were added and stirred for 2 days. Hydrochloric acid (1 M) was added until pH 7 was reached. The neutral solution was extracted three times with dichloromethane, dried over $MgSO_4$, filtered and concentrated. Column chromatography (EtOAc) furnished 209 mg (97%) of a colorless solid. This compound (148 mg, 0.263 mmol) was dissolved in methanol (15 ml), $Pd(OH)_2/C$ (15 mg) was added and the Ar atmosphere replaced by a H_2 atmosphere. The reaction mixture was stirred for 12 h. Filtration through a plug of celite removed all solids. Column chromatography using silica gel and dichloromethane/methanol (1:1) in order to remove traces of Pd yielded 75 mg (98%) of **I** as a colorless wax: $[\alpha]_D$: -10.0 (c = 0.13, H_3COH). ^1H NMR (CD_3OD, 300 MHz) δ 0.89 (b, 3H), 1.29-1.37 (b, 10H), 1.61 (m, 2H), 3.47-3.53 (m, 4H), 3.73 (m, *J* = 6.0 Hz, 2H), 3.84 (s, b, 1H), 3.88 (m, 1H), 4.20 (d, *J* = 6.6 Hz, 1H). ^{13}C NMR (CD_3OD, 75 MHz) δ

14.0, 23.3, 26.7, 30.0, 30.2, 30.4, 32.6, 62.0, 69.8, 70.4, 72.1, 74.6, 76.1, 104.5. MALDI-MS: *m/z* [M + Na]$^+$ calculated 315.1778, observed 315.1775.

***n*-Octyl 3,4-di-*O*-benzyl-α-L-rhamnopyranosyl-(1→3)-4,6-di-*O*-benzyl-β-D-galactopyranoside (g)-**Compound **c** (185 mg, 0.332 mmol) and rhamnosyl trichloroacetimidate (228 mg, 0.431 mmol) were azeotroped three times with toluene and dried *in vacuo*. Dichloromethane (6 ml) were added and the solution cooled to 0 °C. TMSOTf (11 mg, 9 µl, 0.05 mmol) was added. The mixture was stirred for 1 h and quenched by addition of some drops of pyridine. Afterwards the solvent was removed *in vacuo*. The resulting crude product was purified by column chromatography (4:1, hexane/EtOAc) to afford a mixture of disaccharide and rearranged trichloroacetimidate. Also, further attempts to separate the compounds failed. Therefore, the mixture was dissolved in THF (35 ml), LiOH solution (1 M) (19 ml) and hydrogen peroxide solution (12 ml) were added at -5 °C. The reaction mixture was stirred for 30 h while warming to room temperature. Then methanol (17 ml) and potassium hydroxide (3 M, 35 ml) were added and stirred for 3 d. Hydrochloric acid (1 M) was added until pH 7 was reached. The neutral solution was extracted three times with dichloromethane, dried over $MgSO_4$, filtered and concentrated. Column chromatography (3:1, hexane/EtOAc) furnished 145 mg (55%) of **g** as a colorless solid: $[\alpha]_D$: -41.2 (c = 0.25, $CHCl_3$). IR (thin film, $CHCl_3$): 3005, 2923, 1600, 1492, 1451, 1082 cm^{-1}. ^{1}H NMR ($CDCl_3$, 300 MHz) δ 0.87 (t, *J* = 6.9 Hz, 3H), 1.33 (m, 15H), 1.61 (m, 2H), 2.33 (s, 1H), 2.52 (s, 1H), 3.48 (m, 2H), 3.62 (m, 4H), 3.78-3.92 (m, 7H), 4.20 (m, 2H), 4.46 (m, 2H), 4.52-4.66 (m, 4H), 4.83 (d, *J* = 11.7 Hz, 1H), 4.90 (d, *J* = 11.1 Hz, 1H), 5.38 (s, 1H), 7.23-7.37 (m, 20H). ^{13}C NMR ($CDCl_3$, 75 MHz) δ 14.0, 18.0, 22.5, 25.8, 29.1, 29.3, 29.4, 31.7, 68.1, 68.3, 68.6, 70.1, 71.7, 72.6, 73.5, 73.7, 74.9, 75.1, 75.9, 77.9, 79.5, 79.7, 99.9, 103.1, 127.4, 127.6, 127.7, 127.8, 128.1, 128.2, 128.3, 128.4, 137.7, 137.8, 138.4, 138.5. MALDI-MS: *m/z* [M + Na]$^+$ calculated 821.4235, observed 821.4247.

***n*-Octyl α-L-rhamnopyranosyl-(1→3)-β-D-galactopyranoside (II)-**Disaccharide **g** (103 mg, 0.129 mmol) was dissolved in a mixture of methanol and dichloromethane (30 ml, 2:1), $Pd(OH)_2$/C was added and the Ar atmosphere was replaced by a H_2 atmosphere. The reaction mixture was stirred for 12 h. Filtration through a plug of celite removed all solids. Column chromatography using silica gel and dichloromethane/methanol (10:1 → 3:1) in order to remove traces of Pd yielded 50 mg (quant.) of (**II**) as a colorless solid: $[\alpha]_D$: -32.3 (c = 0.13, H_3COH). ^{1}H NMR (CD_3OD, 300 MHz) δ 0.89 (b, 3H), 1.24-1.36 (b, 13H), 1.61 (m, 2H), 3.31 (s, 2H), 3.39 (m, 1H), 3.51 (m, 2H), 3.63 (m, 1H), 3.71 (m, 3H), 3.92 (m, 3H), 4.23 (d, *J* = 7.5 Hz, 1H), 5.05 (s, 1H). ^{13}C NMR (CD_3OD, 75 MHz) δ 14.0, 17.6, 23.3, 26.7, 30.0, 30.1, 30.4, 32.6, 61.8, 69.6, 69.7, 70.5, 71.7, 73.6, 76.0, 81.2, 103.4, 104.5. MALDI-MS: *m/z* [M + Na]$^+$ calculated 461.2357, observed 461.2357.

***n*-Octyl 2-*O*-acetyl-3,4-di-*O*-benzyl-α-L-rhamnopyranosyl-(1→3)-2-*O*-acetyl-4-*O*-benzyl-α-L-rhamnopyranosyl-(1→3)-4,6-di-*O*-benzyl-2-*O*-pivaloyl-β-D-galactopyranoside (h)-**Galactose building block **c** (100 mg, 0.179 mmol) and rhamnose disaccharide building block **f** (209 mg, 0.258 mmol) were azeotroped three times with toluene and dried *in vacuo*. Dichloromethane (3.5 ml) were added and the solution cooled to 0 °C. TMSOTf (9 mg, 7 µl, 0.04 mmol) was added. The mixture was stirred for 120 min and quenched by addition of some drops of pyridine. Afterwards the solvent was removed *in vacuo*. The resulting crude product was purified by column chromatography (4:1, hexane/EtOAc) to afford 164 mg (76%) of **h** as a colorless oil: $[\alpha]_D$: -25.8 (c = 0.33, $CHCl_3$). IR (thin film, $CHCl_3$): 2933, 2871, 1739, 1369, 1072 cm^{-1}. ^{1}H NMR ($CDCl_3$, 300 MHz) δ 0.88 (t, *J* = 6.9 Hz, 3H), 1.24-1.34 (m, 25H), 1.58 (m, 2H), 2.10 (s, 3H), 2.13 (s, 3H), 3.41 (m, 4H), 3.64 (ps, 3H), 4.78-4.88 (m, 5H), 4.07 (dd, *J* = 10.7, 2.2 Hz, 1H), 4.38 (d, *J* = 10.1 Hz, 1H), 4.40-4.68 (m, 7H), 4.80 (d, *J* = 12.3 Hz, 1H), 5.91 (m, 3H), 5.17 (ps, 1H), 5.21 (ps, 2H), 7.24-7.39 (m, 25H). ^{13}C NMR ($CDCl_3$, 75 MHz) δ 14.2, 18.0, 18.2, 21.1, 21.1, 22.8, 26.1, 27.2, 29.3, 29.5, 29.7, 31.9, 38.9, 68.6, 68.7, 68.8, 69.0, 69.7, 71.5, 71.7, 73.5, 73.7, 74.8, 75.0, 75.1, 76.1, 76.2, 77.6, 78.2, 79.8, 80.3, 98.7, 99.1, 101.6, 127.4, 127.4, 127.5, 127.6, 127.7, 127.7, 127.8, 127.8, 128.2, 128.2, 128.3, 128.3, 137.7, 137.8, 137.9, 138.1, 138.5, 169.3, 169.9, 176.5. MALDI-MS: *m/z* [M + Na]$^+$ calculated 1225.6070, observed 1225.6050.

***n*-Octyl α-L-rhamnopyranosyl-(1→3)-α-L-rhamnopyranosyl-(1→3)-β-D-galactopyranoside (III)-**Trisaccharide **h** (135 mg, 0.112 mmol) was dissolved in THF (13 mL), LiOH solution (1 M, 7 ml) and hydrogen peroxide solution (5 ml) were added at -5 °C. The reaction mixture was stirred for 3 d while warming to room temperature. Then methanol (7 ml) and potassium hydroxide (3 M, 13 ml) were added and stirred for 2 d. Hydrochloric acid (1 M) was added until pH 7 was reached. The

neutral solution was extracted three times with dichloromethane, dried over $MgSO_4$, filtered and concentrated. Mass spectrometry still shows a pivaloyl group in the molecule. Therefore, the crude product was dissolved in methanol (10 ml) and NaOMe (500 mg) were added. The mixture was stirred for 12 h and the solvent removed *in vacuo*. Column chromatography (EtOAc) furnished 102 mg (88%) of a colorless solid. This compound (95 mg, 0.092 mmol) was dissolved in a mixture of methanol (10 ml) and dichloromethane (15 mL), $Pd(OH)_2/C$ was added and the Ar atmosphere replaced by a H_2 atmosphere. The reaction mixture was stirred for 12 h. Filtration through a plug of celite removed all solids. Column chromatography using silica gel and dichloromethane/methanol (10:1 $\rightarrow$ 4:1) was performed in order to remove traces of Pd. 38 mg (quant.) of (**III**) as a colorless highly viscous oil were obtained: $[\alpha]_D$: -60.5 (c = 1.25, H_3COH). ^{1}H NMR (CD_3OD, 300 MHz) δ 0.89 (b, 3H), 1.24-1.33 (b, 16H), 1.61 (b, 2H), 3.38 (s, 1H), 3.47-3.57 (m, 4H), 3.62-3.92 (m, 9H), 3.99 (s, b, 1H), 4.07 (s, b, 1H), 4.23 (d, *J* = 7.5 Hz, 1H), 5.02 (s, 1H), 5.04 (s, 1H). ^{13}C NMR (CD_3OD, 75 MHz) δ 14.0, 17.6, 17.6, 23.3, 26.7, 30.0, 30.1, 30.4, 32.6, 61.8, 69.6, 69.7, 70.0, 70.5, 71.5, 71.7, 72.8, 73.7, 76.0, 79.4, 81.1, 103.4, 103.6, 104.6. MALDI-MS: *m/z* $[M + Na]^+$ calculated 607.2936, observed 607.2925.

References included in supplementary methods:

26. Fürstner, A., Müller, T. (1999) *J. Am. Chem. Soc.* **121**, 7814-7821
27. Love, K. R., Seeberger, P. H. (2004) *Angew. Chem. Int. Ed.* **43**, 602-605
28. Werz, D. B., Seeberger, P. H. (2005) *Angew. Chem. Int. Ed.* **44**, 6315-6318

[1]The abbreviations used are: Ac, acetyl; Ar, argon; Bn, benzyl; CAN, cerium ammonium nitrate; EtOAc, ethyl acetate; Fmoc, fluorenylmethoxycarbonyl; H_2, hydrogen; IR, IR spectroscopy; *J*, coupling constant; MP, methoxyphenyl; NaOMe, sodium methoxide; OBu, *O*-butyl; Pd, palladium; Piv, pivaloyl; THF, tetrahydrofuran; TMSOTf, trimethylsilyl trifluoromethanesulfonate.

SUPPLEMENTARY RESULTS

Sequence comparison of the transferases WsaC, WsaD, WsaE and WsaF–WsaC shows homology to the α-1,3-L-rhamnosyltransferases from *Streptococcus pyogenes* MGAS10394 (62) (protein accession number YP_059939) and MGAS6180 (63) (YP_280052) as well as to RgpB from *Lactococcus lactis* and *Streptococcus thermophilus*, RgpBc from *Streptococcus mutans*, and Cps2F from *Streptococcus pneumoniae*, and several other enzymes, all of which have been classified as rhamnosyltransferases involved in cell wall biosynthesis (Fig. S2). WsaD shows high homology to the putative rhamnosyltransferase JexE from *Paenibacillus jamilae*, the glycosyltransferases from *Clostridium beijerincki* (protein accession number ZP_00911019) and *Methanosphaera stadtmanae*, as well as some homology to rhamnosyltransferases from *Shigella dysenteriae* (37) and *Shigella flexneri* (64).

The N-terminal portion of WsaE (aa 70 to 150) revealed homology to methyltransferases *e.g.*, to COG2226.2, UbiE, a menaquinone biosynthesis methyltransferase from *Methanosarcina acetivorans* (65), and to PFAM 08241.1 and PFAM 08242.1 methyltransferase family 11 and 12, respectively, both of which are SAM-dependent methyltransferases (Fig. S2). The central and C-terminal portions contain two glycosyltransferase domains, which are homologous *e.g.*, to the O-antigen biosynthesis protein from *Planctomyces maris* DSM 8797 (ZP_01855298) and *Xanthomonas oryzae* (ABI93188), or to the GT-2 glycosyltransferases from *Burkholderia cenocepacia* MC0-3 (ZP_01563358) and from *Enterococcus faecalis* (66) (ORFde16, AAC35930) (Fig. S3). The first glycosyltransferase domain contains motifs that are typical of inverting glycosyltransferases and contains a conserved DD motif (aa 638-639) and, 53 aa downstream, the DXDD motif (DHDD, aa 691-694). The second glycosyltransferase domain contains ED motifs (aa 970-971 and 981-982) and a DXE motif (aa 1062-1064).

Highest homology for WsaF was found to conserved hypothetical proteins from *Planctomyces maris* (ZP_01855299; 430 aa; E-value: $2e^{-89}$) and from *Streptococcus pneumoniae* (CAI34499; 414 aa, E-value: $8e^{-70}$). Homologies to putative glycosyltransferases, for instance, from *Anabaena variabilis* (protein accession number ABA22956; 406 aa, E-value: $6e^{-20}$) and WbbX (421 aa, E-value: $2e^{-08}$) from *Yersinia enterocolitica* were found as well.

References included in supplementary results:

37. Morona, R., Macpherson, D. F., Van Den, B. L., Carlin, N. I., and Manning, P. A. (1995) *Mol. Microbiol.* **18,** 209-223
62. Banks, D. J., Porcella, S. F., Barbian, K. D., Beres, S. B., Philips, L. E., Voyich, J. M., DeLeo, F. R., Martin, J. M., Somerville, G. A., and Musser, J. M. (2004) *J. Infect. Dis.* **190,** 727-738
63. Green, N. M., Zhang, S., Porcella, S. F., Nagiec, M. J., Barbian, K. D., Beres, S. B., LeFebvre, R. B., and Musser, J. M. (2005) *J. Infect. Dis.* **192,** 760-770
64. Klena, J. D. and Schnaitman, C. A. (1993) *Mol. Microbiol.* **9,** 393-402
65. Galagan, J. E., Nusbaum, C., Roy, A., Endrizzi, M. G., Macdonald, P., FitzHugh, W., Calvo, S., Engels, R., Smirnov, S., Atnoor, D., Brown, A., Allen, N., Naylor, J., Stange-Thomann, N., DeArellano, K., Johnson, R., Linton, L., McEwan, P., McKernan, K., Talamas, J., Tirrell, A., Ye, W., Zimmer, A., Barber, R. D., Cann, I., Graham, D. E., Grahame, D. A., Guss, A. M., Hedderich, R., Ingram-Smith, C., Kuettner, H. C., Krzycki, J. A., Leigh, J. A., Li, W., Liu, J., Mukhopadhyay, B., Reeve, J. N., Smith, K., Springer, T. A., Umayam, L. A., White, O., White, R. H., Conway, d. M., Ferry, J. G., Jarrell, K. F., Jing, H., Macario, A. J., Paulsen, I., Pritchett, M., Sowers, K. R., Swanson, R. V., Zinder, S. H., Lander, E., Metcalf, W. W., and Birren, B. (2002) *Genome Res.* **12,** 532-542
66. Xu, Y., Murray, B. E., and Weinstock, G. M. (1998) *Infect. Immun.* **66,** 4313-4323

SUPPLEMENTARY TABLES

Table S1: Bacterial strains and plasmids used in this work.

Strains or plasmids	Genotype and relevant characteristics	Sources or references
E. coli DH5α™	F$^-$ φ80*lac*ZM*15* (*lacZYA-argF*) *U169 deoR recA1 endA1 hsdR17* (*rk*$^-$, *mk*$^+$) *phoA supE44 thi-1 gyrA96 relA1 λ*$^-$	Invitrogen
E. coli BL21 Star (DE3)	F$^-$ *ompT hsdSB* (*rB*$^-$*mB*$^-$) *gal dcm rne131*	Invitrogen
E. coli C43(DE3)	F$^-$ *ompT hsdSB* (*rB*$^-$*mB*$^-$) *gal dcm* (DE3) C43	Lucigen (67)
pET28a(+)	*E. coli* expression vector; Km^R	Novagen
pNGB220	pET28a-WsaC; pET28a(+) expressing WsaC (aa 1-324) from *G. stearothermophilus* NRS 2004/3a with N-terminal His$_6$-tag; Km^R	This study
pNGB221	pET28a-WsaC_I; pET28a(+) expressing WsaC from *G. stearothermophilus* NRS 2004/3a devoid of the C-terminal transmembrane domain (aa 1-280) with N-terminal His$_6$-tag; Km^R	This study
pNGB230	pET28a-WsaD; pET28a(+) expressing WsaD (aa 1-289) from *G. stearothermophilus* NRS 2004/3a with N-terminal His$_6$-tag; Km^R	This study
pNGB231	pET28a-WsaD_I; pET28a(+) expressing WsaD from *G. stearothermophilus* NRS 2004/3a devoid of the C-terminal transmembrane domain (aa 1-254) with N-terminal His$_6$-tag; Km^R	This study
pNGB240	pET28a-WsaE; pET28a expressing WsaE (aa 1-1127) from *G. stearothermophilus* NRS 2004/3a with N-terminal His$_6$-tag; Km^R	This study
pNGB241	pET28a-WsaE_M; pET28a expressing WsaE_M (aa 1-170) from *G. stearothermophilus* NRS 2004/3a with N-terminal His$_6$-tag; Km^R	This study
pNGB242	pET28a-WsaE_N; pET28a expressing WsaE_N (aa 1-368) from *G. stearothermophilus* NRS 2004/3a with N-terminal His$_6$-tag; Km^R	This study
pNGB243	pET28a-WsaE_B; pET28a expressing WsaE_B (aa 368-1127) from *G. stearothermophilus* NRS 2004/3a with N-terminal His$_6$-tag; Km^R	This study
pNGB244	pET28a-WsaE_C; pET28a expressing WsaE_C (aa 765-1127) from *G. stearothermophilus* NRS 2004/3a with N-terminal His$_6$-tag; Km^R	This study
pNGB245	pET28a-WsaE_A; pET28a expressing WsaE_A (aa 368-863) from *G. stearothermophilus* NRS 2004/3a with N-terminal His$_6$-tag; Km^R	This study
pNGB250	pET28a-WsaF; pET28a(+) expressing WsaF (aa 1-413) from *G. stearothermophilus* NRS 2004/3a with N-terminal His$_6$-tag; Km^R	This study
pNGB200	pET28a-WsaP; pET28a(+) expressing WsaP (aa 1-471) from *G. stearothermophilus* NRS 2004/3a with N-terminal His$_6$-tag; Km^R	(22)
pNGB261	pET28a-RmlB; pET-28a(+) expressing RmlB from *G. stearotermophilus* NRS 2004/3a with N-terminal His$_6$-tag; Km^R	This study
pNGB262	pET28a-RmlC; pET-28a(+) expressing RmlC from *G. stearothermophilus* NRS 2004/3a with N-terminal His$_6$-tag; Km^R	This study
pNGB263	pET28a-RmlD; pET-28a(+) expressing RmlD from *G. stearotermophilus* NRS 2004/3a with N-terminal His$_6$-tag; Km^R	This study

Table S2: PCR primers used for the amplification of WsaD, WsaC, WsaE, WsaF, RmlB, RmlC and RmlD from *G. stearothermophilus* NRS 2004/3a for the design of different expression constructs.

Primer	Nucleotide sequence (5' → 3')	Orientation
pET-WsaD_for	AATCACCATATGGATATTAGCATTATTATCGTGAAT	forward
pET-WsaD_rev	ATAAGAATCTCGAGTTAACCACCTATTTTTCGAAAAGTGT	reverse
pET-WsaD_I_rev	AATCACTCGAGCTAATGCTTCCTATGGAATAAAAACATC	reverse
pET-WsaC_for	AATCACCATATGGAGATGCCATTGGTTT	forward
pET-WsaC_rev	ATAAGAATCTCGAGCTAATATTTTAACTTTTTAAAAATCCATATTG	reverse
pET-WsaC_I_rev	AATCACTCGAGCTATTGACTAAATTTTTTTCTAGCTTCCG	reverse
pET-WsaE_for	AATCAGCTAGCATGGAGCGTTGTAGAATGAATAA	forward
pET-WsaE _B_for	AATCAGCTAGCATGCGTATTAAGAATAGATTAAAAA	forward
pET-WsaE _C_for	AATCAGCTAGCATGGGCTTTCGAAAAGGTTTTG	forward
pET-WsaE_rev	ATAAGAATCTCGAGCTACGACCTTATTGTTGGAATCAAA	reverse
pET-WsaE_A_rev	ATAAGAATCTCGAGCTAAACCAATGGATAATCCTCTG	reverse
pET-WsaE_N_rev	ATAAGAATCTCGAGCTACATAGACTCAGCTTGGTTTTGC	reverse
pET-WsaE_M_rev	ATAAGAATCTCGAGCTATATCTTCCGTTTCCTCTAGG	reverse
pET-WsaF_for	GGGGTACCCCATATGgTTCAAAAATTAATACAGATATTAAG	forward
pET-WsaF_rev	GGGGTACCCCGAGCTCGAAAATAAACCTACGAATAGAGTCATCA	reverse
RmlB_for	AATCAGCTAGCATGAAAGTATTGATTACCGGC	forward
RmlB_rev	AATCACTCGAGCCTAACTGCCCGTTTGC	reverse
RmlC_for	AATCACCATATGgAAATTATTGAGACTAAGTTTAGTAATG	forward
RmlC_rev	AATCACTCGAGCTGCATTTCCTTCCCTTAATAAG	reverse
RmlD_for	AATCACCATATGgAAATTGTTGTTACGGGGG	forward
RmlD_rev	AATCACTCGAGGATCTATTGTAAATATATCACTTCAAATC	reverse

Triplets corresponding to the initiation and termination codons in the primer sequence are boxed. Lowercase letters indicate changes in the original nucleotides sequence. Artificial restriction sites are underlined.

Table S3: ESI-QTOF MS analysis of octyl-linked products of *in vitro* activity assays.

Sample	$[M+Na]^+_{exp}$	$[M+Na]^+_{theor}$	assignment
substrate (**II**)	461.22	461.24	Rha-Gal-octyl
substrate (**III**)	607.25	607.29	Rha-Rha-Gal-octyl
WsaF + (**II**) + dTDP-β-L-Rha	607.26	607.29	Rha-Rha-Gal-octyl (**IV**)
WsaC + (**III**) + dTDP-β-L-Rha	753.33	753.35	Rha-Rha-Rha-Gal-octyl (**V**)
WsaC + (**III**) + dTDP-β-L-Rha	899.37	899.41	Rha-Rha-Rha-Rha-Gal-octyl (**VI**)
WsaE + (**III**) + dTDP-β-L-Rha	753.33	753.35	Rha-Rha-Rha-Gal-octyl (**VII**)

References included in supplementary tables:

22. Steiner, K., Novotny, R., Patel, K., Vinogradov, E., Whitfield, C., Valvano, M. A., Messner, P., and Schäffer, C. (2007) *J. Bacteriol.* **189,** 2590-2598
67. Miroux, B. and Walker, J. E. (1996) *J. Mol. Biol.* **260,** 289-298

SUPPLEMENTARY FIGURE LEGENDS

Fig. S1. **Reaction scheme** for the synthesis of β-D-Gal-(1→*O*)-octyl (**I**), α-L-Rha-(1→3)-β-D-Gal-(1→*O*)-octyl (**II**) and α-L-Rha-(1→3)-α-L-Rha-(1→3)-β-D-Gal-(1→*O*)-octyl (**III**) used as acceptors in this study.

Fig. S2. **Multiple sequence alignment of the N-terminal region of WsaC** from *G. stearothermophilus* NRS 2004/3a with putative rhamnosyltransferases of two different *Streptococcus pyogenes* strains, RgbB from *Lactococcus lactis* and RgpBc from *Streptococcus mutans*. Conserved amino acids described for inverting transferases are highlighted.

Fig. S3. **Sequence alignment** of the N-terminal region of WsaE from *G. stearothermophilus* NRS 2004/3a, containing the putative methyltransferase domain, with menaquinone biosynthesis methyltransferase (UbiE) from *Methanosarcina acetivorans* C2A.

Fig. S4. **Sequence alignment** of the C-terminal region of WsaE from *G. stearothermophilus* NRS 2004/3a, containing the putative rhamnosyltransferase domains, with O-antigen biosynthesis proteins from *Planctomyces maris* DSM 8797 and (protein accession number ZP_01855298) and *Xanthomonas oryzae* (ABI93188) or the GT-2 glycosyltransferases from *Burkholderia cenocepacia* MC0-3 (ZP_01563358) and *Enterococcus faecalis* (AAC35930).

Fig. S5. A: **Western immunoblot analysis of the expression of WsaC, WsaD and WsaF** in *E. coli* BL21 Star (DE3). Proteins were detected with anti-His-tag antibody. Lanes 1, 6 and 8: Precision Plus Protein™ Standard All Blue (Biorad); lane 2: WsaC (37 kDa); lane 3: WsaC_I (36.5 kDa); lane 4: WsaD (30 kD); lane 5: WsaD_I (31.6 kDa); lane 7: WsaF (50.5 kDa). B: **Western Immunoblot analysis of the expression of WsaE and truncated forms thereof** in *E. coli* BL21 Star (DE3). Proteins were detected with anti-His-tag antibody. Lane 1: Precision Plus Protein™ Standard All Blue (Biorad); lane 2: full length WsaE (135.0 kDa); lane 3: WsaE_A (60.5 kDa; aa 368-863, first rhamnosyltransferase domain); lane 4: WsaE_B (90.8 kDa; aa 368-1127, both rhamnosyltransferase domains); lane 5: WsaE_C (44.1 kDa; aa 765-1127, C-terminal rhamnosyltransferase domain) lane 6: WsaE_M (22.7 kDa; aa 1-170, UbiE domain); lane 7: WsaE_N (46.8 kDa; aa 1-368, UbiE domain plus 200 aa downstream). For details about the truncated forms see Figure 1.

Fig. S6. **TLC pattern of rhamnosyltransferase activity assays using octyl-linked oligosaccharides.** For the reactions crude extract (WsaC and WsaD) or purified enzyme (WsaC_I, WsaE and WsaF) were used and the products were separated on Silica TLC plates with chloroform/methanol/water 65:25:4 as solvent. Lanes 1-7: WsaD; lanes 12-18: WsaC; lanes 20-23: WsaC_I; lanes 24-27: WsaE; lanes 28-31: WsaF; lanes 1, 12, 20, 24 and 28: β-D-Gal-(1→*O*)-octyl (**I**) and dTDP-β-L-Rha; lanes 2, 13, 21, 25 and 29: α-L-Rha-(1→3)-β-D-Gal-(1→*O*)-octyl (**II**) and dTDP-β-L-Rha; lanes 3, 14, 22, 26 and 30: α-L-Rha-(1→3)-α-L-Rha-(1→3)-β-D-Gal-(1→*O*)-octyl (**III**) and dTDP-β-L-Rha; lanes 4, 8 and 15: (**I**); lanes 5, 9 and 16: (**II**); lanes 6, 10 and 17: (**III**); lanes 7, 11, 18, 23, 27 and 31: dTDP-β-L-Rha. Staining was performed with thymol.

Fig. S7. **(+)ESI-QTOF MS^2 spectrum of the singly charged ions at *m/z* 899.40** corresponding to the product (**VI**) of the *in vitro* assay of WsaC; the loss of rhamnoses can be deduced from the mass difference of ~146 between neighboring Y-ions.

Figure S1

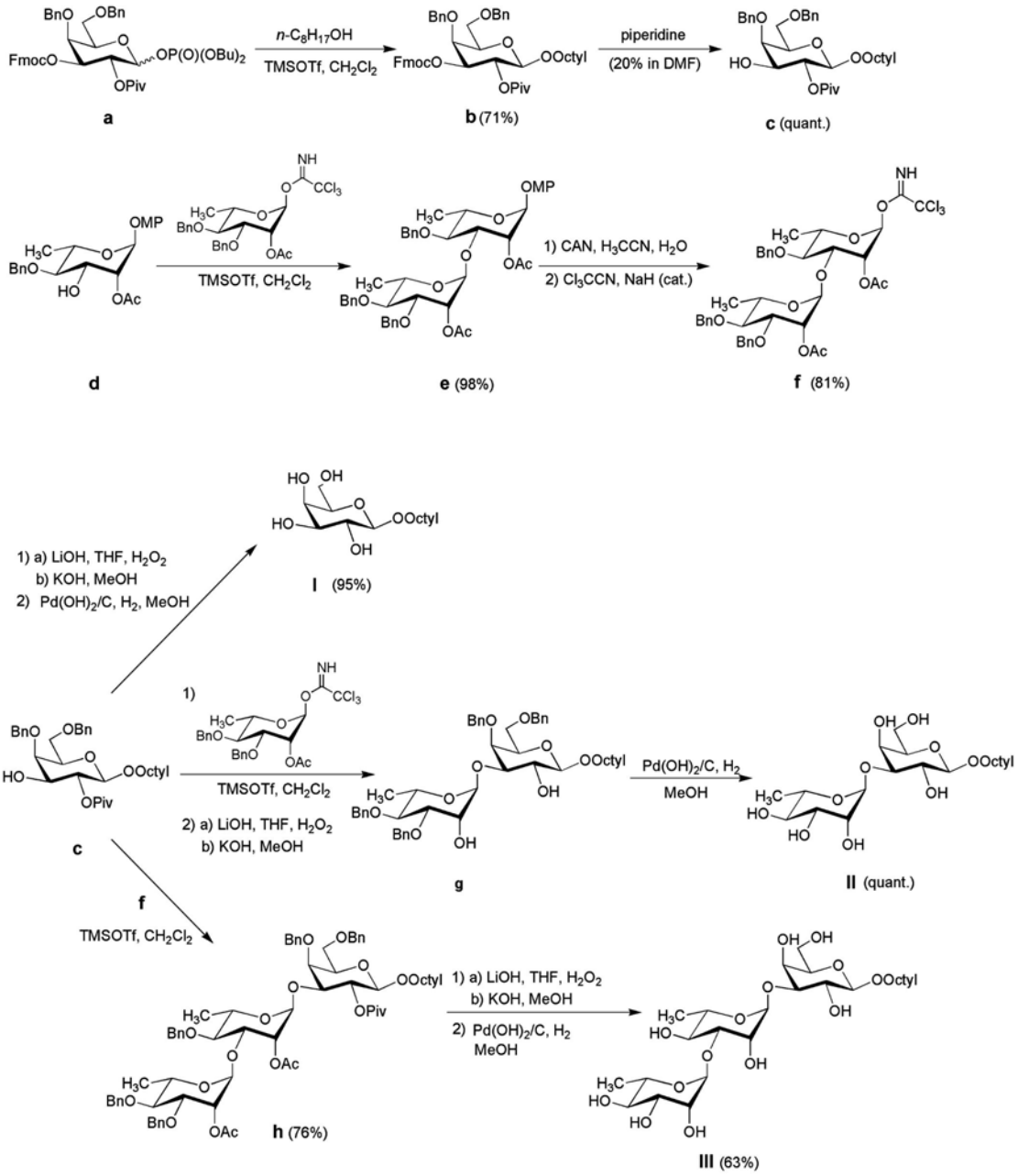


Figure S2


StrepPyo10394 ---MNINILLSTYN-GERFLAEQIQSIQRQTVNDWTLLIRDDGSTDGTQDIIRTFVK-ED 55
StrepPyo6180 ---MNINILLSTYN-GERFLAEQIQSIQRQTVNDWTLLIRDDGSTDGTQDIIRTFVK-ED 55
RgpBc ---MKVNILMSTYN-GQEFIAQQIQSIQKQTFENWNLLIRDDGSSDGTPKIIADFAK-SD 55
RgbBLac ---MRVNILMSTYN-GEKFVADQIESIQKQTYTDWNLIIRDDGSSDRTCEIVDDFVS-KD 55
WsaC MEMPLVSIVVATYFPRTDFFEKQLQSLNNQTYENIEIIICDDSANDAEYEKVKKMVENII 60
DXDD
SP10394 KRIQWINEGQTENLGVIKNFYTLLR-HQKADVYFFSDQDDIWLDNKLEVTLLEAQKHEMT 114
SP6180 KRIQWINEGQTENLGVIKNFYTLLK-HQKADVYFFSDQDDIWLDNKLEVTLLEAQKHEMT 114
RgpBc ARIRFINADKRENFGVIKNFYTLLK-YEKADYYFFSDQDDVWLPQKLELTLASVEKENNQ 114
RgbBLac NRIKLIR---AENVGVIKSFHELVTDSNNADFYFFADQDDYWLPEKLSVMLEETKKHDNS 112
WsaC SRFPCKVIRNEKNVGSNKTFERLTQ-EANGDYICYCDQDDIWLSEKVERLVNHITKHHCT 119
SP10394 APLLVYTDLKVVTQHLAICHDSMIKT----QSGHANTSLLQELTENTVTGGTMMITHALA 170
SP6180 APLLVYTDLKVVTQHLAICHDSMIKT----QSGHANTSLLQELTENTVTGGTMMITHALA 170
RgpBc IPLMVYTDLTVVDRDLQVLHDSMIKT----QSHHANTSLLEELTENTVTGGTMMVNHCLA 170
RgbBLac KPVMYYTDLKVTDRNLNVTSESMIRG----QSDHANTKLVQELTENTVTGGASMINHELA 168
WsaC ---LVYSDLSLIDENDRIIHKSFKRSNFRLKHVHGDNTFAHLINRNSVTGCAMMIRADVA 176
SP10394 EEWTTC--DGLLMHDWYLALLASATGKLVYLDIPTELYRQHDANVLGARTWS--KRMKNW 226
SP6180 EEWTTC--DGLLMHDWYLALLASATGKLVYLDIPTELYRQHDANVLGARTWS--KRMKNW 226
RgpBc KQWKQCY-DDLIMHDWYLALLAASLGKLIYLDETTELYRQHESNVLGARTWS--KRLKNW 227
RgbBLac QLWQST--NDIIMHDWYLAIVAAALGELVYIDQPTHLYRQHDSNVLGARTLS--KRIKKW 224
WsaC KSAIPFPDYDEFVHDHWLAIHAAVKGSLGYIKEPLVWYRIHLGNQIGNQRLVNITNINDY 236

Figure S3

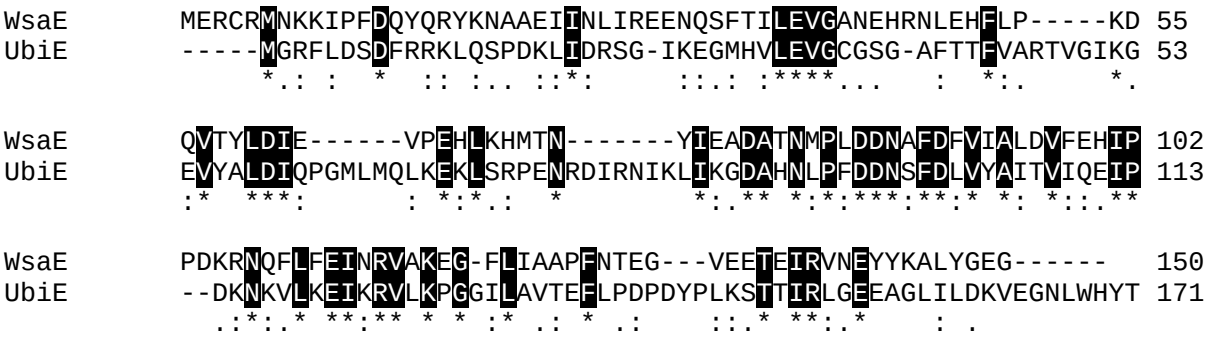

WsaE MERCRMNKKIPFDQYQRYKNAAEIINLIREENQSFTILEVGANEHRNLEHFLP-----KD 55
UbiE -----MGRFLDSDFRRKLQSPDKLIDRSG-IKEGMHVLEVGCGSG-AFTTFVARTVGIKG 53
WsaE QVTYLDIE------VPEHLKHMTN-------YIEADATNMPLDDNAFDFVIALDVFEHIP 102
UbiE EVYALDIQPGMLMQLKEKLSRPENRDIRNIKLIKGDAHNLPFDDNSFDLVYAITVIQEIP 113
WsaE PDKRNQFLFEINRVAKEG-FLIAAPFNTEG---VEETEIRVNEYYKALYGEG------ 150
UbiE --DKNKVLKEIKRVLKPGGILAVTEFLPDPDYPLKSTTIRLGEEAGLILDKVEGNLWHYT 171

Figure S4

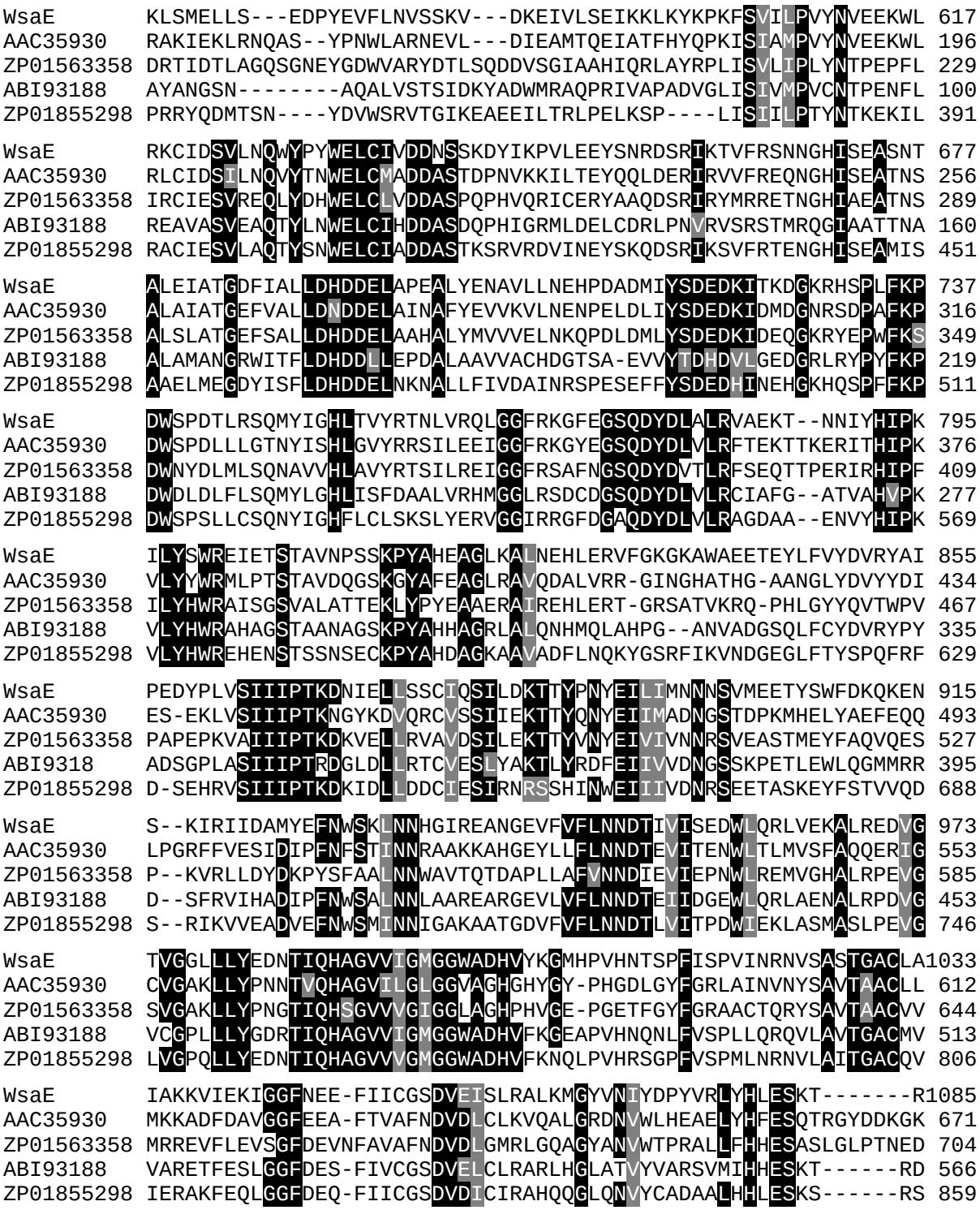

WsaE KLSMELLS---EDPYEVFLNVSSKV---DKEIVLSEIKKLKYKPKFSVILPVYNVEEKWL 617
AAC35930 RAKIEKLRNQAS--YPNWLARNEVL---DIEAMTQEIATFHYQPKISIAMPVYNVEEKWL 196
ZP01563358 DRTIDTLAGQSGNEYGDWVARYDTLSQDDVSGIAAHIQRLAYRPLISVLIPLYNTPEPFL 229
ABI93188 AYANGSN--------AQALVSTSIDKYADWMRAQPRIVAPADVGLISIVMPVCNTPENFL 100
ZP01855298 PRRYQDMTSN----YDVWSRVTGIKEAEEILTRLPELKSP----LISIILPTYNTKEKIL 391
WsaE RKCIDSVLNQWYPYWELCIVDDNSSKDYIKPVLEEYSNRDSRIKTVFRSNNGHISEASNT 677
AAC35930 RLCIDSILNQVYTNWELCMADDASTDPNVKKILTEYQQLDERIRVVFREQNGHISEATNS 256
ZP01563358 IRCIESVREQLYDHWELCLVDDASPQPHVQRICERYAAQDSRIRYMRRETNGHIAEATNS 289
ABI93188 REAVASVEAQTYLNWELCIHDDASDQPHIGRMLDELCDRLPNVRVSRSTMRQGIAATTNA 160
ZP01855298 RACIESVLAQTYSNWELCIADDASTKSRVRDVINEYSKQDSRIKSVFRTENGHISEAMIS 451
WsaE ALEIATGDFIALLDHDDELAPEALYENAVLLNEHPDADMIYSDEDKITKDGKRHSPLFKP 737
AAC35930 ALAIATGEFVALLDNDDELAINAFYEVVKVLNENPELDLIYSDEDKIDMDGNRSDPAFKP 316
ZP01563358 ALSLATGEFSALLDHDDELAAHALYMVVVELNKQPDLDMLYSDEDKIDEQGKRYEPWFKS 349
ABI93188 ALAMANGRWITFLDHDDLLEPDALAAVVACHDGTSA-EVVYTDHDVLGEDGRLRYPYFKP 219
ZP01855298 AAELMEGDYISFLDHDDELNKNALLFIVDAINRSPESEFFYSDEDHINEHGKHQSPFFKP 511
WsaE DWSPDTLRSQMYIGHLTVYRTNLVRQLGGFRKGFEGSQDYDLALRVAEKT--NNIYHIPK 795
AAC35930 DWSPDLLLGTNYISHLGVYRRSILEEIGGFRKGYEGSQDYDLVLRFTEKTTKERITHIPK 376
ZP01563358 DWNYDLMLSQNAVVHLAVYRTSILREIGGFRSAFNGSQDYDVTLRFSEQTTPERIRHIPF 409
ABI93188 DWDLDLFLSQMYLGHLISFDAALVRHMGGLRSDCDGSQDYDLVLRCIAFG--ATVAHVPK 277
ZP01855298 DWSPSLLCSQNYIGHFLCLSKSLYERVGGIRRGFDGAQDYDLVLRAGDAA--ENVYHIPK 569
WsaE ILYSWREIETSTAVNPSSKPYAHEAGLKALNEHLERVFGKGKAWAEETEYLFVYDVRYAI 855
AAC35930 VLYYWRMLPTSTAVDQGSKGYAFEAGLRAVQDALVRR-GINGHATHG-AANGLYDVYYDI 434
ZP01563358 ILYHWRAISGSVALATTEKLYPYEAAERAIREHLERT-GRSATVKRQ-PHLGYYQVTWPV 467
ABI93188 VLYHWRAHAGSTAANAGSKPYAHHAGRLALQNHMQLAHPG--ANVADGSQLFCYDVRYPY 335
ZP01855298 VLYHWREHENSTSSNSECKPYAHDAGKAAVADFLNQKYGSRFIKVNDGEGLFTYSPQFRF 629
WsaE PEDYPLVSIIIPTKDNIELLSSCIQSILDKTTYPNYEILIMNNNSVMEETYSWFDKQKEN 915
AAC35930 ES-EKLVSIIIPTKNGYKDVQRCVSSIIEKTTYQNYEIIMADNGSTDPKMHELYAEFEQQ 493
ZP01563358 PAPEPKVAIIIPTKDKVELLRVAVDSILEKTTYVNYEIVIVNNRSVEASTMEYFAQVQES 527
ABI9318 ADSGPLASIIIPTRDGLDLLRTCVESLYAKTLYRDFEIIVVDNGSSKPETLEWLQGMMRR 395
ZP01855298 D-SEHRVSIIIPTKDKIDLLDDCIESIRNRSSHINWEIIIVDNRSEETASKEYFSTVVQD 688
WsaE S--KIRIIDAMYEFNWSKLNNHGIREANGEVFVFLNNDTIVISEDWLQRLVEKALREDVG 973
AAC35930 LPGRFFVESIDIPFNFSTINNRAAKKAHGEYLLFLNNDTEVITENWLTLMVSFAQQERIG 553
ZP01563358 P--KVRLLDYDKPYSFAALNNWAVTQTDAPLLAFVNNDIEVIEPNWLREMVGHALRPEVG 585
ABI93188 D--SFRVIHADIPFNWSALNNLAAREARGEVLVFLNNDTEIIDGEWLQRLAENALRPDVG 453
ZP01855298 S--RIKVVEADVEFNWSMINNIGAKAATGDVFVFLNNDTLVITPDWIEKLASMASLPEVG 746
WsaE TVGGLLLYEDNTIQHAGVVIGMGGWADHVYKGMHPVHNTSPFISPVINRNVSASTGACLA1033
AAC35930 CVGAKLLYPNNTVQHAGVILGLGGVAGHGHYGY-PHGDLGYFGRLAINVNYSAVTAACLL 612
ZP01563358 SVGAKLLYPNGTIQHSGVVVGIGGLAGHPHVGE-PGETFGYFGRAACTQRYSAVTAACVV 644
ABI93188 VCGPLLLYGDRTIQHAGVVIGMGGWADHVFKGEAPVHNQNLFVSPLLQRQVLAVTGACMV 513
ZP01855298 LVGPQLLYEDNTIQHAGVVVGMGGWADHVFKNQLPVHRSGPFVSPMLNRNVLAITGACQV 806
WsaE IAKKVIEKIGGFNEE-FIICGSDVEISLRALKMGYVNIYDPYVRLYHLESKT-------R1085
AAC35930 MKKADFDAVGGFEEA-FTVAFNDVDLCLKVQALGRDNVWLHEAELYHFESQTRGYDDKGK 671
ZP01563358 MRREVFLEVSGFDEVNFAVAFNDVDLGMRLGQAGYANVWTPRALLFHHESASLGLPTNED 704
ABI93188 VARETFESLGGFDES-FIVCGSDVELCLRARLHGLATVYVARSVMIHHESKT------RD 566
ZP01855298 IERAKFEQLGGFDEQ-FIICGSDVDICIRAHQQGLQNVYCADAALHHLESKS------RS 859
WsaE DSFIPERDFELSAKYYSP-YREIGDPYYNQNLSYNHLIPTIRS 1127
AAC35930 KKKRFEQEKVMMEEKWGP-LIEN-DPFYNPNLTRDIP------ 706
ZP01563358 RRRQFLEECDNFRRIWAD-VIRN-DPFYNPNLTISGGDFRPNF 762
ABI93188 PREIPESDFVRSAQAYSPYREE-GDPFFSPNLDYMASSPRLRG 611
ZP01855298 S-FIPKQDFLMSEIRYAPYRNDKGDPYFNENLDLMSTMPRMLT 903

Figure S5

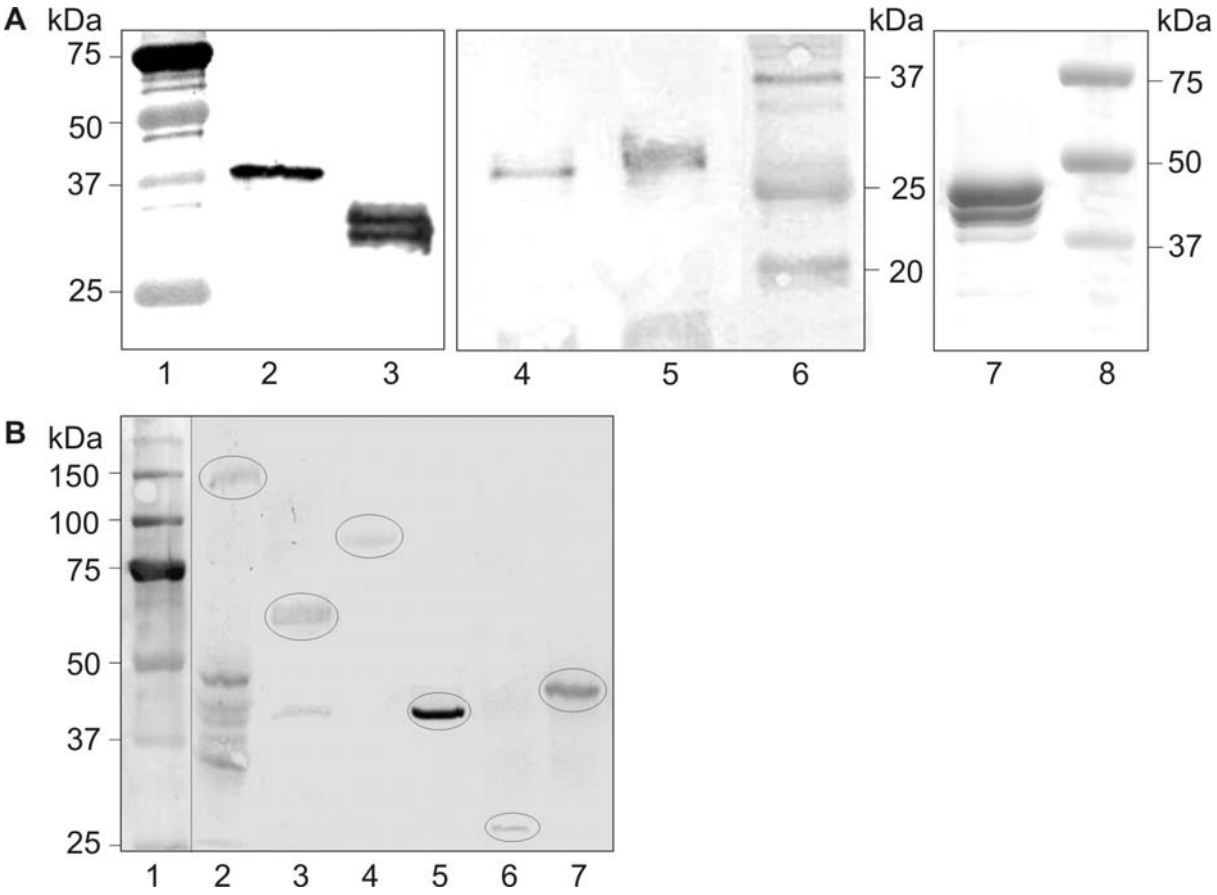

A
kDa
75
50
37
25
1
2
3
kDa
37
25
20
4
5
6
kDa
75
50
37
7
8
B
kDa
150
100
75
50
37
25
1
2
3
4
5
6
7

Figure S6

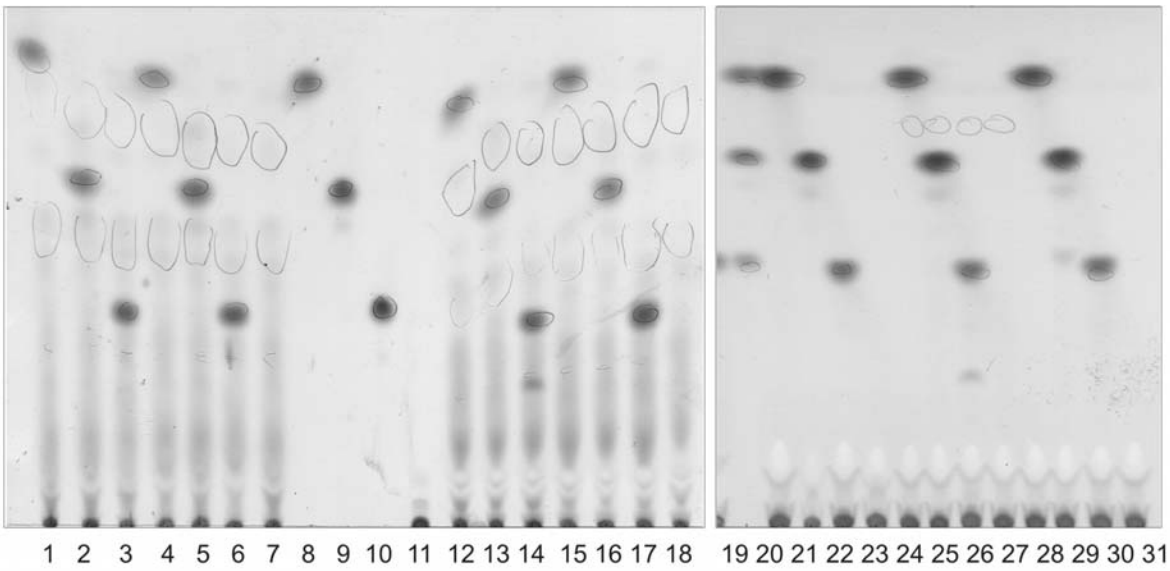
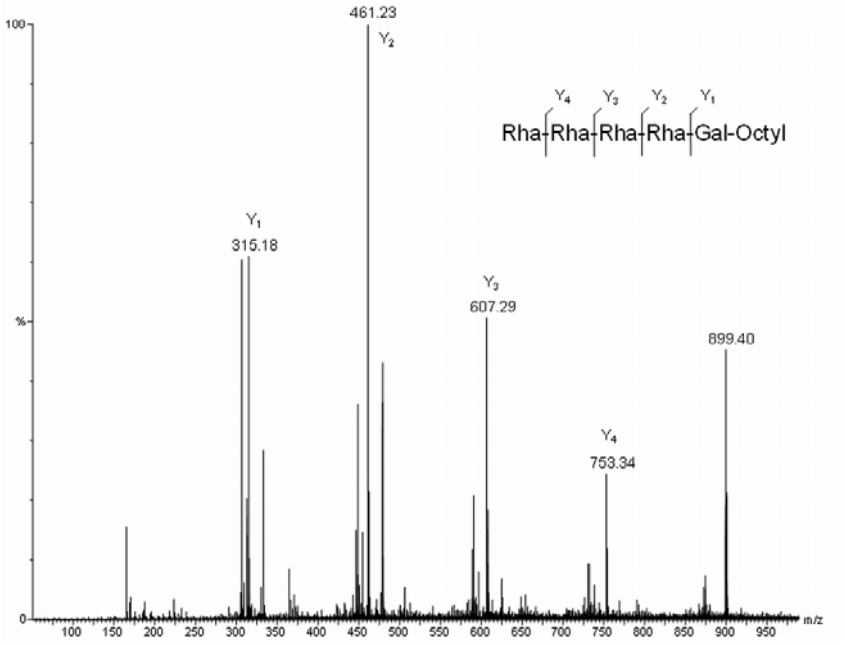

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31

Figure S7


461.23
Y2
Y4 Y3 Y2 Y1
Rha-Rha-Rha-Rha-Gal-Octyl
Y1
315.18
Y3
607.29
899.40
Y4
753.34
100
%
0
m/z
100 150 200 250 300 350 400 450 500 550 600 650 700 750 800 850 900 950

Eurocarb 15 - 15th European Carbohydrate Symposium, Vienna, Austria, August 19-24, 2009.

Proposal of a pathway of the surface layer protein glycosylation in *Paenibacillus alvei* CCM 2051^T enabled by a glycan mutant approach.

Kristof Zarschler, Bettina Janesch, Christina Schäffer, Paul Messner

The cell surface of the Gram-positive, mesophilic organism *Paenibacillus alvei* CCM 2051^T is completely covered with a crystalline two-dimensional protein coat, representing the outer-most cell wall layer (1). This surface (S-) layer is formed by non-covalent, entropy-driven assembly of identical glycoprotein protomers constituting a regular, closed lattice with an oblique (p2) symmetry (a=10.0, b=7.9 nm). The S-layer *O*-glycan is a polymer of [→3)-β-D-Gal*p*-(1→)-[α-D-Glc*p*-(1→6)]→4)-β-D-Man*p*NAc-(1→] repeating units that is linked by an adaptor of -[GroA-2→OPO_2→4-β-D-Man*p*NAc-(1→4)]→3)-α-L-Rha*p*-(1→3)-α-L-Rha*p*-(1→3)-α-L-Rha*p*-(1→3)-β-D-Gal*p*-(1→ to specific tyrosine residues of the S-layer protein (2, 3).

For *P. alvei* CCM 2051^T, a ~24 kb S-layer glycosylation (*slg*) gene cluster coding for more then a dozen of open reading frames possessing high homology with enzymes involved in LPS biosynthesis has been identified by data bank comparison. The cluster encodes proteins responsible for the biosynthesis of nucleotide-activated sugars, for the assembly of the glycan chain, its export across the cell wall and final ligation of the elongated glycan chain to the S-layer protein. Gene disruption of these open reading frames by use of a modified TargeTron^{TM} Gene Knockout System shed light on the role of the corresponding enzymes in the S-layer glycosylation pathway. With this approach, the UDP-Gal:phosphoryl-polyprenol Gal-1-phosphate transferase WsaP, a WbaP homolog, was identified as the initiation enzyme of the glycosylation process (4). Our current model starts with the transfer of galactose from its nucleotide-activated form to the membrane-associated lipid carrier undecaprenol-P at the cytoplasmic face of the plasma membrane by WsaP. Chain extension continues in the cytoplasm by processive addition of monosaccharide residues to the non-reducing terminus of the lipid-linked glycan chain by several glycosyltransferases until chain growth is terminated. The export of the complete glycan chain is

accomplished by an ABC transporter and finally it is transferred by an oligosaccharyl::protein transferase on the S-layer protein.

1. **Sára, M., K. Moser-Thier, U. Kainz, and U. B. Sleytr.** 1990. Characterization of S-layers from mesophilic *Bacillaceae* and studies on their protective role towards muramidases. Archives of Microbiology **153:**209-214.
2. **Altman, E., J. R. Brisson, P. Messner, and U. B. Sleytr.** 1991. Structure of the glycan chain from the surface layer glycoprotein of *Bacillus alvei* CCM 2051. Biochemistry and Cell Biology **69:**72-78.
3. **Messner, P., R. Christian, C. Neuninger, and G. Schulz.** 1995. Similarity of "core" structures in two different glycans of tyrosine-linked eubacterial S-layer glycoproteins. Journal of Bacteriology **177:**2188-2193.
4. **Zarschler, K., B. Janesch, S. Zayni, C. Schäffer, and P. Messner.** 2009. Construction of a gene knockout system for application in *Paenibacillus alvei* CCM 2051^{T}, exemplified by the S-layer glycan biosynthesis initiation enzyme WsfP. Applied and Environmental Microbiology **75:**3077-3085.

PROPOSAL OF A PATHWAY OF THE SURFACE LAYER PROTEIN GLYCOSYLATION IN *PAENIBACILLUS ALVEI* CCM 2051^T ENABLED BY A GLYCAN MUTANT APPROACH

VIBT Vienna Institute of BioTechnology
BOKU
Department of NanoBiotechnology

Kristof Zarschler, Bettina Janesch, Sonja Zayni, Christina Schäffer, Paul Messner
Department of NanoBiotechnology, University of Natural Resources and Applied Life Sciences, Vienna, Austria

Introduction

The cell surface of the Gram-positive, mesophilic organism *Paenibacillus alvei* CCM 2051^T is completely covered with a crystalline two-dimensional glycoprotein coat, representing the outermost cell wall layer [1]. This surface (S-) layer is formed by non-covalent, entropy-driven assembly of identical glycoprotein subunits constituting a regular, closed lattice. The glycan chain is *O*-glycosidically linked via galactose to certain tyrosine residues of the protein [2, 3].

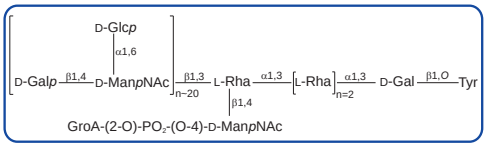


Fig. 1. Structure of the S-layer *O*-glycan of *P. alvei* CCM 2051^T [2, 3].

Results

By chromosome walking, starting from the *rmlACBD* genes encoding enzymes involved in dTDP-L-rhamnose biosynthesis, a large gene cluster responsible for the S-layer glycan biosynthesis (*slg* gene cluster) has been identified and entirely sequenced [4]. The ~24.3-kb cluster encodes 18 open reading frames whose translation products are involved in the biosynthesis of nucleotide-activated monosaccharides and in assembly, export as well as *en bloc* transfer of the polysaccharide to the S-layer protein (Fig. 2).

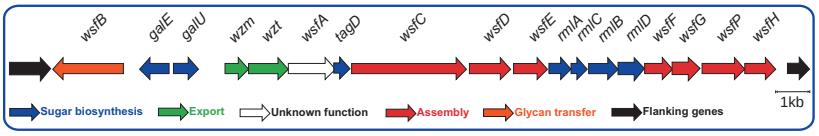


Fig. 2. Genetic organisation of the *slg* gene cluster of *P. alvei* CCM 2051^T [4].

In order to determine the function of individual enzymes in the process of S-layer glycosylation and to analyse the glycosylation process *in vivo*, corresponding genes were disrupted by applying a newly-developed gene knockout system (Fig. 3). This system bases on the commercially available bacterial mobile group II intron LI.LtrB of *Lactococcus lactis* in combination with the broad host-range S-layer gene promoter of *sgsE* from *G. stearothermophilus* NRS 2004/3a and the *Geobacillus-Bacillus-E. coli* shuttle vector pNW33N [5].

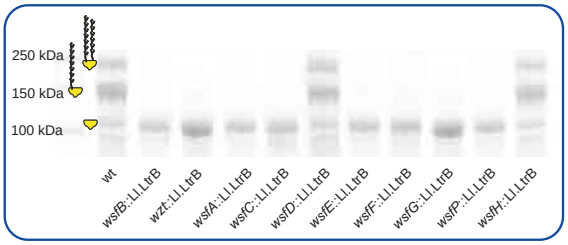


Fig. 3. Effect of insertional inactivation of genes from the *P. alvei* CCM 2051^T *slg* gene cluster on the *O*-glycosylation profile. An aliquote of biomass from various *P. alvei* CCM 2051^T mutant strains was analyzed by SDS-PAGE followed by Coomassie staining. The mutant strains carry the LI.LtrB insertion as indicated [4].

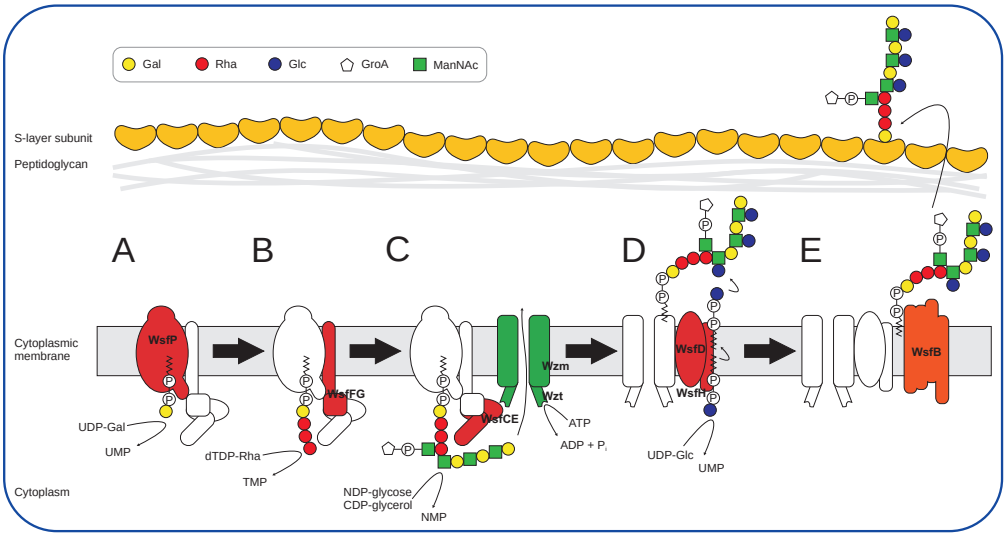


Fig. 4. Model of S-layer glycan biosynthesis in *P. alvei* CCM 2051^T [4]. The initial transfer of a Gal residue to a lipid carrier is catalyzed by WsfP (A). The adaptor saccharide is formed by the 1,3-linkage of a rhamnose residue to the linkage sugar Gal performed by WsfG, followed by the transfer of two additional 1,3-linked rhamnoses by the action of WsfF (B). The glycan chain is elongated by the activity of the aminosugar transferase WsfE and the three domain transferase WsfC. The glycan chain is recognized and exported by the ABC transporter system through the cytoplasmic membrane (C). The transfer of cytoplasmic Glc to the lipid carrier is carried out by WsfH and, after reorientation, is used in the periplasm by WsfD for 1,6-linkage of the Glc residues to ManNAc residues of the repeating units (D). The final transfer of the completed S-layer glycan to the S-layer protein is catalyzed by the oligosaccharyl transferase WsfB (E).

Detailed protein database comparison and analysis of the different *slg* knockout mutants resulted in a proposal of a model for S-layer glycan biosynthesis in *P. alvei* CCM 2051^T. The identification of an ABC transporter system (Wzm and Wzt) and the loss of S-layer glycosylation in the *wzt*::LI.LtrB mutant corroborate the assumption that the glycan chain is synthesized in a process comparable to the ABC-transporter-dependent pathway of the LPS O-polysaccharide biosynthesis [6]. According to this pathway, the cytoplasmic glycan chain extension is achieved by processive addition of sugar residues to the nonreducing terminus of the lipid-linked growing chain. After chain length termination the polymer is exported through the cytoplasmic membrane by an ABC transporter for *en bloc* transfer from the lipid carrier to certain amino acid residues of the target protein by an oligosaccharyl transferase.

Conclusion & Outlook

The current report describes the identification, annotation and characterization of the *slg* gene cluster involved in the biosynthesis of the S-layer *O*-glycan in *P. alvei* CCM 2051^T as well as the analysis of the corresponding ABC transporter system and the oligosaccharyl transferase. This data marks a starting point for further studies and applications, possibly for the future design of functional glycans and controlled surface display of heterologous glycoproteins.

Reference:
[1] Sára, M.; Moser-Thier, K.; Kainz, U.; Sleytr, U. B. *Archives of Microbiology* **1990**, *153*, 209-214.
[2] Altman, E.; Brisson, J.R.; Messner, P.; Sleytr, U.B. *Biochem Cell Biol* **1991**, *69*, 72-78.
[3] Messner, P.; Christian, R.; Neuninger, C.; Schulz, G. *J Bacteriol* **1995**, *177*, 2188-2193.
[4] Zarschler, K.; Janesch, B.; Pabst, M.; Altmann, F.; Messner, P.; Schäffer, C. *Manuscript in preparation.*
[5] Zarschler, K.; Janesch, B.; Zayni, S.; Schäffer, C.; Messner, P. *Appl. Envion. Microbiol* **2009**, *75*, 3077-3085.
[6] Whitfield, C. *Trends Microbiol* **1995**, *3*, 175-185.

Acknowledgements:
We thank our collaborators Dr. Martin Pabst and Prof. Dr. Friedrich Altmann for mass spectrometrical analysis. We acknowledge the excellent technical support by Andrea Scheberl, whose contributions to the advancement of S-layer nanoglycobiology by far exceeded routine work. Financial support came from the Austrian Science Fund (Project Number P18013-B10 and P20745-B11 to P.M.) and the City of Vienna (Hochschuljubiläumsstiftung Project Number H-02229-2007 to K.Z.)

Contact:
bettina.janesch@boku.ac.at
kristof.zarschler@boku.ac.at

Life Sciences 2008 - Gemeinsame Jahrestagung der Österreichischen Gesellschaft für Biochemie und Molekularbiologie, der Österreichischen Gesellschaft für Genetik und Gentechnik, der Österreichischen Gesellschaft für Biotechnologie und des Austrian Networks für Gentherapie, Graz, Austria, September 22-24, 2008.

The glycosylation mechanism of the surface (S-) layer protein of *Paenibacillus alvei.*

Kristof Zarschler, Bettina Janesch, Christina Schäffer, Paul Messner

The cell surface of the Gram-positive, mesophilic organism *Paenibacillus alvei* CCM 2051^{T} is completely covered with a crystalline two-dimensional protein coat, representing the outermost cell wall layer. This S-layer is formed by non-covalent, entropy-driven assembly of identical glycoprotein protomers constituting a regular, closed lattice. Once detached from the cell wall by treatment with chaotropic agents, the subunits characteristically self-assemble into monomolecular, arrays with oblique symmetry, which are identical with those observed on the native bacteria. Recently, for *P. alvei* CCM 2051 the gene *spaA* encoding the S-layer protein SpaA, an S-layer glycosylation (*slg*) gene cluster and a secondary cell wall (SCWP) biosynthesis gene locus have been identified. From this data, a model for the glycosylation mechanism of SpaA concerning the cytoplasmic biosynthesis of the glycan chain, the transport across the cell wall and final ligation to the S-layer protein emerged.

Acknowledgement:
FWF (P18013-B10, P20745-B11; to P.M.)
Hochschuljubiläumsstiftung der Stadt Wien (H-02229-2007; to K.Z.)

First insights into the surface layer glycosylation of *Paenibacillus alvei*

Kristof Zarschler, Bettina Janesch, Christina Schäffer and Paul Messner

Center for NanoBiotechnology, University of Natural Resources and Applied Life Sciences, Vienna

S-layer proteins

The cell surface of the Gram-positive, mesophilic organism *Paenibacillus alvei* CCM 2051^T is completely covered with a crystalline two-dimensional protein coat, representing the outermost cell wall layer [1]. This surface (S-) layer is formed by non-covalent, entropy-driven assembly of identical glycoprotein protomers constituting a regular, closed lattice. Once detached from the cell wall by treatment with chaotropic agents, the subunits characteristically self-assemble into monomolecular arrays with oblique symmetry, which are identical with those observed on the native bacterium [2].

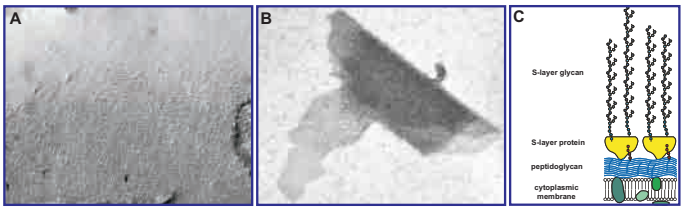


Fig.1: Electron micrograph of a freeze-etched preparation of an intact cell of *P. alvei* CCM 2051^T showing an oblique S-layer lattice (A) and of a negatively stained self-assembly product composed of the S-layer glycoprotein SpaA of *P. alvei* CCM 2051^T (B). Schematic drawing of the cell envelope architecture of S-layer glycoprotein possessing Gram-positive bacteria (C).

S-layer glycan

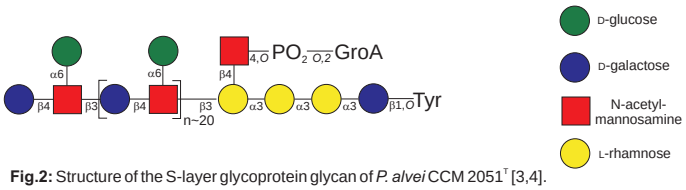
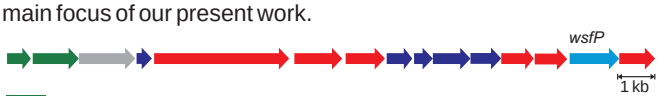


Fig.2: Structure of the S-layer glycoprotein glycan of *P. alvei* CCM 2051^T [3,4].

S-layer glycosylation (*slg*) gene cluster

For *P. alvei* CCM 2051^T, a large *slg* gene cluster coding for more then a dozen open reading frames possessing high homology with enzymes involved in LPS biosynthesis as shown by data bank analysis has been identified. The cluster encodes proteins responsible for the biosynthesis of nucleotide activated sugars and for the assembly and export of the elongated glycan chain. The functional characterisation of the cluster's putative enzymes is the main focus of our present work.



ABC transporter for export of the glycan chain

Enzymes for biosynthesis of nucleotide activated glycan building blocks

Glycosyltransferases for transfer of sugar on a substrate

Enzyme for the transfer of sugar on a lipid anchor

Unknown

Fig.3: Genetic organisation of the *slg* gene cluster of *P. alvei* CCM 2051^T

Targetron-mediated gene disruption of *wsfP*

The 468-aa enzyme WsfP of *P. alvei* CCM 2051^T shows high similarity to the UDP-Gal:phosphoryl-polyprenol Gal-1-phosphate transferase WsaP of the thermophilic organism *Geobacillus stearothermophilus* NRS 2004/3a, which there functions as initiation enzyme of S-layer glycan biosynthesis [5].

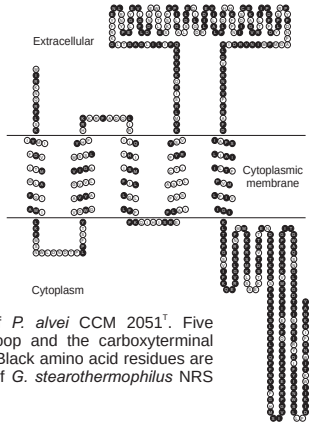


Fig.4: Predicted topology of the WsfP protein of *P. alvei* CCM 2051^T. Five transmembrane helices, the central extracellular loop and the carboxyterminal cytosolic tail containing the catalytic site are shown. Black amino acid residues are identical with corresponding amino acids in WsaP of *G. stearothermophilus* NRS 2004/3a.

Disruption of the chromosomal *wsfP* gene in *P. alvei* CCM 2051^T using a bacterial mobile group II intron (Fig. 5) results in the loss of S-layer glycan biosynthesis and, thus, in complete abolishment of SpaA glycosylation. This effect could be fully restored by the expression of plasmid-encoded *wsfP* in the *wsfP* mutant (Fig. 6). These observations demonstrate, that WsfP catalyses the transfer of galactose from its nucleotide-activated form to phosphoryl-polyprenol and thus is responsible for the initiation reaction of S-layer glycan biosynthesis in *P. alvei* CCM 2051^T.

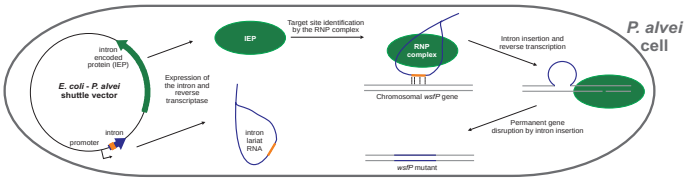
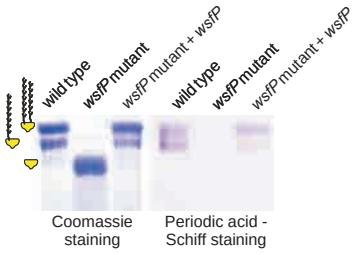


Fig.5: Principle of gene disruption mediated by bacterial mobile group II introns [6].

Fig.6: SDS-PAGE showing the *O*-glycosylation profile of *P. alvei* CCM 2051^T wild type, *wsfP* mutant and reconstituted *wsfP* mutant cells. Unglycosylated, mono-glycosylated and di-glycosylated S-layer protein SpaA is indicated. UDP-Gal:phosphoryl-polyprenol Gal-1-phosphate transferase activity was reconstituted by plasmid-based expression of *wsfP*.



Conclusions

From the adaptation of the targetron-mediated gene disruption system in *P. alvei* CCM 2051^T, an essential tool for elucidating the entire S-layer glycosylation process and for the specific modification of the glycan towards functional S-layer *neo*-glycoproteins has evolved.

References

[1] Sara, M., Moser-Thier, K., Kainz, U., and Sleytr, U.B. (1990), Arch. Microbiol 153: 209-214.
[2] Sleytr, U.B., and Beveridge, T.J. (1999), Trends Microbiol 7: 253-260.
[3] Altman, E., Brisson, J.R., Messner, P., and Sleytr, U.B. (1991), Biochem Cell Biol 69: 72-78.
[4] Messner, P., Christian, R., Neuninger, C., and Schulz, G. (1995), J Bacteriol 177: 2188-2193.
[5] Steiner, K., Novotny, R., Patel, K., Vinogradov, E., Whitfield, C., Valvano, M.A., Messner, P., and Schäffer, C. (2007), J Bacteriol 189: 2590-2598.
[6] Lambowitz, A.M., and Zimmerly, S. (2004), Annu Rev Genet 38: 1-35.

Acknowledgements

We acknowledge the excellent technical support by Sonja Zayni and Andrea Scheberl, whose contributions to the advancement of S-layer nanoglycobiology by far exceeded routine work.
Financial support came from the Austrian Science Fund (Project Number P18013-B10 and P20745-B11 to P.M.) and the City of Vienna (Hochschuljubiläumsstiftung Project Number H-02229-2007 to K.Z.)

Contact

Kristof.Zarschler@boku.ac.at
http://www.nano.boku.ac.at/znb.html

FEBS Special Meeting: ATP-Binding Cassette (ABC) Proteins from Multidrug Resistance to Genetic Diseases, Innsbruck, Austria, March 4-10, 2006.

Is an ABC-transporter responsible for the export of the cytoplasmatically synthesized surface layer (S-layer) glycan to the cell surface *Geobacillus stearothermophilus* NRS 2004/3a?

Kristof Zarschler, Renè Novotny, Kerstin Steiner, Christina Schäffer, Paul Messner

The Gram-positive thermophilic organism *Geobacillus stearothermophilus* NRS 2004/3a possesses an oblique S-layer composed of glycosylated subunits as the outermost cell wall component. Recently the structure of the S-layer glycan as well as the sequence of the polycistronic S-layer glycan biosynthesis gene cluster (*slg*) have been elucidated (1, 2). The glycan consists of, on average, 15 [→2)-α-L-Rha*p*-(1→3)-β-L-Rha*p*-(1→2)-α-L-Rha*p*-(1→] trisaccharide repeating units and is *O*-glycosidically linked to the S-layer protein via a core region of two L-rhamnoses and a β-D-galactose residue. More than ten enzymes responsible for nucleotide sugar biosynthesis, assembly, export of the homo-polymeric glycan chain and its transfer to the S-layer protomer are encoded by a ~16.5 kb *slg* gene cluster. Within this cluster, two open reading frames were identified, whose putative translational products reveal high similarity to proteins of the ABC-2 transporter family. Whereas one ORF encodes a Wzm-like protein with six transmembrane domains, the putative protein encoded by another ORF shows more than 60% similarity to common Wzt proteins. The occurrence of this ABC-2-type transporter system supports the notion, that the glycan chain is built up by successive addition of nucleotide-activated monosaccharides at the non-reducing end of the glycan chain in a process similar to the so called “ABC-transporter-dependent” pathway described for the biosynthesis and assembly of linear *O*-polysaccharide structures of Gram-negative bacteria (3). Our model of S-layer glycoprotein biosynthesis includes the cytoplasmic synthesis of the complete glycan chain, its ATP-dependent export to the cell surface via the membrane-spanning Wzm-like protein and, finally, the glycosidic linkage to the S-layer protein protomer by a specific ligase reaction.

This project (P18013-B10) is supported by the Austrian Science Fund (FWF).

1. **Schäffer, C., T. Wugeditsch, H. Kählig, A. Scheberl, S. Zayni, and P. Messner.** 2002. The surface layer (S-layer) glycoprotein of *Geobacillus stearothermophilus* NRS 2004/3a. Analysis of its glycosylation. Journal of Biological Chemistry **277:**6230-6239.
2. **Novotny, R., C. Schäffer, J. Strauss, and P. Messner.** 2004. S-layer glycan-specific loci on the chromosome of *Geobacillus stearothermophilus* NRS 2004/3a and dTDP-L-rhamnose biosynthesis potential of *G. stearothermophilus* strains. Microbiology **150:**953-965.
3. **Raetz, C. R. H., and C. Whitfield.** 2002. Lipopolysaccharide endotoxins. Annual Review of Biochemistry **71:**635-700.

Is an ABC-transporter responsible for the export of the S-layer glycan to the cell surface of *G. stearothermophilus* NRS 2004/3a?

Kristof Zarschler, René Novotny, Kerstin Steiner, Christina Schäffer and Paul Messner

Center for NanoBiotechnology, University of Natural Resources and Applied Life Sciences, Vienna

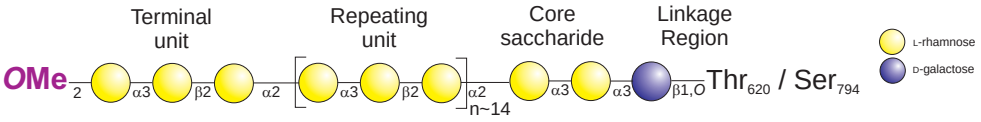
S-(Surface) layer Proteins

S-layer proteins are two-dimensional crystalline arrays composed of identical protein or glycoprotein subunits that form the outermost envelope layer of bacterial cells.
S-layer monomers have the ability to self-assemble into regular arrays in suspension and to recrystallise on solid supports, showing a wide application potential in the field of nanobiotechnology and biomedicine.
The S-layer protein SgsE of the Gram-positive, thermophilic, non pathogenic, wild-type strain *G. stearothermophilus* NRS 2004/3a was one of the first bacterial glycoproteins to be described [1].

Electron micrograph of a freeze-etched preparation of an intact cell of *Geobacillus stearothermophilus* NRS 2004/3a showing an oblique S-layer lattice.

S-layer glycan

Structure of the S-layer glycoprotein glycan of *G. stearothermophilus* NRS 2004/3a [2].

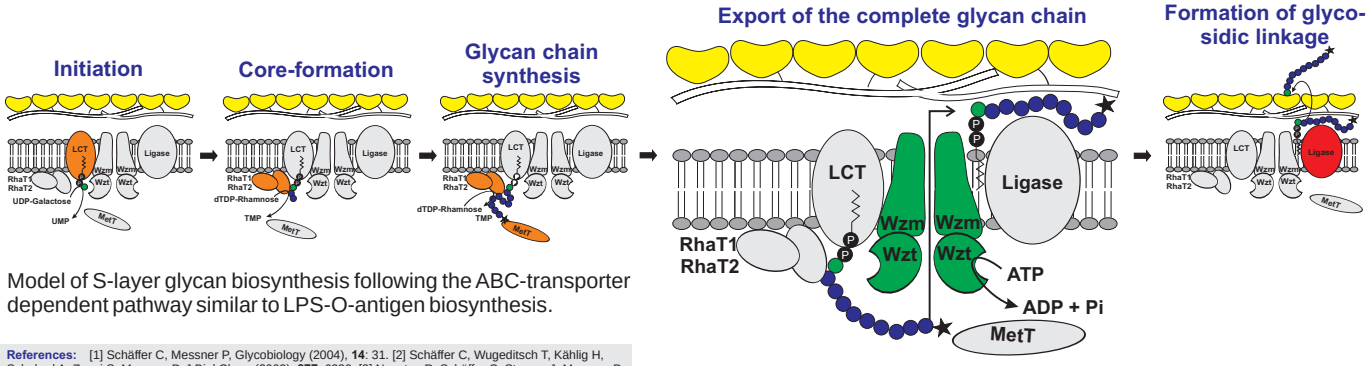


S-layer glycosylation (*slg*) gene cluster

The ~16,5 kb *slg* gene cluster of *G. stearothermophilus* NRS 2004/3a is located immediately downstream of the S-layer structural gene *sgsE* and encodes for 13 open reading frames possessing high homology with enzymes involved in LPS biosynthesis as shown by data bank analysis. The *slg* gene cluster is transcribed as a polycistronic transcriptional unit, while *sgsE* is transcribed monocistronically [3]. The functional characterisation of the cluster's putative enzymes is the main focus of our present work.

S-layer glycan biosynthesis and export

The *slg* gene cluster contains ATP-binding cassette (ABC)-2 export proteins encoding genes (*wzm, wzt*), which have their closest homologs in the export machinery of LPS and CPS, implying that the encoded proteins may be involved in the export of the S-layer glycan chain.



Model of S-layer glycan biosynthesis following the ABC-transporter dependent pathway similar to LPS-O-antigen biosynthesis.

References: [1] Schäffer C, Messner P, Glycobiology (2004), **14**: 31. [2] Schäffer C, Wugeditsch T, Kählig H, Scheberl A, Zayni S, Messner P, J Biol Chem (2002), **277**: 6230. [3] Novotny R, Schäffer C, Strauss J, Messner P, Microbiology (2004), **150**: 953.

Acknowledgements: Supported by the Austrian Science Fund (Project Number P18013-B10)

MOLECULAR BASIS OF S-LAYER GLYCOPROTEIN GLYCAN BIOSYNTHESIS IN *GEOBACILLUS STEAROTHERMOPHILUS**

Kerstin Steiner[‡,1], René Novotny[‡,2], Daniel B. Werz[§,3], Kristof Zarschler[‡], Peter H. Seeberger[§], Andreas Hofinger[¶], Paul Kosma[¶], Christina Schäffer[‡,4], and Paul Messner[‡,4]

SUPPLEMENTARY METHODS

Materials-All chemicals used were reagent grade and used as supplied except where noted. Dichloromethane (CH_2Cl_2) was purchased from JT Baker and purified by a Cycle-Tainer Solvent Delivery System. Pyridine was refluxed over calcium hydride and distilled prior to use. Analytical thin-layer chromatography was performed on E. Merck silica gel 60 F_{254} plates (0.25 mm). Compounds were visualized by dipping the plates in a cerium sulfate ammonium molybdate solution or a sulfuric acid/methanol solution followed by heating. Liquid chromatography was performed using forced flow of the indicated solvent on Sigma H-type silica (10–40 mm). ^{1}H NMR spectra were obtained on a Varian VXR-300 (300 MHz), Bruker-600 (600 MHz) and are reported in parts per million (δ) relative to $CHCl_3$ (7.26 ppm) or in the case of CD_3OD as solvent relative to TMS (0.00 ppm). Coupling constants (J)[1] are reported in Hertz. ^{13}C NMR spectra were obtained on a *Varian* VXR-300 (75 MHz), Bruker-600 (150 MHz) and are reported in δ relative to $CDCl_3$ (77.0 ppm) as an internal reference or to TMS (0.00 ppm).

Synthesis of acceptor substrate-The synthesis of the β-D-Gal-(1→*O*)-octyl (**I**), α-L-Rha-(1→3)-β-D-Gal-(1→*O*)-octyl (**II**) and α-L-Rha-(1→3)-α-L-Rha-(1→3)-β-D-Gal-(1→*O*)-octyl (**III**) acceptors is shown in Fig. S1. Building block dibutyl 4,6-di-*O*-benzyl-3-*O*-fluorenylmethoxycarbonyl-2-*O*-pivaloyl-β-D-galactopyranosyl phosphate **a** (27) was reacted using common glycosylation conditions with 1-octanol to afford **b**. Fmoc deprotection furnished compound **c** (*n*-octyl 4,6-di-*O*-benzyl-2-*O*-pivaloyl-β-D-galactopyranoside), which served as a starting point for the synthesis of all three substrates. For the synthesis of the trisaccharide (**III**) an (2+1) approach was envisioned. For this approach the dirhamnose building block **f** was synthesized. The MP protected rhamnose unit **d** (28) was glycosylated at 0 °C with another rhamnose fragment to furnish **e**. A removal of the MP group using cerium ammonium nitrate (CAN) in aqueous solution and the subsequent reaction with trichloroacetonitrile yielded the dirhamnose building block **f** in good yield.

A removal of the pivaloyl group of compound **c** using strong basic conditions and the global deprotection of all the benzyl groups afforded β-D-Gal-(1→*O*)-octyl (**I**). For the synthesis of disaccharide (**II)**, compound **c** was glycosylated using a rhamnose building block with Ac in position 2 to ensure α-selectivity. Column chromatography to purify the reation products yielded an inseparable mixture of the desired compound together with (Overman)-rearranged trichloroacetimidate. Therefore, this mixture was subjected to basic deprotection removing acetyl (Ac) and pivaloyl (Piv) moieties. Using this approach disaccharide **g** was obtained as a pure compound. Global deprotection using $Pd(OH)_2$/C and H_2 in methanol/dichloromethane afforded disaccharide (**II**).

Glycosylating agent **f** as well as octyl galactoside **c** were utilized for the assembly of the trisaccharide **h**. A deprotection of all ester moieties followed by a hydrogenation in order to remove all benzyl groups afforded the completely deprotected trisaccharide (**III**). All the deprotected trisaccharide was purified by silica gel chromatography using mixtures of dichloromethane/methanol in order to remove all traces of palladium which could cause problems in biological experiments.

***n*-Octyl 4,6-Di-*O*-benzyl-2-*O*-pivaloyl-β-D-galactopyranoside (c)**-Compound **a** (960 mg, 1.14 mmol) (27) was azeotroped three times with toluene and dried *in vacuo*. Dichloromethane (25 ml) were added and the solution cooled to -40 °C. 1-Octanol (296 mg, 359 µl, 2.28 mmol) and afterwards TMSOTf (253 mg, 206 µl, 1.14 mmol) were added. The mixture was stirred for 1 h and quenched by addition of some drops of pyridine before the solvent was removed *in vacuo*. The resulting crude product was purified by flash chromatography (5:1 hexane/EtOAc) to afford 631 mg (71%) of **b** as a colorless oil. Compound **b** (585 mg, 0.75 mmol) was dissolved in 8 ml of DMF. Piperidine (2 ml) was added and the mixture stirred for 90 min at room temperature. The solvent was removed *in vacuo* and the residue purified by column chromatography (4:1, hexane/EtOAc) to afford 413 mg (99%) of **c** as a

colorless oil: $[\alpha]_D$: +17.0 (c = 0.41, $CHCl_3$). IR (thin film, $CHCl_3$): 3000, 1742, 1440, 1260, 1077 cm^{-1}. ^1H NMR ($CDCl_3$, 300 MHz) δ 0.90 (t, *J* = 6.4 Hz, 3H), 1.25 (s, 9H), 1.28 (b, 10H), 1.55 (b, 2H), 2.39 (d, *J* = 9.6 Hz, 2H), 3.45 (quart, *J* = 6.9 Hz, 1H), 3.65 (m, 4H), 3.86 (m, 2H), 4.38 (d, *J* = 7.8 Hz, 1H), 4.52 (m, 2H), 4.71 (s, 2H), 5.00 (dd, *J* = 10.2, 7.8 Hz, 1H), 7.25-7.39 (m, 10H). ^{13}C NMR ($CDCl_3$, 75 MHz) δ 14.2, 22.8, 26.1, 27.2, 29.3, 29.5, 29.7, 31.9, 38.9, 68.3, 69.8, 73.4, 73.5, 75.4, 76.5, 101.0, 127.8, 128.0, 128.1, 128.4, 128.4, 137.6, 138.0, 178.5. ESI-MS: *m/z* $[M + Na]^+$ calculated 579.3292, observed 579.3283.

***p*-Methoxyphenyl 2-*O*-Acetyl-3,4-di-*O*-benzyl-α-L-rhamnopyranosyl-(1→3)-2-*O*-acetyl-4-*O*-benzyl-α-L-rhamnopyranoside (e)**-Rhamnose building block **d** (237 mg, 0.589 mmol) (28) and rhamnosyl trichloroacetimidate (406 mg, 0.766 mmol) (26) were azeotroped three times with toluene and dried *in vacuo*. Dichloromethane (9 ml) were added and the solution cooled to 0 °C. TMSOTf (20 mg, 11 µl, 0.09 mmol) was added. The mixture was stirred for 90 min and quenched by addition of some drops of pyridine. Afterwards the solvent was removed *in vacuo*. The resulting crude product was purified by column chromatography (7:3, hexane/EtOAc) to afford 448 mg (98%) of **e** as a colorless oil: $[\alpha]_D$: -45.3 (c = 1.08, $CHCl_3$). IR (thin film, $CHCl_3$): 3005, 2933, 1739, 1585, 1503, 1456, 1369, 1226, 1097, 1046 cm^{-1}. ^1H NMR ($CDCl_3$, 300 MHz) δ 1.32 (pt, *J* = 7.2 Hz, 6H), 2.15 (s, 3H), 2.17 (s, 3H), 3.47 (t, *J* = 9.6 Hz, 1H), 3.54 (t, *J* = 9.6 Hz, 1H), 3.77 (s, 3H), 3.84 (m, 1H), 3.92 (m, 2H), 4.34 (dd, *J* = 9.3 Hz, 3.3 Hz, 1H), 4.48 (d, *J* = 11.4 Hz, 1H), 4.62-4.68 (m, 3H), 4.85 (d, *J* = 10.8 Hz, 1H), 4.93 (d, *J* = 11.1 Hz, 1H), 5.11 (ps, 1H), 5.32 (m, 1H), 5.37 (ps, 1H), 5.49 (m, 1H), 6.81 (pd, *J* = 9.3 Hz, 2H), 6.97 (pd, *J* = 9.3 Hz, 2H), 7.25-7.39 (m, 15H). ^{13}C NMR ($CDCl_3$, 75 MHz) δ 18.0, 18.1, 21.4, 21.2, 55.7, 68.4, 68.8, 69.1, 71.7, 72.1, 75.3, 75.5, 77.6, 79.8, 80.1, 95.9, 99.8, 114.5, 117.6, 127.6, 127.8, 127.8, 127.9, 127.9, 128.3, 128.4, 137.8, 138.3, 149.9, 154.9, 170.0, 170.1. MALDI-MS: *m/z* $[M + Na]^+$ calculated 793.3195, observed 793.3182.

2-*O*-Acetyl-3,4-di-*O*-benzyl-α-L-rhamnopyranosyl-(1→3)-2-*O*-acetyl-4-*O*-benzyl-α-L-rhamnopyranosyl trichloroacetimidate (f)-Disaccharide **e** (420 mg, 0.545 mmol) was suspended in a mixture of acetonitrile and water (20 ml, 1:1). The mixture was stirred for 2 h at room temperature until TLC control experiments showed that no starting material was left. The mixture was poured onto brine and extracted 2x with EtOAc. The combined organic phases were washed twice with water, dried over Na_2SO_4 and concentrated. The resulting crude product was purified by column chromatography (2:1, hexane/EtOAc) to afford 334 mg (92%) of the hemiacteal as a colorless oil. This hemiacetal (285 mg, 0.429 mmol) was dissolved in dichloromethane (5 ml) and trichloroacetonitrile (5 ml). NaH (5 mg) was added and the mixture was stirred for 90 min at room temperature. The solvent was removed *in vacuo*, column chromatography (3:1, hexane/EtOAc) furnished 305 mg (88%) of **f** as a slightly yellow oil: $[\alpha]_D$: -29.0 (c = 0.70, $CHCl_3$). IR (thin film, $CHCl_3$): 2995, 1744, 1672, 1615, 1451, 1369, 1092 cm^{-1}. ^1H NMR ($CDCl_3$, 300 MHz) δ 1.29 (d, *J* = 6.3 Hz, 3H), 1.36 (d, *J* = 6.0 Hz, 3H), 2.14 (s, 3H), 2.17 (s, 3H), 3.46 (t, *J* = 9.3 Hz, 1H), 3.59 (t, *J* = 9.6 Hz, 1H), 3.82 (m, 1H), 3.88-3.99 (m, 2H), 4.25 (dd, *J* = 9.6 Hz, 3.3 Hz, 1H), 4.51 (d, *J* = 11.1 Hz, 1H), 4.62-4.70 (m, 3H), 4.84 (d, *J* = 11.1 Hz, 1H), 4.92 (d, *J* = 11.1 Hz, 1H), 5.09 (ps, 1H), 5.31 (m, 1H), 5.50 (m, 1H), 6.21 (ps, 1H), 7.26-7.39 (m, 15H), 8.71 (s, 1H). ^{13}C NMR ($CDCl_3$, 75 MHz) δ 18.0, 18.1, 21.1, 21.1, 68.9, 69.0, 70.7, 70.7, 71.7, 75.2, 75.7, 76.1, 79.7, 79.8, 90.8, 94.3, 99.7, 127.6, 127.6, 127.9, 128.0, 128.3, 128.3, 128.4, 137.5, 137.8, 138.3, 159.9, 169.7, 169.9. MALDI-MS: *m/z* $[M + Na]^+$ calculated 830.1872, observed 830.1858.

***n*-Octyl β-D-galactopyranoside (I)**-Monosaccharide **c** (255 mg, 0.459 mmol) was dissolved in THF (20 ml), LiOH solution (1 M, 25 ml) and hydrogen peroxide solution (15 ml) were added at -5 °C. The reaction mixture was stirred for 2 days while warming to room temperature. Then methanol (15 ml) and potassium hydroxide (3 M, 30 ml) were added and stirred for 2 days. Hydrochloric acid (1 M) was added until pH 7 was reached. The neutral solution was extracted three times with dichloromethane, dried over $MgSO_4$, filtered and concentrated. Column chromatography (EtOAc) furnished 209 mg (97%) of a colorless solid. This compound (148 mg, 0.263 mmol) was dissolved in methanol (15 ml), $Pd(OH)_2/C$ (15 mg) was added and the Ar atmosphere replaced by a H_2 atmosphere. The reaction mixture was stirred for 12 h. Filtration through a plug of celite removed all solids. Column chromatography using silica gel and dichloromethane/methanol (1:1) in order to remove traces of Pd yielded 75 mg (98%) of **I** as a colorless wax: $[\alpha]_D$: -10.0 (c = 0.13, H_3COH). ^1H NMR (CD_3OD, 300 MHz) δ 0.89 (b, 3H), 1.29-1.37 (b, 10H), 1.61 (m, 2H), 3.47-3.53 (m, 4H), 3.73 (m, *J* = 6.0 Hz, 2H), 3.84 (s, b, 1H), 3.88 (m, 1H), 4.20 (d, *J* = 6.6 Hz, 1H). ^{13}C NMR (CD_3OD, 75 MHz) δ

14.0, 23.3, 26.7, 30.0, 30.2, 30.4, 32.6, 62.0, 69.8, 70.4, 72.1, 74.6, 76.1, 104.5. MALDI-MS: *m/z* [M + Na]$^+$ calculated 315.1778, observed 315.1775.

***n*-Octyl 3,4-di-*O*-benzyl-α-L-rhamnopyranosyl-(1→3)-4,6-di-*O*-benzyl-β-D-galactopyranoside (g)-**Compound **c** (185 mg, 0.332 mmol) and rhamnosyl trichloroacetimidate (228 mg, 0.431 mmol) were azeotroped three times with toluene and dried *in vacuo*. Dichloromethane (6 ml) were added and the solution cooled to 0 °C. TMSOTf (11 mg, 9 µl, 0.05 mmol) was added. The mixture was stirred for 1 h and quenched by addition of some drops of pyridine. Afterwards the solvent was removed *in vacuo*. The resulting crude product was purified by column chromatography (4:1, hexane/EtOAc) to afford a mixture of disaccharide and rearranged trichloroacetimidate. Also, further attempts to separate the compounds failed. Therefore, the mixture was dissolved in THF (35 ml), LiOH solution (1 M) (19 ml) and hydrogen peroxide solution (12 ml) were added at -5 °C. The reaction mixture was stirred for 30 h while warming to room temperature. Then methanol (17 ml) and potassium hydroxide (3 M, 35 ml) were added and stirred for 3 d. Hydrochloric acid (1 M) was added until pH 7 was reached. The neutral solution was extracted three times with dichloromethane, dried over $MgSO_4$, filtered and concentrated. Column chromatography (3:1, hexane/EtOAc) furnished 145 mg (55%) of **g** as a colorless solid: $[\alpha]_D$: -41.2 (c = 0.25, $CHCl_3$). IR (thin film, $CHCl_3$): 3005, 2923, 1600, 1492, 1451, 1082 cm^{-1}. ^{1}H NMR ($CDCl_3$, 300 MHz) δ 0.87 (t, *J* = 6.9 Hz, 3H), 1.33 (m, 15H), 1.61 (m, 2H), 2.33 (s, 1H), 2.52 (s, 1H), 3.48 (m, 2H), 3.62 (m, 4H), 3.78-3.92 (m, 7H), 4.20 (m, 2H), 4.46 (m, 2H), 4.52-4.66 (m, 4H), 4.83 (d, *J* = 11.7 Hz, 1H), 4.90 (d, *J* = 11.1 Hz, 1H), 5.38 (s, 1H), 7.23-7.37 (m, 20H). ^{13}C NMR ($CDCl_3$, 75 MHz) δ 14.0, 18.0, 22.5, 25.8, 29.1, 29.3, 29.4, 31.7, 68.1, 68.3, 68.6, 70.1, 71.7, 72.6, 73.5, 73.7, 74.9, 75.1, 75.9, 77.9, 79.5, 79.7, 99.9, 103.1, 127.4, 127.6, 127.7, 127.8, 128.1, 128.2, 128.3, 128.4, 137.7, 137.8, 138.4, 138.5. MALDI-MS: *m/z* [M + Na]$^+$ calculated 821.4235, observed 821.4247.

***n*-Octyl α-L-rhamnopyranosyl-(1→3)-β-D-galactopyranoside (II)-**Disaccharide **g** (103 mg, 0.129 mmol) was dissolved in a mixture of methanol and dichloromethane (30 ml, 2:1), $Pd(OH)_2$/C was added and the Ar atmosphere was replaced by a H_2 atmosphere. The reaction mixture was stirred for 12 h. Filtration through a plug of celite removed all solids. Column chromatography using silica gel and dichloromethane/methanol (10:1 → 3:1) in order to remove traces of Pd yielded 50 mg (quant.) of (**II**) as a colorless solid: $[\alpha]_D$: -32.3 (c = 0.13, H_3COH). ^{1}H NMR (CD_3OD, 300 MHz) δ 0.89 (b, 3H), 1.24-1.36 (b, 13H), 1.61 (m, 2H), 3.31 (s, 2H), 3.39 (m, 1H), 3.51 (m, 2H), 3.63 (m, 1H), 3.71 (m, 3H), 3.92 (m, 3H), 4.23 (d, *J* = 7.5 Hz, 1H), 5.05 (s, 1H). ^{13}C NMR (CD_3OD, 75 MHz) δ 14.0, 17.6, 23.3, 26.7, 30.0, 30.1, 30.4, 32.6, 61.8, 69.6, 69.7, 70.5, 71.7, 73.6, 76.0, 81.2, 103.4, 104.5. MALDI-MS: *m/z* [M + Na]$^+$ calculated 461.2357, observed 461.2357.

***n*-Octyl 2-*O*-acetyl-3,4-di-*O*-benzyl-α-L-rhamnopyranosyl-(1→3)-2-*O*-acetyl-4-*O*-benzyl-α-L-rhamnopyranosyl-(1→3)-4,6-di-*O*-benzyl-2-*O*-pivaloyl-β-D-galactopyranoside (h)-**Galactose building block **c** (100 mg, 0.179 mmol) and rhamnose disaccharide building block **f** (209 mg, 0.258 mmol) were azeotroped three times with toluene and dried *in vacuo*. Dichloromethane (3.5 ml) were added and the solution cooled to 0 °C. TMSOTf (9 mg, 7 µl, 0.04 mmol) was added. The mixture was stirred for 120 min and quenched by addition of some drops of pyridine. Afterwards the solvent was removed *in vacuo*. The resulting crude product was purified by column chromatography (4:1, hexane/EtOAc) to afford 164 mg (76%) of **h** as a colorless oil: $[\alpha]_D$: -25.8 (c = 0.33, $CHCl_3$). IR (thin film, $CHCl_3$): 2933, 2871, 1739, 1369, 1072 cm^{-1}. ^{1}H NMR ($CDCl_3$, 300 MHz) δ 0.88 (t, *J* = 6.9 Hz, 3H), 1.24-1.34 (m, 25H), 1.58 (m, 2H), 2.10 (s, 3H), 2.13 (s, 3H), 3.41 (m, 4H), 3.64 (ps, 3H), 4.78-4.88 (m, 5H), 4.07 (dd, *J* = 10.7, 2.2 Hz, 1H), 4.38 (d, *J* = 10.1 Hz, 1H), 4.40-4.68 (m, 7H), 4.80 (d, *J* = 12.3 Hz, 1H), 5.91 (m, 3H), 5.17 (ps, 1H), 5.21 (ps, 2H), 7.24-7.39 (m, 25H). ^{13}C NMR ($CDCl_3$, 75 MHz) δ 14.2, 18.0, 18.2, 21.1, 21.1, 22.8, 26.1, 27.2, 29.3, 29.5, 29.7, 31.9, 38.9, 68.6, 68.7, 68.8, 69.0, 69.7, 71.5, 71.7, 73.5, 73.7, 74.8, 75.0, 75.1, 76.1, 76.2, 77.6, 78.2, 79.8, 80.3, 98.7, 99.1, 101.6, 127.4, 127.4, 127.5, 127.6, 127.7, 127.7, 127.8, 127.8, 128.2, 128.2, 128.3, 128.3, 137.7, 137.8, 137.9, 138.1, 138.5, 169.3, 169.9, 176.5. MALDI-MS: *m/z* [M + Na]$^+$ calculated 1225.6070, observed 1225.6050.

***n*-Octyl α-L-rhamnopyranosyl-(1→3)-α-L-rhamnopyranosyl-(1→3)-β-D-galactopyranoside (III)-**Trisaccharide **h** (135 mg, 0.112 mmol) was dissolved in THF (13 mL), LiOH solution (1 M, 7 ml) and hydrogen peroxide solution (5 ml) were added at -5 °C. The reaction mixture was stirred for 3 d while warming to room temperature. Then methanol (7 ml) and potassium hydroxide (3 M, 13 ml) were added and stirred for 2 d. Hydrochloric acid (1 M) was added until pH 7 was reached. The

neutral solution was extracted three times with dichloromethane, dried over $MgSO_4$, filtered and concentrated. Mass spectrometry still shows a pivaloyl group in the molecule. Therefore, the crude product was dissolved in methanol (10 ml) and NaOMe (500 mg) were added. The mixture was stirred for 12 h and the solvent removed *in vacuo*. Column chromatography (EtOAc) furnished 102 mg (88%) of a colorless solid. This compound (95 mg, 0.092 mmol) was dissolved in a mixture of methanol (10 ml) and dichloromethane (15 mL), $Pd(OH)_2/C$ was added and the Ar atmosphere replaced by a H_2 atmosphere. The reaction mixture was stirred for 12 h. Filtration through a plug of celite removed all solids. Column chromatography using silica gel and dichloromethane/methanol (10:1 $\rightarrow$ 4:1) was performed in order to remove traces of Pd. 38 mg (quant.) of (**III**) as a colorless highly viscous oil were obtained: $[\alpha]_D$: -60.5 (c = 1.25, H_3COH). ^{1}H NMR (CD_3OD, 300 MHz) δ 0.89 (b, 3H), 1.24-1.33 (b, 16H), 1.61 (b, 2H), 3.38 (s, 1H), 3.47-3.57 (m, 4H), 3.62-3.92 (m, 9H), 3.99 (s, b, 1H), 4.07 (s, b, 1H), 4.23 (d, *J* = 7.5 Hz, 1H), 5.02 (s, 1H), 5.04 (s, 1H). ^{13}C NMR (CD_3OD, 75 MHz) δ 14.0, 17.6, 17.6, 23.3, 26.7, 30.0, 30.1, 30.4, 32.6, 61.8, 69.6, 69.7, 70.0, 70.5, 71.5, 71.7, 72.8, 73.7, 76.0, 79.4, 81.1, 103.4, 103.6, 104.6. MALDI-MS: *m/z* $[M + Na]^+$ calculated 607.2936, observed 607.2925.

References included in supplementary methods:

26. Fürstner, A., Müller, T. (1999) *J. Am. Chem. Soc.* **121**, 7814-7821
27. Love, K. R., Seeberger, P. H. (2004) *Angew. Chem. Int. Ed.* **43**, 602-605
28. Werz, D. B., Seeberger, P. H. (2005) *Angew. Chem. Int. Ed.* **44**, 6315-6318

[1]The abbreviations used are: Ac, acetyl; Ar, argon; Bn, benzyl; CAN, cerium ammonium nitrate; EtOAc, ethyl acetate; Fmoc, fluorenylmethoxycarbonyl; H_2, hydrogen; IR, IR spectroscopy; *J*, coupling constant; MP, methoxyphenyl; NaOMe, sodium methoxide; OBu, *O*-butyl; Pd, palladium; Piv, pivaloyl; THF, tetrahydrofuran; TMSOTf, trimethylsilyl trifluoromethanesulfonate.

SUPPLEMENTARY RESULTS

Sequence comparison of the transferases WsaC, WsaD, WsaE and WsaF–WsaC shows homology to the α-1,3-L-rhamnosyltransferases from *Streptococcus pyogenes* MGAS10394 (62) (protein accession number YP_059939) and MGAS6180 (63) (YP_280052) as well as to RgpB from *Lactococcus lactis* and *Streptococcus thermophilus*, RgpBc from *Streptococcus mutans*, and Cps2F from *Streptococcus pneumoniae*, and several other enzymes, all of which have been classified as rhamnosyltransferases involved in cell wall biosynthesis (Fig. S2). WsaD shows high homology to the putative rhamnosyltransferase JexE from *Paenibacillus jamilae*, the glycosyltransferases from *Clostridium beijerincki* (protein accession number ZP_00911019) and *Methanosphaera stadtmanae*, as well as some homology to rhamnosyltransferases from *Shigella dysenteriae* (37) and *Shigella flexneri* (64).

The N-terminal portion of WsaE (aa 70 to 150) revealed homology to methyltransferases *e.g.*, to COG2226.2, UbiE, a menaquinone biosynthesis methyltransferase from *Methanosarcina acetivorans* (65), and to PFAM 08241.1 and PFAM 08242.1 methyltransferase family 11 and 12, respectively, both of which are SAM-dependent methyltransferases (Fig. S2). The central and C-terminal portions contain two glycosyltransferase domains, which are homologous *e.g.*, to the O-antigen biosynthesis protein from *Planctomyces maris* DSM 8797 (ZP_01855298) and *Xanthomonas oryzae* (ABI93188), or to the GT-2 glycosyltransferases from *Burkholderia cenocepacia* MC0-3 (ZP_01563358) and from *Enterococcus faecalis* (66) (ORFde16, AAC35930) (Fig. S3). The first glycosyltransferase domain contains motifs that are typical of inverting glycosyltransferases and contains a conserved DD motif (aa 638-639) and, 53 aa downstream, the DXDD motif (DHDD, aa 691-694). The second glycosyltransferase domain contains ED motifs (aa 970-971 and 981-982) and a DXE motif (aa 1062-1064).

Highest homology for WsaF was found to conserved hypothetical proteins from *Planctomyces maris* (ZP_01855299; 430 aa; E-value: $2e^{-89}$) and from *Streptococcus pneumoniae* (CAI34499; 414 aa, E-value: $8e^{-70}$). Homologies to putative glycosyltransferases, for instance, from *Anabaena variabilis* (protein accession number ABA22956; 406 aa, E-value: $6e^{-20}$) and WbbX (421 aa, E-value: $2e^{-08}$) from *Yersinia enterocolitica* were found as well.

References included in supplementary results:

37. Morona, R., Macpherson, D. F., Van Den, B. L., Carlin, N. I., and Manning, P. A. (1995) *Mol. Microbiol.* **18,** 209-223
62. Banks, D. J., Porcella, S. F., Barbian, K. D., Beres, S. B., Philips, L. E., Voyich, J. M., DeLeo, F. R., Martin, J. M., Somerville, G. A., and Musser, J. M. (2004) *J. Infect. Dis.* **190,** 727-738
63. Green, N. M., Zhang, S., Porcella, S. F., Nagiec, M. J., Barbian, K. D., Beres, S. B., LeFebvre, R. B., and Musser, J. M. (2005) *J. Infect. Dis.* **192,** 760-770
64. Klena, J. D. and Schnaitman, C. A. (1993) *Mol. Microbiol.* **9,** 393-402
65. Galagan, J. E., Nusbaum, C., Roy, A., Endrizzi, M. G., Macdonald, P., FitzHugh, W., Calvo, S., Engels, R., Smirnov, S., Atnoor, D., Brown, A., Allen, N., Naylor, J., Stange-Thomann, N., DeArellano, K., Johnson, R., Linton, L., McEwan, P., McKernan, K., Talamas, J., Tirrell, A., Ye, W., Zimmer, A., Barber, R. D., Cann, I., Graham, D. E., Grahame, D. A., Guss, A. M., Hedderich, R., Ingram-Smith, C., Kuettner, H. C., Krzycki, J. A., Leigh, J. A., Li, W., Liu, J., Mukhopadhyay, B., Reeve, J. N., Smith, K., Springer, T. A., Umayam, L. A., White, O., White, R. H., Conway, d. M., Ferry, J. G., Jarrell, K. F., Jing, H., Macario, A. J., Paulsen, I., Pritchett, M., Sowers, K. R., Swanson, R. V., Zinder, S. H., Lander, E., Metcalf, W. W., and Birren, B. (2002) *Genome Res.* **12,** 532-542
66. Xu, Y., Murray, B. E., and Weinstock, G. M. (1998) *Infect. Immun.* **66,** 4313-4323

SUPPLEMENTARY TABLES

Table S1: Bacterial strains and plasmids used in this work.

Strains or plasmids	Genotype and relevant characteristics	Sources or references
E. coli DH5α™	F$^-$ φ80*lac*ZM*15* (*lacZYA-argF*) *U169 deoR recA1 endA1 hsdR17* (*rk*$^-$, *mk*$^+$) *phoA supE44 thi-1 gyrA96 relA1 λ*$^-$	Invitrogen
E. coli BL21 Star (DE3)	F$^-$ *ompT hsdSB* (*rB*$^-$*mB*$^-$) *gal dcm rne131*	Invitrogen
E. coli C43(DE3)	F$^-$ *ompT hsdSB* (*rB*$^-$*mB*$^-$) *gal dcm* (DE3) C43	Lucigen (67)
pET28a(+)	*E. coli* expression vector; Km^R	Novagen
pNGB220	pET28a-WsaC; pET28a(+) expressing WsaC (aa 1-324) from *G. stearothermophilus* NRS 2004/3a with N-terminal His$_6$-tag; Km^R	This study
pNGB221	pET28a-WsaC_I; pET28a(+) expressing WsaC from *G. stearothermophilus* NRS 2004/3a devoid of the C-terminal transmembrane domain (aa 1-280) with N-terminal His$_6$-tag; Km^R	This study
pNGB230	pET28a-WsaD; pET28a(+) expressing WsaD (aa 1-289) from *G. stearothermophilus* NRS 2004/3a with N-terminal His$_6$-tag; Km^R	This study
pNGB231	pET28a-WsaD_I; pET28a(+) expressing WsaD from *G. stearothermophilus* NRS 2004/3a devoid of the C-terminal transmembrane domain (aa 1-254) with N-terminal His$_6$-tag; Km^R	This study
pNGB240	pET28a-WsaE; pET28a expressing WsaE (aa 1-1127) from *G. stearothermophilus* NRS 2004/3a with N-terminal His$_6$-tag; Km^R	This study
pNGB241	pET28a-WsaE_M; pET28a expressing WsaE_M (aa 1-170) from *G. stearothermophilus* NRS 2004/3a with N-terminal His$_6$-tag; Km^R	This study
pNGB242	pET28a-WsaE_N; pET28a expressing WsaE_N (aa 1-368) from *G. stearothermophilus* NRS 2004/3a with N-terminal His$_6$-tag; Km^R	This study
pNGB243	pET28a-WsaE_B; pET28a expressing WsaE_B (aa 368-1127) from *G. stearothermophilus* NRS 2004/3a with N-terminal His$_6$-tag; Km^R	This study
pNGB244	pET28a-WsaE_C; pET28a expressing WsaE_C (aa 765-1127) from *G. stearothermophilus* NRS 2004/3a with N-terminal His$_6$-tag; Km^R	This study
pNGB245	pET28a-WsaE_A; pET28a expressing WsaE_A (aa 368-863) from *G. stearothermophilus* NRS 2004/3a with N-terminal His$_6$-tag; Km^R	This study
pNGB250	pET28a-WsaF; pET28a(+) expressing WsaF (aa 1-413) from *G. stearothermophilus* NRS 2004/3a with N-terminal His$_6$-tag; Km^R	This study
pNGB200	pET28a-WsaP; pET28a(+) expressing WsaP (aa 1-471) from *G. stearothermophilus* NRS 2004/3a with N-terminal His$_6$-tag; Km^R	(22)
pNGB261	pET28a-RmlB; pET-28a(+) expressing RmlB from *G. stearotermophilus* NRS 2004/3a with N-terminal His$_6$-tag; Km^R	This study
pNGB262	pET28a-RmlC; pET-28a(+) expressing RmlC from *G. stearothermophilus* NRS 2004/3a with N-terminal His$_6$-tag; Km^R	This study
pNGB263	pET28a-RmlD; pET-28a(+) expressing RmlD from *G. stearotermophilus* NRS 2004/3a with N-terminal His$_6$-tag; Km^R	This study

Table S2: PCR primers used for the amplification of WsaD, WsaC, WsaE, WsaF, RmlB, RmlC and RmlD from *G. stearothermophilus* NRS 2004/3a for the design of different expression constructs.

Primer	Nucleotide sequence (5' → 3')	Orientation
pET-WsaD_for	AATCACCATATGGATATTAGCATTATTATCGTGAAT	forward
pET-WsaD_rev	ATAAGAATCTCGAGTTAACCACCTATTTTTCGAAAAGTGT	reverse
pET-WsaD_I_rev	AATCACTCGAGCTAATGCTTCCTATGGAATAAAAACATC	reverse
pET-WsaC_for	AATCACCATATGGAGATGCCATTGGTTT	forward
pET-WsaC_rev	ATAAGAATCTCGAGCTAATATTTTAACTTTTTAAAAATCCATATTG	reverse
pET-WsaC_I_rev	AATCACTCGAGCTATTGACTAAATTTTTTTCTAGCTTCCG	reverse
pET-WsaE_for	AATCAGCTAGCATGGAGCGTTGTAGAATGAATAA	forward
pET-WsaE _B_for	AATCAGCTAGCATGCGTATTAAGAATAGATTAAAAA	forward
pET-WsaE _C_for	AATCAGCTAGCATGGGCTTTCGAAAAGGTTTTG	forward
pET-WsaE_rev	ATAAGAATCTCGAGCTACGACCTTATTGTTGGAATCAAA	reverse
pET-WsaE_A_rev	ATAAGAATCTCGAGCTAAACCAATGGATAATCCTCTG	reverse
pET-WsaE_N_rev	ATAAGAATCTCGAGCTACATAGACTCAGCTTGGTTTTGC	reverse
pET-WsaE_M_rev	ATAAGAATCTCGAGCTATATCTTCCGTTTCCTCTAGG	reverse
pET-WsaF_for	GGGGTACCCCATATGgTTCAAAAATTAATACAGATATTAAG	forward
pET-WsaF_rev	GGGGTACCCCGAGCTCGAAAATAAACCTACGAATAGAGTCATCA	reverse
RmlB_for	AATCAGCTAGCATGAAAGTATTGATTACCGGC	forward
RmlB_rev	AATCACTCGAGCCTAACTGCCCGTTTGC	reverse
RmlC_for	AATCACCATATGgAAATTATTGAGACTAAGTTTAGTAATG	forward
RmlC_rev	AATCACTCGAGCTGCATTTCCTTCCCTTAATAAG	reverse
RmlD_for	AATCACCATATGgAAATTGTTGTTACGGGGG	forward
RmlD_rev	AATCACTCGAGGATCTATTGTAAATATATCACTTCAAATC	reverse

Triplets corresponding to the initiation and termination codons in the primer sequence are boxed. Lowercase letters indicate changes in the original nucleotides sequence. Artificial restriction sites are underlined.

Table S3: ESI-QTOF MS analysis of octyl-linked products of *in vitro* activity assays.

Sample	$[M+Na]^+_{exp}$	$[M+Na]^+_{theor}$	assignment
substrate (**II**)	461.22	461.24	Rha-Gal-octyl
substrate (**III**)	607.25	607.29	Rha-Rha-Gal-octyl
WsaF + (**II**) + dTDP-β-L-Rha	607.26	607.29	Rha-Rha-Gal-octyl (**IV**)
WsaC + (**III**) + dTDP-β-L-Rha	753.33	753.35	Rha-Rha-Rha-Gal-octyl (**V**)
WsaC + (**III**) + dTDP-β-L-Rha	899.37	899.41	Rha-Rha-Rha-Rha-Gal-octyl (**VI**)
WsaE + (**III**) + dTDP-β-L-Rha	753.33	753.35	Rha-Rha-Rha-Gal-octyl (**VII**)

References included in supplementary tables:

22. Steiner, K., Novotny, R., Patel, K., Vinogradov, E., Whitfield, C., Valvano, M. A., Messner, P., and Schäffer, C. (2007) *J. Bacteriol.* **189,** 2590-2598
67. Miroux, B. and Walker, J. E. (1996) *J. Mol. Biol.* **260,** 289-298

SUPPLEMENTARY FIGURE LEGENDS

Fig. S1. **Reaction scheme** for the synthesis of β-D-Gal-(1→*O*)-octyl (**I**), α-L-Rha-(1→3)-β-D-Gal-(1→*O*)-octyl (**II**) and α-L-Rha-(1→3)-α-L-Rha-(1→3)-β-D-Gal-(1→*O*)-octyl (**III**) used as acceptors in this study.

Fig. S2. **Multiple sequence alignment of the N-terminal region of WsaC** from *G. stearothermophilus* NRS 2004/3a with putative rhamnosyltransferases of two different *Streptococcus pyogenes* strains, RgbB from *Lactococcus lactis* and RgpBc from *Streptococcus mutans*. Conserved amino acids described for inverting transferases are highlighted.

Fig. S3. **Sequence alignment** of the N-terminal region of WsaE from *G. stearothermophilus* NRS 2004/3a, containing the putative methyltransferase domain, with menaquinone biosynthesis methyltransferase (UbiE) from *Methanosarcina acetivorans* C2A.

Fig. S4. **Sequence alignment** of the C-terminal region of WsaE from *G. stearothermophilus* NRS 2004/3a, containing the putative rhamnosyltransferase domains, with O-antigen biosynthesis proteins from *Planctomyces maris* DSM 8797 and (protein accession number ZP_01855298) and *Xanthomonas oryzae* (ABI93188) or the GT-2 glycosyltransferases from *Burkholderia cenocepacia* MC0-3 (ZP_01563358) and *Enterococcus faecalis* (AAC35930).

Fig. S5. A: **Western immunoblot analysis of the expression of WsaC, WsaD and WsaF** in *E. coli* BL21 Star (DE3). Proteins were detected with anti-His-tag antibody. Lanes 1, 6 and 8: Precision Plus Protein™ Standard All Blue (Biorad); lane 2: WsaC (37 kDa); lane 3: WsaC_I (36.5 kDa); lane 4: WsaD (30 kD); lane 5: WsaD_I (31.6 kDa); lane 7: WsaF (50.5 kDa). B: **Western Immunoblot analysis of the expression of WsaE and truncated forms thereof** in *E. coli* BL21 Star (DE3). Proteins were detected with anti-His-tag antibody. Lane 1: Precision Plus Protein™ Standard All Blue (Biorad); lane 2: full length WsaE (135.0 kDa); lane 3: WsaE_A (60.5 kDa; aa 368-863, first rhamnosyltransferase domain); lane 4: WsaE_B (90.8 kDa; aa 368-1127, both rhamnosyltransferase domains); lane 5: WsaE_C (44.1 kDa; aa 765-1127, C-terminal rhamnosyltransferase domain) lane 6: WsaE_M (22.7 kDa; aa 1-170, UbiE domain); lane 7: WsaE_N (46.8 kDa; aa 1-368, UbiE domain plus 200 aa downstream). For details about the truncated forms see Figure 1.

Fig. S6. **TLC pattern of rhamnosyltransferase activity assays using octyl-linked oligosaccharides.** For the reactions crude extract (WsaC and WsaD) or purified enzyme (WsaC_I, WsaE and WsaF) were used and the products were separated on Silica TLC plates with chloroform/methanol/water 65:25:4 as solvent. Lanes 1-7: WsaD; lanes 12-18: WsaC; lanes 20-23: WsaC_I; lanes 24-27: WsaE; lanes 28-31: WsaF; lanes 1, 12, 20, 24 and 28: β-D-Gal-(1→*O*)-octyl (**I**) and dTDP-β-L-Rha; lanes 2, 13, 21, 25 and 29: α-L-Rha-(1→3)-β-D-Gal-(1→*O*)-octyl (**II**) and dTDP-β-L-Rha; lanes 3, 14, 22, 26 and 30: α-L-Rha-(1→3)-α-L-Rha-(1→3)-β-D-Gal-(1→*O*)-octyl (**III**) and dTDP-β-L-Rha; lanes 4, 8 and 15: (**I**); lanes 5, 9 and 16: (**II**); lanes 6, 10 and 17: (**III**); lanes 7, 11, 18, 23, 27 and 31: dTDP-β-L-Rha. Staining was performed with thymol.

Fig. S7. **(+)ESI-QTOF MS^2 spectrum of the singly charged ions at *m/z* 899.40** corresponding to the product (**VI**) of the *in vitro* assay of WsaC; the loss of rhamnoses can be deduced from the mass difference of ~146 between neighboring Y-ions.

Figure S1

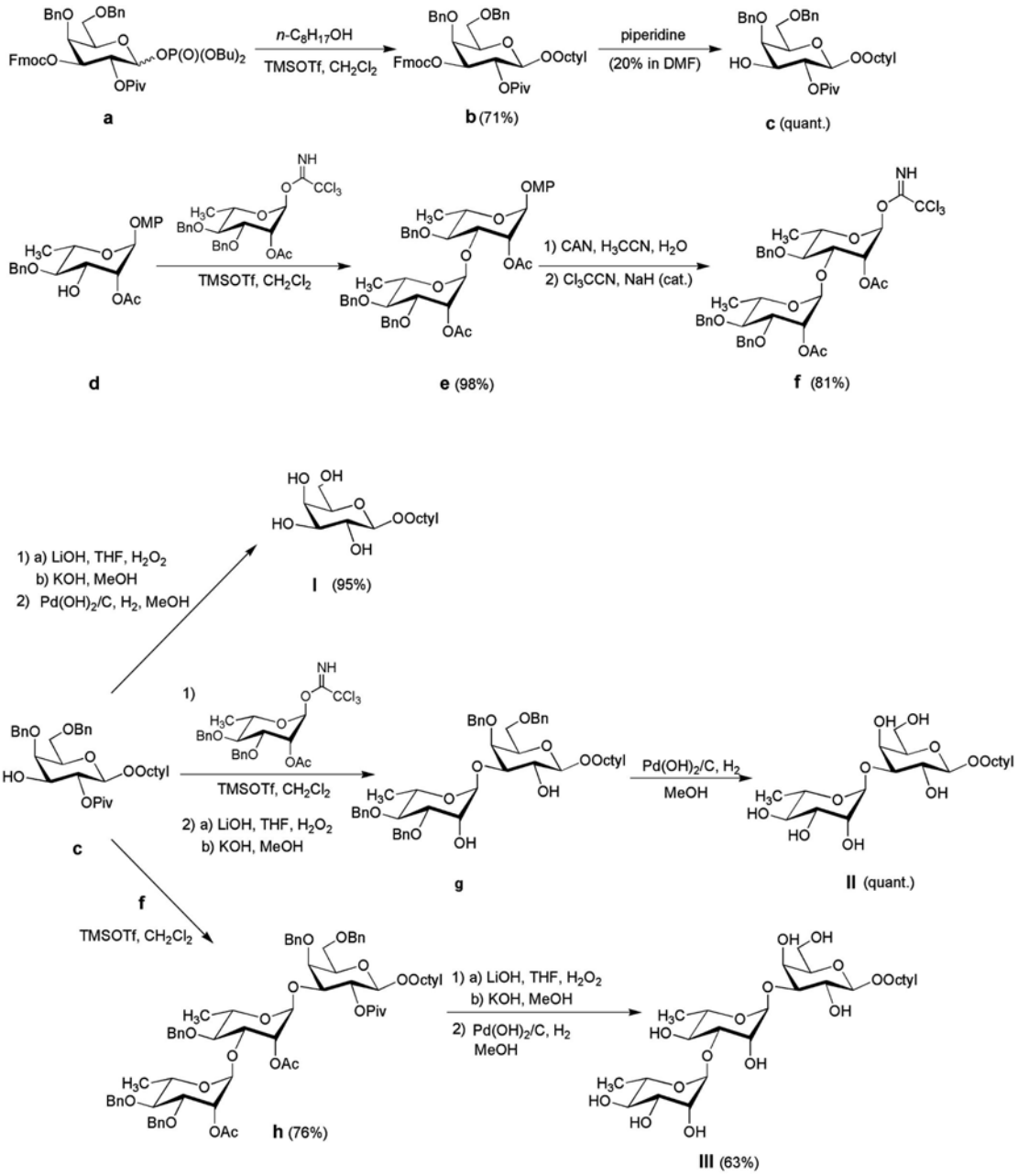


Figure S2


StrepPyo10394 ---MNINILLSTYN-GERFLAEQIQSIQRQTVNDWTLLIRDDGSTDGTQDIIRTFVK-ED 55
StrepPyo6180 ---MNINILLSTYN-GERFLAEQIQSIQRQTVNDWTLLIRDDGSTDGTQDIIRTFVK-ED 55
RgpBc ---MKVNILMSTYN-GQEFIAQQIQSIQKQTFENWNLLIRDDGSSDGTPKIIADFAK-SD 55
RgbBLac ---MRVNILMSTYN-GEKFVADQIESIQKQTYTDWNLIIRDDGSSDRTCEIVDDFVS-KD 55
WsaC MEMPLVSIVVATYFPRTDFFEKQLQSLNNQTYENIEIIICDDSANDAEYEKVKKMVENII 60
DXDD
SP10394 KRIQWINEGQTENLGVIKNFYTLLR-HQKADVYFFSDQDDIWLDNKLEVTLLEAQKHEMT 114
SP6180 KRIQWINEGQTENLGVIKNFYTLLK-HQKADVYFFSDQDDIWLDNKLEVTLLEAQKHEMT 114
RgpBc ARIRFINADKRENFGVIKNFYTLLK-YEKADYYFFSDQDDVWLPQKLELTLASVEKENNQ 114
RgbBLac NRIKLIR---AENVGVIKSFHELVTDSNNADFYFFADQDDYWLPEKLSVMLEETKKHDNS 112
WsaC SRFPCKVIRNEKNVGSNKTFERLTQ-EANGDYICYCDQDDIWLSEKVERLVNHITKHHCT 119
SP10394 APLLVYTDLKVVTQHLAICHDSMIKT----QSGHANTSLLQELTENTVTGGTMMITHALA 170
SP6180 APLLVYTDLKVVTQHLAICHDSMIKT----QSGHANTSLLQELTENTVTGGTMMITHALA 170
RgpBc IPLMVYTDLTVVDRDLQVLHDSMIKT----QSHHANTSLLEELTENTVTGGTMMVNHCLA 170
RgbBLac KPVMYYTDLKVTDRNLNVTSESMIRG----QSDHANTKLVQELTENTVTGGASMINHELA 168
WsaC ---LVYSDLSLIDENDRIIHKSFKRSNFRLKHVHGDNTFAHLINRNSVTGCAMMIRADVA 176
SP10394 EEWTTC--DGLLMHDWYLALLASATGKLVYLDIPTELYRQHDANVLGARTWS--KRMKNW 226
SP6180 EEWTTC--DGLLMHDWYLALLASATGKLVYLDIPTELYRQHDANVLGARTWS--KRMKNW 226
RgpBc KQWKQCY-DDLIMHDWYLALLAASLGKLIYLDETTELYRQHESNVLGARTWS--KRLKNW 227
RgbBLac QLWQST--NDIIMHDWYLAIVAAALGELVYIDQPTHLYRQHDSNVLGARTLS--KRIKKW 224
WsaC KSAIPFPDYDEFVHDHWLAIHAAVKGSLGYIKEPLVWYRIHLGNQIGNQRLVNITNINDY 236

Figure S3

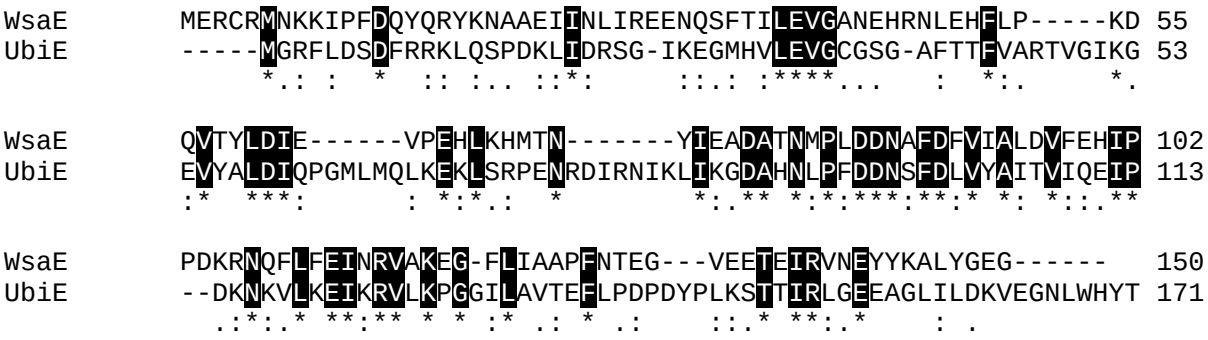

WsaE MERCRMNKKIPFDQYQRYKNAAEIINLIREENQSFTILEVGANEHRNLEHFLP-----KD 55
UbiE -----MGRFLDSDFRRKLQSPDKLIDRSG-IKEGMHVLEVGCGSG-AFTTFVARTVGIKG 53
WsaE QVTYLDIE------VPEHLKHMTN-------YIEADATNMPLDDNAFDFVIALDVFEHIP 102
UbiE EVYALDIQPGMLMQLKEKLSRPENRDIRNIKLIKGDAHNLPFDDNSFDLVYAITVIQEIP 113
WsaE PDKRNQFLFEINRVAKEG-FLIAAPFNTEG---VEETEIRVNEYYKALYGEG------ 150
UbiE --DKNKVLKEIKRVLKPGGILAVTEFLPDPDYPLKSTTIRLGEEAGLILDKVEGNLWHYT 171

Figure S4

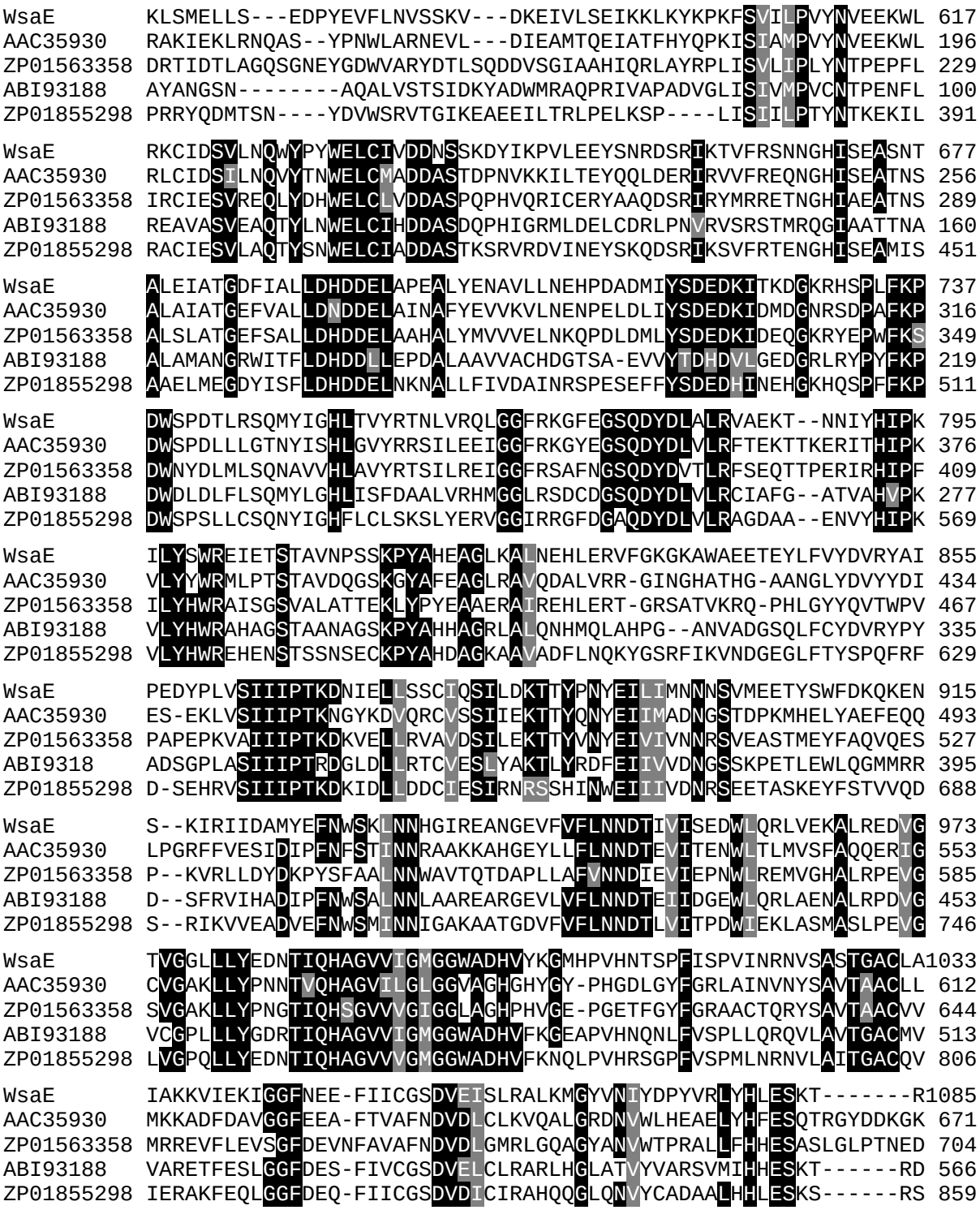

WsaE KLSMELLS---EDPYEVFLNVSSKV---DKEIVLSEIKKLKYKPKFSVILPVYNVEEKWL 617
AAC35930 RAKIEKLRNQAS--YPNWLARNEVL---DIEAMTQEIATFHYQPKISIAMPVYNVEEKWL 196
ZP01563358 DRTIDTLAGQSGNEYGDWVARYDTLSQDDVSGIAAHIQRLAYRPLISVLIPLYNTPEPFL 229
ABI93188 AYANGSN--------AQALVSTSIDKYADWMRAQPRIVAPADVGLISIVMPVCNTPENFL 100
ZP01855298 PRRYQDMTSN----YDVWSRVTGIKEAEEILTRLPELKSP----LISIILPTYNTKEKIL 391
WsaE RKCIDSVLNQWYPYWELCIVDDNSSKDYIKPVLEEYSNRDSRIKTVFRSNNGHISEASNT 677
AAC35930 RLCIDSILNQVYTNWELCMADDASTDPNVKKILTEYQQLDERIRVVFREQNGHISEATNS 256
ZP01563358 IRCIESVREQLYDHWELCLVDDASPQPHVQRICERYAAQDSRIRYMRRETNGHIAEATNS 289
ABI93188 REAVASVEAQTYLNWELCIHDDASDQPHIGRMLDELCDRLPNVRVSRSTMRQGIAATTNA 160
ZP01855298 RACIESVLAQTYSNWELCIADDASTKSRVRDVINEYSKQDSRIKSVFRTENGHISEAMIS 451
WsaE ALEIATGDFIALLDHDDELAPEALYENAVLLNEHPDADMIYSDEDKITKDGKRHSPLFKP 737
AAC35930 ALAIATGEFVALLDNDDELAINAFYEVVKVLNENPELDLIYSDEDKIDMDGNRSDPAFKP 316
ZP01563358 ALSLATGEFSALLDHDDELAAHALYMVVVELNKQPDLDMLYSDEDKIDEQGKRYEPWFKS 349
ABI93188 ALAMANGRWITFLDHDDLLEPDALAAVVACHDGTSA-EVVYTDHDVLGEDGRLRYPYFKP 219
ZP01855298 AAELMEGDYISFLDHDDELNKNALLFIVDAINRSPESEFFYSDEDHINEHGKHQSPFFKP 511
WsaE DWSPDTLRSQMYIGHLTVYRTNLVRQLGGFRKGFEGSQDYDLALRVAEKT--NNIYHIPK 795
AAC35930 DWSPDLLLGTNYISHLGVYRRSILEEIGGFRKGYEGSQDYDLVLRFTEKTTKERITHIPK 376
ZP01563358 DWNYDLMLSQNAVVHLAVYRTSILREIGGFRSAFNGSQDYDVTLRFSEQTTPERIRHIPF 409
ABI93188 DWDLDLFLSQMYLGHLISFDAALVRHMGGLRSDCDGSQDYDLVLRCIAFG--ATVAHVPK 277
ZP01855298 DWSPSLLCSQNYIGHFLCLSKSLYERVGGIRRGFDGAQDYDLVLRAGDAA--ENVYHIPK 569
WsaE ILYSWREIETSTAVNPSSKPYAHEAGLKALNEHLERVFGKGKAWAEETEYLFVYDVRYAI 855
AAC35930 VLYYWRMLPTSTAVDQGSKGYAFEAGLRAVQDALVRR-GINGHATHG-AANGLYDVYYDI 434
ZP01563358 ILYHWRAISGSVALATTEKLYPYEAAERAIREHLERT-GRSATVKRQ-PHLGYYQVTWPV 467
ABI93188 VLYHWRAHAGSTAANAGSKPYAHHAGRLALQNHMQLAHPG--ANVADGSQLFCYDVRYPY 335
ZP01855298 VLYHWREHENSTSSNSECKPYAHDAGKAAVADFLNQKYGSRFIKVNDGEGLFTYSPQFRF 629
WsaE PEDYPLVSIIIPTKDNIELLSSCIQSILDKTTYPNYEILIMNNNSVMEETYSWFDKQKEN 915
AAC35930 ES-EKLVSIIIPTKNGYKDVQRCVSSIIEKTTYQNYEIIMADNGSTDPKMHELYAEFEQQ 493
ZP01563358 PAPEPKVAIIIPTKDKVELLRVAVDSILEKTTYVNYEIVIVNNRSVEASTMEYFAQVQES 527
ABI9318 ADSGPLASIIIPTRDGLDLLRTCVESLYAKTLYRDFEIIVVDNGSSKPETLEWLQGMMRR 395
ZP01855298 D-SEHRVSIIIPTKDKIDLLDDCIESIRNRSSHINWEIIIVDNRSEETASKEYFSTVVQD 688
WsaE S--KIRIIDAMYEFNWSKLNNHGIREANGEVFVFLNNDTIVISEDWLQRLVEKALREDVG 973
AAC35930 LPGRFFVESIDIPFNFSTINNRAAKKAHGEYLLFLNNDTEVITENWLTLMVSFAQQERIG 553
ZP01563358 P--KVRLLDYDKPYSFAALNNWAVTQTDAPLLAFVNNDIEVIEPNWLREMVGHALRPEVG 585
ABI93188 D--SFRVIHADIPFNWSALNNLAAREARGEVLVFLNNDTEIIDGEWLQRLAENALRPDVG 453
ZP01855298 S--RIKVVEADVEFNWSMINNIGAKAATGDVFVFLNNDTLVITPDWIEKLASMASLPEVG 746
WsaE TVGGLLLYEDNTIQHAGVVIGMGGWADHVYKGMHPVHNTSPFISPVINRNVSASTGACLA1033
AAC35930 CVGAKLLYPNNTVQHAGVILGLGGVAGHGHYGY-PHGDLGYFGRLAINVNYSAVTAACLL 612
ZP01563358 SVGAKLLYPNGTIQHSGVVVGIGGLAGHPHVGE-PGETFGYFGRAACTQRYSAVTAACVV 644
ABI93188 VCGPLLLYGDRTIQHAGVVIGMGGWADHVFKGEAPVHNQNLFVSPLLQRQVLAVTGACMV 513
ZP01855298 LVGPQLLYEDNTIQHAGVVVGMGGWADHVFKNQLPVHRSGPFVSPMLNRNVLAITGACQV 806
WsaE IAKKVIEKIGGFNEE-FIICGSDVEISLRALKMGYVNIYDPYVRLYHLESKT-------R1085
AAC35930 MKKADFDAVGGFEEA-FTVAFNDVDLCLKVQALGRDNVWLHEAELYHFESQTRGYDDKGK 671
ZP01563358 MRREVFLEVSGFDEVNFAVAFNDVDLGMRLGQAGYANVWTPRALLFHHESASLGLPTNED 704
ABI93188 VARETFESLGGFDES-FIVCGSDVELCLRARLHGLATVYVARSVMIHHESKT------RD 566
ZP01855298 IERAKFEQLGGFDEQ-FIICGSDVDICIRAHQQGLQNVYCADAALHHLESKS------RS 859
WsaE DSFIPERDFELSAKYYSP-YREIGDPYYNQNLSYNHLIPTIRS 1127
AAC35930 KKKRFEQEKVMMEEKWGP-LIEN-DPFYNPNLTRDIP------ 706
ZP01563358 RRRQFLEECDNFRRIWAD-VIRN-DPFYNPNLTISGGDFRPNF 762
ABI93188 PREIPESDFVRSAQAYSPYREE-GDPFFSPNLDYMASSPRLRG 611
ZP01855298 S-FIPKQDFLMSEIRYAPYRNDKGDPYFNENLDLMSTMPRMLT 903

Figure S5

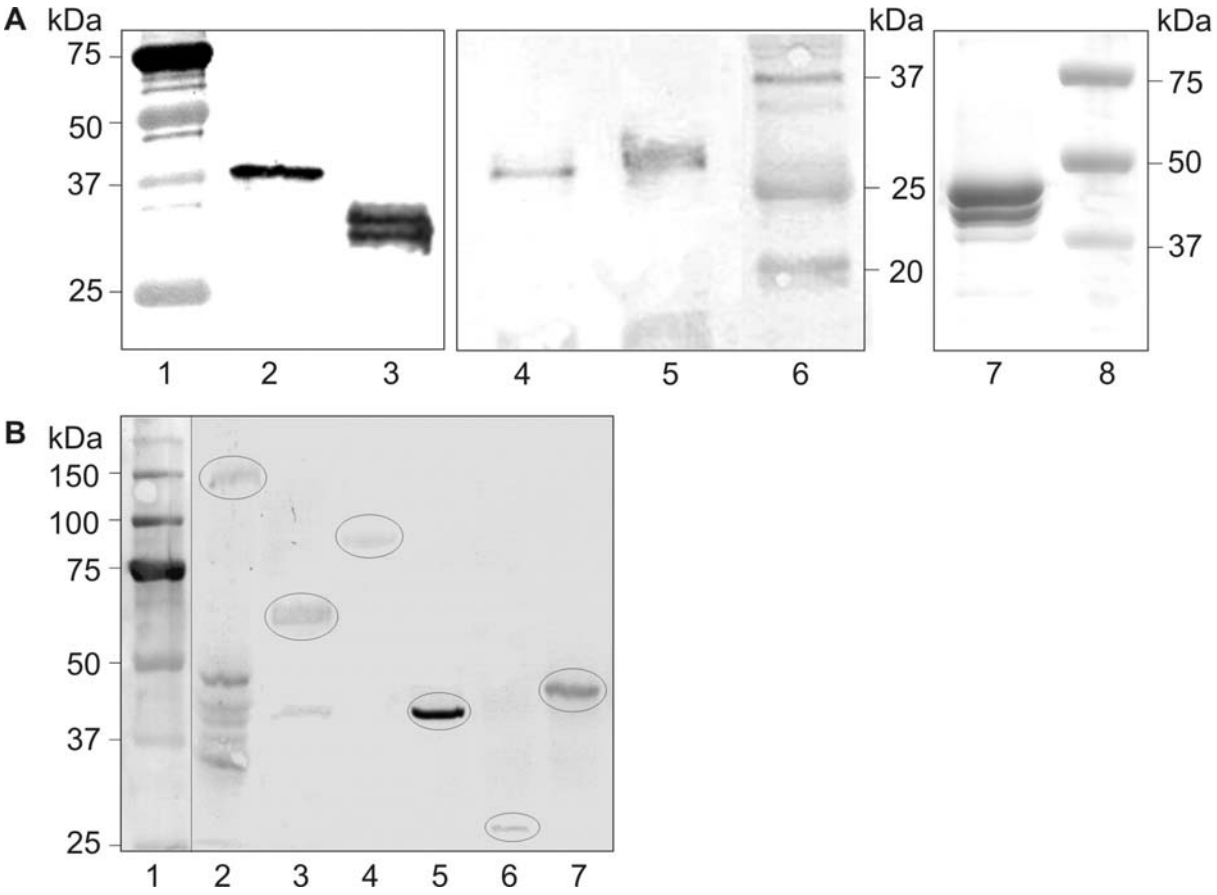

A
kDa
75
50
37
25
1
2
3
kDa
37
25
20
4
5
6
kDa
75
50
37
7
8
B
kDa
150
100
75
50
37
25
1
2
3
4
5
6
7

Figure S6

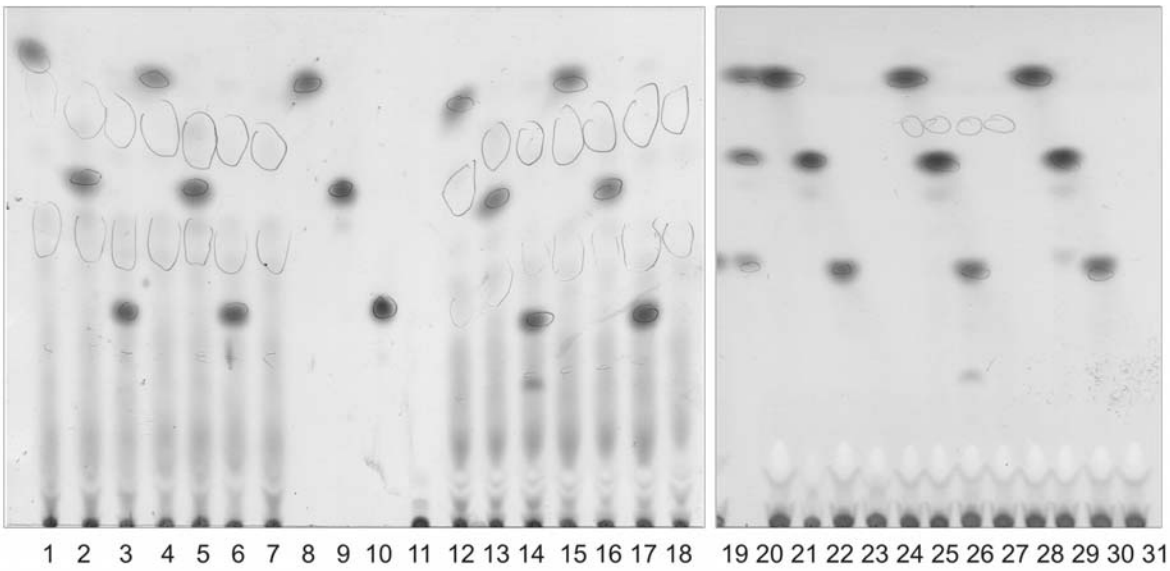
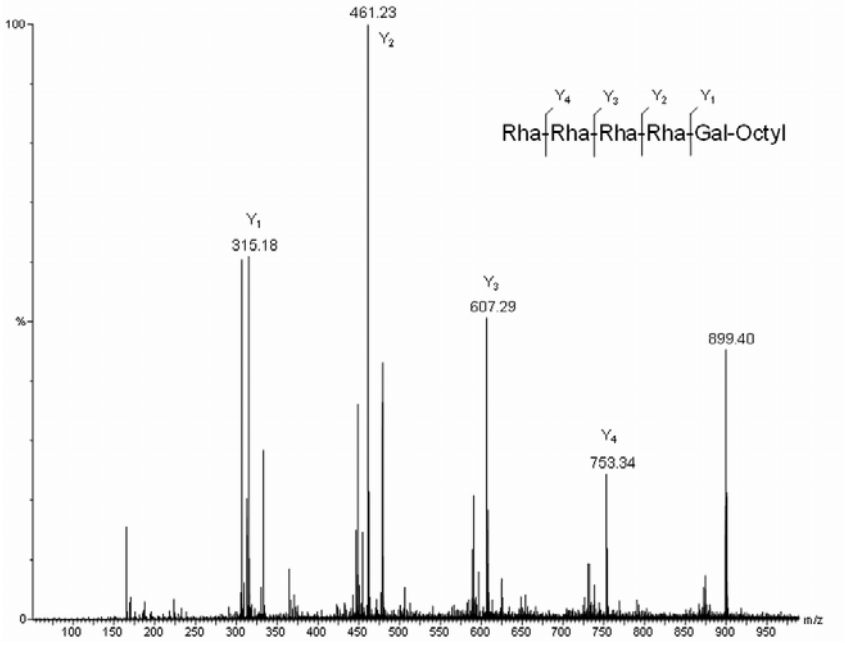

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31

Figure S7


461.23
Y2
Y4 Y3 Y2 Y1
Rha-Rha-Rha-Rha-Gal-Octyl
Y1
315.18
Y3
607.29
899.40
Y4
753.34
100
%
0
m/z
100 150 200 250 300 350 400 450 500 550 600 650 700 750 800 850 900 950