Dissertation

zur Erlangung des akademischen Grades
Doctor rerum naturalium technicarum
an der Universität für Bodenkultur Wien

Functional biochemical analysis and crystallization of essential components from the surface layer anchoring system of *Paenibacillus alvei* CCM 2051^{T}

Eingereicht von

MSc. Arturo López Guzmán

Betreuer: Ao. Univ. Prof. DI. Dr. Paul Messner
Betreuerin: Ao. Univ. Prof. DI. Dr. Christina Schäffer

Durchgeführt am
Department für NanoBiotechnologie, *NanoGlykobiologie* unit
Universität für Bodenkultur Wien
Vorstand: Univ. Prof. Dr. José Luis Toca-Herrera

Wien, Mai 2017

Acknowledgements

In the first place, I would like to express my sincere gratitude to my supervisors, Dr. Paul Messner and Dr. Christina Schäffer for the opportunity to work with you and to be a member of the *Nanoglycobiology* group. Thank you both for all your guidance, advices, discussions, help and support, not only scientific-wise and for the scope of this work, but also for your unconditional help to solve different issues during my stay in Vienna. I really enjoyed my time working on these projects and being a member of your research group.

Secondly, I would also like to convey my thankfulness to Dr. Stephen V. Evans, Dr. Ryan Blackler (Department of Biochemistry and Microbiology, University of Victoria, Victoria, British Columbia, Canada), Dr. Paul Kosma and Gudrun Martinz (Department of Chemistry, Universität für Bodenkultur Wien, Vienna Austria) for your exceptional and valuable effort and contribution to this work. I plentifully learned from this excellent collaboration.

I also thank Dr. Paul Furtmüller (Department of Chemistry, Universität für Bodenkultur Wien, Vienna, Austria) and Dr. Roland Ludwig (Department of Food Science and Technology, Universität für Bodenkultur Wien, Vienna, Austria) for their assistance with the circular dichroism spectrophotometer and the isothermal titration calorimeter equipment.

I would also like to say thank you to all my former and present colleagues from the *Nanoglycobiology* group; Gerald Posch, Julia Anzengruber, Bettina Janesch, Behnam Zolghadr, Andrea Koerdt, Markus Tomek, Susanne Bloch, Cornelia Rath, Matthias Braun, Zoe Megson, Eva Bönisch, Fiona Hager and colleagues from my office; Phillip Spycher, Eva Maria Ladenhauf, Ilaria Pecorari, and Gary Dorken for all your helpfulness, kindness, motivation, useful advice, but above all for the funny, pleasant and amazing moments at

lunch and outside work. To the Bachelor and Master students Mario Minauf and Mathias Polter for their good and diligent work.

To Valentin Friedrich, I would also like to express all my gratitude for your countless help that made my life in Vienna way easier, your extraordinary hospitality, and for all these years of true and faithful friendship.

I also thank the people of the department of Nanobiotechnology for all your friendship, kindness and willingness to always help. Also to Andrea Scheberl and Sonja Zayni for their helpfulness and technical support. To Dieter Jäger for all his valuable help concerning the administrative tasks.

Quiero también agradecer al Dr. Jagoba Iturri, a Alberto Moreno y al Dr. José Toca-Herrera por su disposición de ayudar siempre, por sus buenos consejos y simpáticas ocurrencias.

Agradezco a todos mis amigos latinoamericanos por esas magníficas tardes de "cascara" durante el verano vienés en los diferentes parques de la ciudad, y por su sincera y leal amistad en todos estos años.

Quiero expresar mi más profunda gratitud a mis padres, a mis hermanos Andrés y Alejandra, a mi tía Marilú y a mi primo Gabriel por todo su cariño y apoyo. Muchas gracias por hacerme sentir querido y apoyado a pesar de esa gran distancia que nos separa. En todos estos años que llevo fuera de México nunca dejé de pensar en ustedes. Siempre los tengo en mi mente y mi corazón. Muchas gracias por todo lo bueno que me han brindado y enseñado durante mi existencia. Todas esas enseñanzas me han servido para sobrellevar lo que significa estar fuera de casa.

Quiero igualmente agradecer a todas aquellas personas queridas en México que están siempre al pendiente de mi y me demuestran su cariño en cada una de mis visitas a nuestro hermoso país.

Y finalmente quiero dar las gracias a ellas por brindarme una parte de sus vidas, por mostrarme muchas cosas lindas, y por hacer que mi vida en Viena sea placentera y llevadera.

Abstract

Self-assembling proteinaceous surface (S-) layers are common cell envelope structures found on the outermost surface of many prokaryotes. S-layers are commonly composed of protein or glycoprotein subunits forming an isoporous, 2D crystalline lattice that completely covers the cell surface. Diverse functions have been attributed to S-layers, including action as molecular sieves, determination of cell shape, serving as adhesins involved in cell surface recognition and as antifouling layers. This in combination with the nanometerscale periodicity of S-layer lattices indicates a versiatile application potential of S-layers in nanobiotechnology and biomedicine.

In Gram-positive bacteria, S-layer (glyco)proteins are frequently anchored to the underlying peptidoglycan layer through three tandem repeats of surface layer homology (SLH) domains, which recognize a peptidoglycan-bound pyruvylated secondary cell wall polysaccharide (SCWP). In *Paenibacillus alvei* CCM 2051^T, the SCWP consists of [→3)-β-D-ManpNAc-(1→4)-β-D-GlcpNAc-(1→] disaccharide repeats with the β-ManNAc residue of each repeat modified with 4,6-linked pyruvate ketal. For the binding of SLH domains to SCWP a conserved amino acid motif TRAE within the SLH domains plays a crucial. However, the structural basis for anchoring and how the S-layer proteins attach to the cell wall during cell growth is only poorly understood.

In this work, the crystal structure of the S-layer protein SpaA of the model organism *P. alvei* CCM 2051^T is presented. SpaA possesses within its three N-terminal SLH-domains one TRAE motif and two naturally mutated variants, the TVEE and TRAQ motifs. The interaction of the SLH motifs and SCWP is characterized from a biophysical point of view and at atomic resolution, with co-crystal structures presented of the three consecutive SLH domains of SpaA ($SpaA_{SLH}$) and defined, chemically synthesized SCWP ligands. These data are supported by isothermal titration calorimetry analyses. The $SpaA_{SLH}$ protein

folds in a three-pronged triangular structure that displays binding sites with apparent intramolecular negative cooperativity to bind the SCWP through what could be a mutually exclusive binding mechanism to anchor proteins to the surface of Gram-positive bacteria.

Considering that pyruvylation of SCWP is predicted to be crucial to the binding of SLH domains, the pyruvylation mechanism of the SCWP of *P. alvei* CCM 2051^T was investigated at the biochemical and structural level. The putative pyruvyl transferase CsaB displays high homology at the protein sequence level to different orthologs from diverse members of the *Bacillaceae* family of bacteria, and an overall fold similar to the glycosyltransferase type-B fold found in prokaryotes and eukaryotes. The recombinant protein CsaB showed specificity to recognize the ManNAc moiety as acceptor substrate for the transfer of pyruvate. Furthermore, distinct amino acid residues were identified as potential candidates for prospective mutations to shed light on the reaction mechanism and to modify the substrate selectivity of *P. alvei* CCM 2051^T CsaB.

Together, both aspects of this research can find applications in nanobiotechnology and biomedicine, as pyruvylation of oligosaccharides offers a promising alternative to sialylation to engineer novel glycopeptides with therapeutic applications. Further, disrupting the mechanism of surface protein anchoring may offer an interesting therapeutic strategy against Gram-positive pathogens.

Zusammenfassung

Die Zelloberfläche vieler Prokaryoten ist vollständig mit einer selbst-organisierenden S-Schicht bedeckt. Diese Oberflächenstruktur besteht aus einem isoporösen 2-dimensionalen Kristallgitter, das aus Proteinen oder Glykoproteinen gebildet wird. S-Schichten werden verschiedenste Funktionen zugeschrieben, unter anderm, als Molekularsieb, in der Festlegung und Aufrechterhaltung der Zellform, in Zell-Zell-Interaktionen und als Antifouling-Schichten. Aufgrund ihrer vielseitigen Rollen bieten S-Schichten ein umfangreiches Anwendungspotential in der Nanobiotechnologie und Biomedizin.

In manchen Gram-positiven Bakterien erfolgt die Verankerung der S-Schicht an das darunterliegende Peptidoglykan durch drei Tandem-Wiederholungen der sogenannten Surface Layer Homology (SLH) Domänen, die spezifisch an ein pyruvyliertes sekundäres Zellwand-Polysaccharid (SCWP) binden, welches wiederum mit dem Peptidoglykan verbunden ist. In *Paenibacillus alvei* CCM 2051^T ist das SCWP aus [→3)-β-D-Man*p*NAc-(1→4)-β-D-Glc*p*NAc-(1→] Disaccharid-Wiederholungseinheiten aufgebaut, wobei der β-ManNAc Rest jweils mit einem 4,6-gebundenen Pruvatketal modifiziert ist. In der Interaktion zwischen diesem SCWP und den SLH-Domänen des S-schichtproteins spielt das konservierte Aminosäure-Motiv TRAE eine wesentliche Rolle. Allerdings sind die genauen strukturellen Grundlagen dieser Verankerung oder der Bindung der S-Schichtproteine während des Zellwachstums noch wenig erforscht.

Im Rahmen dieser Arbeit wird die Kristallstruktur des S-Schicht-Proteins SpaA unseres Modellorganismus *Paenibacillus alvei* CCM 2051^T präsentiert. Dieses Protein weist an seinem N-terminus sowohl ein TRAE Motiv als auch zwei natürlich mutierte Varianten, TVEE und TRAQ, auf. Durch die Co-Kristallisation der drei SLH Domänen von SpaA ($SpaA_{SLH}$) mit definierten,

synthetisch hergestellten SCWP Liganden, in Verbindung mit isothermer Titrationskalorimetrie, konnten diese Wechselwirkungen aus biophysikalischer Sicht und mit atomarer Auflösung charakterisiert werden. Die Strukturanalyse von $SpaA_{SLH}$ ließ eine Architektur mit dreifacher Symmetrie erkennen und zeigte Bindungsstellen, die mit dem SCWP offenbar unter intramolekularer negativer Kooperativität inter-agierten. Diese Ergebnisse deuten auf sich gegenseitig ausschließende Bindungs-vorgänge hin, um Proteine an der Oberfläche von Gram-positiven Bakterien zu verankern.

Ein weiterer Aspekt dieser Arbeit war die biochemische und strukturelle Untersuchung des Pyruvylierungs-Mechanismus des SCWP von *P. alvei* CCM 2051^T, da man davon ausgeht, dass die Pyruvylierung des SCWP für seine Ankerfunktion wesentlich ist. Die putative Pyruvyltransferase CsaB verfügt über eine hohe Aminosäure-Sequenzhomologie zu Orthologen in diversen *Bacillaceae*-Spezies und weist außerdem eine Proteinfaltung ähnlich jener der Glykosyltransferase-Typ-B-Faltung aus Prokaryoten und Eukaryoten auf. Rekombinant exprimiertes CsaB zeigte eine spezifische Bindung an ManNAc, dem Akzeptor-Substrat für den Transfer von Pyruvat. Außerdem konnten einige Aminosäuren, die höchstwahrscheinlich an diesem Vorgang beteiligt sind, als potenzielle Kandidaten für zukünftige Mutationstudien zur Aufklärung des Reaktionsmechanismus und der Substratselektivität von *P. alvei* CCM 2051^T CsaB identifiziert werden.

Beide Teilgebiete dieser Arbeit sind für mögliche Anwendungen in Biomedizin und Nanobiotechnologie relevant. Die Pyruvylierung von Oligosacchariden stellt in der Entwicklung neuartiger Glykopeptide zum klinisch-therapeutischen Einsatz eine vielversprechende Alternative zur Sialylierung dar, während ein besseres Verständnis der Verankerung von Oberflächenproteinen interessante Strategien zur Bekämpfung Gram-positiver Krankheitserreger ermöglicht.

A todas las mujeres y hombres de México que han perdido la vida o desaparecido a causa de la violencia

Table of contents

List of abbreviations

ABC transporter	ATP-binding cassette transporter
AFB	American foulbrood
AFM	atomic force microscopy
ATCC	American Type Culture Collection
ATP	adenosine triphosphate
AU	asymmetric unit
Bis-Tris	Bis-(2-hydroxy-ethyl)-amino-tris(hydroxymethyl)-methane
CBB	Coomassie Brilliant Blue G250
CCM	Czech Collection of Microorganisms
CFU	colony-forming unit
CLS	Canadian Light Source
CM	cytoplasmic membrane
C-terminus	carboxy-terminus
CWB	cell wall binding
DAP	diaminopimelic acid
DLS	dynamic light scattering
DMSO	dimethyl sulfoxide
DNA	deoxyribonucleic acid
DNAase	deoxyribonuclease
DTT	dithiothreitol
ECD	electronic circular dichroism
EDTA	ethylenediaminetetraacetic acid
EM	electron microscopy
ERIC	enterobacterial repetitive intergenic consensus
Gal	D-galactose
GDP	guanosine diphosphate
Glc	D-glucose
GlcNAc	*N*-acetyl-D-glucosamine
GroP	glycerol-3-phosphate
GT	glycosyltransferase
HEPES	2-[4-(2-hydroxyethyl)piperazin-1-yl] ethanesulphonic acid

HPLC	high performance liquid chromatography
IPTG	isopropyl-β-D-thiogalactopyranoside
ITC	isothermal titration calorimetry
K_a	association constant
kb	kilobase pair
kDa	kilo Daltons
LB	Luria-Bertani
LP	lipoprotein
LTA	lipoteichoic acid
ManNAc	*N*-acetyl-D-mannosamine
MP	membrane protein
MurNAc	*N*-acetyl-D-muramic acid
NMR	nuclear magnetic resonance
NRPSs	nonribosomal peptide synthases
NTA	nitrilotriacetic acid
N-terminus	amino-terminus
OMP	outer membrane protein
O-OTase	oligosaccharyl transferase
ORF	open reading frame
OD_{600}	optical density at 600 nm
PCR	polymerase chain reaction
PEG	polyethylene glycol
PEG MME	polyethylene glycol monomethyl ether
PEP	phosphoenol pyruvate
PG	peptidoglycan
pI	isolelectric point
PKSs	polyketide synthases
*p*NP	*para*-nitrophenol or 4-nitrophenol
PS	polysaccharide
RNA	ribonucleic acid
RNAase	ribonuclease
RP-HPLC	reverse phase-high performance liquid chromatography
rCsaB	recombinant CsaB
rSpaA	recombinant SpaA
SAD	single-wavelength anomalous diffraction

SAXS	small angle X-ray scattering
SCWP	secondary cell wall polysaccharide
SDS-PAGE	sodium dodecyl sulfate-polyacrylamide gel electrophoresis
S-layer	surface layer
SLH	surface layer homology
slg	S-layer glycosylation
SP	signal peptide
TEM	transmission electron microscopy
TLC	thin-layer chromatography
Tris	tris(hydroxymethyl)aminomethane
UDP	uridine diphosphate
UV	ultraviolet
UV-vis	ultraviolet-visible
X-ray	X radiation
Å	Ångström, 0.1 nm
3D	three-dimensional
15/4 EO/OH	pentaerythritol ethoxylate

One- and three-letter symbols for amino acids:

Alanine=Ala=A	Leucine=Leu=L
Arginine=Arg=R	Lysine=Lys=K
Asparagine=Asn=N	Methionine=Met=M
Aspartic acid=Asp=D	Phenylalanine=Phe=F
Cysteine=Cys=C	Proline=Pro=P
Glutamic acid=Glu=E	Serine=Ser=S
Glutamine=Gln=Q	Threonine=Thr=T
Glycine=Gly=G	Tryptophan=Trp=W
Histidine=His=H	Tyrosine=Tyr=Y
Isoleucine=Ile=I	Valine=Val=V

Mutations are written as: G46A, where glycine 46 is mutated to alanine

1 Introduction

1.1 Structure of the cell wall of Gram-positive bacteria

The bacterial cell envelope contains a wide variety of molecules and serves diverse functions, which are essential for the viability of the cell. It differs significantly between Gram-negative and Gram-positive bacteria. Whereas in Gram-negative bacteria, the cell envelope consists of a cytoplasmic membrane and a thin peptidoglycan (PG) layer that is covered by an outer membrane, the cell envelope of Gram-positive bacteria has no outer membrane and is composed of a thick peptidoglycan layer. The bacterial PG is composed of polysaccharide strands of alternating *N*-acetyl-D-glucosamine (GlcNAc) and *N*-acetyl-D-muramic acid (MurNAc) disaccharide repeating unit and may vary in thickness depending on the bacterial species (Ghuysen and Strominger 1963a, Ghuysen and Strominger 1963b). The D-lactyl moiety of each MurNAc residue is linked to a short peptide chain, via which the polysaccharide strands are, cross-linked, thus creating the three-dimensional meswork that surrounds the bacterial cell and provides the exoskeleton function (Strominger and Ghuysen 1967, Labischinski and Maidhof 1994).

In Gram-positive bacteria, PG is additionally decorated with various compounds, including surface proteins, teichoic and teichuronic acids, lipoteichoic acids, diverse lipoglycans and other anionic or neutral polysaccharides (Navarre and Schneewind 1999, Delcour *et al.* 1999, Weidenmaier and Peschel 2008). Some Gram-positive bacteria synthetize a large polysaccharide capsule or exopolysaccharides (Roberts 1996), whereas others assemble a crystalline layer of surface proteins onto the cell surface (Sleytr 1997). A schematic representation of a Gram-positive cell wall is shown in **Figure 1**.

Teichoic acids, teichuronic acids and lipoteichoic acids are accessory cell wall polysaccharides with a secondary role in cell wall function that have been

classified as “classical” secondary cell wall polymers (SCWPs) (Araki and Ito 1989). Different roles have assigned to SCWPs, as in cell division, in maintenance of cell form in rod-shape bacteria, chelation of divalent cations, in resistance for antibiotics and as binding sites for exoproteins and surface layer proteins (Ward 1981, Navarre and Schneewind 1999, Sára and Sleytr 2000, Schäffer and Messner 2005, Desvaux *et al.* 2006, Brown *et al.* 2013). SCWPs are covalently linked either to the peptidoglycan backbone, such as teichoic acids, teichuronic acids, anionic and neutral polysaccharides or linked to the cytoplasmic membrane via a lipid anchor, like glycolipids and lipoteichoic acid (Ward 1981, Brown *et al.* 2013).

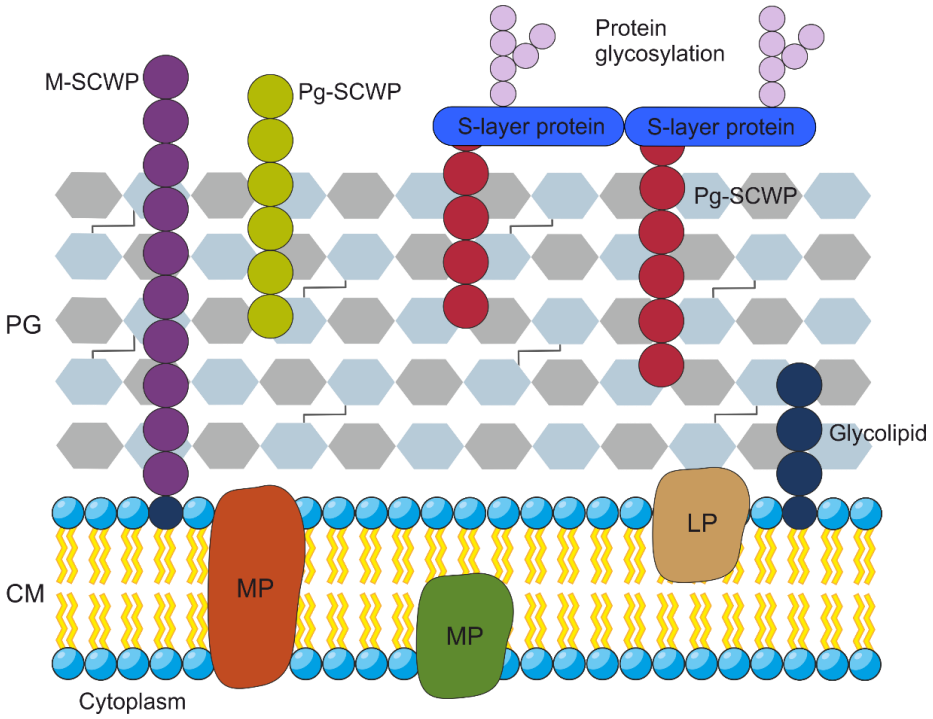


Figure 1. Scheme of the cell envelope of an S-layered Gram-positive bacterium. CM, cytoplasmic membrane; MP, membrane protein; LP, lipoprotein; PG, peptidoglycan; M-SCWP; membrane-linked secondary cell wall polymer; PG-SCWP, peptidoglycan-linked secondary cell wall polymer; S-layer protein, surface layer protein. Capsular or exopoly-saccharides are not shown. Adapted from Weidenmaier and Peschel (2008).

Teichoic acids consist of a highly conserved disaccharide linkage unit and a main polymer chain composed of a phosphodiester-linked polyol repeating unit (Ward 1981, Brown *et al.* 2013). The disaccharide unit is made of a β-D-ManNAc-(1→4)-β-D-GlcNAc-1-phosphate with one or two glycerol-3-phosphate (GroP) units linked to the C4 oxygen of *N*-acetyl-D-mannosamine (ManNAc) (Ward 1981, Kojima *et al.* 1985, Araki and Ito 1989). The phosphodiester-linked polymer chain elongates from the GroP terminus and can consist of polyol (glycerol or ribitol) phosphate, glycosyl polyol phosphate or glycosyl phosphate polyol phosphate repeating units, commonly modified with glucose, GlcNAc or D-alanine residues (Ward 1981, Endl *et al.* 1984, Navarre and Schneewind 1999, Naumova *et al.* 2001). The reducing end of the disaccharide linkage unit is covalently attached to the PG via a phosphodiester bond to the C6 hydroxyl group of MurNAc. Lipoteichoic acids differ from teichoic acids in that they are attached through a glycolipid anchor to the cytoplasmic membrane and are usually composed of GroP repeating units (Ward 1981, Kojima *et al.* 1985, Navarre and Schneewind 1999, Reichmann and Gründling 2011).

Teichuronic acids have a complete different structure, although they are covalently attached to the PG layer; they neither contain a linkage unit nor phosphate groups in their polymer chain. The negative charge in these cell wall polymers arises from the carboxylic groups of glucuronic or mannosamine uronic acid residues (Ward 1981, Araki and Ito 1989, Weidenmaier and Peschel 2008). Teichuronic acids have been described in diverse Gram-positive bacteria, like *Bacillus*, *Geobacillus*, *Streptomyces* and *Micrococcus* species (Sumihiro and Matsushima 1972, Soldo *et al.* 1999, Schäffer *et al.* 1999a, Shashkov *et al.* 2002). It has been shown that in *Bacillus subtilis*, phosphate-limited growth favors the biosynthesis of wall teichuronic acid (Lang *et al.* 1982).

The "non-classical" SCWPs are neutral or anionic polysaccharides (Schäffer and Messner 2005) covalently attached to the PG layer, which cannot be classified as teichoic, lipoteichoic or teichuronic acids (Araki and Ito 1989).

Non-classical SCWPs were initially discovered as by-products during S-layer glycoprotein purification from diverse bacterial sources. S-layers are two-dimensional crystalline proteinaceous arrays on the outermost surface envelope of prokaryotic cells from almost all phylogenetic groups (Sleytr 1976, Sleytr 1978) (see below). Before the performance of detailed structural investigations into the these "non-classical" SCWPs, these were speculated to be a a second glycoconjugate co-purified with the S-layer glycoproteins (Messner *et al.* 1987, Altman *et al.* 1990, Altman *et al.* 1996, Schäffer *et al.* 1999a, Schäffer *et al.* 2000b, Steindl *et al.* 2002).

Later on, it was shown that these compounds were a distinct class of SCWPs and, eventually, SCWP-PG complexes from diverse bacteria were isolated and the linkage of "non-classical" SCWPs to the PG backbone were unambiguously determined (Schäffer and Messner 2005), showing that these polysaccharides are attached to C6 of MurNAc via either a phosphodiester or pyrophosphate group depending on the microorganism (Schäffer *et al.* 1999a, Schäffer *et al.* 2000b, Steindl *et al.* 2002, Steindl *et al.* 2005, Choudhury *et al.* 2006).

Non-classical SCWPs from S-layer carrying *Bacillaceae* were classified in three different groups according to their composition and structure (Schäffer and Messner 2005). These polysaccharides are usually composed of 2-15 linear or branched repeating units, with 2-5 sugars in each repeating unit, with an average molecular mass of 4-22 kDa. Commonly, the polysaccharide chain begins with a GlcNAc residue at the reducing end and finishes with a *manno*- or *gluco*-configured residue. The negative charge of these polymers arises from acidic substituents such as carboxyl, pyruvyl or acetyl groups (**Figure 2**).

In group I, →3)-β-D-Man*p*NAc-(1→4)-β-D-Glc*p*NAc-(1→ disaccharide is the most frequent repeating unit. The entire glycan chain is comprised of several repetitions of this disaccharide. In these SCWPs, ManNAc residues are commonly modified either by the addition of pyruvic acid or by the attachment

of a D-ribofuranose (**Figure 2**) (Altman *et al.* 1990, Altman *et al.* 1996, Ilk *et al.* 1999, Schäffer *et al.* 2000b, Mesnage *et al.* 2000).

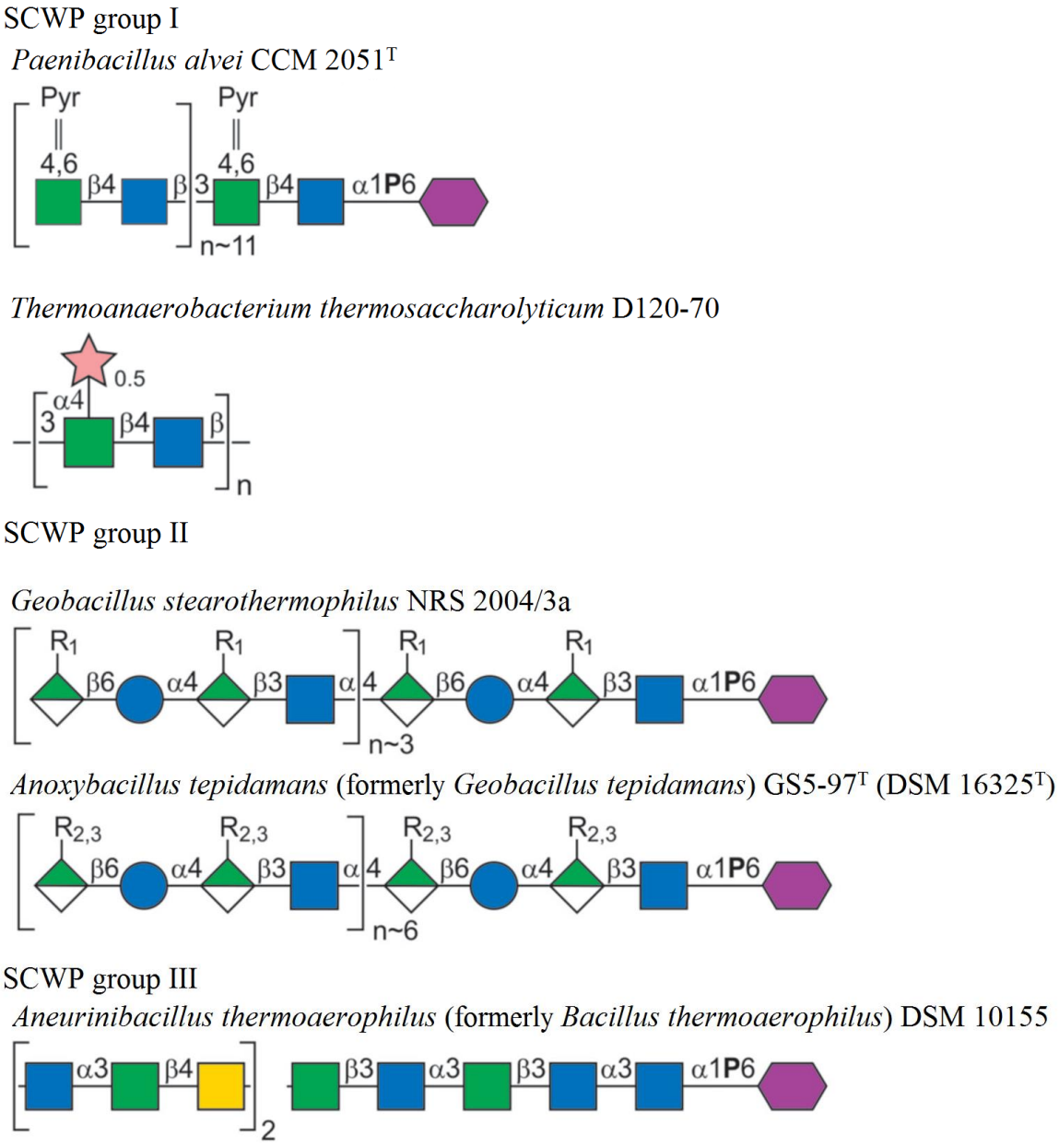


Figure 2. Structures of SCWPs of S-layered Gram-positive bacteria. P, phosphate; Pyr pyruvyl, R1, -COOH; R2, -CONHAc; R3, -CONAc2. Sugar symbols: blue circle, glucose; blue square, *N*acetylglucosamine; pink star, ribose; green square, *N*-acetylmannosamine; green/white diamond, mannosaminuronic acid; yellow square, *N*-acetylgalactosamine; purple hexagon, *N*-acetylmuramic acid. Monosaccharide symbols follow the SNFG (Symbol Nomenclature for Glycans) (Varki *et al.* 2015). Adapted from (Schäffer and Messner 2005) and (Schäffer and Messner 2017).

In group II, the repeating unit is composed of →4)-β-D-Man*p*-2,3-diNAcA-(1→6)-α-D-Glc*p*-(1→4)-β-D-Man*p*-2,3-diNAcA(1→3)-α-D-Glc*p*NAc(1→ tetra-saccharide. The Man-2,3-diNAcA residue can occasionally be modified by the formation of an amide bond, losing its anionic character (**Figure 2**) (Messner *et al.* 1987, Schäffer *et al.* 1999a). In group III, branched repeating units containing neutral sugars and/or amino sugars have been described (**Figure 2**). The biantennary SCWP from the thermophilic bacterium *Aneurinibacillus thermoaerophilus* represents a unique case of group III SCWPs (Steindl *et al.* 2002). Another branched SCWP was found in the bacterium *Geobacillus stearothermophilus* PV72/p2 (Petersen *et al.* 2008). Similar to the SCWPs of group I, the *G. stearothermophilus* SCWP also contains pyruvate modifications on D-ManNAc residues.

It has been found that different strains of *Bacillus anthracis* and *Bacillus cereus* also possess branched pyruvylated SCWPs as integral components of their cell wall (Choudhury *et al.* 2006, Leoff *et al.* 2008, Forsberg *et al.* 2011, Forsberg *et al.* 2012), without further classification of these SCWPs.

The relevance of the "non-classical" SCWPs of some members of the *Bacillaceae* family of bacteria emerged from the extensive research on S-layer (glyco)proteins. "Non-classical" SCWPs have been described as mediators for the attachment of S-layer proteins to the cell surface (Sára *et al.* 1998b, Mesnage *et al.* 2000, Sára 2001, Mader *et al.* 2004, Cava *et al.* 2004, Huber *et al.* 2005, Janesch *et al.* 2013b). The mechanism of interaction between S-layer proteins and SCWPs will be discussed in more detail later on.

1.2 Surface layers (S-layers) of bacteria

The description of S-layers in this section will mainly focus on S-layered Gram-positive bacteria, since the organism used during this research – *Paenibacillus alvei* CCM 2051^T - is a Gram-positive bacterium.

1.2.1 General features, occurrence and structure

Bacterial cell surface (S-) layers are two-dimensional crystalline arrays of (glyco)proteinaceous monomers that shape the outermost cell envelope of diverse prokaryotic organisms (Sleytr 1975, Sleytr 1976), and were first described by Houwink and Le Poole (Houwink and Le Poole 1952, Houwink 1953). They have been identified in Gram-positive and Gram-negative bacteria as well as in archaea – where they constitute an almonst universal feature (Sleytr and Sára 1997, Messner *et al.* 1997, Sára and Sleytr 2000, Sleytr *et al.* 2007, Sleytr *et al.* 2014). S-layers are in most cases composed of single species of S-layer protein or glycoprotein, which self-assemble and fully cover the cell surface, forming the characteristic two-dimensional, regular and porous lattice with oblique (p1, p2), square (p4), or hexagonal (p3, p6) symmetry (Sára and Sleytr 1996, Sleytr *et al.* 1999, Sára and Sleytr 2000, Sleytr *et al.* 2007, Messner *et al.* 2008, Sleytr *et al.* 2014).

The S-layer formation is a dynamic and continuous process of assembly of supramolecular structures during cell growth and division. In fact, S-layer proteins are synthesized in larger amounts than any other type of proteins in the cell, it is estimated that around 500,000 monomers are required to cover a rod-shaped bacterial cell of average size, meaning that around 500 protein monomers per second must be synthetized, transported to the cell surface and incorporated into the nascent S-layer lattice (Sleytr and Glauert 1975, Sleytr 1978, Messner and Sleytr 1991b, Pum *et al.* 1991, Sleytr *et al.* 1999).

Commonly, the identification and characterization of S-layers of different bacteria and archaea have been performed by transmission electron microscopy (TEM) of intact cells or cell wall fragments. Different electron microscopy techniques have been used for this purpose, including thin-sectioning, freeze-etching, freeze-drying, and heavy-metal shadowing or negative-staining. In fact, freeze-etching has been the most suitable electron microscopy technique to identify S-layers on prokaryotes (Thornley *et al.* 1974, Sleytr and Glauert

1975, Sleytr and Messner 1983, Cejka and Baumeister 1987, Sleytr *et al.* 1988, Sleytr *et al.* 1993, Beveridge 1994, Pavkov-Keller *et al.* 2011).

Further structural characterization of S-layer lattices has been done using other microscopical and biophysical methods. The two-dimensional recrystallization of S-layers on solid surfaces has been studied by atomic force microscopy (AFM) (Müller *et al.* 1996, Müller *et al.* 1999, Györvary *et al.* 2003, Ebner *et al.* 2006, Tang *et al.* 2007, Moreno-Flores *et al.* 2008, Chung *et al.* 2010, Lopez *et al.* 2010, Breitwieser *et al.* 2017). In addition, the three-dimensional spatial arrangement of S-layers has been determined by electron crystallography (Baumeister and Engelhardt 1987, Hovmöller *et al.* 1988, Pavkov-Keller *et al.* 2011), and by X-ray and neutron scattering techniques (Weygand *et al.* 1999, Weygand *et al.* 2000, Weygand *et al.* 2002, Horejs *et al.* 2010, Pavkov-Keller *et al.* 2011, Baranova *et al.* 2012).

The unit cell of S-layer lacttices can consists of one, two, three, four or six identical monomer subunits, depending on the type of lattice, with a center to center spacing of 3.5 to 35 nm, approximately. Besides, S-layers lattices have a thickness of 5 to 25 nm, and they display pores of identical size and morphology, ranging from 2 to 8 nm in diameter (Sára and Sleytr 1996, Sleytr and Beveridge 1999, Sleytr *et al.* 1999, Sára and Sleytr 2000, Sleytr *et al.* 2002, Albers and Meyer 2011). It has been described that S-layers have a variety of functions and roles in maintenance of cellular shape, environmental protection, cellular adhesion, molecular sieving and bacterial pathogenicity (Messner *et al.* 1986, Sleytr and Sára 1986, Sára and Sleytr 1987a, Sára and Sleytr 1987b, Pum *et al.* 1991, Sekot *et al.* 2011).

S-layer lattices can be disintegrated into their monomeric subunits by using chaotropic agents (*e.g.*, urea, guanidinium hydrochloride), by varying the concentration of hydrogen ions (H^+) in solution or by using metal-chelating agents (e.g. EDTA), showing that S-layer monomers interact with each other and the underlying cell surface via non-covalent interactions (Messner and

Sleytr 1992, Sára and Sleytr 1996, Sleytr *et al.* 1999, Sára and Sleytr 2000, Sleytr *et al.* 2014). Remarkably, isolated S-layer proteins retain their intrinsic ability to assemble into regular lattices in suspension or on diverse solid supports, upon removal of the disintegrating agent that is required for S-layer (glyco)protein isolation (Sleytr 1997, Sleytr *et al.* 1999, Sára and Sleytr 2000, Sleytr *et al.* 2007, Messner *et al.* 2008, Pum *et al.* 2013, Breitwieser *et al.* 2017). The formation of the self-assembled array depends on the amino acid sequence of the S-layer protein, and, hence, on the folding of the S-layer protein. It has been described that divalent cations, such as Ca^{+2}, stabilize and trigger the process of S-layer self-assembly (Pum and Sleytr 1994, Pum and Sleytr 1995, Korkmaz *et al.* 2011, Baranova *et al.* 2012, Breitwieser *et al.* 2017).

Members of the *Bacillaceae* family are among of the best examples of *in vivo* and *in vitro* research on S-layers. The recrystallization of *Lysinibacillus sphaericus* (formerly *Bacillus sphaericus*) S-layer protein SbpA on hydrophobic and hydrophilic silicon surfaces, on lipid bilayers or on mica supports has been monitored *in situ* by atomic force microscopy, showing that the lattice formation has different kinetic steps, including (i) random adsorption of monomers onto the surface, (ii) condensation and self-assembly of these monomers to form crystalline arrays of folded tetramers, and (iii) growth by new adjacent tetramer formation until the monolayer or bilayer is assembled, which yields an overall square (p4) symmetry of the S-layer lattice (Györvary *et al.* 2003, Chung *et al.* 2010, Lopez *et al.* 2010, Shin *et al.* 2012, Breitwieser *et al.* 2017). Additionally, the self-assembly kinetic process of SbpA in solution under different environmental conditions was studied by dynamic light scattering (DLS) and absorption spectrophotometry (Teixeira *et al.* 2010). The study showed that the self-assembly rate of SbpA is dependent on the protein concentration, ionic strength and temperature.

The amino acid composition of S-layer proteins of organism from different taxonomic phyla are in general similar, although analyses at the genetic level showed that genes from non-related organism share low homology (Sára and

Sleytr 2000, Sleytr *et al.* 2014). Sequence similarity among genes of S-layer proteins is usually found only in closely related species (Boot and Pouwels 1996, Sára and Sleytr 2000). For instance, bacterial S-layer proteins are weakly acidic, moderately hydrophobic and cysteine residues are rarely present (Sára and Sleytr 1996, Sleytr *et al.* 2014). S-layer proteins of Gram-positive bacteria possess an N-terminal secretion signal peptide which is cleaved off after translocation through the plasma membrane.

However, S-layer proteins must possess common structural principles, including a region capable of attaching the protein to the bacterial cell envelope, and a region involved in the two-dimensional lattice assembly. In Gram-negative bacteria, there is evidence that S-layer proteins bind to the outer membrane via the lipopolysaccharide (Walker *et al.* 1994, Ford *et al.* 2007). In Gram-positive bacteria, the S-layer proteins interact via "non classical" SCWPs with the PG layer. Archaea are a special case, since they lack a common cell envelope structure. The S-layer proteins consequently attach directly to the cytoplasmic membrane, playing the role of a pseudo cell wall component or pseudomurein in this type of organism (Sleytr *et al.* 1993, Sára and Sleytr 1996, Sleytr *et al.* 1999, Sára and Sleytr 2000, Pum *et al.* 2013).

Glycosylation is the most common post-translational modification of S-layer proteins in bacteria and archaea, and S-layer glycosylation is among the best characterized models of protein glycosylation in prokaryotes (Sleytr and Thorne 1976, Mescher and Strominger 1976, Sára *et al.* 1989, Sumper *et al.* 1990, Konrad and Eichler 2002, Eichler and Adams 2005, Messner *et al.* 2009, Albers and Meyer 2011). The glycan chains, linear or branched homo- or heterosaccharides, are composed of 20 to 50 repeating units, consisting of neutral hexoses, pentoses, heptoses, 6-deoxyhexoses and amino sugars (Lechner and Wieland 1989, Messner and Sleytr 1991a, Messner 1997, Sára and Sleytr 2000).

The glycan chains of S-layer proteins of bacteria and archaea differ significantly from those of eukaryotes with regard to composition and structure, and are, architecture-wise similar to the lipopolysaccharide of Gram-negative bacteria (Sleytr *et al.* 2002, Schäffer and Messner 2004, Messner *et al.* 2009). Commonly, S-layer proteins are either *N*- or *O*-glycosylated, and in rare cases both types of modification can be found on the same S-layer protein (Sumper *et al.* 1990, Albers and Meyer 2011). However, *N*-glycosylation seems to be more common in archaea than in bacteria, and bacterial S-layer proteins are typically *O*-glycosylated (if glycosylated at all). S-layer protein *O*-glycosylation in bacteria varies between 2% to 10%, the glycan moieties are covalently linked to selected serine, threonine and tyrosine residues, and frequently S-layer proteins are polyglycosylated (Messner *et al.* 2009, Messner *et al.* 2013).

A new type of post-translational modification of S-layer proteins has been lately found in archaea. The S-layer glycoprotein of *Haloferax volcanii* is processed and covalently linked to a membrane-embedded lipid (Szabo and Pohlschroder 2012, Eichler and Maupin-Furlow 2013). A group of enzymes named archaeosortases, homologues of bacterial exosortases, is predicted to participate in anchoring of proteins to membrane embedded lipids (Haft *et al.* 2012). One example is the enzyme ArtA, which is predicted to cleave proteins bearing a C-terminal membrane-spanning domain. The *H. volcanii* S-layer glycoprotein is transferred to a lipid moiety on the cell surface following cleavage of the C-terminal membrane-spanning domain (Eichler and Maupin-Furlow 2013). It has been demonstrated that *H. volcanii* possesses two classes of S-layer glycoproteins, a lipid-modified, chelating agent-solubilized class and a second, detergent-solubilized class, which preserves the hydrophobic C-terminal transmembrane domain (Kandiba *et al.* 2013).

Structural information about S-layer proteins is limited to truncated, nanobody-assist crystallized or assembled negatively-stained proteins, regardless of their biological importance and abundance (Pavkov *et al.* 2003,

Norville *et al.* 2007, Pavkov *et al.* 2008, Fagan *et al.* 2009, Kern *et al.* 2011, Baranova *et al.* 2012). Secondary structure analyses of S-layer proteins by far-UV circular dichroism (CD) spectroscopy revealed that α-helical structures are often located at the N-terminal part, while β-sheet content is mainly present or higher at the C-terminal region (Rünzler *et al.* 2004, Kern *et al.* 2011). These findings are in accordance with structure predictions based on protein sequences (Sára and Sleytr 2000). On the contrary, determination of the tertiary structure of S-layer proteins has been hindered by their high molecular masses, being unsuitable for nuclear magnetic resonance (NMR) analyses, and mainly by their intrinsic ability to assemble in two-dimensional arrays rather than to form three-dimensional crystals in solution. Thus, the structure of only four bacterial S-layer proteins using X-ray crystallography has been determined, three S-layer proteins were crystallized as truncated derivatives and one full-size S-layer protein was crystallized by the aid of nanobodies as crystallization chaperones (Pavkov *et al.* 2008, Fagan *et al.* 2009, Kern *et al.* 2011, Baranova *et al.* 2012).

Still, additional strategies have been applied to further elucidate the tertiary structure of S-layer proteins and to gain insight into the topology of the assembled S-layers. For example, the surface accessibility and location of single residues in the monomeric and assembled forms of the S-layer protein SbsB from *Geobacillus stearothermophilus* PV72/p2 were screened by single cysteine mutations followed by targeted chemical modification and by insertion mutagenesis to generate assembly-compromised mutants (Howorka *et al.* 2000, Kinns and Howorka 2008, Kinns *et al.* 2010). Moreover, an alternative approach to provide additional structural information of full-size S-layer proteins is the small-angle X-ray scattering (SAXS) technique. SAXS has been employed to estimate the particle dimensions of the S-layer proteins SbsC of *G. stearothermophilus* and of the low-molecular-weight SlpA of *Clostridium difficile* in solution, as well as for imaging of surface layers on intact bacteria (Pavkov *et al.* 2008, Fagan *et al.* 2009, Sekot *et al.* 2013).

Finally, the three-dimensional structure of the S-layer protein SbsB of *G. stearothermophilus* was predicted by using molecular dynamics simulations, and the obtained structure was validated by utilizing the mean force method (Horejs *et al.* 2008).

1.2.2 Attachment of S-layer proteins to the cell envelope

In the course of evolution, Gram-positive bacteria have developed different mechanisms to display cytoplasmically synthesized proteins on their surface. Thus, surface displayed proteins can be classified either as membrane-associated proteins or cell wall-associated proteins. The strategies to display cell wall-associated proteins include covalent anchoring of proteins to the peptidoglycan via an LPXTG motif or non-covalent anchoring via cell wall-binding domains (Navarre and Schneewind 1999, Desvaux *et al.* 2006). Five types of non-covalent binding have been described: The choline-binding domain, also named cell wall-binding (CWB) domain Type 1, the CWBD Type 2, the GW modules, containing tandem repeats of the dipeptide Gly (G)-Trp (W), the lysin motif domains and the surface layer homology (SLH) domains (Desvaux *et al.* 2006, Scott and Barnett 2006).

S-layer proteins attach to the bacterial cell surface via non-covalent interactions between a cell wall-binding domain and bacterial cell wall components (**Figure 1**). The binding mechanism has been gradually elucidated on a molecular basis, with different types of mechanisms having been identified. In one mechanism, the SLH domain recognizes specific epitopes on the PG sacculus. SLH domains were first described by Lupas and co-workers (Lupas *et al.* 1994). Usually, they consist of one to three consecutive domains, where each domain is composed of 50 to 55 amino acids among which 10 to 15 are conserved. SLH domains are located either at the N- or C-terminus of the polypeptide chain (**Figure 3**) and are predicted to be rich in α-helical structures (Lupas *et al.* 1994, Olabarria *et al.* 1996, Engelhardt and Peters 1998).

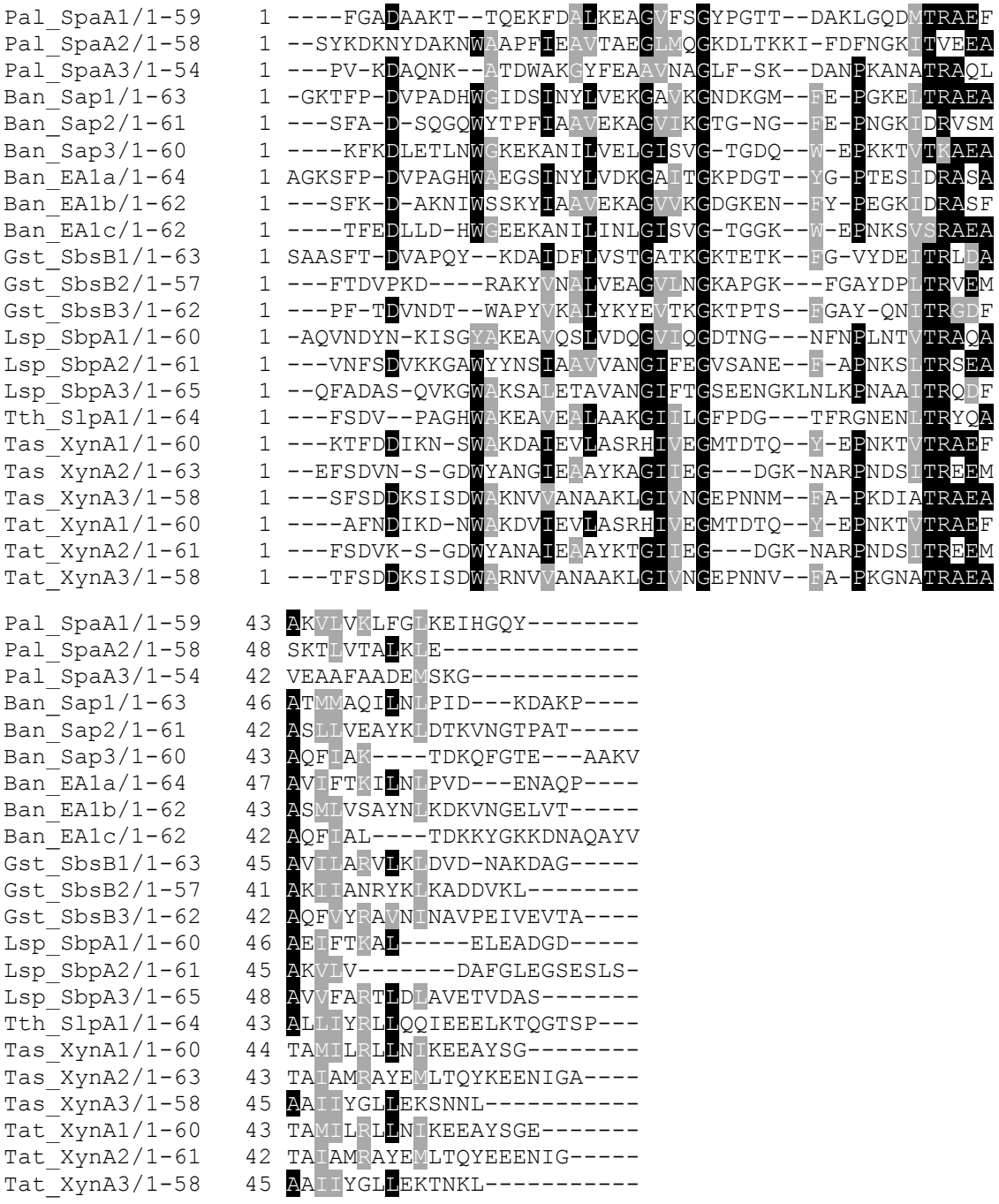


Figure 3. Sequence alignment of SLH domains (Pfam entry: Pf00395) arising from S-layer proteins and glycoside hydrolases. The SLH domain repeats are shown in six S-layer proteins and two glycoside hydrolases from the xylanase family GH-10. The abbreviations of proteins are: Pal, *Paenibacillus alvei*; Ban, *Bacillus anthracis*; Gst, *Geobacillus stearothermophilus*; Lsp, *Lysinibacillus sphaericus*; Tth, *Thermusr thermophilus*; Tas, *Thermoanaerobacterium saccharolyticum*; Tat, *Thermoanaerobacter thermosulfurigenes*. The residues conserved at least 50% level are shaded in black by using the BOXSHADE program.

Besides S-layer proteins, SLH domains are found in other cell surface displayed proteins, such as extracellular enzymes or proteins involved in degradation of plant cell wall polysaccharides (**Figure 3**) (Engelhardt and Peters 1998, Bayer *et al.* 2004, Zona and Janeček 2005).

Further investigations with the XynA protein, a xylanase from *Thermoanaerobacterium thermosulfurigenes* with three C-terminal SLH domains (Brechtel *et al.* 1999), and the S-layer protein SlpA from *Thermus thermophilus* HB8, which has only one SLH domain, reinforced the hypothesis that SLH domains are necessary to attach proteins to the cell surface. A truncated version of XynA without SLH domains or with only one intact domain could not bind to PG sacculi, indicating that three domains may be essential for cell wall anchoring (Brechtel and Bahl 1999). In *T. thermophilus,* a defective SlpA protein lacking the SLH domain was incapable of binding to the cell surface *in vivo* and to form the natural S-layer lattice with hexagonal symmetry (Olabarria *et al.* 1996).

Subsequently, it was demonstrated that the SLH domains interact with a cell wall component that could be extracted from a PG preparation with hydrofluoric acid. This treatment removes SCWPs covalently linked to the PG without hydrolyzing the SCWP chain, albeit with affecting acid-labile groups. Studies with the oxygen-induced strain variant *G. stearothermophilus* PV72/p2 showed that the N-terminal region of the S-layer protein SbsB, which contains three typical SLH domains, recognizes the cognate SCWP of the bacterium (Ries *et al.* 1997). Later on, it was demonstrated that SbsB had within the N-terminus an alternative region able to interact with the PG deprived of SCWP (Ries *et al.* 1997, Sára *et al.* 1998b).

Further evidence highlighted the relevance of SCWPs for cell wall binding of SLH domain proteins in *Bacillus anthracis* for its two S-layer proteins SapA and EA1 (Chauvaux *et al.* 1999, Mesnage *et al.* 1999), in *Clostridium thermocellum* for the S-layer protein SlpA (Lemaire *et al.* 1998, Chauvaux *et al.* 1999) and for the extracellular enzyme XynA in *T. thermosulfurigenes*

(Brechtel and Bahl 1999). A step forward in understanding at a molecular basis the interaction between SLH domains and SCWPs was established when the involvement of pyruvate groups in S-layer protein binding was discovered (Mesnage *et al.* 2000). Mesnage and co-workers demonstrated that the SCWP of *B. anthracis* is modified by the action of the *csaB* gene product. *csaB* is part of the two-gene operon *csaAB*, and *csaB* homologues were also found in many other bacteria that synthesize SLH domain-carrying proteins (Mesnage *et al.* 2000). It was shown that in *B. anthracis*, CsaB is involved in the cell wall metabolism, specifically in the pyruvylation of its SCWP and, that CsaB function is required for cell surface localization of EA1 and Sap proteins, since mutants lacking *csaB* were deprived of pyruvylated SCWP (Mesnage *et al.* 2000).

Moreover, mutants of *T. thermophilus* defective in the *csaB* gene were unable to attach the S-layer protein SlpA to the cell surface during cell growth (Cava *et al.* 2004). It is now well known that pyruvylated SCWPs serve as attachment sites for many S-layer proteins of S-layered bacteria. For example, the $SbsB_{SLH}$ protein, a truncated variant of the S-layer protein SbsB, has specific affinity for the pyruvylated SCWP of *G. stearothermophilus* PV72/p2 (Mader *et al.* 2004). Pyruvylated SCWPs were also found on the PG-containing sacculi of *Lysinibacillus sphaericus* CCM 2177 (Ilk *et al.* 1999) and *P. alvei* CCM 2051^T (Schäffer *et al.* 2000b), which is the focus of the present study. Both, the *L. sphaericus* CCM 2177 S--layer protein SbpA and the *P. alvei* CCM 2051^T S-layer protein SpaA, possess SLH domains and recognize pyruvylated SCWPs as anchoring molecule for cell wall binding (Huber *et al.* 2005, Janesch *et al.* 2013b). However, it was demonstrated that *P. alvei* SpaA could also bind to PG deprived of SCWP (Janesch *et al.* 2013b), probably with a similar mechanism as in *G. stearothermophilus* PV72/p2, the S-layer protein of which also binds to PG (Sára *et al.* 1998b).

All these findings confirm that pyruvylated SCWPs and SLH domains represent a well-conserved mechanism to non-covalently attach proteins to the

cell wall in prokaryotes. Although SLH domains share limited sequence identity, 10-12 amino acid residues are usually highly conserved. Particularly a group of four residues, designated the TRAE motif for its consensus sequence, has been described to be crucial for binding of SLH domain proteins to the cell wall (**Figure 3**). Site-directed mutagenesis of the positively charged arginine (R) and the negatively charged glutamic acid (E) residues showed that a functional TRAE motif is needed for the functionality of an SLH domain (May *et al.* 2006, Janesch *et al.* 2013b).

The crystal structure of the three consecutive SLH domains of the S-layer protein Sap from *B. anthracis* (**Figure 4A**) revealed that the SLH domains adopt a three-pronged spindle fold (**Figure 4B**) (Kern *et al.* 2011). Each SLH domain contributes with one α-helix at the center, whereas a second α-helix (rotated by 90°) and the connecting loop forms a prong, eventually adopting a three-pronged spindle (**Figure 4B**). The three-pronged spindle fold of the SLH domain generates three almost identical grooves, where the conserved cationic and anionic amino acid residues are harbored (**Figure 4B**) (Kern *et al.* 2011). Therefore, the three SLH domains provide residues to the surface structure of each groove, which may create a functional binding pocket to interact with the SCWP. Modelling experiments confirmed that the predicted grooves may bind *B. anthracis* SCWP (Kern *et al.* 2011).

It needs to be mentioned here that not all bacterial S-layer proteins possess SLH domains. Thus, another cell wall binding mechanism has been hypothesized for these S-layer proteins that do not possess SLH domains (**Figure 4C**). This alternative mechanism utilizes basic residues that are frequently present at the N- or C-terminus of the S-layer protein and interact with negatively charged SCWPs – or even neutral ones - (Schäffer and Messner 2005) via electrostatic interactions and hydrogen bonds. For instance, in the Gram-positive bacterium *Aneurinibacillus thermoaerophilus* it was found that the cell wall anchoring ligand is a neutral biantennary SCWP (Steindl *et al.* 2002).

The N-terminal region of the S-layer proteins SbsA and SbsC of *G. stearothermophilus* PV72/p6 and ATCC 12980, respectively, possess highly conserved arginine, lysine and tyrosine residues and interact with negatively charged mannuronic acid-containing SCWPs (Egelseer *et al.* 1998, Jarosch *et al.* 2000). Truncated variants of the SbsC protein confirmed that the N-terminal region is involved in cell wall anchoring and folds independently of the self-assembling C-terminus (Jarosch *et al.* 2001, Ferner-Ortner *et al.* 2007). The crystal structure of the variant $SbsC_{31\text{-}443}$, revealed a high content of α-helices in the N-terminal region and a homogenous distribution of positive charges along the predicted cell wall binding domain (**Figure 4D**) (Pavkov *et al.* 2008).

In *Lactobacillus* S-layer proteins, so far no SLH domains have been identified. However, a positively charge region within the S-layer protein interacts with a negatively charged SCWP to mediate cell wall binding. For instance, the C-terminal region of the S-layer protein from *Lactobacillus acidophilus* ATCC4356 and *Lactobacillus crispatus* JCM 5810, S_A and CbsA, respectively, is positively charged and mediates binding to the cell wall (Smit *et al.* 2001, Antikainen *et al.* 2002). The S-layer proteins of *L. acidophilus* and *L. crispatus* bind to SCWPs, whereby S_A and CbsA recognize teichoic and lipoteichoic acids as cell wall receptors (Smit and Pouwels 2002, Antikainen *et al.* 2002).

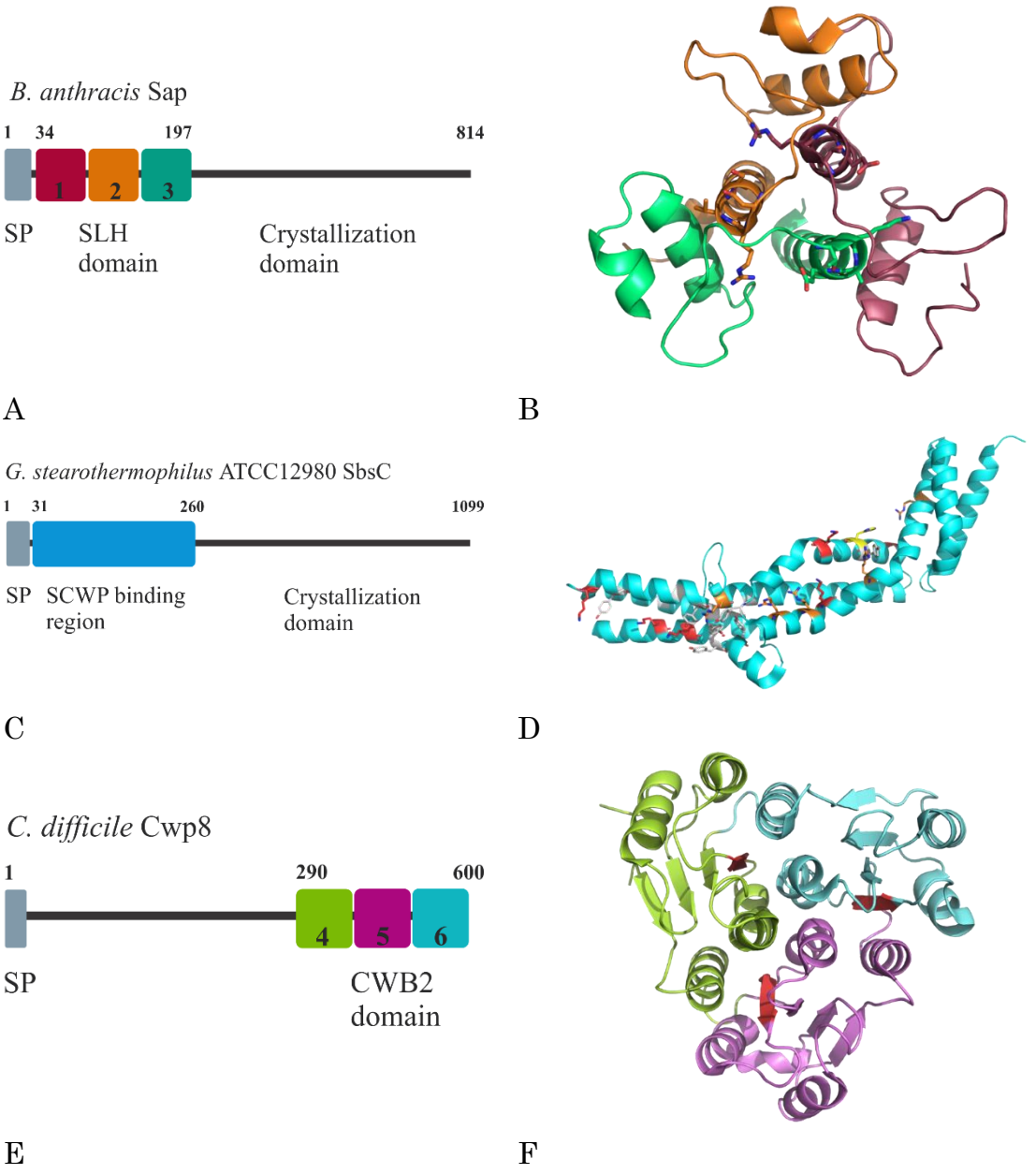


Figure 4. Different protein modules to attach proteins to the cell wall of Gram-positive bacteria. Structurally derived domains are colored consistently throughout the figure. (A) Schematic depiction of the full-length S-layer protein Sap. SLH domains (1-3) are showed in different colors and the signal peptide (SP) in gray. (B) Crystal structure of $Sap_{(34\text{-}197)}$ shown from the N-terminus (Kern *et al.* 2011). Residues of the TRAE motifs are shown as sticks. (C) Full-length representation of the S-layer protein SbsC. The SCWP binding region is shown in blue, while the SP is colored in gray. (D) Cartoon representation of the $SbsC_{(31\text{-}260)}$ protein consisting of three α-helical bundles (Pavkov *et al.* 2008). Putative residues involved in SCWP binding are shown as sticks (Lys, red; Arg, orange; His, yellow; Tyr, gray). (E) Full-length representation of the *C. difficile* Cwp8 protein. The CWB2 (cell wall binding type 2) motifs (4-6) are depicted in different colors, while the SP is in gray. (F) Cartoon representation of the CWB2 domains of Cwp8 shown from the N-terminus (Usenik *et al.* 2017).

A third cell wall anchoring mechanism of S-layer proteins has been found within a family of 29 cell wall proteins (CWPs) in *Clostridium difficile*, including its S-layer protein SlpA, the adhesin Cwp66 and protein Cwp2. All contain three tandem CWB2 motifs at the N-terminus, the C-terminus or in the middle of the polypeptide chain (Fagan and Fairweather 2011, Willing *et al.* 2015). This family of proteins all attach to the cell wall via non-covalent interactions. The cell wall of *C. difficile* is decorated with three different anionic SCWPs: PSI, PSII and PSIII (Ganeshapillai *et al.* 2008, Reid *et al.* 2012). PSI and PSII are PG-bound polymers and contain penta- and hexaglycosyl phosphate repeating units, respectively (Ganeshapillai *et al.* 2008). PSIII is a lipid-bound glycosyl phosphate polymer similar to lipoteichoic acids (Reid *et al.* 2012). Recently, it was shown that the PSII polymer serves as a ligand for binding of *C. difficile* CWB2 domain-carrying proteins and that the three tandem CWB2 motifs are required for cell wall attachment (Willing *et al.* 2015). The CWB2 motif is around 100 amino acid residues long and is present in one to three tandem repeats (**Figure 4E**).

Although sequence conservation within the CWB2 motifs in the CWPs of *C. difficile* is relatively low, a group of four residues define the PILL region, because of its consensus sequence, which is highly conserved among this family of proteins. Mutation of the PILL residues to four alanine residues resulted in the release of the Cwp66 protein into the culture supernatant, thus, confirming the importance of this region within the CWB2 motifs for protein anchoring (Willing *et al.* 2015). Recently, the crystal structures of the Cwp8 and Cwp6 proteins from *C. difficile* revealed that the CWB2 domain adopts an arrangement of a triangular trimer (**Figure 4F**) (Usenik *et al.* 2017). Each motif consists of a central four-stranded parallel β-sheet flanked by two pairs of α-helices on either side of the sheet (**Figure 4F**). The three motifs come together in the center of the trimer to form a three-helical bundle core (Usenik *et al.* 2017). Interestingly, the triangular trimer fold of CWB2 is similar to the SLH domain three-pronged spindle fold of S-layer proteins of other bacteria,

suggesting a common evolutionary mechanism to bind surface proteins on the cell wall of Gram-positive bacteria (**Figure 4B, F**).

Thus, the interaction between S-layer proteins of Gram-positive bacteria and SCWPs seems to be highly specific and similar to lectin-carbohydrate interactions (Sára 2001), where electrostatic interactions play a significant role.

1.3 Paenibacilli and their S-layer proteins

1.3.1 Occurrence and general features of Paenibacillus *S-layer proteins*

Paenibacilli are mesophilic, Gram-positive, rod-shaped, spore-forming microorganisms with implications in agriculture, apiculture, industry and medicine. S-layer proteins have been detected on some *Paenibacillus* species. Among these is *Paenibacillus larvae*; it is the causative agent of American foulbrood (AFB) disease and a primary pathogen of honey bees (Genersch *et al.* 2006, Genersch 2010). *Paenibacillus larvae* only affects the larval stage of honeybees and its spores are the only infectious form of this microorganism (Genersch *et al.* 2006). Larvae are infected upon oral uptake of spore-contaminated larval food. Spores passes through the alimentary canal and germinate in the midgut after ingestion, then the vegetative bacteria colonize the midgut and proliferate there, and finally they begin to penetrate into the midgut epithelium until the bacteria invade the haemocoel and kill the host. *Paenibacillus larvae* continues degrading the larval corpse to a ropy phase to escape while still in the vegetative form to ensure sporulation and further transmission in the bee colony (Yue *et al.* 2008, Fünfhaus and Genersch 2012).

Within the species *P. larvae*, four different genotypes, *P. larvae* ERIC I-IV have been described, based on genomic fingerprinting analyses with repetitive element-PCR using enterobacterial repetitive intergenic consensus (ERIC) primers (Genersch *et al.* 2006). Interestingly, it was found that these four subspecies differ in phenotype and in virulence, with being ERIC I less virulent

and ERIC II-IV being highly virulent on larval level. The ERIC I and ERIC II genotypes are the most frequent isolates from AFB infected colonies (Genersch *et al.* 2006, Fünfhaus *et al.* 2009). Therefore ERIC I and ERIC II are the most relevant and investigated genotypes within the species *P. larvae*. Genomic analyses of *P. larvae* ERIC I and ERIC II showed that this bacterium is able to produce different peptide antibiotics and virulence factors. For instance, *P. larvae* possesses gene clusters that encode nonribosomal peptide synthases (NRPSs) and polyketide synthases (PKSs), which participate in the biosynthesis of nonribosomal peptides and polyketides (Fünfhaus *et al.* 2009, Müller *et al.* 2015). Such biosynthesis products are the nonribosomal peptide paenilamicin and the tripeptide sevadicin. *In vitro* assays demonstrated that paenilamicin and sevacidin have antibacterial and antifungal activities. It is conceivable that *P. larvae* utilizes these secondary metabolites to suppress potential competitors present in the larval gut microbiome and in the brood cell (Garcia-Gonzalez *et al.* 2014b, Müller *et al.* 2014, Garcia-Gonzalez *et al.* 2014a).

Proteomic analysis of *P. larvae* evidenced a putative S-layer protein in this bacterium. Fünfhaus and Genersch identified an S-layer protein precursor homologous to the S-layers proteins from *Bacillaceae* in the proteome of *P. larvae* ERIC II, which was not present in the proteome of *P. larvae* ERIC I. The putative S-layer protein SplA is relatively acidic with an isoelectric point is 4.97, and has a theoretical molecular mass of 100 kDa (Fünfhaus and Genersch 2012). Further sequence analyses on genome level showed that both genotypes, ERIC I and ERIC II have a gene to encode a putative S-layer protein. In *P. larvae* ERIC I, the presence of an additional adenine nucleobase at position 894 produces a reading frame shift causing an interruption of *splA* translation due to a stop codon TAA. This insertion is not present in *P. larvae* ERIC II, which, hence, holds a functional gene for the expression of a putative S-layer protein (Poppinga *et al.* 2012).

Moreover, bioinformatic analyses of the putative S-layer proteins showed that SplA has two SLH domains at the N-terminal region, belonging to the SLH

domain superfamily. These SLH domains were predicted to line up from residues 117-164 (SLH_1) and 188-220 (SLH_2) and to be homologous to SLH domains of other *Bacillaceae* (Poppinga *et al.* 2012). To confirm the S-layer protein status of SplA protein, *in vitro* self-assembly studies with recombinant SplA (rSlpA) were performed. Purified rSplA was dialyzed against a 10 mM $CaCl_2$ solution to promote self-assembly of the protein. Self-assembly products were negatively-stained and visualized by a transmission electron microscope (TEM). TEM micrographs confirmed the formation of monolayered, cylindrical self-assembly products with 2 µm length and 180 µm width. Thus, SplA is a true S-layer protein and forms two- square dimensional crystalline lattices displaying lattice parameters of 10.0 nm x 15.4 nm and $\gamma = 90°$ (Poppinga *et al.* 2012).

Paenibacillus larvae mutants deficient in SplA were generated to investigate whether SplA is involved in bacterial adhesion to the midgut of larval cells. The *splA* gene in *P. larvae* ERIC II was disrupted by insertion of a *splA*-specific targeted intron, resulting in the loss of expression of SplA. The *P. larvae* ERIC II Δ*splA* strain showed differences in colony and cell morphology when compared to the wild-type *P. larvae* ERIC II strain. In fact, the cell morphology of *P. larvae* ERIC II Δ*splA* resembles more the natural S-layer-deficient *P. larvae* ERIC I genotype (Poppinga *et al.* 2012). Moreover, cell adhesion assays indicated that *P. larvae* ERIC II Δ*splA* has a lower cell adhesion capacity in comparison to the wild-type strain *P. larvae* ERIC II, but its adhesion capacity is similar to that of the of *P. larvae* ERIC I strain. However, *P. larvae* ERIC II Δ*splA* was still able to kill larvae, implicating that SplA is not an essential factor for the pathogenicity of *P. larvae* (Poppinga *et al.* 2012). The studies of Poppinga and co-workers demonstrated that the S-layer protein SplA of *P. larvae* is a cell morphology determining factor and an important virulence mediator for cell adhesion in the early stage of infection, and is a significant indicator to elucidate the molecular pathogenesis of *P. larvae*.

1.3.2 The model organism Paenibacillus alvei *CCM 2051^T*

A second example of paenibacilli carrying an S-layer protein is the species *Paenibacillus alvei*, which used to belong to the genera *Bacillus* but was reclassified in 1994 into a distinct genus named *Paenibacillus* (Ash *et al.* 1993). *Paenibacllus alvei* is a Gram-positive, saprophytic, aerobic bacterium, which has been described as a second invader during the European foulbrood honeybee disease caused by the bacterium *Melissococcus plutonius* (Forsgren 2010). *Paenibacillus alvei* does not grow in the gut of healthy larvae, but its spores germinate in dead bees. Although *P. alvei* is not known to cause human disease, it has been rarely described as a causative agent of human infections (Reboli *et al.* 1989).

Paenibacillus alvei, like other members of the *Bacillaceae* family, is a spore forming bacterium and is able to excrete indole into the environment, as a byproduct of the tryptophan metabolism (Hoch and Demoss 1965). Sporulation in bacteria is a defense mechanism to adverse environmental conditions and nutritional limitations, generating a dormant and reduced form of the bacterial cell, known as endospore. Endospores are highly resistant to diverse chemicals, desiccation and high temperatures. Studies performed by Kim and co-workers showed the role of indole as an environmental signal in *P. alvei*, demonstrating that indole and indole derivatives influence the spore maturation in this bacterium. They showed that *P. alvei* cells exposed to indole and 3-indolylacetonitrile produced endospores with an irregular coat (Kim *et al.* 2011). Moreover, indole and 3-indolylacetonitrile had an adverse effect on the heat-resistant colony-forming unit (CFU) of *P. alvei*, suggesting that the low heat resistant CFU arises as a consequence of defects on the spore structure. Therefore, indole and its derivatives may act as spore inhibitors during spore germination, affecting the resistance of *P. alvei* to environmental stress (Kim *et al.* 2011).

Paenibacillus alvei CCM 2051^T, our model organism, is a flagellated, motile bacterium that can strenuously swarm on semi-solid surfaces (**Figure 5**)

(Cohen *et al.* 2000, Djukic *et al.* 2012, Janesch *et al.* 2013a). Swarming is a surface translocation mechanism produced by the continuous movement of a flagellum and can also be significantly influenced by the growth conditions. For instance, the collective migration of *P. alvei* clusters on semi-solid surfaces arises from swarming (Cohen *et al.* 2000). Also, the *P. alvei* colonial development has showed different branching patterns depending on the agar and nutrient concentrations or on the presence of antibiotics (Cohen *et al.* 2000).

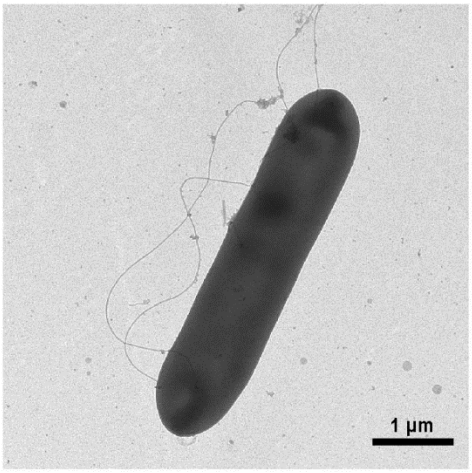


Figure 5. Electron micrograph of a *P. alvei* CCM 2051^T wild-type cell. The cell is ~1 µm wide and ~4 µm long. The flagella are located at the polar regions of the cell. Adapted from Janesch *et al.* (2013a).

Recently, it was demonstrated that *P. alvei* CCM 2051^T has polar flagella embedded into its surface (S-) layer lattice and that the biofilm life-style of the bacterium is crucially influenced by the flagellin protein Hag (Janesch *et al.* 2013a). Flagellin Hag, with an apparent molecular mass of ~30 kDa, self-assembles into cylindrical filaments to form the *P. alvei* CCM 2051^T flagella (Janesch *et al.* 2016). The Hag protein is *O*-glycosylated at positions Thr^{192}, either Ser^{140} or Thr^{141}, and a third glycosylation site was also proposed, although the exact position could not be determined (Janesch *et al.* 2016). Janesch and co-workers showed that the flagellin glycosylation system is independent of the S-layer glycosylation machinery; it is composed of genes that encode the Hag flagellin protein, flagellum biosynthesis enzymes, glycan biosynthesis precursors, two putative glycosyltranferases (PAV2c_01630 and PAV2c_01640) and proteins of unknown function. Furthermore, *P. alvei* CCM

2051^T mutants lacking either of the putative glycosyltranferases were unable to produce flagella or to swarm, confirming the involvement of these putative enzymes in flagellin glycosylation and the pivotal role of glycosylation for flagellum production *in vivo* (Janesch *et al.* 2016).

Paenibacillus alvei CCM 2051^T cells are covered by a proteinaceous S-layer with oblique lattice symmetry composed of a single glycosylated S-layer protein subunit (Sára *et al.* 1990, Altman *et al.* 1991, Messner *et al.* 1993). The gene *spaA* codes for a protein of 983 amino acids, with a 24-amino acids signal peptide. Mature SpaA is relatively acidic with an isoelectric point of 5.83 and a theoretical molecular mass of 105.95 kDa (Zarschler *et al.* 2010a). SpaA has at the N-terminus a domain belonging to the SLH domain superfamily, composed by three consecutive SLH domain repeats, which is implicated in anchoring SpaA to the bacterial cell wall, and a C-terminal self-assembling domain involved in the formation of the two-dimensional crystalline S-layer lattice. The SLH domains align from residues 25-65 (SLH_1), residues 82-129 (SLH_2) and residues 140-181 (SLH_3). Amino acid sequence analyses of SpaA revealed that the *P. alvei* CCM 2051^T S-layer protein is related to S-layer proteins of different *Bacillaceae* and more distant-related to surface proteins of different *Clostridiaceae* (Zarschler *et al.* 2010a).

The S-layer protein SpaA is *O*-glycosylated at least at two glycosylation sites by the addition of a glycan consisting of 23 branched trisaccharide repeating units, with the structure [→3)-β-D-Gal*p*-(1[α-D-Glc*p*-(1→6)]→4)-β-D-Man*p*NAc-(1→], linked via an adaptor with the structure [→3)-[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 the hydroxyl group of tyrosine 47 and tyrosine 155 (**Figure 6**) (Altman *et al.* 1991, Messner *et al.* 1995). SDS-PAGE analyses of *P. alvei* CCM 2051^T wild-type cells showed that the S-layer protein is separated into three main bands with apparent molecular masses of 240, 160 and 105 kDa, respectively.

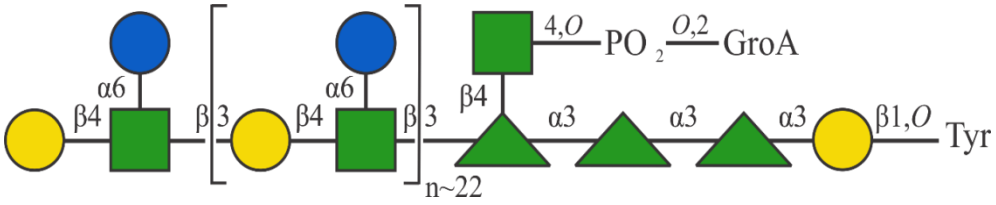


Figure 6. Schematic representation of the S-layer glycoprotein glycan of *P. alvei* CCM 20551^T. Sugar symbols: blue circle, glucose; yellow circle, galactose; green square, *N*-acetylmannosamine; green triangle, L-rhamnose. Monosaccharide symbols follow the SNFG (Symbol Nomenclature for Glycans) (Varki *et al.* 2015). Adapted from Zarschler *et al.* (2010b).

The bands with 240 and 160 kDa react positively with the periodic acid Schiff reagent, indicating, according to our interpretation, di- and monoglycosylated S-layer proteins, respectively (Zarschler *et al.* 2009).

The glycosylation of the S-layer SpaA as well as the underlying molecular mechanisms for S-layer glycan biosynthesis have already been explored (Hartmann *et al.* 1993, Zarschler *et al.* 2009, Zarschler *et al.* 2010b, Ristl *et al.* 2012). Initially, it was proposed that the biosynthesis of the SpaA glycan starts with the formation of a nucleotide-activated trisaccharide based on the inveolvement of GDP-ManNAc, UDP-Gal and UDP-GlcNAc. The GDP-activated trisaccharide was further modified by the addition of a Glc residue from a UDP-Glc molecule, forming a branched tetrasaccharide. This branched oligosaccharide has a similar composition as the trisaccharide repeating unit of SpaA glycan plus an extra GlcNAc residue. Then, the GDP-activated tetrasaccharide would be transferred to a dolichyl-C_{55} molecule, generating a lipid-inked precursor which contained two more GlcNAc residues (Hartmann *et al.* 1993). However the SpaA glycan does not carry a single GlcNAc residue.

Later on, Zarschler and co-workers revealed the presence of an S-layer glycosylation (*slg*) gene cluster of about 24 kb on the *P. alvei* CCM 2051^T chromosome. The gene cluster encodes enzymes that participate in the initiation, linker extension, glycan chain synthesis, translocation of the glycan, and transfer of the complete oligosaccharide chain to the S-layer protein SpaA

(Zarschler *et al.* 2010b). This gene cluster showed high homology to genes involved in the biosynthesis of surface polysaccharides of different *Bacillaceae* (Novotny *et al.* 2004, Messner *et al.* 2008). According to our current model, SpaA glycan biosynthesis is initiated by the transfer of a galactose residue from a UDP-galactose to a phosphorylated lipid carrier by a UDP-Gal-phosphoryl-polyprenol Gal-1-phosphate transferase, WsfP. WsfP catalyzes the first step of the biosynthesis of the linker saccharide of the SpaA glycan at the cytoplasmic surface of the plasma membrane (Zarschler *et al.* 2009).

A homologous enzyme has also been described in the Gram-positive bacterium *Geobacillus stearothermophilus* NRS 2004/3a (Steiner *et al.* 2007). The glycan of the S-layer protein SgsE of *G. stearothermophilus* NRS 2004/3a has a similar linker saccharide structure, and its biosynthesis begins with an analogous reaction catalyzed by the UDP-Gal-phosphoryl-polyprenol Gal-1-phosphate transferase WsaP (Schäffer *et al.* 2002, Steiner *et al.* 2006, Steiner *et al.* 2007). Disruption of the chromosomal *wsfP* gene by insertion of an intron RNA into *wsfP* resulted in *P. alvei* CCM 2051^T cells deficient in S-layer glycosylation. Yet, glycosylation of the S-layer was fully restored when the *P. alvei* CCM 2051^T *wsfP* deficient mutant cells were transformed with a plasmid carrying the *wsfP* gene (Zarschler *et al.* 2009), thus, confirming the role of the WsfP enzyme in S-layer glycan biosynthesis.

The SpaA glycan linker chain extension most likely continues in the cytoplasm by addition of three L-rhamnose residues from dTDP-rhamnose to the non-reducing end of the lipid-linked D-galactose molecule. Linker biosynthesis proceeds with the transfer of a D-ManNAc residue from a UDP-ManNAc to the third rhamnose residue to form a β-1,4-linkage and finishes with the addition of a glycerol phosphate from CDP-glycerol to the last ManNAc residue. The glycan chain is then elongated by incorporation of a second D-ManNAc sugar to the third rhamnose residue to form a α-1,3-linkage and a D-galactose to form a β-1,4-linkage to the ManNAc residue of the repeating units. The glycan chain would be later transported across the cytoplasmic membrane by an adenosine

triphosphate (ATP)-binding cassette transporter (ABC transporter) system. Glycan biosynthesis terminates with the transfer of the completed oligosaccharide chain to specific tyrosine residues of SpaA by the enzyme WsfB upon membrane translocation of SpaA (Zarschler *et al.* 2010b).

The *wsfB* gene is predicted to encode an integral membrane protein with a conserved Wzy_C motif, distinctive of the transfer of undecaprenyl-pyrophosphate-linked oligosaccharide to a specific protein (Zarschler *et al.* 2010b). WsfB, thus, belongs to the *O*-oligosaccharyl-protein transferase (*O*-OTase) family, essential for the formation of the *O*-glycosidic linkage between the glycan chain and the SpaA S-layer protein via a β-D-Gal→tyrosine bond. Tyrosine *O*-glycosylation is a rare type of *O*-glycosidic linkage in S-layer proteins and has only been described in the thermophilic microorganisms *Thermoanaerobacter thermohydrosulfuricus*, *Thermoanaerobacterium thermo-saccharolyticum* and in the mesophilic bacterium *Paenibacillus alvei* CCM 2051^T (Bock *et al.* 1994, Messner *et al.* 1995, Schäffer *et al.* 2000a). On the other hand, serine and threonine residues have been identified as the most prevalent sites at specific consensus sequences for *O*-glycosylation in S-layer proteins of diverse prokaryotes, making SpaA *O*-glycosylation of special interest (Schäffer *et al.* 1999b, Steiner *et al.* 2006, Posch *et al.* 2011, Anzengruber *et al.* 2014).

To confirm the role of WsfB in SpaA glycan biosynthesis, a knockout mutant of *P. alvei* CCM 2051^T was constructed. Coomassie Brilliant Blue-stained SDS gels of the *wsfB* deficient mutant showed a single thick protein band corresponding to the non-glycosylated S-layer protein (Zarschler *et al.* 2010b). In addition, structural studies performed on WsfB proposed a transmembrane topology model of this enzyme. The study revealed that WsfB possesses 13 membrane spanning helices, four large loops at the outer side of the membrane and six small loops plus a C-terminal tail at the cytoplasm. The Wzy_C motif of WsfB is located in the third periplasmic loop and shows a similar degree of sequence conservation as the *O*-OTases PilO and PglL, involved in pilin glycosylation in the Gram-negative bacteria *Pseudomonas aeruginosa* and

Neisseria meningitides, respectively (Ristl *et al.* 2012). This finding is noteworthy since the transfer of the glycan to the S-layer protein is a membrane associated process and presumably takes place at the outer side of the cytoplasmic membrane.

Interestingly, point mutations or deletions within the Wzy_C motif resulted in loss of S-layer glycosylation. Periodic acid Schiff-stained SDS gels of these mutants exhibited various glycosylated proteins, except for the mono- and diglycosylated forms of SpaA, restricting the role of WsfB to the glycosylation pathway of the S-layer protein (Ristl *et al.* 2012). These studies contributed to a better understanding of S-layer protein glycosylation systems in Gram-positive bacteria and represent an advance toward the production of tailor made glycoproteins.

1.3.3 Molecular mechanism of S-layer protein SpaA cell wall anchoring in P. alvei *CCM 2051^T*

Paenibacillus alvei CCM 2051^T has been used as a model to study S-layer protein-SCWP interactions. One on hand, the structure of its SCWP was elucidated, being composed of the repeating unit [(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→, where a single chain is tethered to the C-6 hydroxyl group of MurNAc residues of the PG layer via a phosphodiester bond (**Figure 2**) (Schäffer *et al.* 2000b). On the other hand, the S-layer protein SpaA possesses three typical SLH domains at the N-terminus. Interestingly, only one SLH domain of the SpaA protein has the conserved TRAE motif, whereas the other two SLH domains possess a naturally mutated variant of the TRAE motif: TRAE in SLH_1, TVEE in SLH_2 and TRAQ in SLH_3 (**Figure 7A**) (Zarschler *et al.* 2010a). Distinctive is the lack of a positively charged Arg residue in the TVEE variant, since the ubiquity of a positive charge has been described to be crucial to PG binding in the xylanase XynA of *T. thermosulfurigenes* (May *et al.* 2006).

May and co-workers observed that substitution of the positively charged Arg residue to a small non-polar Ala residue drastically reduced the binding to PG. The position of Arg within the motif is also crucial for the SLH domain functionality (May *et al.* 2006). Similarly, mutation of the TRAE motifs in the *P. alvei* CCM 2051^{T} SpaA protein unveiled an unequal contribution of these three motifs to PG binding, where mutation of each motif to TAAA reduced PG binding to 37%, 88% and 51% for SLH domain 1 to 3, respectively (Janesch *et al.* 2013b).

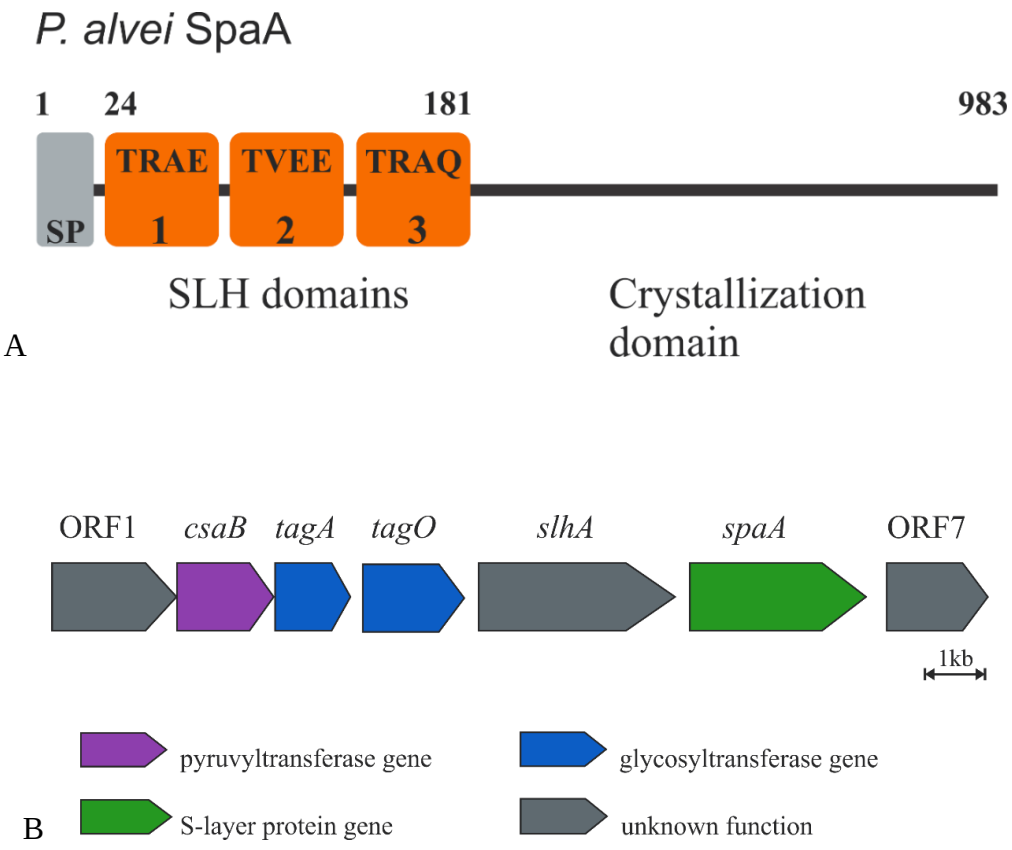


Figure 7. Schematic representation of the S-layer protein SpaA, showing the three SLH domains and the four-residue motif (A) and the genetic organization of the *P. alvei* CCM 2051^{T} SCWP biosynthesis locus (B). SP, signal peptide; SLH, surface layer homology; kb, kilobase pair. Adapted from Zarschler *et al.* (2010a).

As described for *Bacillus anthracis* (Mesnage *et al.* 2000) and *Thermus thermophilus* (Cava *et al.* 2004), binding of SLH domain-carrying proteins depends on the activity of CsaB, a pyruvyltranferase synthesized by many S-layered Gram-positive bacteria with SLH domains. In fact, *T. thermophilus csaB* deficient mutants contain a substantially reduced amount of pyruvic acid (<0.1 µg/mg) in cell wall fractions compared to wild-type *T. thermophilus* (3.8 µg/mg), confirming CsaB implication in pyruvylation of SCWP (Cava *et al.* 2004).

In *P. alvei* CCM 2051^T a homologue of *csaB* has been identified, which shows highly homology to genes coding for pyruvyltransferases (CsaB) in many *Bacillus* species (Zarschler *et al.* 2010a). The *csaB* gene is located upstream of the S-layer protein gene *spaA* in a locus predicted to be involved in *P. alvei* CCM 2051^T SCWP biosynthesis (**Figure 7B**). The *csaB* gene is adjacent to an *orf1* region and is followed by the *tagO* and *tagA* genes in the predicted SCWP biosynthesis gene locus. The gene product of *orf1* is predicted to be a membrane protein and to function as a transporter of polysaccharides across the cytoplasmic membrane, similar to CsaA of *B. anthracis* (Zarschler *et al.* 2010a).

The integral membrane protein nature of CsaA suggests its involvement in the synthesis of SCWP, however in *B. anthracis* and *T. thermophilus*, CsaA does not have a significant role in the synthesis of SCWP (Mesnage *et al.* 2000, Acosta *et al.* 2012). According to bioinformatic analyses, the products of *tagA* and *tagO* show high similarity to the glycosyltranferases TagA and TagO of diverse *Bacillaceae*. Members of the Tag-family proteins participate in the synthesis of classical SCWPs, such as teichoic acids in Gram-positive bacteria (Zhang *et al.* 2006, Ginsberg *et al.* 2006, Formstone *et al.* 2008, Brown *et al.* 2013). In *Bacillus subtilis*, synthesis of teichoic acids starts when TagO transfers a GlcNAc residue from UDP-GlcNAc to a membrane-embedded lipid undecaprenyl-pyrophosphate carrier, followed by the addition of a ManNAc residue from UDP-ManNAc catalyzed by TagA, to produce the lipid-linked GlcNAc-ManNAc disaccharide (Zhang *et al.* 2006, Ginsberg *et al.* 2006). Thus, it is plausible that TagA and TagO participate in the synthesis of *P. alvei* CCM

2051^T SCWP, since the backbone of this polysaccharide is composed of a ManNAc-GlcNAc disaccharide unit. Finally, CsaB should pyruvylate the disaccharide at ManNAc residues of each subunit to produce the final pyruvylated SCWP. Therefore, the presence of a *csaB* homologue and an S-layer composed of a protein with SLH domains in *P. alvei* CCM 2051^T, confirms the well-conserved strategy to display surface proteins in Gram-positive bacteria. However, a structural explanation to unravel the binding mechanism between SLH domains and pyruvylated SCWP has not been described thus far.

2 Aim of the thesis

Paenibacillus alvei CCM 2051^T is a Gram-positive, mesophilic, aerobic, flagellated bacterium that is completely covered by a glycosylated surface (S-) layer. The *P. alvei* CCM 2051^T S-layer is composed of a single protein species, SpaA, which is *O*-glycosylated at least at two specific tyrosine residues. The genetic organization of the S-layer glycosylation (*slg*) gene cluster involved in the glycosylation process is known and its translation products have gradually been investigated, which makes this bacterium a prospective candidate to produce tailor-made, glycosylated S-layer proteins.

Furthermore, *P. alvei* CCM 2051^T is an unsurpassed model to investigate in detail how the SHL domain containing S-layer proteins attach to the underlying cell wall layer, as the structure of its surface anchoring ligand, the secondary cell wall polysaccharide (SCWP) was elucidated; SpaA possesses three tandem repeats of surface layer homology (SLH) domains at its N-terminus. The current knowledge on S-layer protein anchoring to the cell surface describes that the SLH domains of SpaA recognize both the peptido-glycan and pyruvylated SCWP of *P. alvei* CCM 2051^T, where the highly conserved amino acid motif TRAE and its two variants TVEE and TRAQ are essential for recognition. In addition, the genetic organization of the *P. alvei* CCM 2051^T SCWP biosynthesis gene locus is known.

The aim of the present work was, firstly to investigate the influence of native *P. alvei* CCM 2015^T SCWP on the water-solubility of the self-assembly products of recombinant S-layer protein SpaA to obtain a stable SpaA-SCWP complex suitable for three-dimensional crystallization of the full-size S-layer protein. Secondly, this work also pursued to elucidate the mechanism of interaction between SpaA and the SCWP through the SLH domains using a biophysical approach and at atomic resolution. For this purpose a truncated SpaA version containing the three SLH domains should be created and three-dimensionally

crystallized in cooperation with the laboratory of Dr. Stephen V. Evans (Uniersit of Vicrota, Canada). Additionally, small oligosaccharides of the repeating unit of *P. alvei* CCM 2051^{T} SCWP should be synthesized in cooperation with Dr. Paul Kosma (Universität für Bodenkultur, Wien) to obtained co-crystal structures of SpaA with defined SCWP oligomers, and to investigate the binding interaction of SLH domains and the SCWP in solution. Finally, the third goal of this work was to understand the mechanism of *P. alvei* CCM 2051^{T} SCWP pyruvylation, which is predicted to be crucial for an SLH-SCWP interaction to take place. To accomplish that goal, the primary sequence of the pyruvyl transferase CsaB was investigated at a structural point of view. Besides, CsaB was heterologously expressed and biochemically characterized *in vitro*.

3 Materials and methods

3.1 General methods

3.1.1 Bacterial strains and cultivation conditions

Paenibacillus alvei CCM 2051^T was obtained from the Czech Collection of Microorganisms (CCM; Brno, Czech Republic) and was grown aerobically at 37 °C and 200 rpm in Luria-Bertani (LB) medium (Sigma-Aldrich, Vienna, Austria) or on LB agar plates. *Escherichia coli* DH5α cells (Life Technologies, Vienna, Austria) were used as the host strain for cloning and for expression of *spaA*, *spaA$_{SLH}$*, *spaA$_{SLH}$/G109A*, *spaA$_{SLH}$/G46A/G109A, spaA$_{SLH}$/TAAA$_{12}$, spaA$_{SLH}$/TAAA$_{13}$* and *csaB* derivatives. Overexpression of the genes encoding the aforementioned polypeptides was performed in *E. coli* BL21 (DE3-Star) cells (Life Technologies). *Escherichia coli* was routinely grown in LB medium at 37°C and 200 rpm or on LB agar plates supplemented with ampicillin (100 µg/mL) or kanamycin (50 µg/mL), if required. Protein overexpression was induced at an optical density at 600 nm (OD_{600}) of 0.5 - 0.8 with isopropyl-β-D-thiogalactopyranoside (IPTG) at a final concentration of 0.6 mM. All strains and plasmids used in the course of this work are summarized in **Table 1**.

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

Strain or plasmid	Genotype and/or relevant characteristic	Source
P. alvei CCM 2051^T	Wild-type isolate	Czech Collection of Microorganisms (CCM)
Escherichia coli DH5α	F−φ80dlacZ M15 (lacZYA-argF)U169 deoR recA1 endA1 hsdR17 (rK−mK−) phoA supE44 thi-1 gyrA96 relA1−	Life Technologies
Escherichia coli BL21 (DE)	F−, ompT, hsdS (rB−mB−), gal, dcm (DE3)	Life Technologies
pET22b	Expression vector with a His_6-tag, Amp^r	Novagen
pET28a_SpaA_ 6His	pET28a carrying His-tagged *spaA* gene of *P. alvei* CCM 2051^T; Kan^r	Janesch *et al.*, (2013)

Strain or plasmid	Genotype and/or relevant characteristic	Source
pETSMut4H	pET28a carrying His_6-tagged *spaA* mutated in TRAE and TVEE motifs	Janesch *et al.*, (2013)
pETSMut5H	pET28a carrying His_6-tagged *spaA* mutated in TRAE and TRAQ motifs	Janesch *et al.*, (2013)
pET22b_spaA$_{SLH}$	pET22b carrying the His_6-tagged *spaA$_{SLH}$* gene of *P. alvei* CCM 2051^T, Amp^r	This study
pET22b_spaA$_{SLH}$/ G109A	pET22b carrying the His_6-tagged *spaA$_{SLH}$* alanine-mutated at residue glycine 109	This study
pET22b_spaA$_{SLH}$/ G46A/G109A	pET22b carrying the His_6-tagged *spaA$_{SLH}$* alanine-mutated at residues glycine 46 and glycine 109	This study
pET22b_spaA$_{SLH}$/ $TAAA_{12}$	pET22b carrying the His_6-tagged *spaA$_{SLH}$* carrying mutated TRAE and TVEE motifs in SLH_1 and SLH_2, respectively	This study
pET22b_spaA$_{SLH}$/ $TAAA_{13}$	pET22b carrying the His_6-tagged *spaA$_{SLH}$* carrying mutated TRAE and TRAQ motifs in SLH_1 and SLH_3, respectively	This study
pET22b_csaB	pET22b carrying the His_6-tagged *csaB* gene of *P. alvei* CCM 2051^T, Amp^r	This study

3.1.2 Molecular biology methods

Genomic DNA of *P. alvei* CCM 2051^T was isolated from 2 mL of a bacterial suspension as described previously (Cheng and Jiang 2006). All restriction enzymes were supplied by Fermentas (Waltham, MA, USA). The GeneJET™ Gel Extraction Kit (Fermentas) was employed to purify digested plasmids from agarose gels, DNA fragments and oligonucleotides. Plasmid DNA from transformed cells was purified with the GeneJET™ Plasmid Miniprep Kit (Fermentas). Agarose gel electrophoresis was performed as described elsewhere (Sambrook *et al.* 1989). Primers for PCR and DNA sequencing were supplied by Life Technologies (**Table 2**). PCR was performed using the Phusion® High-Fidelity DNA Polymerase (Fermentas) and a My CyclerTM (Bio-Rad, Vienna, Austria) thermal cycler. Transformation of chemically competent *Escherichia coli* DH5α and BL21 (DE3-Star) cells was carried out

according to the manufacturer's protocol (Life Technologies). *E. coli* transformant colonies were screened by PCR using the RedTaq Ready Mix PCR mixture (Sigma-Aldrich). Recombinant plasmids were analyzed by restriction mapping and by DNA sequencing (Microsynth, Wolfurt, Austria) to confirm the presence of the desired gene and mutations.

Table 2: Oligonucleotide primers used for PCR amplification and DNA sequencing.

Primer	Sequence (5'→3') [a,b]	Comment
$SpaA_{SLH}$-NdeI	ggctagCATATGGTAGCTTTCGGTGCT GACGCAGCAAAAACAACTCAAGAGA AATTTGATGC	PCR amplification
$SpaA_{SLH}$-SacI	cgcgtcGAGCTCTTAGTGGTGGTGGTG GTGGTGAGAACCAGAACCTTTAGAC ATTTCGTCAGCTGCAAATGCTGC	PCR amplification
Fwd-mut-G109A	GGCTTGATGCAA*GCC*AAAGATTTGA CTAAG	Overlap extension PCR to mutate G109
Rev-mut-G109A	CTTAGTCAAATCTTT*GGC*TTGCATC AAGCC	Overlap extension PCR to mutate G109
Fwd-mut-G46A	GCTGGTGTATTCTCC*GCC*TATCCAG GAACTACT	Overlap extension PCR to mutate G46
Rev-mut-G46A	AGTAGTTCCTGGATA*GGC*GGAGAAT ACACCAGC	Overlap extension PCR to mutate G46
CsaB-NdeI	ctgtggCATATGGCGTCCAAAGCTACA AGAATAGTACTTTCCGGATATTACG GATTC	PCR amplification
CsaB-XhoI	gacctaCTCGAGCGCCTTATGACGCAG CCACTTCACAATTTGTTGCGCTGGC TGTTCTGCTT	PCR amplification

[a] Restriction sites are underlined. Lowercase letters indicate artificially introduced bases to improve restriction enzyme cutting.
[b] Mutated codons are underlined and shown in italics.

3.1.3 Analytical methods

Sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) was performed according to a standard method using a Protean II electrophoresis apparatus (Bio-Rad) (Laemmli 1970). Protein bands were visualized with Coomassie Brilliant Blue G250 (CBB) staining reagent. A Mini Trans-Blot Cell (Bio-Rad) was used for western immunoblotting of proteins onto a nitrocellulose membrane (Peqlab, Erlangen, Germany). Detection of the His_6-tag fused to rSpaA, $rSpaA_{SLH}$, $rSpaA_{SLH}$/G109A, $rSpaA_{SLH}$/G46A/G109A, $rSpA_{SLH}/TAAA_{12}$, $rSpaA_{SLH}/TAAA_{13}$ and rCsaB was carried out with the Li-Cor Odyssey Infrared Imaging System (Li-Cor Biosciences, Lincoln, NE, USA) using a mouse anti-His_6 anti-body (Roche, Basel, Switzerland) in combination with a goat anti-mouse IgG-IRDye 800 CW conjugate (Li-Cor).

Peptidoglycan containing cell wall sacculi with or without SCWP were analyzed for their amino sugars, amino acids and neutral sugar content. For neutral and amino sugars, samples were hydrolyzed with 2.2 M trifluoroacetic acid (TFA) at 110 °C for four hours. For amino sugars and amino acids, sample hydrolysis was carried out with 6 N HCl with 0.2% thioglycolic acid at 110°C for four hours (Ries *et al.* 1997). Acid was removed by a gentle N_2 flow and samples were washed with MilliQ H_2O. Dried samples were dissolved in MilliQ H_2O or in 0.2 M sodium citrate buffer pH 2.2 and analyzed with a HPAEC-PAD Dionex ICS3000 (Dionex, Sunnyvale, CA, USA) apparatus and a Biochrom 30 amino acid analyzer (Biochrom, Cambourne, Cambridge, UK).

3.2 Isolation of peptidoglycan-containing cell wall sacculi and glycosylated S-layer protein SpaA from *Paenibacillus alvei* CCM 2051^T

S-layer protein and peptidoglycan-containing cell wall sacculi (PG-SCWP) of *P. alvei* CCM 2051^T were prepared as described by (Schäffer *et al.* 1999a) with minor modifications. *Paenibacillus alvei* CCM 2051^T was grown in continuous fermentation for 25 hours at 37 °C until an OD_{600} of ~5.0. After cultivation cells were harvested by continuous centrifugation at 17,000 x *g* at 4.0°C and the

biomass was stored at -20°C until use. 50 g (wet weight) of washed cells were resuspended in 150 mL of cold distilled water and disrupted by ultrasonication at maximum output on ice for four minutes. Intact cells and cell fragments were centrifuged at 48,000 x *g* for 25 minutes at 4°C. The upper, whitish layer consisting of cell wall fragments, was separated from the lower layer, which consisted of unbroken cells, by gentle shaking with 150 mM NaCl. Intact cells were resuspended with cold distilled water and sonicated two more times as described before, in each round the upper, whitish layer was separated from the pellet. The collected cell wall suspension was centrifuged at 20,000 x *g* for 20 minutes. The glycosylated S-layer protein was isolated from the PG-containing cell wall sacculi by incubating the cell wall fragments for one hour in 100 mL of 6 M urea with constant stirring. Solubilized S-layer protein was separated from the cell wall fragments by centrifugation at 48,000 x *g* for 30 minutes at 4 °C. The supernatant containing the S-layer protein was extensively dialyzed against distilled water to remove urea, and the self-assembly products were centrifuged at 40,000 x *g* for 20 minutes. Self-assembly products were stored at -20°C.

Cell wall fragments were resuspended in 1 M NaCl and dialyzed against the same reagent to remove the excess of urea. Subsequently, the cell wall fragments were pelleted and lyophilized. The lyophilized material was resuspended in 100 mL of distilled water, incubated at 100°C for 20 minutes, afterward harvested and stored at -20°C.

The thawed cell wall suspension was mixed with 200 mL of 1 M sodium acetate, pH 5.9 , incubated with DNAse (25 µg/mL) and RNAse (25 µg/mL) (Roche) for three hours at 37°C with shaking and then centrifuged for one hour at 18,000 x *g*. The pellet was resuspended in 200 mL of pre-warmed 50 mM phosphate buffer, pH 7.5, supplemented with 20 mg of trypsin and incubated for two hours at 37°C to degrade any contaminant proteins. Harvested cell wall fragments were incubated in 250 mL of 0.4% SDS at room temperature with

stirring for one hour, centrifuged at 48,000 x *g* for 30 minutes, washed eight to ten times with distilled water and lyophilized.

Lyophilized material, approximately 0.5 g, was resuspended in 50 mL of distilled water containing 4.2 g of $NaHCO_3$ and 2.5 mL of acetic acid anhydride, and incubated overnight at 4°C to re-*N*-acetylate amino groups of partial de-*N*-acetylated amino sugars. The cell wall material was centrifuged, washed with distilled water and lyophilized. Next, *N*-acetylated cell wall fragments were digested at 37°C for 48 hours in 200 mL of 50 mM Tris-HCl, pH 7.2, containing 80 mg of lysozyme (Sigma-Aldrich) and 0.1% sodium azide (Amano *et al.* 1977). Following incubation, the lysozyme digest was concentrated to 50 mL using a rotavapor and dialyzed against five L of distilled water. Dialysis was repeated six times with a total duration of 72 hours. Insoluble material was separated from the non-dialyzable fraction by centrifugation, the supernatant was collected and lyophilized giving the PG-containing cell wall sacculi.

The lyophilized, soluble fraction (0.1 g) was dissolved in 5 mL of 50 mM NH_4HCO_3 and applied onto a Sephadex® G-50 (GE Healthcare Life Sciences, Vienna, Austria) column (2.6 x 100 cm). The column was eluted with 50 mM NH_4HCO_3 at a flow rate of 1.5 mL/min, and the elution pattern was monitored by measurement of the refraction index and UV absorbance at 280 nm. The total elution volume was 600 mL and fractions of 6 mL were collected. Fractions were examined for *N*-acetylhexosamines and hexosamines content with thymol-sulfuric acid stain (Adachi 1965). Those giving a positive color reaction were pooled, concentrated and further purified. Pooled fractions, 5 mL, was applied onto a Bio-Gel® P-10 polyacrylamide gel column (Bio-Rad) (1.5 x 100 cm). The column was eluted with 0.1 mM NaCl at a flow rate of 0.15 mL/min. The sample elution was monitored by measurement of the refraction index and UV absorbance at 280 nm. The total elution volume was 180 mL and fractions of 3 mL were collected. Fractions were analyzed on Silica gel TLC plates (Sigma-Aldrich) and stained with thymol-sulfuric acid reagent, and those giving a positive reaction were pooled, dialyzed against distilled

water and lyophilized. Chemical analysis of *P. alvei* CCM 2051^T PG-containing cell wall sacculi (PG-SCWP) was performed as described above in the general and analytical methods section.

3.3 Expression and purification of recombinant full-size *P. alvei* CCM 2051^T S-layer protein SpaA

Studies of the interaction between *P. alvei* CCM 2051^T S-layer protein self-assembly products and native SCWP were carried out with recombinant full-size SpaA and purified PG-containing cell wall sacculi. Recombinant His_6-tagged SpaA was expressed in *E. coli* BL21 (DE3-Star) cells by the addition of IPTG at a final concentration of 0.6 mM. The bacterial strain carrying the plasmid pET28a_SpaA_6His was kindly provided by Bettina Janesch (Janesch *et al.* 2013b). After incubation, cells were harvested at 6,000 x *g* for 20 minutes. Cell pellets were resuspended in lysis buffer (50 mM NaH_2PO_4, pH 8.0, 200 mM NaCl, 10 mM imidazole and 4 M urea) and disrupted by sonication (duty cycle 50%, output 8, for three minutes with five seconds pause each 20 seconds) on ice. The cell lysate was clarified by ultracentrifugation at 126,000 x *g* for 30 minutes and applied to a chromatography column loaded with Ni-NTA matrix (QIAGEN®, Hilden, Germany) and equilibrated with lysis buffer. The column was washed with 10 column volumes of washing buffer (50 mM NaH_2PO_4, pH 8.0, 200 mM NaCl, 40 mM imidazole and 4 M urea).

His_6-tagged SpaA was eluted with three volumes of elution buffer (50 mM NaH_2PO_4, pH 8.0, 250 mM imidazole and 4 M urea). Fractions were analyzed by SDS-PAGE and proteins were visualized by CBB staining. Elution fractions were pooled and dialyzed against elution buffer without imidazole to remove the latter. Next, the protein concentration was determined by UV-VIS light spectroscopy using a molar extinction coefficient of 76,670 M^{-1} cm^{-1} at 280 nm. Finally, the rSpaA solution was dialyzed against distilled water to remove urea and promote formation of self-assembly products. Suspension of rSpaA self-assembly products were stored at 4 °C until use.

3.4 Binding studies of *P. alvei* CCM 2051^T S-layer protein self-assembly products and its native SCWP *in vitro*

To investigate the influence of *P. alvei* CCM 2051^T SCWP on the water-solubility of rSpaA self-assembly products, different ratios of rSpaA self-assembly products to PG-SCWP (wt/wt) were tested. Specifically, 0.5 mg of rSpaA self-assembly products were mixed with different amounts of native, lyophilized PG-containing cell wall sacculi (5 to 0.05 mg) in a final volume of 400 µL. The suspensions were stirred and incubated at room temperature for one hour. After incubation, the samples were centrifuged at 20,000 x *g* for 20 minutes. The supernatants were removed and the pellets were washed once with distilled water. All the fractions (supernatant, wash solution and pellet) were subjected to SDS-PAGE analysis. To estimate the amount of rSpaA that became water-soluble (percentage of solubilized S-layer protein) by association or interaction with PG-associated SCWP, the protein content in the supernatants, the wash solutions and the pellets was determined by the Bradford protein assay (Bradford 1976).

3.5 Recombinant production, expression and purification of a His_6-tagged truncated variant $SpaA_{SLH}$ for crystallization and *in vitro* studies

The $SpaA_{SLH}$ protein comprising amino acids 21 to 193, including a Ser-Gly-Ser linker preceding a C-terminal His_6-tag was created by standard PCR amplification from *P. alvei* CCM 2051^T genomic DNA. Primers used for *spaA$_{SLH}$* amplification are listed in **Table 2**. The PCR product was subsequently digested and ligated into linearized pET-22b vector (Novagen, Vienna, Austria) using the NdeI and SacI restriction enzymes. The resulting plasmid, pET-22b-*spaA$_{SLH}$*, was transformed into *E. coli* BL21 (DE3-Star) cells and plated on LB agar plates containing 100 µg/mL of ampicillin. Single colonies were used to inoculate 5 mL of LB medium supplemented with antibiotic and grown overnight at 37°C with shaking at 180 rpm. This culture was transferred into four 2-L Erlenmeyer flasks containing 0.5 L of ampicillin-supplied LB medium.

Protein expression was induced at an OD_{600} of 0.5-0.8 by the addition of IPTG at a final concentration of 0.6 mM and cells were further shaken at 180 rpm and incubated for three hours at 37°C. Cells were harvested by centrifugation at 6,000 x *g* for 20 minutes.

Pelleted cells were resuspended in lysis buffer (50 mM NaH_2PO_4 or 25 mM Tris-HCl, pH 8.0, 200 mM NaCl, 5 mM imidazole) and disrupted by sonication (duty cycle 50%, output 8, for three minutes with five seconds pause each 20 seconds) on ice. The lysate was clarified by ultracentrifugation at 126,000 x *g* for 30 minutes and applied to a chromatography column loaded with Ni-NTA matrix (QIAGEN®) and equilibrated with lysis buffer. The column was washed with 10 column volumes of buffer A (50 mM NaH_2PO_4 or 25 mM Tris-HCl, pH 8.0, 200 mM NaCl, 20 mM imidazole), followed by 10 volumes of buffer B (50 mM NaH_2PO_4 or 25 mM Tris-HCl pH 8.0, 200 mM NaCl, 50 mM imidazole). His_6-tagged $SpaA_{SLH}$ was eluted with two volumes of elution buffer (50 mM NaH_2PO_4 or 25mM Tris-HCl, pH 8.0, 250 mM imidazole). Further purification of $SpaA_{SLH}$ was performed by size exclusion chromatography on a Superdex™ 75 HiLoad™ 16/60 or Superdex™ 200 HiLoad™ 26/60 (GE Healthcare Life Sciences) column equilibrated with 20 mM HEPES, pH 7.5, containing 100 mM NaCl. Protein elution was performed with an isocratic flow and fractions were analyzed by SDS-PAGE. Proteins were detected by CBB.

3.6 Construction of $SpaA_{SLH}$/G109A (single mutant) and $SpaA_{SLH}$/G46A/G109A (double mutant) by site-directed mutagenesis

Site-directed mutagenesis of residues glycine 109 and glycine 46 of SpaA to alanine were performed by overlapping extension PCR. The upstream and downstream part of the mutation site in $SpaA_{SLH}$ were amplified separetly using as templates the purified, recombinant plasmids pET22b-*spaA*$_{SLH}$ for the single mutant and pET22b-*spaA*$_{SLH}$/*G109A* for the double mutant. The forward primer of the downstream part and the reverse primer of the upstream part were overlapping and included the point mutations that should be introduced

in both elongations. These two amplicons were mixed and amplified in a second round of PCR. Linear PCR fragments hybridize, generating a nicked, circular plasmidic DNA with the desired mutation. Template DNA was degraded with the DpnI restrcition enzyme, which specifically digests methylated DNA. Nicked plasmids were then transformed into *E. coli* DH5α cells. The presence of G46A and G109A mutations was confirmed by direct DNA sequencing. Expression and purification of the single and double mutant proteins were performed as described for the $SpaA_{SLH}$ variant. Primers for site-directed mutagenesis are listed in **Table 2**.

3.7 Construction of $SpaA_{SLH}/TAAA_{12}$ and $SpaA_{SLH}/TAAA_{13}$ double TAAA mutants

$SpaA_{SLH}/TAAA_{12}$ and $SpaA_{SLH}/TAAA_{13}$ mutants were generated by standard PCR amplification with the primers $SpaA_{SLH}$-NdeI and $SpaA_{SLH}$-SacI utilized to generate $SpaA_{SLH}$, using as templates the plasmid pETSMut4H (pET28a carrying His_6-tagged *spaA* mutated in TRAE and TVEE motifs) for $SpaA_{SLH}/TAAA_{12}$ and the plasmid pETSMut5H (pET28a carrying His_6-tagged *spaA* mutated in TRAE and TRAQ motifs) for $SpaA_{SLH}/TAAA_{13}$ (**Table 1**) (Janesch *et al.* 2013b). The PCR products were digested and ligated into linearized pET-22b vector using the NdeI and SacI restriction enzymes. The presence of the TAAA mutations in $SpaA_{SLH}$/TAAA variants was confirmed by direct DNA sequencing. Expression and purification of $SpaA_{SLH}/TAAA_{12}$ and $SpaA_{SLH}/TAAA_{13}$ mutants were carried out as described for the $SpaA_{SLH}$ variant.

3.8 Secondary structure analysis for $SpaA_{SLH}$ by electronic circular dichroism (ECD) spectroscopy

Far-UV ECD for secondary structure analyses of $SpaA_{SLH}$ was performed on a Chirascan ECD instrument (Applied Photophysics, Leatherhead, UK) equipped with a thermostatic cell holder and a Peltier element for temperature

control. The instrument was flushed with a nitrogen flow at a rate of 5 L/min and measurements were performed in the far-UV region (180-260 nm). The instrument parameters were as follows: path length, 1.0 mm; spectral bandwidth, 3.0 nm; step size, 1.0 nm; scan period, 10 seconds. Spectra were baseline-corrected to remove birefringence of the quartz cell. Three scans were recorded and the mean value was calculated. $SpaA_{SLH}$ was analyzed at a concentration of 10 µM in 20 mM KH_2PO_4, pH 7.0 at 20°C. To analyze the thermal stability of $SpaA_{SLH}$, single wavelength scans were performed at 208 nm using the same protein solution as above, with a scan time per point of 10 seconds and stepwise temperature increase rate of 1°C/min over an interval from 20-90°C, using the identical instrument parameters as described above.

3.9 Chemical synthesis of SCWP ligands

The mono- and disaccharide SCWP ligands used in this study were synthesized by G. Martinz and S. Krauter in the laboratory of Dr. P. Kosma (Universität für Bodenkultur Wien). For details see Supplementary methods.

3.10 Protein-ligand interactions analyzed by isothermal titration calorimetry (ITC)

ITC was performed using a MCS titration calorimeter (MicroCal, Northampton, MA USA). Concentrations of the monosaccharide and disaccharide ligands were determined based on dry weight and confirmed by high performance anion exchange chromatography with pulsed amperometric detection on a PA-1 column (ThermoFisher Scientific, Vienna, Austria) (Ries *et al.* 1997). Protein concentration was determined by UV-vis light spectroscopy using a molar extinction coefficient of 18450 M^{-1} cm^{-1} at 280 nm. To avoid sample-related artifacts, protein solutions were freshly prepared prior to each set of experiments by dialysis of $SpaA_{SLH}$, $SpaA_{SLH}$/G109A and $SpaA_{SLH}$/G46A/G109A against 20 mM KH_2PO_4, pH 7.0, used for the titration experiments. The buffer dialysate was used for concentration adjustments and

blank titrations, the ITC measurements were done at 20 °C in 20 mM KH_2PO_4, pH 7.0. Protein, ligand and buffer solutions were degassed prior to run the titrations. The sample cell was filled with 0.02 mM $SpaA_{SLH}$ solution (0.25 mL), the reference cell was filled with buffer and the injection syringe contained 20 times concentrated ligand solutions. The injection sequence consisted of an initial injection of 0.5 μL of ligand solution to prevent low binding enthalpies arising from the filling of the syringe, followed by injection of 1 μL of ligand solution at 150 seconds intervals, each, until complete saturation of $SpaA_{SLH}$ binding sites had been reached. $SpaA_{SLH}$/G109A and $SpaA_{SLH}$/G46A/G109A 0.02 mM solutions were titrated with a 20 times concentrated monosaccharide ligand solution, following a similar titration sequence as with the $SpaA_{SLH}$ variant. A blank sample was run in the absence of protein, to determine the heat of dilution of the ligand. Data analysis was done with Origin software (Microcal) by fitting a single-site binding isotherm. The obtained thermodynamic parameters were enthalpy of binding $\Delta H°$, entropy of binding $\Delta S°$ and the association constant K_A. Measurements were performed at least in triplicates.

3.11 Protein crystallization and ligand soaking

Recombinant, purified $SpaA_{SLH}$ was concentrated to 30 mg/mL using Amicon (Billerica, MA, USA) ultra centrifugal filter units (10 KDa MW cut-off). Crystal screens were prepared using an Art Robbins Instruments (Sunnyvale, CA, USA) crystal gryphon robot and Hampton 96-well intelli plates (Hampton Research, Aliso Viejo, CA, USA). Crystals were obtained with the Hampton Index screen in condition number 54 (50 mM calcium chloride dihydrate, 0.1 M bis-Tris, pH 6.5, and 30% v/v PEG MME 550), and optimized by hanging drop vapor diffusion in 35 x 10 mm tissue culture dishes in the same condition at 16°C with higher protein-to-reservoir ratios. Crystals grew within one to three days. 4,6-Pyr-β-D-ManNAc-*O*-Me was soaked into existing crystals of $SpaA_{SLH}$ at a concentration of 10 mM. 4,6-Pyr-β-D-ManNAc-*O*-Me was also co-crystallized with $SpaA_{SLH}$ in the same crystallization condition at 5 mM and

27 mg/mL, respectively. Additional co-crystals with 4,6-Pyr-β-D-ManNAc-*O*-Me were obtained from the Hampton Index screen, with the crystals that provided the C2 structure grown in condition 66 (0.2 M ammonium sulfate, 0.1 M bis-Tris pH 5.5 and 25% (w/v) PEG 3350), the $P2_12_12_1$ structure in condition 41 (0.1 M sodium acetate trihydrate pH 4.5, 25% w/v PEG 3350), and the P1 structure in condition 87 (0.2 M sodium malonate pH 7.0, 20% w/v PEG 3350). Co-crystals of $SpaA_{SLH}$ and β-D-GlcNAc-(1→3)-4,6-Pyr-β-D-ManNAc-*O*-Me were obtained in Hampton Index screen condition number 57 (50 mM ammonium sulfate, 50 mM bis-Tris pH 6.5 and 30% (v/v) pentaerythritolethoxylate [15/4 EO/OH]) with initial protein and ligand concentrations of 27 mg/mL and 5 mM, respectively. Crystals of $SpaA_{SLH}$/G109A were obtained with the Hampton Index screen in condition number 81 (0.2 M ammonium acetate, 0.1 M Tris pH 8.5 and 25% (w/v) PEG 3350). Co-crystals of $SpaA_{SLH}$/G109A with 4,6-Pyr-β-D-ManNAc-*O*-Me were obtained with the Hampton Index screen in condition number 95 (0.1 M potassium thiocyanate, pH 6.8, and 30% PEG MME 2000). Co-crystals of $SpaA_{SLH}$/G46A/G109A with 4,6-Pyr-β-D-ManNAc-*O*-Me were obtained with the Hampton Index screen in condition number 47 (0.1 M bis-Tris, pH 6.5, and 28% PEG MME 2000).

3.12 Data collection, structure determination and refinement

Crystals of unliganded $SpaA_{SLH}$ were soaked overnight at 16°C in mother liquor with the addition of 0.2 M KI. X-ray diffraction data were collected on a Rigaku R-AXIS IV++ area detector at a distance of 95 mm and exposure time of two min for 0.5° oscillations. X-rays were produced by an MM-003 generator at a wavelength of 1.5418 Å (Rigaku Americas, The Woodlands, TX, USA) coupled to Osmic "Blue" confocal X-ray mirrors with power levels of 30W (Osmic, Troy, MI, USA). The data were scaled, averaged and integrated using HKL2000 (Otwinowski and Minor 1997). PhenixAutoSol (Adams *et al.* 2010) was used to solve SAD phases from 14 iodide ions and to generate initial models in space group $P3_221$ with two molecules in the asymmetric unit (AU). A second data set was collected from a native crystal on beamline CMCF-ID at the

Canadian Light Source (CLS) synchrotron (Saskatoon, SK, Canada) at a wavelength of 0.9795 Å using a Marmosaic CCD300 detector at a distance of 250 mm and exposure time of 2.5 seconds for 0.5° oscillations. Data were processed using HKL2000. The structure was solved in space group $P3_221$ with two molecules in the AU by molecular replacement with *Phaser* (McCoy *et al.* 2007) using the lower resolution KI derivative structure as a search model. Data sets from $SpaA_{SLH}$ crystals soaked and co-crystallized with synthetic 4,6-Pyr-β-D-ManNAc-*O*-Me were collected on a Rigaku R-AXIS IV++ area detector as described above. Data from co-crystals of $SpaA_{SLH}$ with β-D-GlcNAc-(1→3)-4,6-Pyr-β-D-ManNAc-*O*-Me were collected on CLS beamline CMCF-ID as described above, with a detector distance of 250 mm and an exposure time of one second for 0.5° oscillations. Complex structures were solved by molecular replacement with *Phaser* using the native unliganded $SpaA_{SLH}$ structure as a search model. Data sets from $SpaA_{SLH}$/G109A crystals were collected at on CLS beamline CMCF-ID as described above, with a detector distance of 125 mm and an exposure time of 0.5 seconds for 0.5° scillations. All model building and refinement was carried out using *Coot* (Emsley *et al.* 2010) and *Refmac5* through the *CCP4* interface (Murshudov *et al.* 2011, Winn *et al.* 2011).

3.13 Recombinant production, expression and purification of a putative His_6-tagged *P. alvei* CCM 2051^T pyruvyltransferase CsaB

The *csaB* gene was amplified from genomic DNA of *P. alvei* CCM 2051^T with a fused C-terminal His_6-tag by standard PCR using the primer pair CsaB-NdeI/-XhoI (**Table 2**). The amplification products were then digested and ligated into linearized pET-22b (Novagen) vector using the NdeI and XhoI restriction enzymes (Fermentas). The resulting plasmids, pET-22b-*csaB* was transformed in *E.coli* BL21 (DE3-Star) cells and plated on LB agar plates containing 100μg/mL of ampicillin. A single colony was transferred to 5 mL of ampicillin-LB medium and grown overnight at 37°C with shaking at 180 rpm. The overnight culture was used to inoculate 2 L of LB medium with ampicillin

(100μg/mL) distributed in four 2-L Erlenmeyer flasks. Protein expression was induced at an OD_{600} of 0.5-0.8 by the addition of IPTG at a final concentration of 0.6 mM. The bacterial culture was further incubated at 37°C for three hours. Cells were pelleted by centrifugation at 6,000 x *g* for 20 minutes and pellets were stored at -20 °C.

Cells were resuspended in lysis buffer (50 mM NaH_2PO_4 or 25 mM Tris-HCl, pH 8.0, 200 mM NaCl, 5 mM imidazole) and disrupted by sonication (for three minutes with five seconds pause each 20 seconds) on ice. The cell lysate was ultracentrifuged at 126,000 x *g* for 0.5 h. The clear supernatant was applied to a chromatography column loaded with Ni-NTA matrix (QIAGEN®) and equilibrated with lysis buffer. The column was washed with 10 column volumes of buffer A (50 mM NaH_2PO_4 or 25 mM Tris-HCl, pH 8.0, 200 mM NaCl, 20 mM imidazole), followed by a second wash with 10 volumes of buffer B (50 mM NaH_2PO_4 or 25 mM Tris-HCl pH 8.0, 200 mM NaCl, 50 mM imidazole). His_6-tagged protein was eluted with two volumes of elution buffer (50 mM NaH_2PO_4 or 25mM Tris-HCl, pH 8.0, 250 mM imidazole) and fractions were analyzed by SDS-PAGE. Elution fractions containing the recombinant protein were pooled and dialyzed overnight against phosphate or Tris-HCl buffer to remove imidazole.

3.14 Amino acid sequence analysis, secondary structure determination and fold recognition of the putative *P. alvei* CCM 2051^T pyruvyl transferase CsaB

A sequence homology search and a domain prediction of CsaB was used to identify conserved motifs or domains that are structurally or functionally important. In addition, secondary structure analysis of rCsaB was performed by far-UV (180 – 260 nm) ECD spectroscopy as described above for $SpaA_{SLH}$. Briefly, rCsaB was analyzed at a concentration of 4 and 6 μM in 20 mM NaH_2PO_4, pH 7.5, at 25°C. Three scans were recorded and the mean value was calculated. The instrument parameters were as follows, path length: 10 mm;

spectral bandwidth: 3.0 nm; step size: 1.0 nm with a scan time per point of 10 seconds.

Attempts to crystallize recombinant *P. alvei* CCM 2051^T CsaB have been unsuccessful so far, therefore a three-dimensional structure of CsaB was modeled using the web tool *Phyre2* for protein modeling, prediction and analysis (Kelley *et al.* 2015).

3.15 Studies on pyruvyl tranferase activity of recombinant CsaB

The pyruvyl transferase activity of rCsaB was investigated in 200-µL reaction mixtures containing 1.0 mM phosphoenolpyruvate (PEP; Sigma-Aldrich), 1.0 mM *p*NP-β-ManNAc (kindly provided by Dr. Stephen G. Withers, University of British Columbia, Vancouver, BC Columbia, Canada) or *p*NP-β-GlcNAc (Sigma-Aldrich) and 11 to 15 µM rCsaB in 25 mM Tris-HCl, pH 7.5. To account for the putative necessity of divalent cations, the reactions were performed in the presence of 5 mM of $MgCl_2$. The enzyme reactions were incubated at 37°C for one to 12 hours. After incubation, the enzyme was removed from the reaction mixture using an Amicon™ ultracentrifugal filter prior to HPLC analysis, and the filtrates were analyzed by RP-HPLC. Separation of *p*NP-β-*N*-acetylhexosamines was performed using a Dionex UltiMate 3000 HPLC system equipped with a UV spectrophotometer. Fractionation was performed using a reverse-phase column NUCLEOSIL C-18 (ThermoFisher Scientific) with a flow rate of 0.6 mL/min at 25°C. The column was equilibrated with buffer A (0.3% ammonium formate, pH 7.4, 13% acetonitrile) used as mobile phase. 10 µL of each filtrate sample were loaded onto the column and *p*NP-β-ManNAc or *p*NP-β-GlcNAc were detected at 265 nm as described by (Yoritsune *et al.* 2013).

The predicted pyruvyl transferase activity of rCsaB was also investigated through the release of phosphate from PEP (Sigma-Aldrich) using the malachite green spectrophotometric method for the determination of inorganic phosphate. The malachite green method is based on the reaction at low pH

between ammonium molybdate, polyvinyl alcohol and malachite green (Lanzetta *et al.* 1979, Motomizu and Li 2005). The orthophosphate (Pi) released from PEP was detected utilizing the Pi ColorLock™ Gold reagent, an improved malachite green formulation, as recommended by the manufacturer (Innova Biosciences, Babraham, Cambridge, UK).

The reactions with rCsaB were performed in 96-well formats in a final volume of 200 µL, where 2 µM rCsaB in 50 mM Tris-HCl, pH .5, was incubated with varying concentrations of PEP (6 to 400 µM). The reactions were incubated at 37°C for 30 to 60 minutes. In a second set of reactions, 2 to 4 µM of rCsaB were incubated with ManNAc or *p*NP-β-ManNAc (0.5 to 1 mM) together with PEP (1:1) in the same buffer at 37°C for 60 minutes. The reactions were stopped by the addition of the malachite green reagent, and the samples were incubated at room temperature for 30 minutes. Blank reactions were prepared in a similar manner as the samples, but omitting the addition of rCsaB. The amount of free Pi released from PEP was determined by measuring the absorbance of the solutions at 595 nm using a TECAN Infinite F200 plate well reader (TECAN Austria, Salzburg, Austria).

3.16 Visualization and graphics

Figures 3 and 21 were produced using the BOXSHADE software to display multiple aligments (http://www.ch.embnet.org/software/BOX_form.html). Figure 15 was produced using the PROSITE sequence logo for domain profile (Crooks *et al.* 2004, Sigrist *et al.* 2013). Figures 4, 10, 11, 13, 14, 16, 18, 19, 23 and 27 were produced with PyMOL Molecular Graphics System, Version 1.8 Schrödinger, LLC, New York, NY USA. Electron density figures were produced with SetoMac, an unpublished development of SETOR (Evans 1993).

3.17 Data availability

The atomic coordinates and structures factors (**Table 4**) have been deposited in the Protein Data Bank under the accession codes 5JK1, 5JKY, 5JKZ, 5JL0, 5JL1, 5JL2, 5JL3 and 5TMJ.

4 Results

4.1 Analysis of purified SCWP from *P. alvei* CCM 2051^T

The structure of *P. alvei* CCM 2051^T SCWP has been characterized as [(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→, linked to muramic acid residues of the PG layer via a phosphate group (Schäffer *et al.* 2000b). Since the ketal pyruvyl modification is acid-labile, SCWP was isolated from wet cells by mechanical and enzymatic treatment to avoid hydrolysis of the pyruvyl groups. The PG-containing cell wall sacculi were further purified by size-exclusion chromatography on a Bio-Gel P10 column, eluting as a single peak (not shown) and recovered material was subjected to chemical analysis (Ries *et al.* 1997). The HPLC analysis revealed the following composition: ManNAc:GlcNAc:DAP:Glu:Ala in a molar ratio of 1:1.8:0.4:0.4:0.6. These data suggest that a considerable part of PG is still linked to the SCWP.

4.2 Purification of *P. alvei* CCM 2051^T recombinant, full-size S-layer protein SpaA

Overexpression of full-size, recombinant SpaA in *E. coli* BL21 star cells was induced by the addition of IPTG. Cells before and after addition of IPTG were harvested and their crude extracts were subjected to SDS-PAGE analysis. An additional protein band was observed in *E. coli* BL 21 cultures, indicative of the expression of the S-layer protein SpaA (**Figure 8A**). The apparent molecular mass of 108 kDa agrees with the theoretical value of *P. alvei* CCM 2051^T SpaA. The identity of His$_6$-tagged rSpaA was confirmed by Western immunoblotting using an anti-His tag antibody. (**Figure 8B**). After incubation cells were harvested and disrupted by sonication. The rSpaA protein was purified from clarified cell lysates by metal-affinity chromatography, and its purity was assessed by SDS-PAGE analysis (**Figure 8A**).

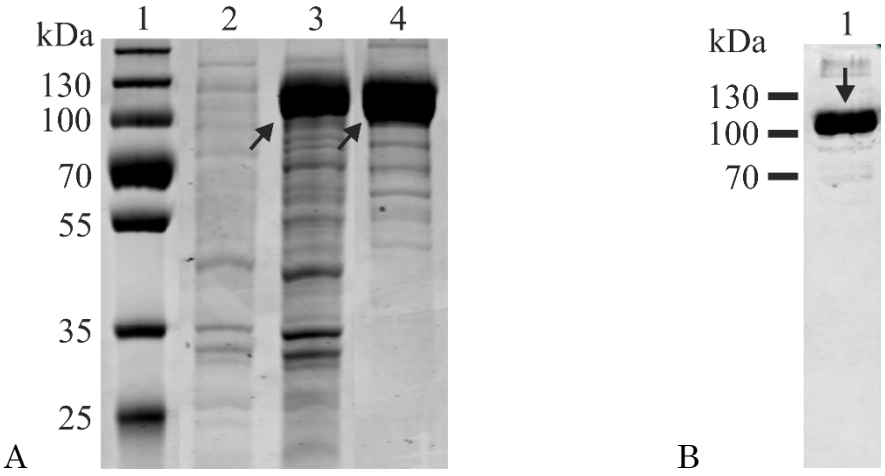


Figure 8. SDS-PAGE analysis of *P. alvei* CCM 2051^T recombinant, full-size S-layer protein SpaA. (A) SDS-PAGE (12%) gel of rSpaA after CBB staining. Lane 1, protein molecular mass ladder; Lane 2 and 3, cell crude extract of *E. coli* BL21 cells before and after induction of protein expression, respectively; Lane 4, rSpaA electrophoretic profile after metal-affinity chromatography. The arrow indicates the recombinant protein. (B) Western immunoblot analysis with fluorescence detection of rSpaA using an anti-His6 tag antibody. The arrow indicates the detected His-tagged protein.

4.3 *In vitro* binding studies of recombinant *P. alvei* CCM 2051^T SpaA self-assembly products and native SCWP, towards three-dimensional crystallization of full-size rSpaA

Crystallization of S-layer proteins into three-dimensional crystals has been hindered so far because of their intrinsic ability to spontaneously self-assemble into two-dimensional monomolecular arrays. Therefore, their solubilization in aqueous environments is only achieved by the use of chelating or chaotropic agents such as EDTA, urea or guanidine hydrochloride, respectively. Nonetheless, it has been described that SCWPs can influence or inhibit the self-assembly of S-layer proteins. For instance, native, purified SCWP inhibited the formation of self-assembly products in suspension of the S-layer protein of *G. stearothermophilus* PV72/p2 (Sára *et al.* 1998a). The S-layer protein from the Gram-positive bacterium *Aneurinibacillus thermoaerophilus* DSM 10155, was obtained in a water-soluble state after isolation from its natural source

because of S-layer protein-SCWP interactions. The self-assembling property of the protein was immediately restored when the SCWP was dissociated from the S-layer protein (Steindl *et al.* 2002).

Therefore, solubility of rSpaA self-assembly products by the influence of SCWP was investigated aiming at the subsequent three-dimensional crystallization of the rSpaA/SCWP complex. Different amounts of isolated *P. alvei* CCM 2051^T SCWP were added to rSpaA self-assembly products, and samples were incubated with constant stirring. Water-solubility of rSpaA self-assembly products seemed to be dependent on the amount of added SCWP (**Table 3**). When the ratio rSpaA:SCWP was 1:10 (w/w), solubilization of the self-assembly products took place almost instantly, with an overall yield of solubilized S-layer protein of 72%. The sample containing rSpaA:SCWP in a ratio of 1:5 (w/w) showed similar solubility, and a yield of solubilization of 78% was observed. On the other hand, with a ratio of 1:1 (w/w) solubilization required 20 to 40 minutes to take place. Besides, the solubility of self-assembly products decreased significantly, a yield of 54% was observed in these samples. Still, with rSpaA:SCWP ratios of 1:0.2 and 1:0.1 (w/w) solubility occurred to some extent. Although no significant difference was observed with this amount of SCWP, yields of solubilization of 50% and 42% were obtained, respectively.

The molar ratios were calculated on the basis of the molar mass of rSpaA of 108kDa and an average molar mass of isolated SCWP of ~7000 Da. The average mass of the SCWP comes from the uncertainty of the length variation from the different repeats of the disaccharide units forming the glycan chain, the still attached PG moieties at the reducing end and if some pyruvyl groups were hydrolyzed. The water-solubility experiments were performed with this heterogeneous SCWP. The approximate molar ratios are shown in **Table 3**. This water-soluble rSpaA/SCWP complex was used to attempt a three-dimensional crystallization of full-size recombinant SpaA. However, three dimensional crystals of the rSpaA/SCWP complex did not diffracted sufficiently to collect data and crystals were extremely difficult to reproduce (personal

communication with Dr. Stephen V. Evans and Ryan L. Blackler, data not shown).

Table 3. Influence of the SCWP on water-solubility of recombinant SpaA self-assembly products of *P. alvei* CCM 2051^T.

Isolated SCWP (mg)[a]	rSpaA:SCWP ratio (wt/wt)	rSpaA:SCWP Molar ratio[b]	(%)[c]
5	1:10	1:150	72
2.5	1:5	1:75	78
0.5	1:1	1:15	54
0.1	1:0.2	1:3	50
0.05	1:0.1	1:1.5	42

[a] Different amounts of isolated SCWP were added to rSpaA self-assembly products in suspension.
[b] The molar ratio was calculated with an average molecular mass of *P. alvei* CCM 2051^T isolated SCWP.
[c] The yield of solubilization was calculated with a control sample without SCWP, in which the self-assembly products were dissolved with 4 M urea, representing 100% of solubilization.

4.4 Recombinant expression, purification and secondary structure determination of a truncated His$_6$-tagged SpaA$_{SLH}$ variant

Since three-dimensional crystallization of rSpaA/SCWP complex was un-successful, a truncated variant of *P. alvei* CCM 2051^T S-layer protein SpaA comprising its three N-terminal SLH domains was created. The truncated variant, termed SpaA$_{SLH}$ is based on the published structure of the SLH domains from the *B. anthracis* S-layer protein Sap (Kern *et al.* 2011), and was produced as a C-terminal His$_6$-tagged protein in *E. coli* BL21 (DE3-Star) (**Figure 9A**).

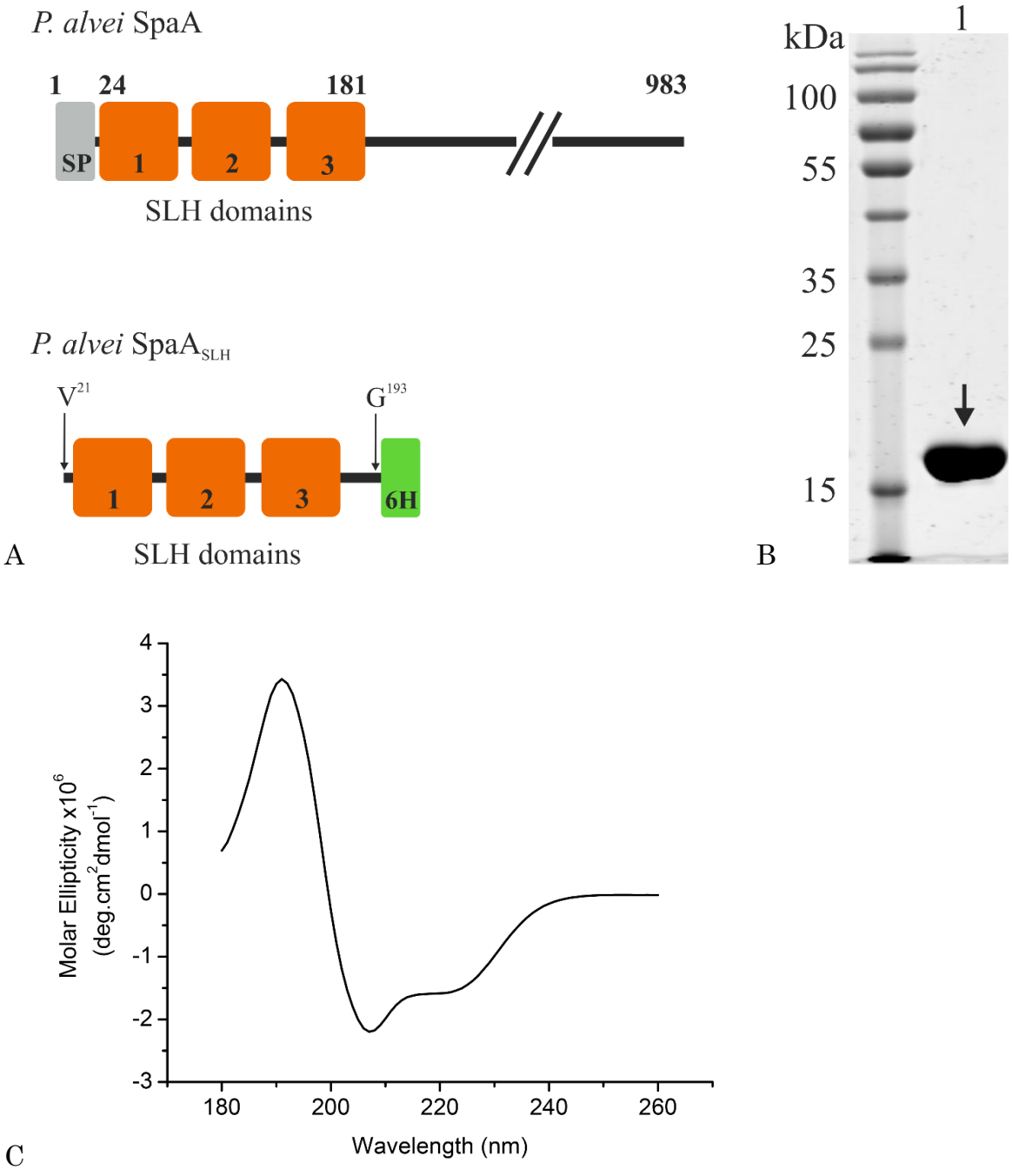


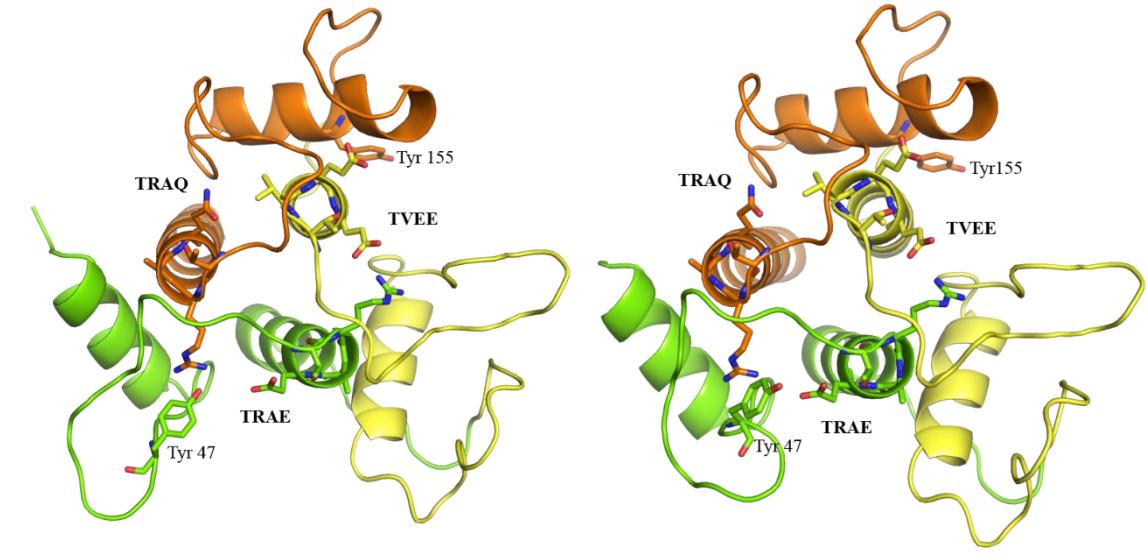
Figure 9. Schematic depiction of the primary structure of mature *P. alvei* CCM 2051^T SpaA and SpaA$_{SLH}$ and SDS-PAGE analysis of recombinant, purified SpaA$_{SLH}$. (A) SpaA including its signal peptide (SP), the SLH domains and the self-assembling domain in comparison to SpaA$_{SLH}$ used in this study. (B) Recombinant His$_6$-tagged SpaA$_{SLH}$ was purified by metal-affinity and size exclusion chromotagraphy and analyzed by an SDS-PAGE gel upon CBB staining. The arrow indicates the purified recombinant protein. (C) Circular dichroism spectrum of purified *P. alvei* CCM 2051^T SpaA$_{SLH}$.

The protein was recovered from the cytoplasmic fraction in soluble form and purified to homogeneity from the cleared cell lysate on a Ni-NTA-agarose column at a concentration of 250 mM imidazole, followed by Superdex 75 size exclusion chromatography in 20 mM HEPES buffer. Purity of recombinant $SpaA_{SLH}$ was confirmed by SDS-PAGE analysis upon staining with CBB (**Figure 9B**).

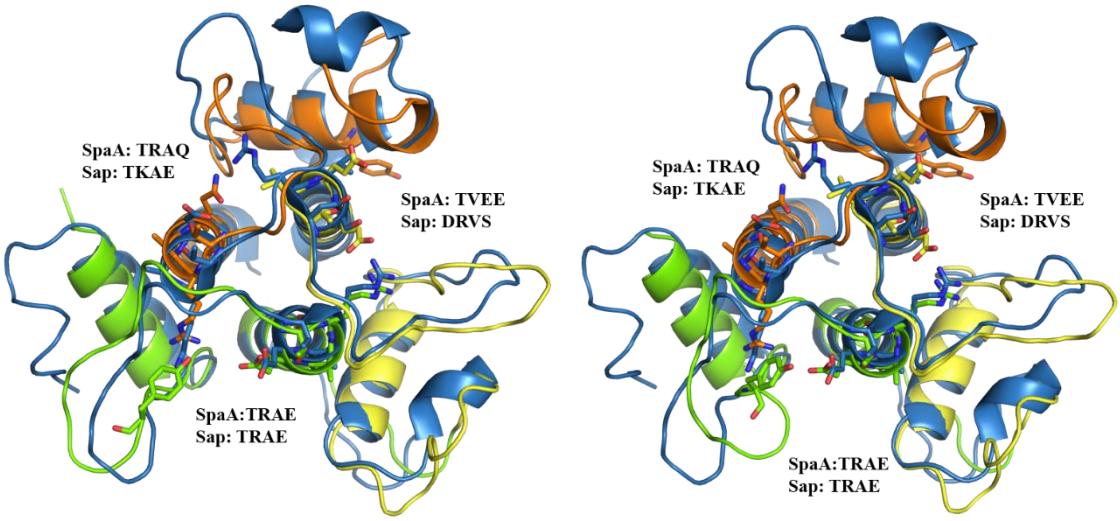
In addition, the secondary structure of the purified $SpaA_{SLH}$ was determined by ECD. The far-UV spectrum of $SpaA_{SLH}$ measured between 180 and 260 nm exhibited a maximum at 190 nm and minima between 208 and 222 nm (**Figure 9C**). The ECD spectrum showed a typical pattern for α-helical proteins. This result is in agreement with the overall helical conformation described for the truncated S-layer proteins Sap_{SLH} of *B. anthracis* (Kern *et al.* 2011) and $SbsB_{SLH}$ of *G. stearothermophilus* PV72/p2 (Rünzler *et al.* 2004).

4.5 Crystal structure of *P. alvei* CCM 2051^T $SpaA_{SLH}$

$SpaA_{SLH}$ was initially crystallized by the Evans laboratory in a space group $P3_221$, and two molecules were found in the asymmetric unit (AU) (Blackler RB, López-Guzmán A, Messner P, Schäffer C and Evans SV, manuscript in preparation). The structure was solved to a resolution of 2.50 Å by the single-wavelength anomalous diffraction (SAD) method from KI-soaked crystals. Then, a native crystal was solved to a resolution of 1.80 Å by molecular replacement using the initial solved SAD structure for data completeness. Refinement statistics are shown in **Table 4A**. The full polypeptide chain of both molecules in the AU displayed excellent stereochemistry and electron density, except for the N-terminal residues 22-28, which were not observed in either molecules in the AU, and the residues 48-52, which were disordered in molecule B. The complete structure of both molecules is similar to the three-pronged spindle fold described for the three consecutive SLH domains of *B. anthracis* surface array protein Sap (**Figure 10**) (Kern *et al.* 2011).



A



B

Figure 10. (A) Crystal structure of SpaA$_{SLH}$ three-pronged spindle trimer shown from the N-terminus. SLH_1 green, SLH_2 yellow and SLH_3 orange. Side chains of the conserved four-residue motifs are shown as sticks, likewise the known SpaA *O*-glycosylation sites Tyr^{47} and Tyr^{155}. (B) Stereo image of superposed *B. anthracis* Sap$_{SLH}$ (blue) and *P. alvei* CCM 2051^T SpaA$_{SLH}$ crystal structures.

Each SLH domain contributes with one α-helix to a triple-helix core, whereas a second α-helix rotated by approximately 90° from the center, forms a prong, adopting a three-pronged spindle fold (**Figure 10A**). The two molecules in the AU superimpose almost perfectly when omitting residues 47-52 from the connecting loop between the two helices of domain SLH_1 that are observed in two different conformations. Molecule A displayed an excellent electron density for these residues, whereas molecule B displayed fragmented electron density which refined to higher general temperature factors. Moreover, *P. alvei* CCM 2051^T $SpaA_{SLH}$ superimposes well with *B. anthracis* Sap_{SLH} structure at all helices, except for the connecting loops, where notable differences are observed, particularly in the connecting loop of SLH_1, where the glycosylation site Tyr^{46} of *P. alvei* CCM 2051^T SpaA is located (**Figure 10B**). In this context it is important to consider that the *B. anthracis* Sap protein is not glycosylated. Besides, a slightly different spatial orientation of the conserved residues inside the predicted binding grooves in *B. anthracis* is observed (**Figure 10 B**). Both structures have three prongs and three grooves (named here G1, G2 and G3), which are the potential binding pockets for the *P. alvei* CCM 2051^T SCWP.

4.6 Presentation of the conserved motifs and the *O*-glycosylation sites of the S-layer protein SpaA

The TRAE, TVEE and the TRAQ motifs are located at the N-termini of the core helices adjacent to each groove. The side chain of the second residue of each motif, that is Arg^{61}, Val^{125} and Arg^{177}, thread under the connecting loop of the nearby SLH domain into the neighboring groove (**Figure 10A**). An exception is Val^{125} of TVEE motif, which short side chain does not protrude into the nearby G3. Nevertheless, the side chain of Arg^{61} and Arg^{177} contribute to a positively charged surface into G2 and G1, respectively (**Figure 11A** to **C**). The side chains of the fourth residue of each motif, that is Glu^{63}, Glu^{127} and Gln^{179}, lie adjacent to their parent helix, directly next to the protruding arginine

residues 61 and 177 (**Figure 10A**). Electron density for a sulfate ion was observed in G2 of both molecules in the AU.

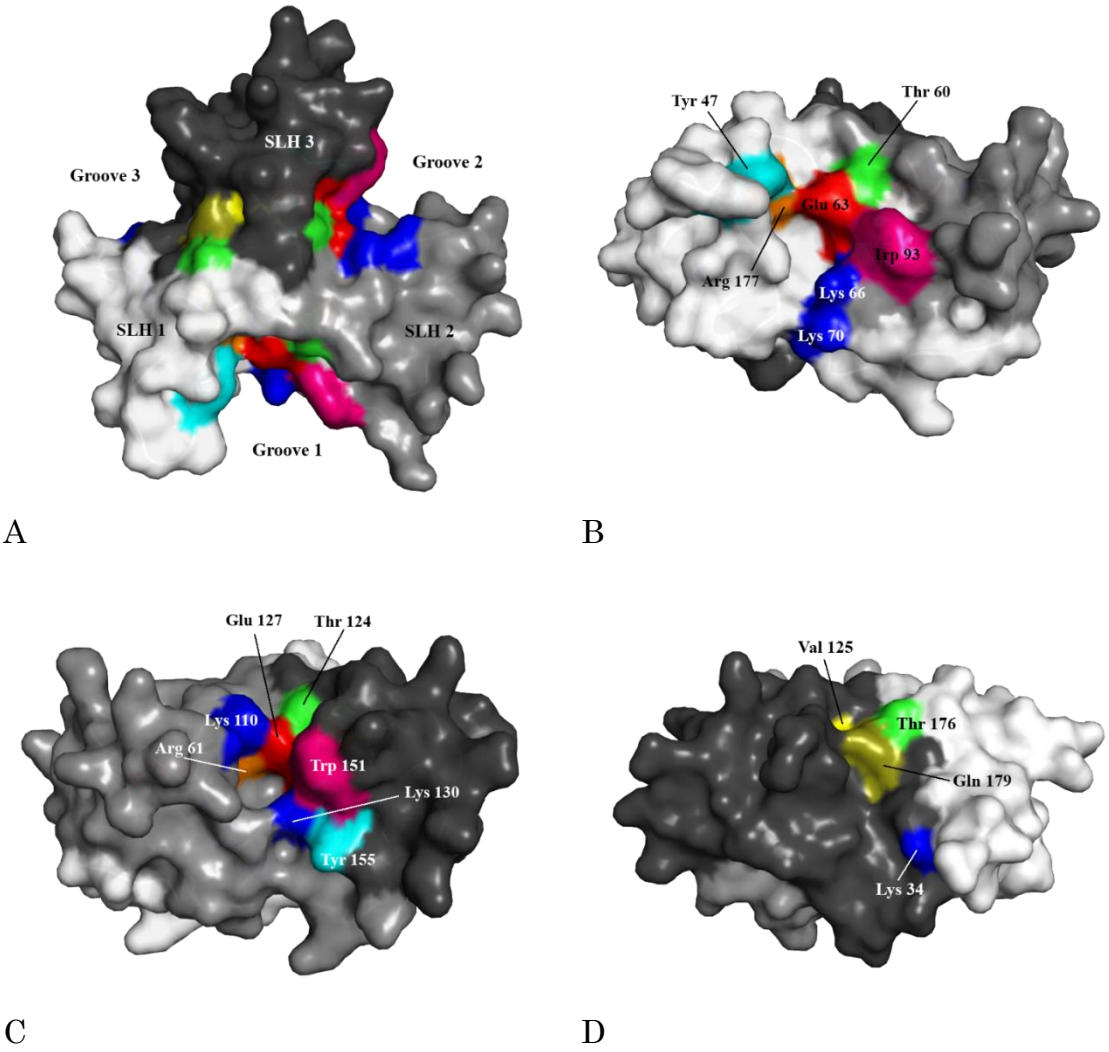


Figure 11. Surface representation of SpaA$_{SLH}$ with each domain depicted individually to show their groove architecture: SLH$_1$ white; SLH$_2$ gray; SLH$_3$ dark gray. (A) N-terminus view. (B) Interface at groove 1. (C) Interface at groove 2. (D) Interface at groove 3. Residues of interest are colored: Arg, orange; Gln, olive green; Glu, red; Lys, blue; Thr, green; Trp, pink; Tyr, cyan; Val, yellow.

The two known glycosylation sites of SpaA, Tyr^{47} and Tyr^{155}, are located adjacent to the grooves G1 and G2, respectively (**Figure 11A** to **C**). Interestingly, Tyr^{47} is within the SLH_1 loop that was observed in two different conformations (**Figure 10A** and **Figure 11A**). Therefore, the presence of large oligosaccharides next to these pockets would influence their structure or binding properties.

4.7 Binding studies of $SpaA_{SLH}$ with *P. alvei* CCM 2051^T synthetic SCWP ligands

Binding and affinity of $SpaA_{SLH}$ to *P. alvei* CCM 2051^T synthetic SCWP ligands 4,6-Pyr-β-D-ManNAc-*O*-Me (monosaccharide) and β-D-GlcNAc-(1→3)-4,6-Pyr-β-D-ManNAc-*O*-Me (disaccharide) were analyzed by isothermal titration microcalorimetry (ITC).

The average stoichiometry of binding between 4,6-Pyr-β-D-ManNAc-*O*-Me and $SpaA_{SLH}$ is 0.91 with an equilibrium association constant (K_A) of 3.48 x 10^7 M^{-1} for the protein-ligand complex. The binding isotherm displayed that $SpaA_{SLH}$ has a high affinity for the synthetic monosaccharide, since the slope of the saturation curve of the protein is very steep, reaching a plateau and showing a binding stoichiometry of 1:1 (**Figure 12A**). This result correlates with the determined binding stoichiometry between *G. stearothermophilus* PV72/p2 $SbsB_{SLH}$ and its SCWP (Rünzler *et al.* 2004). Upon binding of the synthetic monosaccharide, an entropic penalty of $-T\Delta S$ = 45.74 kJ/mol was observed, which, however, was compensated by an enthalpy gain of ΔH = -87.85 kJ/mol. This is revealed by a strong exothermic binding reaction (**Figure 12A**). The protein-ligand interaction exhibited a free energy of binding (ΔG) of -42.10 kJ/mol, indicative of a favorable process dominated by an enthalpic contribution.

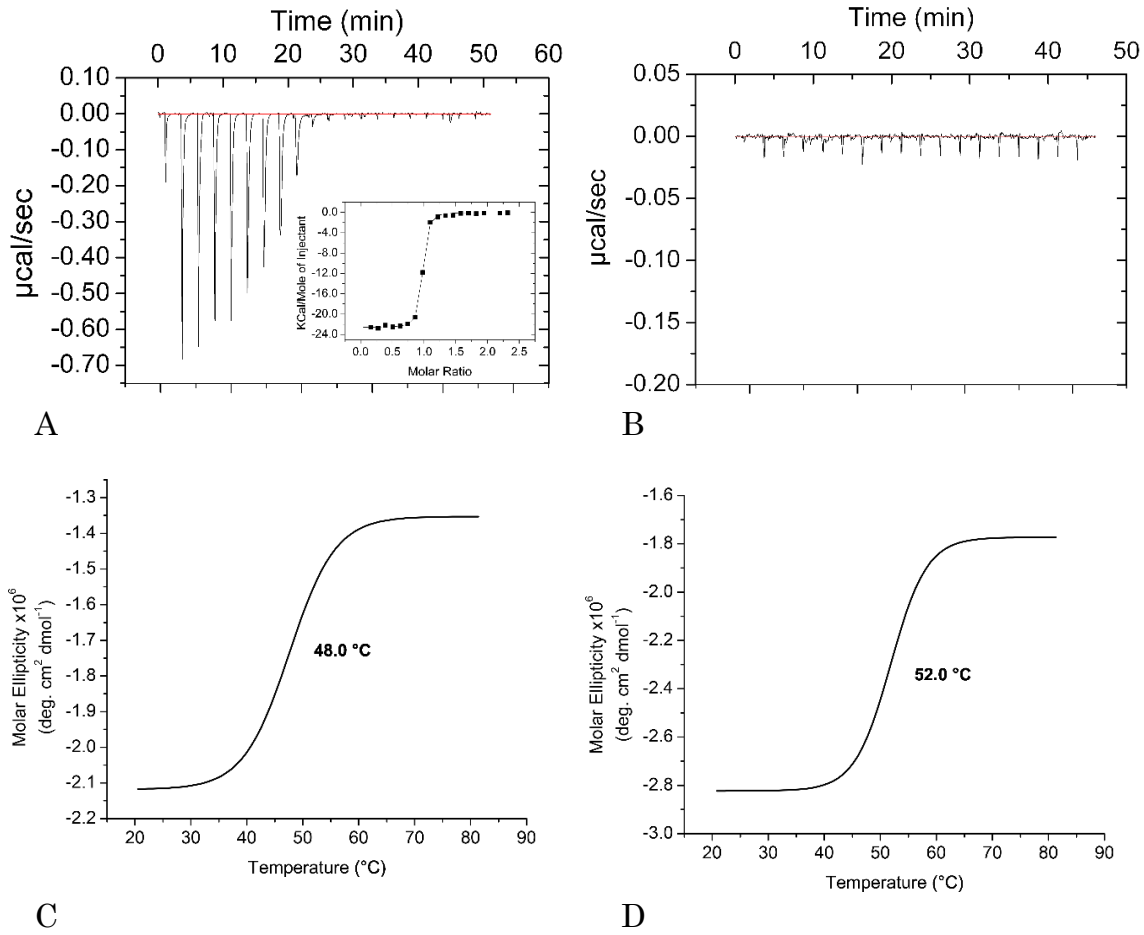


Figure 12. Binding performance of synthetic SCWP ligands to SpaA$_{SLH}$. Calorimetric titration of SpaA$_{SLH}$ (A) with synthetic monosaccharide and (B) with synthetic disaccharide ligands. The panels show the incremental heat release upon each 1-µL injection of ligand into protein solution. The insets show the binding isotherms fitted to a 1:1 binding model. (C) Far-UV thermal denaturation spectrum at 208 nm of unliganded SpaA$_{SLH}$ and (D) for complexed SpaA$_{SLH}$, after heating from 20°C to 90°C. Transition temperatures, Tm, are shown in the plots.

In contrast, upon titration of SpaA$_{SLH}$ with the synthetic disaccharide β-D-GlcNAc-(1→3)-4,6-Pyr-β-D-ManNAc-*O*-Me, a very small amount of heat was recorded after each injection (**Figure 12B**), corresponding only to heat of dilution of disaccharide into protein solution.

Furthermore, a notable difference in thermal stability was observed between unliganded and monosaccharide-complexed SpaA$_{SLH}$. Unfolding of unliganded SpaA$_{SLH}$ took place between 34.5°C and 57.6°C, with T_m= 48.0°C (**Figure 12C**),

whereas unfolding of complexed $SpaA_{SLH}$ occurred between 39.7°C and 61.5°C, with T_m= 52.0°C (**Figure 12D**), suggesting that $SpaA_{SLH}$ would be stabilized upon its interaction with the 4,6-Pyr-β-D-ManNAc-*O*-Me ligand. A similar effect has been observed for the S-layer protein SbsC of *G. stearothermophilus* ATCC 12980, where a truncated variant comprising the SCWP binding region experiences a substantial stabilization upon binding to its isolated native SCWP (Pavkov *et al.* 2008).

In either case, the unfolding process was irreversible, as evident from the failure of the proteins to adopt their native state conformation after cooling down to 20°C. However, a significant fraction of residual secondary structure was still observed (not shown).

4.8 Soaking and co-crystallization of 4,6-Pyr-β-D-ManNAc-*O*-Me with $SpaA_{SLH}$

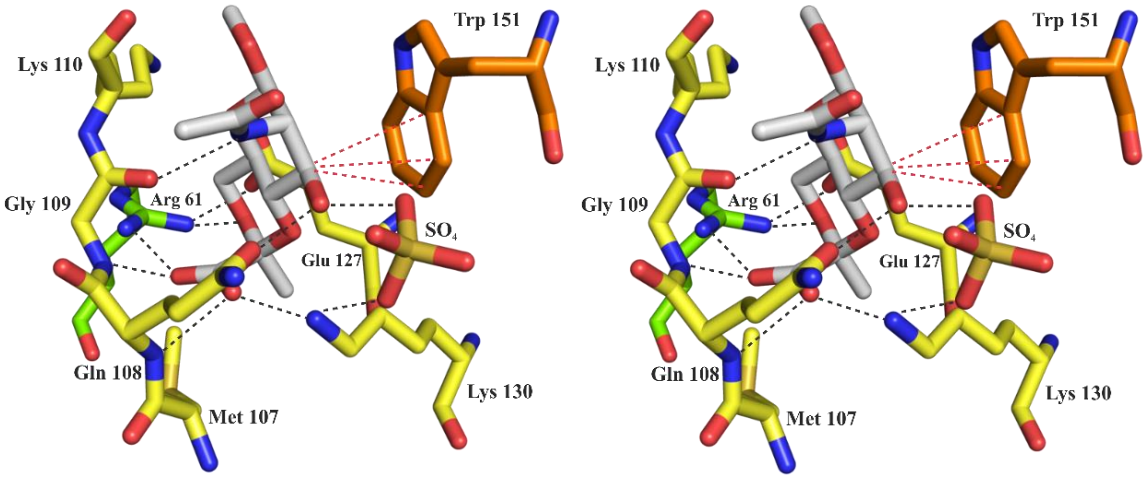
The complex structure of $SpaA_{SLH}$ and the synthetic monosaccharide 4,6-Pyr-β-D-ManNAc-*O*-Me was solved through both soaking (to 2.10 Å) and co-crystallization (to 2.50 Å; not shown) (Blackler RB, López-Guzmán A, Martinz G, Kosma P, Messner P, Schäffer C and Evans SV, manuscript in preparation). Initial soaking and co-crystallized structures with the monosaccharide were obtained in the same space group as the unliganded structure, $P3_221$, and displayed ligand binding only in G2, as well as the same two conformations and disorder in the 47-52 residue region next to G1. The effect of this disorder and crystal packing forces on potential ligand binding in G1 was investigated by solving additional co-crystal structures from different crystallization conditions in space groups $P2_12_12_1$ (2.77 Å resolution), P1 (2.00 Å resolution), and C2 (2.40 Å resolution) (**Table 4A**). All structures displayed the same regions of disorder and the two conformations of residues 47-52, and excellent ligand density in the G2 pocket. Only the C2 structure displayed a density consistent with a weakly bound monosaccharide ligand in the G1 pocket, and in only one

of two molecules in the AU. The remainder of the structures and ligand binding in G2 was identical in all cases.

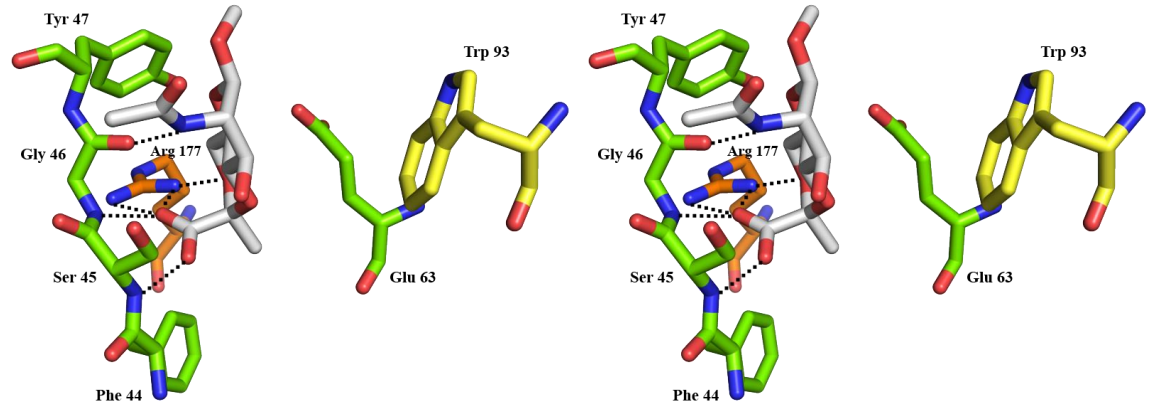
The synthetic ligand 4,6-Pyr-β-D-ManNAc-*O*-Me was observed bound in G2 in all complex structures, in a narrow pocket formed by residues from all three SLH domains: Arg^{61} from SLH_1; Met^{107}, Gln^{108}, Gly^{109}, Lys^{110}, Glu^{127} and Lys^{130} from SLH_2; and Trp^{151} from SLH_3 (**Figures 13**).

The pyruvate moiety is bound deep in the pocket and makes salt bridge interactions from carboxyl oxygens of the pyruvyl group to Arg^{61} Nη2 and to Lys^{130} Nζ, and hydrogen bonds to Gln^{108} N and Gly^{109} N. Arg^{61} makes a bidentate interaction with the sugar, also forming a bond to O6 of ManNAc from Nη1. The hydrophobic face of the ManNAc ring is stacked against Trp^{151}, centered about ManNAc C3. Gln^{108} makes a hydrogen bond to O3 of ManNAc, and both of these interact with a nearby sulfate ion that is also coordinated by Lys^{130}. Lastly, the ManNAc *N*-acetyl nitrogen hydrogen bonds to Gly^{109} O (**Figure 13A**).

The weakly bound 4,6-Pyr-β-D-ManNAc-*O*-Me in G1 of the C2 structure displays similar interactions (**Figure 13B**). The pyruvate group is coordinated by a salt bridge from Arg^{177} and by hydrogen bonds from Ser^{45} N and Gly^{46} N, equivalent to Arg^{61}, Gln^{108} and Gly^{109} described above for binding in G2. Gly^{46} O also makes a hydrogen bond to the ManNAc *N*-acetyl nitrogen. In this case, no stacking is observed from Trp^{93}, as observed for Trp^{151} in G2.



A



B

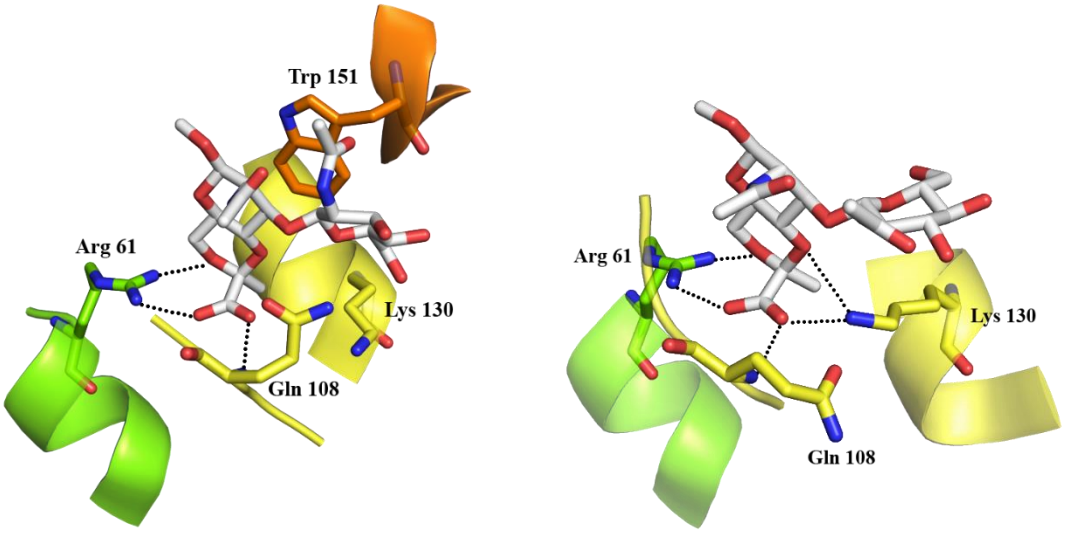
Figure 13. Stereo images of monosaccharide ligand bound in grooves G1 and G2 of SpaA$_{SLH}$. Stereo image of 4,6-Pyr-β-D-ManNAc-*O*-Me bound in SpaA$_{SLH}$ G2 (A) and G1 (B). SpaA$_{SLH}$ backbone and amino acid side chains are colored according to SLH domain (SLH_1 green, SLH_2 yellow and SLH_3 orange). Hydrogen bonds and salt bridges are depicted as black dashed lines. Pi-stacking is shown as red dashed lines.

4.9 Co-crystallization of β-D-GlcNAc-(1→3)-4,6-Pyr-β-D-ManNAc-*O*-Me with $SpaA_{SLH}$

The complex structure of $SpaA_{SLH}$ with β-D-GlcNAc-(1→3)-4,6-Pyr-β-D-ManNAc-*O*-Me was solved through co-crystallization to 2.05 Å in space group $P3_221$ with two molecules in the AU (Blackler RB, López-Guzmán A, Martinz G, Kosma P, Messner P, Schäffer C and Evans SV, manuscript in preparation). Disordered regions again include residues 22-28 (not observed) and residues 47-52 with the same conformations as observed for molecule A, and completely disordered in molecule B. In addition, residues 139-152 beside G2 were completely disordered in molecule B.

The synthetic disaccharide β-D-GlcNAc-(1→3)-4,6-Pyr-β-D-ManNAc-*O*-Me was observed bound in the G2 pocket in both molecules in the AU, with the 4,6-Pyr-β-D-ManNAc-*O*-Me moiety bound in each case as described above. However, minor differences were observed: The side chain of Gln^{108} does not interact with ManNAc O3 in this case, but is rotated away from the ligand and hydrogen bonds to the opposite side of Lys^{130} in one molecule and is partially disordered in the other (**Figure 14A, B**). In addition, electron density for the GlcNAc moiety is fragmented in both molecules and displayed two different conformations that do not suggest any additional hydrogen bonds to the protein. The adopted conformation of GlcNAc in molecule B collides with the residue region 139-152, where Trp^{151} is located, thus its stacking interactions to the hydrophobic region of ManNAc ring are not observed (**Figure 14B**).

The bound SCWP ligands are oriented such that the non-reducing ends leads towards the C-terminus of $SpaA_{SLH}$, which would continue into the self-assembling domain of the protein. Therefore, the non-reducing end points away the PG, consistent with the directionality of SCWP extending from the cell wall in a natural context.



A B

Figure 14. Stereo images of the disaccharide ligand bound in groove G2 of SpaA$_{SLH}$. Stereo image of β-D-GlcNAc-(1→3)-4,6-Pyr-β-D-ManNAc-*O*-Me bound with two different conformations in SpaA$_{SLH}$ G2. (A) Disaccharide bound in molecule A, amino acid side chains are colored according to SLH domain; SLH$_1$ green, SLH$_2$ yellow and SLH$_3$ orange. (B) Disaccharide bound in molecule B, Trp^{151} in this molecule is disordered, including residues 139-152.

Table 4A. Data collection and refinement statistics[a]

Protein	**SpaA$_{SLH}$**	**SpaA$_{SLH}$**	**SpaA$_{SLH}$**	**SpaA$_{SLH}$**	**SpaA$_{SLH}$**
Ligand	None	Monosacch.[b]	Monosacch.[b]	Monosacch.[b]	Disacch.[b]
PDB code	5JK1	5JKY	5JKZ	5JL0	5JL1
Data collection					
Space group	$P3_221$	$P3_221$	P1	C2	$P3_221$
Resolution (Å)	40.0-1.90	40.0-2.25	40.0-2.00	40.0-2.15	40.0-2.15
Cell dimensions					
a (Å)	72.20	72.00	46.22	138.57	72.26
b (Å)	72.20	72.00	72.24	45.94	72.26
c (Å)	126.63	125.56	72.31	94.92	126.14
α (°)	90.00	90.00	86.73	90.00	90.00
β (°)	90.00	90.00	71.29	131.56	90.00
γ (°)	120.00	120.00	71.40	90.00	120.00
R_{sym}	0.072 (0.521)	0.047 (0.653)	0.044 (0.350)	0.046 (0.446)	0.073 (0.545)
R_{pim}	0.020 (0.135)	0.016 (0.222)	0.035 (0.273)	0.025 (0.265)	0.021 (0.152)
$CC_{1/2}$	(0.953)	(0.871)	(0.934)	(0.947)	(0.970)
I/σ(I)	33.3 (5.6)	44.8 (3.3)	19.9 (2.5)	27.1 (2.0)	34.3 (5.8)
Completeness (%)	99.6 (99.5)	100 (100)	95.0 (79.8)	97.1 (96.0)	99.9 (100)
Redundancy	14.7 (15.5)	9.4 (9.5)	2.4 (2.2)	4.3 (3.4)	13.2 (13.6)
Unique reflections	30690	18478	53737	23687	21339
Refinement					
Resolution	40.0-1.90	40.0-2.25	40.0-2.00	40.0-2.15	40.0-2.15
No. reflections	29054	17170	39929	17214	20168
R_{work} (%)	16.2	21.2	18.5	19.7	21.5
R_{free} (%)	21.4	25.4	23.5	23.7	25.5
No. atoms					
Protein	2544	2540	5096	2535	2393
Ligand	8 (Cl, SO_4)	42 (ligand[b])	84 (ligand[b])	73 (SO_4, ligand[b])	70 (ligand[b])
Water	212	90	350	126	83
B factors					
Protein	26.0	43.4	33.8	42.7	36.1
Ligand	46.1	42.6	24.7	42.5	53.0
Water	29.4	34.9	31.7	33.7	37.9
r.m.s. bonds (Å)	0.019	0.016	0.016	0.015	0.015
r.m.s. angles (°)	1.852	1.678	1.605	1.604	1.581

[a] Unpublished data (Blackler RB, López-Guzmán A, Martinz G, Kosma P, Messner P, Schäffer C and Evans SV, manucript in revision).

[b] Monosaccharide = 4,6-Pyr-β-D-ManNAc-*O*-Me; Disaccharide β-D-GlcNAc-(1→3)-4,6-Pyr-β-D-ManNAc-*O*-Me

[c] Values in parenthesis represent the highest resolution shell.

Table 4B. Data collection and refinement statistics[a]

Protein	**SpaA$_{SLH}$/G109A**	**SpaA$_{SLH}$/G109A**	**SpaA$_{SLH}$/G46A/G109A**
Ligand	None	Monosaccharide[b]	Monosaccharide[b]
PDB code	5JL2	5JL3	5TMJ
Data collection			
Space group	$P2_12_12_1$	C2	$P2_12_12_1$
Resolution (Å)	50.0-1.15	50.0-1.53	30.0-1.24
Cell dimensions			
a (Å)	34.35	93.38	32.54
b (Å)	65.67	36.25	65.83
c (Å)	73.17	56.90	71.92
α (°)	90.00	90.00	90.00
β (°)	90.00	103.74	90.00
γ (°)	90.00	90.00	90.00
R_{sym}	0.043 (0.745)	0.044 (0.155)	0.040 (0.693)
R_{pim}	0.018 (0.303)	0.021 (0.097)	0.020 (0.473)
$CC_{1/2}$	(0.942)	(0.962)	(0.660)
I/σ(I)	38.4 (2.5)	28.6 (7.4)	27.2 (1.7)
Completeness (%)	96.4 (93.5)	98.6 (84.8)	99.2 (97.3)
Redundancy	7.3 (6.9)	4.3 (3.4)	4.1 (2.8)
Unique reflections	57497	27852	44339
Refinement			
Resolution	50.0-1.15	50.0-1.53	30.0-1.24
No. reflections	51418	26441	42050
R_{work} (%)	13.1	11.3	14.7
R_{free} (%)	15.9	15.1	17.9
No. atoms			
Protein	1337	1265	1284
Ligand	1 (Cl)	37 (MPD, ligand[b])	21 (ligand[b])
Water	185	176	166
B factors			
Protein	19.4	15.7	17.5
Ligand	26.4	33.0	17.7
Water	30.7	29.7	29.2
r.m.s. bonds (Å)	0.024	0.021	0.012
r.m.s. angles (°)	2.158	2.048	1.481

[a] Unpublished data (Blackler RB, López-Guzmán A, Martinz G, Gagnon SML, Haji-Ghassemi O, Kosma P, Messner P, Schäffer C and Evans SV, un manucript in revision).

[b] Monosaccharide = 4,6-Pyr-β-D-ManNAc-*O*-Me; Disaccharide β-D-GlcNAc-(1→3)-4,6-Pyr-β-D-ManNAc-*O*-Me

[c] Values in parenthesis represent the highest resolution shell.

4.10 Creation of *P. alvei* CCM 2051^T SpaA$_{SLH}$/G109A and SpaA$_{SLH}$/G46A/G109A mutants

In addition to the TRAE motif, a second conserved short sequence has also been identified within the SLH domains. This motif, denoted GIIxG, is higly conserved in different S-layer proteins and glycoside hydrolases of Gram-positive bacteria (Engelhardt and Peters 1998, Zona and Janeček 2005). The GIIxG motif, with the second glycine residue being the most conserved (**Figure 15**), has been scarcely investigated and no function has been attributed yet. *P. alvei* CCM 2051^T SpaA contains this sequence in SLH$_1$ and SLH$_2$, with the sequences G^{41}VFSG^{46} and G^{105}LMQG^{109}, respectively.

The crystal structures have shown these motifs within the loops forming one side of the grooves G1 and G2 (**Figure 13**). In G2, the Gln^{108}-Gly^{109} peptide bond flips upon ligand binding, so that the pyruvyl carboxyl O of 4,6-Pyr-β-D-ManNAc-*O*-Me hydrogen bonds to the N of Gly^{109}, replacing a hydrogen bond between the carbonyl O of Gln^{108} and the Nη2 of Arg^{61} (**Figure 16A**).

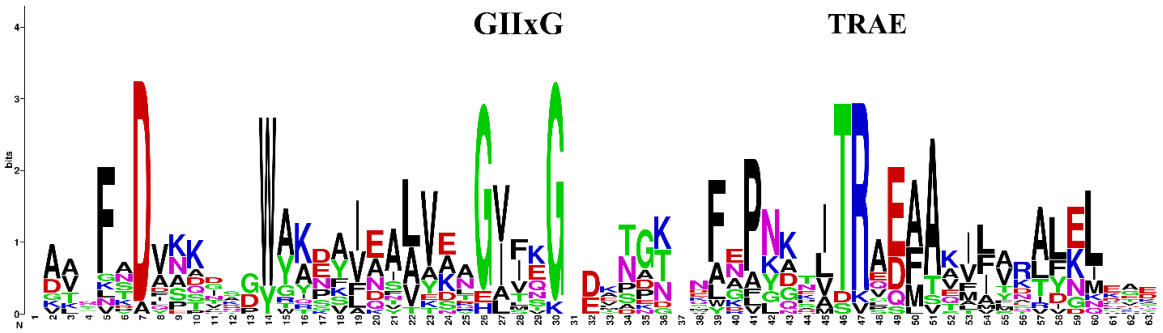
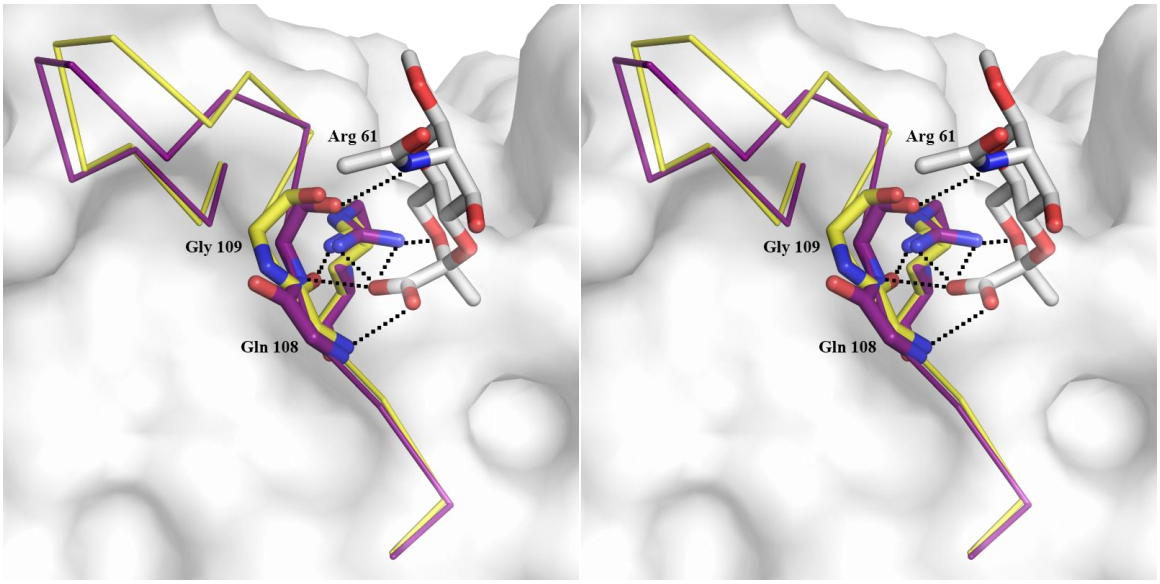
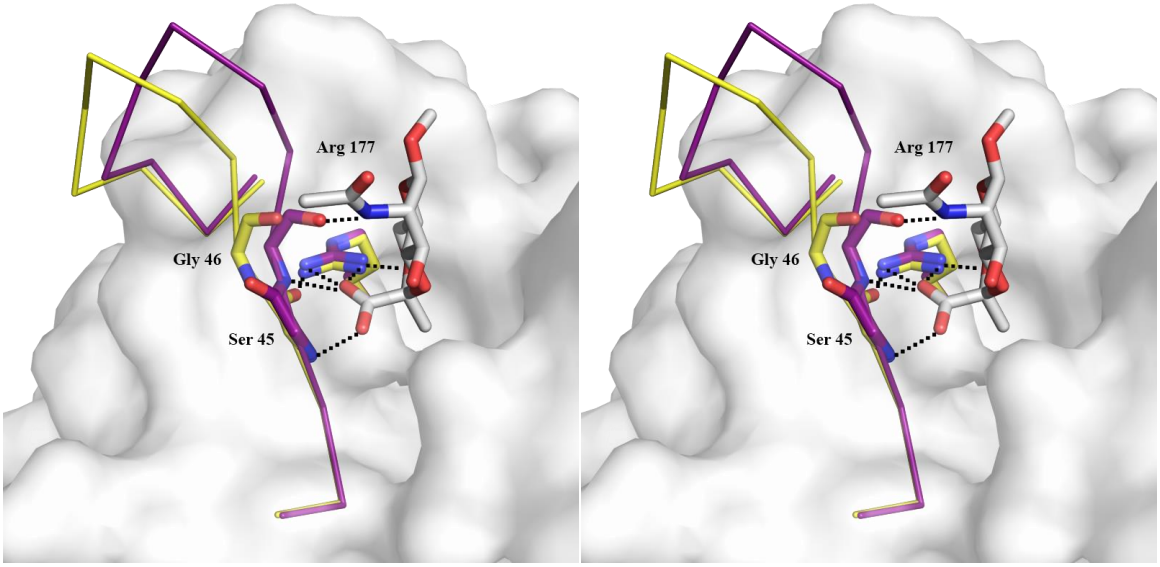


Figure 15. Sequence logo of SLH domain profile from PROSITE (Crooks *et al.* 2004). The overall height of the stack indicates the sequence conservation at that position, whereas the height of symbols within the stack indicates the relative frequency of each residue at that position. The GIIxG and TRAE motifs are shown above the alignment.



A



B

Figure 16. Comparison of the conformation of the GIIxG motif before and after ligand binding. A comparison of the conformations of SpaA$_{SLH}$ residues (A) 105-119 in G2 and (B) 42-54 in G1 between free protein (yellow) and protein in complex with the monosaccharide ligand (purple). Main-chain atoms are shown for Ser^{45}, Gly^{46}, Gln^{108} and Gly^{109}, whereas side chains for Arg^{61} and Arg^{177}, to illustrate the peptide bond flip and hydrogen bonding upon ligand binding. The rest of the protein is shown as a white surface.

The same flip is observed in groove G1 upon weak ligand binding, in this case being involved Ser^{45} and Gly^{46}. Both G^{46} and G^{109} stand for the second, highly conserved glycine residue of the GIIxG motif (**Figure 16B**). In SpaA, the only SLH domain to lack this Gly is the SLH_3 (GLFSK), which has not displayed ligand binding so far. Therefore, mutation of G^{46} and G^{109} to alanine was performed to elucidate the biological role of this conserved motif.

4.11 Binding studies of $SpaA_{SLH}$/G109A and $SpaA_{SLH}$/G46A/G109A mutants with 4,6-Pyr-β-D-ManNAc-*O*-Me ligand

To investigate the effect of the Gly^{46} and Gly^{109} mutations on SCWP binding, *P. alvei* CCM 2051^T $SpaA_{SLH}$/G109A and $SpaA_{SLH}$/G46A/G109A mutants were created. Both mutant proteins were produced and purified as described above for the $SpaA_{SLH}$ protein, and were titrated with the 4,6-Pyr-β-D-ManNAc-*O*-Me ligand. The binding of the monosaccharide ligand to the single and double mutants was analyzed by ITC.

$SpaA_{SLH}$/G109A showed a K_A of 4.48 x 10^6 M^{-1}, with an average stoichiometry of 0.92 for the monosaccharide ligand, indicating again one ligand molecule per protein molecule. The binding isotherm of the G109A mutant also displayed a ligand to protein molar ratio of 1:1, with a strong affinity for the ligand (**Figure 17A**), although the K_A of $SpaA_{SLH}$/G109A for 4,6-Pyr-β-D-ManNAc-*O*-Me is approximately 7-fold smaller than that of $SpaA_{SLH}$. A comparison of the thermodynamic parameters between $SpaA_{SLH}$ and $SpaA_{SLH}$/G109A is summarized in **Table 5**. Likewise, the enthalpy (ΔH) and the entropy ($-T\Delta S$) of binding, -50.64 kJ/mol and 13.33 kJ/mol respectively, are smaller for the G109A mutant. Besides, there is no a significant differences in the free energy (ΔG) of binding for the monosaccharide between $SpaA_{SLH}$ (-42.10 kJ/mol) and $SpaA_{SLH}$/G109A (-37.31 kJ/mol), meaning that complex formation is a favorable and spontaneous process in both cases. Upon binding of 4,6-Pyr-β-D-ManNAc-*O*-Me to either $SpaA_{SLH}$ or $SpaA_{SLH}$/G109A, the entropy of the system decreases, meaning that an entropic penalty occurs, which is compensated by

an enthalpy gain, indicative of an enthalpy driven process (**Table 5**). This entropy/enthalpy compensation is comparable for $SpaA_{SLH}$ and $SpaA_{SLH}$/G109A, as the free energy of binding is similar in both complex formation processes.

On the contrary, titration of $SpaA_{SLH}$/G46A/G109A with the monosaccharide ligand did not display heat coming from a binding event. The intensity of the peaks was extremely low, resembling more peaks of ligand dilution into the protein solution (**Figure 17B**). This result is similar to the titration of $SpaA_{SLH}$ with the disaccharide ligand (**Figure 12B**), and suggests that the double mutant may not bind 4,6-Pyr-β-D-ManNAc-*O*-Me anymore.

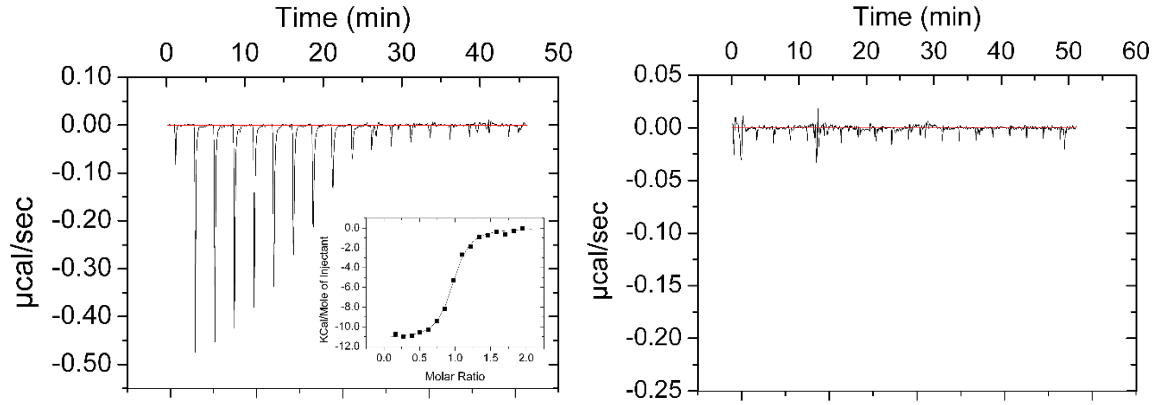


Figure 17. Binding performance of synthetic monosaccharide ligand to $SpaA_{SLH}$ glycine mutants. Calorimetric titration of $SpaA_{SLH}$/G109A (A) and $SpaA_{SLH}$/G46A/G109A (B) with 4,6-Pyr-β-D-ManNAc-*O*-Me ligand. The panels show the incremental heat release upon each 1-µL injection of ligand into protein solution. The insets show the binding isotherms fitted to a 1:1 binding model.

Table 5. Isotermal titration calorimetry (ITC) analyses of SCWP ligand binding[a]

	-TΔS (kJ/mol)	*ΔH* (kJ/mol)	*ΔG* (kJ/mol)	Stoichio-metry	K_A (M^{-1})	K_d (nM)
$SpaA_{SLH}$	45.74 ± 16.01[b]	-87.85 ± 15.62	-42.10 ± 0.65	0.91 ± 0.04	$3.48x10^7$ ± 0.36	29
$SpaA_{SLH}$/G109A	13.33 ± 7.0	-50.64 ± 7.29	-37.31 ± 0.33	0.92 ± 0.04	$4.48x10^6$ ± 0.62	226

[a] Thermodynamic parameters for recombinant proteins binding to 4,6-Pyr-β-D-ManNAc-*O*-Me measured by isothermal titration calorimetry.
[b] Data are shown as means ± standard deviations from at least three independent replicas.

4.12 Crystal structure of *P. alvei* CCM 2051^T $SpaA_{SLH}$/G109A mutant and complex $SpaA_{SLH}$/G109A with 4,6-Pyr-β-D-ManNAc-*O*-Me

As discussed above, given the flip of the Gln^{108}-Gly^{109} peptide bond upon ligand binding, which places Gly^{109} in a conformation that would be unfavorable for any other amino acid, Gly^{109} was mutated to alanine to investigate the effect of this change at a structural level.

Therefore, *P. alvei* CCM 2051^T $SpaA_{SLH}$/G109A mutant was crystallized in space group $P2_12_12_1$ and the structure was solved to 1.15 Å resolution by molecular replacement using the unliganded $SpaA_{SLH}$ structure as search model (**Table 4B**) (Blackler RB, López-Guzmán A, Martinz G, Kosma P, Messner P, Schäffer C and Evans SV, manuscript in preparation). The structure contained one molecule in the AU that displayed excellent stereochemistry and electron density, excluding the residues 22-28 at the N-terminus, which were not observed.

The structure of $SpaA_{SLH}$/G109A is almost identical to $SpaA_{SLH}$ (Cα, rmsd of 0.34 Å, excluding the residues 44-54). The G109A mutation is placed with no change in the position of Gln^{108} and Ala^{109} backbone atoms in comparison with Gln^{108} and Gly^{109} in $SpaA_{SLH}$ structure, however a minor displacement of residues 111-114 is observed in the mutant protein (**Figure 18A**). The main-chain O of Ala^{109} hydrogen bonds to Arg^{61}, similar to the unliganded $SpaA_{SLH}$

structure, but interrupted by a coordinated sulfate ion in that case (**Figure 18A**).

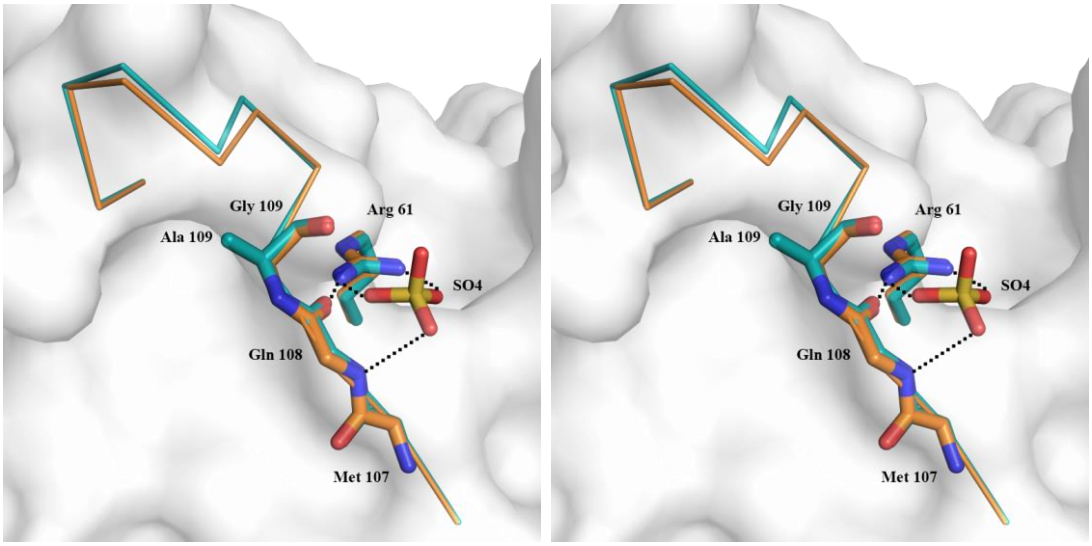
To investigate the G109A mutation on ligand binding, 4,6-Pyr-β-D-ManNAc-*O*-Me was initially soaked into $SpaA_{SLH}$/G109A crystals (in both $P2_12_12_1$ and $P3_221$), however no ligand density was observed in the resulting structures (data not shown, personal communication with Blackler RB and Dr. Evans SV). Similarly, co-crystallization of the monosaccharide with $SpaA_{SLH}$/G109A produced poor diffracttion-quality crystals. High resolution diffraction data was achieved from a co-crystal that took several months to grow, however the structure could not be solved by molecular replacement either with the solved $SpaA_{SLH}$ or $SpaA_{SLH}$/G109A structures as search models. A second crystal from the same crystallization drop was soaked in KI and this time the structure was solved by SAD. This structure was then utilized to solve the original data set by molecular replacement to a 1.50 Å resolution (Blackler RB, López-Guzmán A, Messner P, Schäffer C and Evans SV, manucript in revsion).

The structure of complex $SpaA_{SLH}$/G109A with 4,6-Pyr-β-D-ManNAc-*O*-Me showed a large structural change in comparison to $SpaA_{SLH}$ or unliganded $SpaA_{SLH}$/G109A. Excellent electron density for the monosaccharide was observed in the groove G1, which experienced a rearrangement of residues, becoming narrower relative to previous structures to optimize ligand binding. On the contrary, the groove G2, which contains the G109A mutation, displayed no electron density for ligand and is widened compared to $SpaA_{SLH}$ or unliganded $SpaA_{SLH}$/G109A.

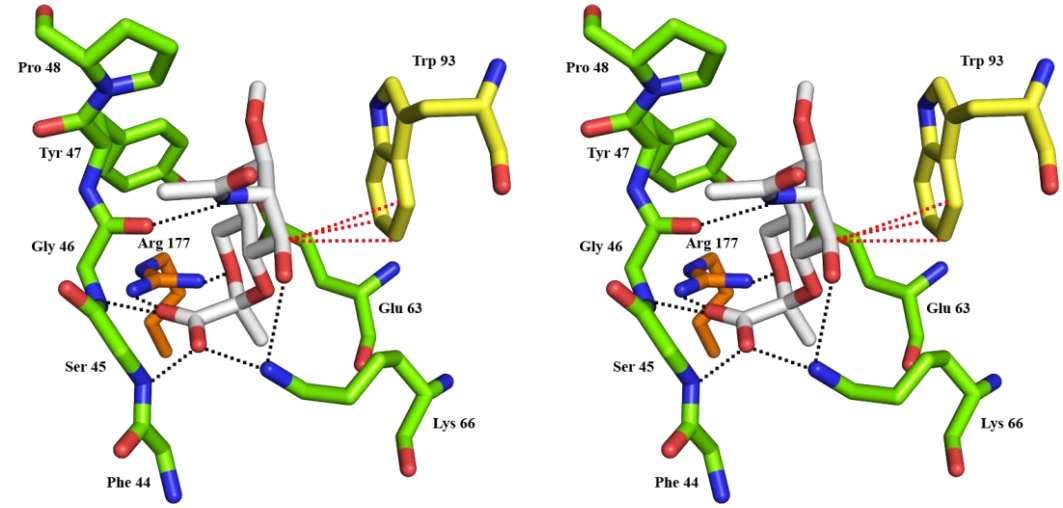
The structural change produced a widening of G2 from 9.1 to 11.8 Å as measured between Trp^{151} Cη2 and Gly^{109} main-chain N or Ala^{109} main-chain O, whereas G1 narrows from 10.6 to 9.2 Å as measured between Trp^{93} Cη2 and Gly^{46} main-chain N, compared to the C2 structure of liganded $SpaA_{SLH}$ with the 4,6-Pyr-β-D-ManNAc-*O*-Me (**Figure 18C**). The overall Cα rmsd between liganded and unliganded $SpaA_{SLH}$/G109A is 1.73 Å or 1.52 Å when excluding

residues 44-55. This important structural change clarifies the failure to soak 4,6-Pyr-β-D-ManNAc-*O*-Me into $SpaA_{SLH}$/G109A crystals and the failure to solve the liganded structure by molecular replacement (personal communication with Blackler RB and Dr. Evans SV).

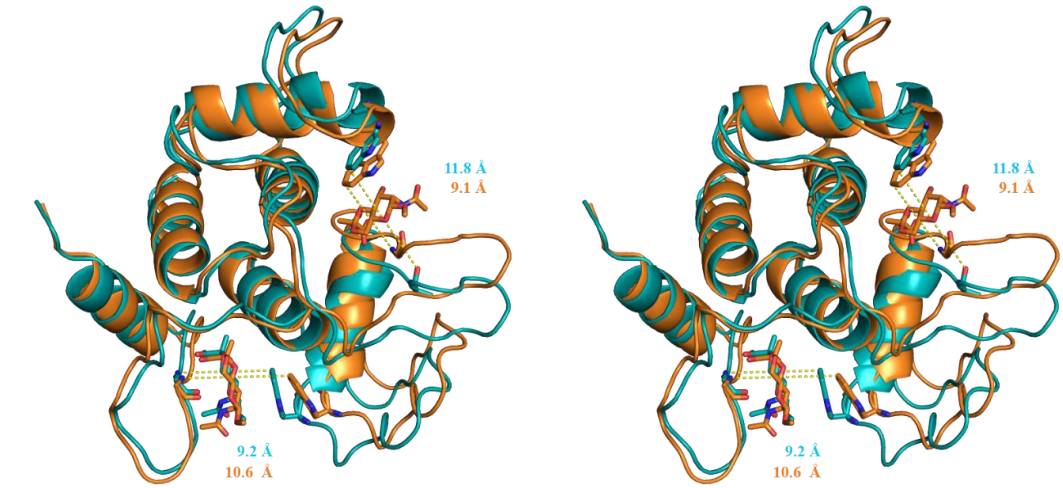
In $SpaA_{SLH}$/G109A structure, Gln^{108} carbonyl O hydrogen bonds to Arg^{61} Nη2 in the G2 pocket in the unliganded structure. In the G1 pocket, Ser^{45}-Gly^{46} peptide bond has flipped upon ligand binding, so that a hydrogen bond between Ser^{45} carbonyl O and Arg^{177} Nη2 is replaced by a hydrogen bond between Gly^{46} main-chain N and a pyruvyl carboxyl O of ligand (**Figure 18B**). This structure shows the same conformation of loop residues 44-54 next to G1 as observed in molecule A of the C2 structure of linganded $SpaA_{SLH}$ with monosaccharide in G1. In addition, Trp^{93} is stacked against the hydrophobic face of the ManNAc ring, centered about ManNAc C3, similar to $SpaA_{SLH}$ with bound monosaccharide in G2.


Gly 109
Arg 61
Ala 109
SO4
Gln 108
Met 107
Gly 109
Arg 61
Ala 109
SO4
Gln 108
Met 107

A


Pro 48
Trp 93
Tyr 47
Gly 46
Arg 177
Glu 63
Ser 45
Lys 66
Phe 44
Pro 48
Trp 93
Tyr 47
Gly 46
Arg 177
Glu 63
Ser 45
Lys 66
Phe 44

B


11.8 Å
9.1 Å
9.2 Å
10.6 Å
11.8 Å
9.1 Å
9.2 Å
10.6 Å

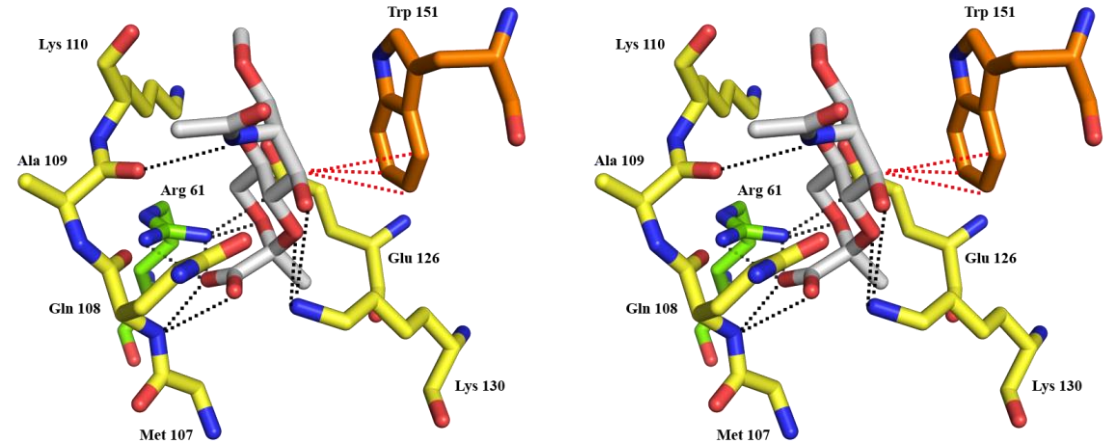
C

Figure 18. (A) Comparison of the conformations of residues 111-114 in G2 of $SpaA_{SLH}$ (cyan) and $SpaA_{SLH}$/G109A (orange) to exemplify the effect of G109A mutation. (B) Stereo image of 4,6-Pyr-β-D-ManNAc-*O*-Me bound in G1 of $SpaA_{SLH}$/G109A of C2 structure. $SpaA_{SLH}$/G109A carbon atoms are colored according to SLH domains (SLH_1 green, SLH_2 yellow, SLH_3 orange). (C) Ribbon diagram of superposed $SpaA_{SLH}$ (orange) and $SpaA_{SLH}$/G109A (cyan) C2 structures in complex with 4,6-Pyr-β-D-ManNAc-*O*-Me. Alignment was done only with the residues comprising the core three-helices to emphasize the structural change of G1 and G2. The width of G1 and G2 are measured between Trp^{93} Cη2 and Gly^{46} N and Trp^{151} Cη2 and Gly^{109} N or Ala^{109} carbonyl O, respectively.

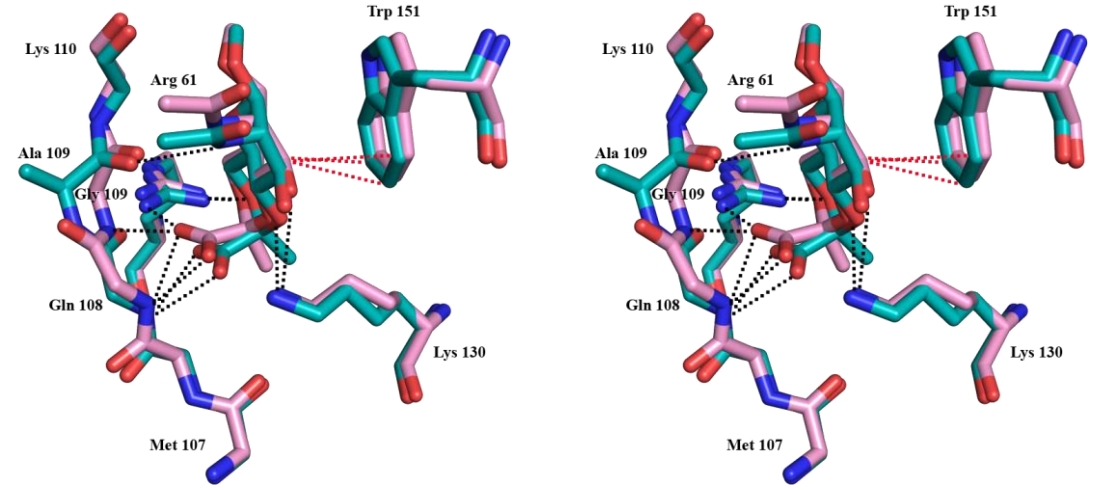
4.13 Crystal structure of *P. alvei* CCM 2051^T $SpaA_{SLH}$/G46A/G109A mutant

The $SpaA_{SLH}$/G46A/G109A mutant was generated to support the importance of the glycine flip to SCWP binding. High resolution diffraction data was collected from a co-crystal of $SpaA_{SLH}$/G46A/G109A with 4,6-Pyr-β-D-ManNAc-*O*-Me with an initial 6-fold molar excess of ligand. The structure was solved to 1.24 Å resolution by molecular replacement using the solved structure of unliganded $SpaA_{SLH}$/G109A as a search model (**Table 4B**) (Blackler RB, López-Guzmán A, Martinz G, Gagnon SML, Haji-Ghassemi O, Kosma P, Messner P, Schäffer C and Evans SV, manuscript in preparation). This structure shows the groove G2 narrower than G1, similar to $SpaA_{SLH}$ and unliganded $SpaA_{SLH}$/G109A.

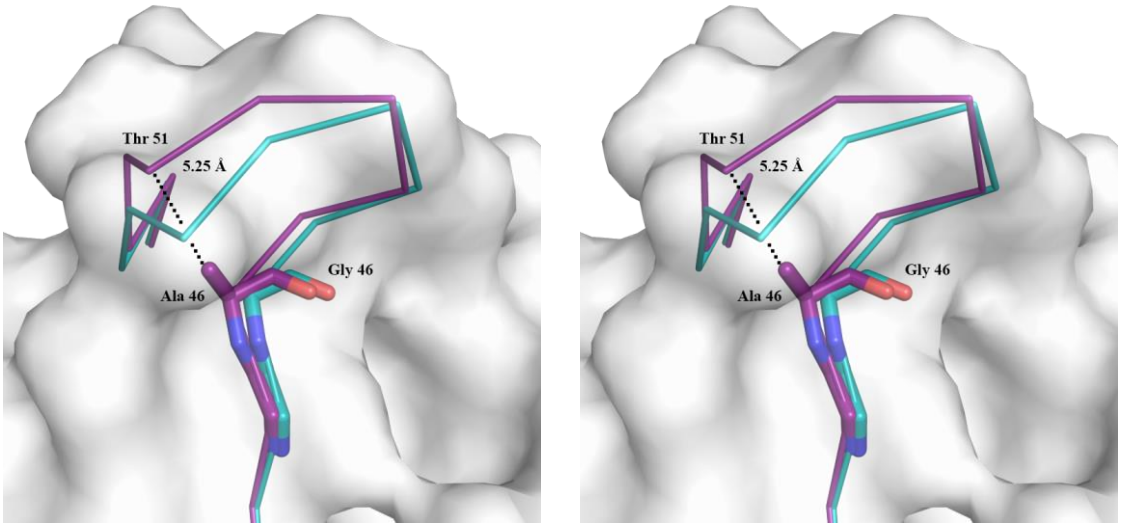
Excellent electron density for the monosaccharide was observed in G2 pocket, the Gln^{108}-Ala^{109} peptide bond is not flipped, similar to the unliganded conformations. The ligand is bound with similar interactions as observed in the complex with $SpaA_{SLH}$, except for Ala^{109} main-chain N that does not hydrogen bond to the pyruvyl carboxyl oxygen of ligand, there is a minor displacement of ligand away from the Gln^{108} main-chain O (**Figure 19A**). The Arg^{61} side chain is also away from Gln^{108} main-chain O compared to the unliganded structure of $SpaA_{SLH}$/G109A (3.4 Å *vs.* 3.1 Å).


Trp 151
Lys 110
Ala 109
Arg 61
Glu 126
Gln 108
Lys 130
Met 107
Trp 151
Lys 110
Ala 109
Arg 61
Glu 126
Gln 108
Lys 130
Met 107

A


Trp 151
Lys 110
Arg 61
Ala 109
Gly 109
Gln 108
Lys 130
Met 107
Trp 151
Lys 110
Arg 61
Ala 109
Gly 109
Gln 108
Lys 130
Met 107

B


Thr 51
5.25 Å
Gly 46
Ala 46
Thr 51
5.25 Å
Gly 46
Ala 46

C

Figure 19. (A) Stereo image of 4,6-Pyr-β-D-ManNAc-*O*-Me bound in G2 of SpaA$_{SLH}$/G46A/G109A of $P2_12_12_1$ structure. Carbon atoms are colored according to SLH domains (SLH_1 green, SLH_2 yellow, SLH_3 orange). (B) Comparison of 4,6-Pyr-β-D-ManNAc-*O*-Me bound in G2 of SpaA$_{SLH}$ (pink) and SpaA$_{SLH}$/G46A/G109A (blue), to illustrate the different conformation of residue 108-109 peptide bond. (C) Comparison of the 43-54 residues in G1 between unliganded SpaA$_{SLH}$/G109A (cyan) and liganded SpaA$_{SLH}$/G46A/G109A (purple) to show the effect of the G46A mutation.

Excellent electron density is also observed for the G46A mutation in groove G1. Residues 46-55 within the loop besides G1 are in a similar conformation as in the unliganded SpaA$_{SLH}$/G109A structure, but with a minor displacement of Ala^{46}, the distance between the Cα of residue 46 to Thr^{51} Cα increases from 3.96 Å to 5.25 Å (**Figure 19C**). **Table 6** shows the rotation of backbone dihedral angles phi (φ) and psi (ψ) of Ser^{45} and Gly^{46} as well as Gln^{108} and Gly^{109} residues in groove 1 and groove 2, respectively, in free SpaA$_{SLH}$ and free SpaA$_{SLH}$/G109A mutant, and upon SCWP ligand binding.

Table 6. Dihedral angles of SLH-residues in GIIxG motif

	Ser^{45} or Gln^{108} [a]		**Gly^{46} or Gly^{109} [a]**	
Structure	**φ (°)**	**ψ (°)**	**φ (°)**	**ψ (°)**
Unliganded SpaA G1 (A)[b]	-123	160	-106	-150
Unliganded SpaA G1 (B)	-121	169	-54	163
Unliganded SpaA G2 (A)	-135	176	-70	164
Unliganded SpaA G2 (B)	-135	178	-72	163
Unliganded SpaA/G109A G1	-116	169	-62	160
Unliganded SpaA/G109A G2	-141	171	-55	149
Liganded SpaA G2 (A)[c]	-129	-19	86	-175
Liganded SpaA G2 (B)	-138	-5	89	-173
Liganded SpaA/G109A G1	-136	-8	62	-162

[a] Ser^{45} and Gly^{46} are located in G1, whereas Gln^{108} and Gly^{109} in G2.
[b] Letters in parenthesis correspond to PDB chain ID
[c] Dihedral angles are from the structure of SpaA$_{SLH}$ in complex with monosaccharide in space group C2 (PDB 5JL0).

In addition, SpaA$_{SLH}$/G46A/G109A and 4,6-Pyr-β-D-ManNAc-*O*-Me were co-crystallized with 3:1 and 0.6:1 molar ratio of ligand to protein. The structures were solved to 1.95 and 2.10 Å resolution, respectively. Both structures were

isomorphous to the structure at 6:1 molar ratio and displayed good electron density for ligand in groove G2 (data not shown, personal communication with Blackler RB and Dr. Evans SV).

4.14 The conserved TRAE motifs have a critical role in *P. alvei* CCM 2051^T SpaA$_{SLH}$ three-pronged trimer folding

To abolish binding of 4,6-Pyr-β-D-ManNAc-*O*-Me to SpaA$_{SLH}$, the mutants SpaA$_{SLH}$/TAAA$_{12}$ (where the TRAE in SLH$_1$ and TVEE in SLH$_2$ are both mutated to TAAA) and SpaA$_{SLH}$/TAAA$_{13}$ (where the TRAE in SLH$_1$ and TRAQ in SLH$_3$ are both mutated to TAAA) were created and overexpressed in *E. coli* cells. Overexpression of the SpaA$_{SLH}$/TAAA mutants was performed as described for the SpaA$_{SLH}$, SpaA$_{SLH}$/G109A and SpaA$_{SLH}$/G46A/G109A proteins, however, this time overexpression resulted in insoluble protein (**Figure 20**). These results are in contrast to previous studies performed by Janesch and co-workers (Janesch *et al.* 2013b) that produced TAAA double and triple mutants of full-size SpaA. In that case, the full-size of SpaA might have supported the stability, and hence solubility of the S-layer protein.

On the other hand, May and co-workers have also reported a reduction in the solubility of XynA SLH protein from *T. thermosulfurigenes* overexpressed in *E. coli*, when the TRAE motifs were mutated to TAAA simultaneously in the different SLH domains, in comparison to the wild-type XynA SLH protein (May *et al.* 2006).

These results suggest that the TRAE motif and its variants are not only important for SCWP binding, but also for the SLH domain three-pronged trimer folding. It is important to highlight the location of these conserved motifs in the SpaA$_{SLH}$ structure. The TRAE motifs are located at the N-terminal part of each core helix, interconnecting the SLH domains and contributing to the formation of the binding grooves (**Figure 10**). Therefore,

mutation of the conserved motifs may destabilize the sterical requirements of the SLH-trimer structure.

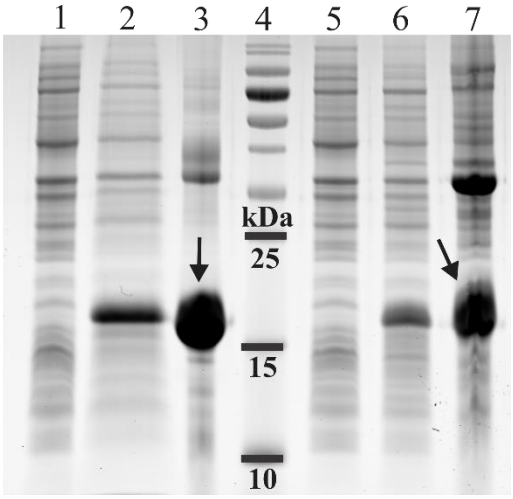


Figure 20. SDS-PAGE analysis of *P. alvei* CCM 2051^T SpaA$_{SLH}$/TAAA mutants. SDS-PAGE gel of SpaA$_{SLH}$/TAAA$_{13}$ (lanes 1 to 3) and SpaA$_{SLH}$/TAAA$_{12}$ (lanes 5 to 7) overexpression in *E. coli*. Lanes 1 and 5, *E. coli* cells before induction. Lanes 2 and 6, protein overexpression after addition of IPTG. Lanes 3 and 7, pellet fraction after centrifugation of crude cell extract. Lane 4; Protein molecular mass ladder. The arrows indicate the insoluble recombinant proteins.

4.15 Sequence analysis and domain prediction of structurally relevant regions of *P. alvei* CCM 2051^T CsaB

Sequence homology analyses of CsaB amino acid sequence were performed using the BLAST tool (Altschul *et al.* 1997) at the website of the National Center for Biotechnology Information, (https://www.ncbi.nlm.nih.gov/). Sequences homologies were found to other CsaB proteins from different Firmicutes, although no tertiary structure has been elucidated yet.

Regarding the sequence similarity among the diverse CsaB proteins, individual and groups of conserved residues were found (**Figure 21**). The latter might represent conserved motifs that are structurally or functionally important for activity.

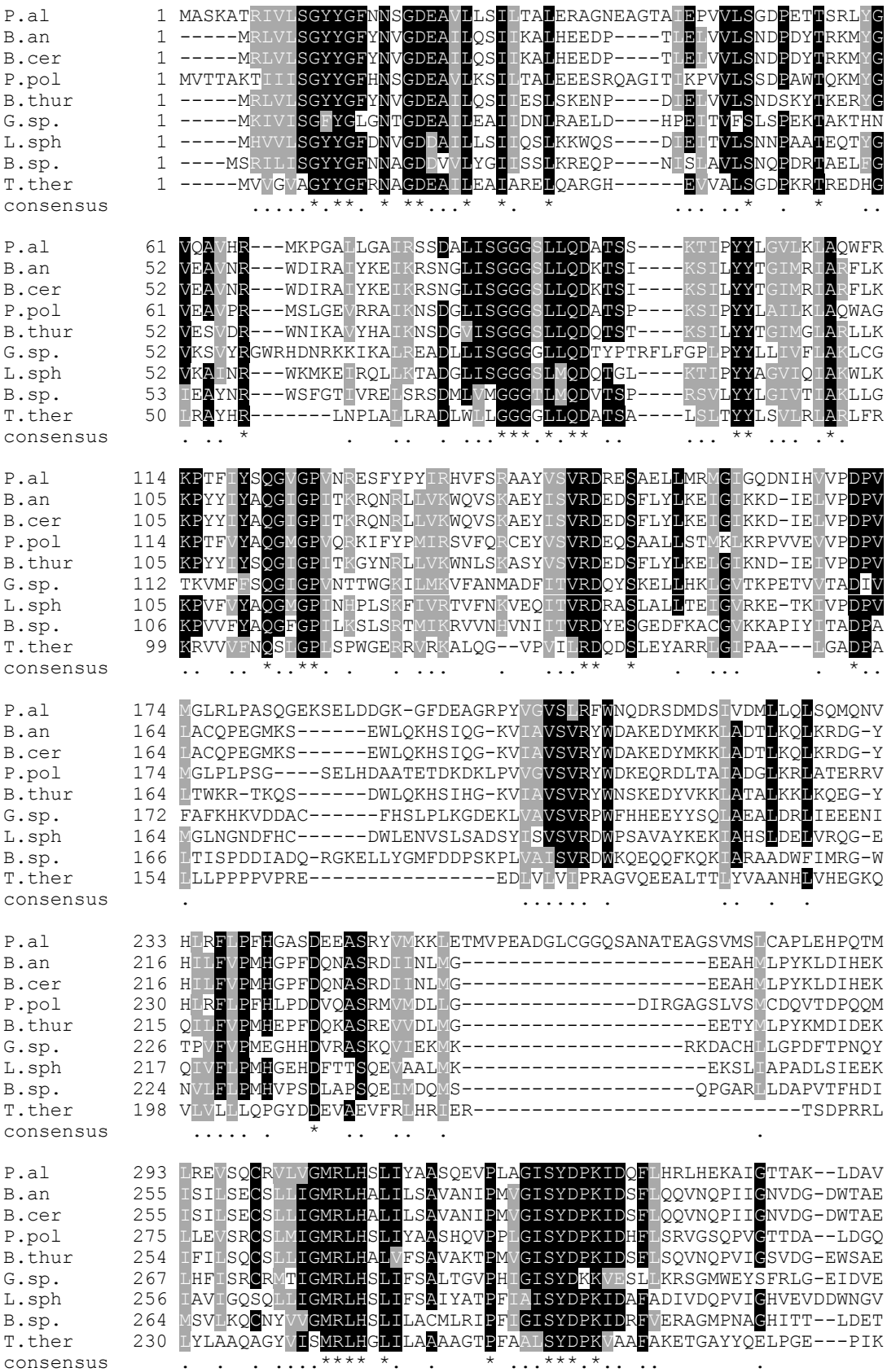


Figure 21. Sequence alignment of CsaB sequences from different Firmicutes. Putative polysaccharide pyruvyl transferases from *Paenibacillus alvei* CCM 2051^T (P.al), *Bacillus anthracis* (B.an), *Bacillus cereus* (B.cer), *Paenibacillus polymyxa* (P.pol), *Bacillus thuringiensis* (B.thur), *Geobacillus* sp. WCH70 (G.sp.), *Lysinibacillus sphaericus* (L.sph), *Brevibacillus* sp. (B.sp) and *Thermus thermophilus* HB27 (T.ther) were aligned using ClustalW2 program. The residues conserved at least at 80% level are shaded in black, whereas 100% conserved residues are marked with an asterisk. Shading was done by BOXSHADE program.

The presence of conserved domains was analyzed by using the following web tools: Conserved Domain Database (CDD) blast (Marchler-Bauer *et al.* 2011, Marchler-Bauer *et al.* 2015) and the Pfam database (Finn *et al.* 2014). The servers predicted a polysaccharide (PS) pyruvyl transferase domain in *P. alvei* CCM 2051^T CsaB, ranging from residue 18 to residue 326 (**Figure 22**), which transfers a pyruvyl group to peptidoglycan-linked polysaccharides.

Besides, the pyruvyl transferase domain was classified in the glycosyl-transferase GT-B type superfamily. Glycosyltransferases catalyze the transfer of saccharide moieties from nucleotide-activated sugar donors to a specific acceptor that can be a lipid, a protein or another carbohydrate molecule. Members of the GT-B family display a common topology that consists of N- and C-terminal domains separated and linked by a flexible loop, forming a deep cleft that includes the catalytic center. The two domains exhibit a typical Rossmann-fold which consists of two sets of β-α-β-α-β units, forming a single parallel sheet flanked by several α-helices. The Rossmann-fold was first discovered in the lactate dehydrogenase enzyme from *Squalus acanthus* and found in several nucleotide-binding proteins (Rossmann *et al.* 1974, Lesk 1995).

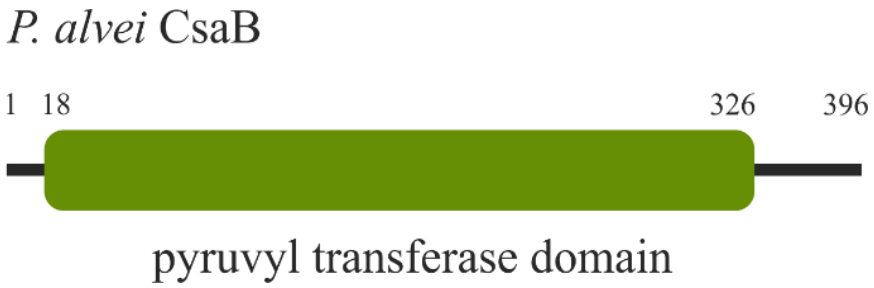


Figure 22. Sketch of the domain prediction for the *P. alvei* CCM 2051^T CsaB protein. The predicted polysaccharide pyruvyl transferase domain, glycosyltransferase GT-B fold superfamily is located in the central region of the polypeptide chain and is shown in green.

4.16 Fold recognition of *P. alvei* CCM 2051^T CsaB

Fold recognition was performed using the Phyre2 web portal (http://www.sbg.bio.ic.ac.uk/phyre2/) (Kelley *et al.* 2015). Fold recognition is a theoretical approach which compares a protein sequence to a large database of sequences, generating a suitable profile of that sequence. Next, this sequence is compared to a large database of sequences of known structures.

The analysis results in an alignment of a sequence of unknown structure and one of known structure to construct a model (Godzik 2005). **Figure 23** shows the three-dimensional (3D) models of CsaB generated by *Phyre2*.

The four models consist of β-sheets, α-helices and coils, and have almost identical lengths. Each model was constructed using published crystal structures based on secondary structure prediction with a confidence value between 99-100%. The secondary structure of these solved structures closely matches the predicted secondary structure of CsaB, which all are GT-B type fold glycosyltransferases, conforming also the prediction of the domain classification. The four models are composed of two separated domains with a connecting linker region and a cleft, probably containing the active site, located between both domains. Every domain consists of parallel β-sheets flanked by α-helices, thus resembling the Rossmann-fold domain.

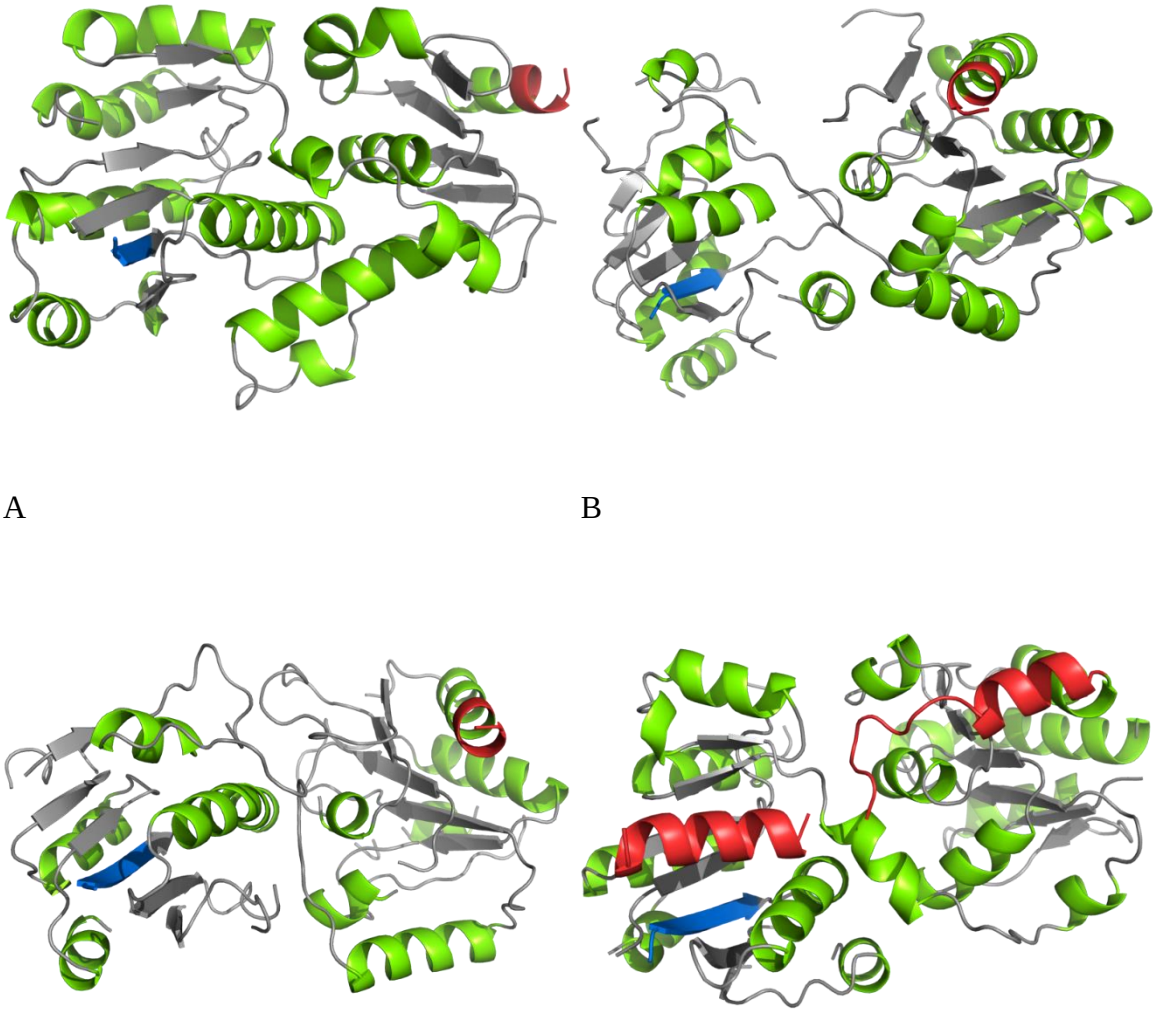


Figure 23. Three-dimensional models of *P. alvei* CCM 2051^T CsaB by the fold recognition approach using the *Phyre2* tool. The models show the typical GT-B fold, the N-and the C-terminus are colored in blue and red, respectively. (A) CsaB model based on the yeast pyruvyl transferase Pvg1p structure (PDB: 5AX7). (B) Model based on UDP-glycosyltransferase, ADP-heptose LPS heptosyltransferase II structure (PDB: 1PSW). (C) CsaB model based on glycosyltransferase family 9 structure (PDB: 3TOV). (D) Model based on the isomerase UDP-*N*-acetylglucosamine 2-epimerase structure (PDB: 3OT5).

The first model, taking into account *P. alvei* CCM 2051^T CsaB residues 5-361, was created using as a template the crystal structure of the pyruvyl transferase Pvg1p from the yeast *Schizosaccharomyces pombe* (**Figure 23A**). Pvg1p shares

low primary sequence similarity with predicted bacterial pyruvyl transferases, and is the only biochemically characterized and crystallized eukaryotic ketal pyruvyl transferase thus far (Yoritsune *et al.* 2013, Higuchi *et al.* 2016). The second and third models, taking into account *P. alvei* CCM 2051^T CsaB residues 6-364 and 4-362 respectively, were constructed based on glycosyltransferases structures. The templates used were an UDP-glycosyltranferase/glycogen phosphorylase / ADP-heptose LPS heptosyltransferase II for the second model (**Figure 23B**) and a glycosyltranferase of family 9 from *Veilonella parvula* DSM 2008 for the third one (**Figure 23C**). The last model, taking into account *P. alvei* CCM 2051^T CsaB residues 4-394, was generated utilizing an UDP-*N*-acetylglucosamine 2-epimerase structure from *Listeria monocytogenes* as a template (**Figure 23D**). The four structures have different orientations of their lobes but all begin with a β-sheet corresponding to RIVL residues at the N-terminus (shown in blue) and all of them terminate after a D or N residue at the exit of an α-helix, except for the fourth model, whose C-terminus extends with a large α-helix tail opposite to the cleft (hypothetical active site) between the two lobes (**Figure 23D**).

4.17 Expression and secondary structure determination of *P. alvei* CCM 2051^T recombinant CsaB

Overexpression of recombinant *P. alvei* CCM 2052^T CsaB in *E. coli* BL21 star cells was induced by the addition of IPTG, after the incubation time cells were harvested. The apparent molecular mass of 44 kDa concurs with the theoretical value of *P. alvei* CCM 2051^T CsaB (**Figure 24A**). The recombinant protein was purified by metal-affinity chromatography, and its purity was assessed by SDS-PAGE analysis. The identity of rCsaB was confirmed by Western immuno-blotting (**Figure 24B**). The far-UV CD spectra of rCsaB were measured from 190 nm to 260 nm. The recombinant protein exhibited a maximum band at 192 nm and a minimum broad band between 208 and 221 nm, indicative of a high content of α-helical and β-sheet structures (**Figure 24C**). Deconvolution

of CD spectra of rCsaB was performed by the algorithms SELCON, CDSSTR and CONTIN, and showed a content of 36-38% of α-helical structures, 14.5-16% of β-sheets structures and 46-47% of β-turns and random coils. These values are in accordance with the secondary structure prediction.

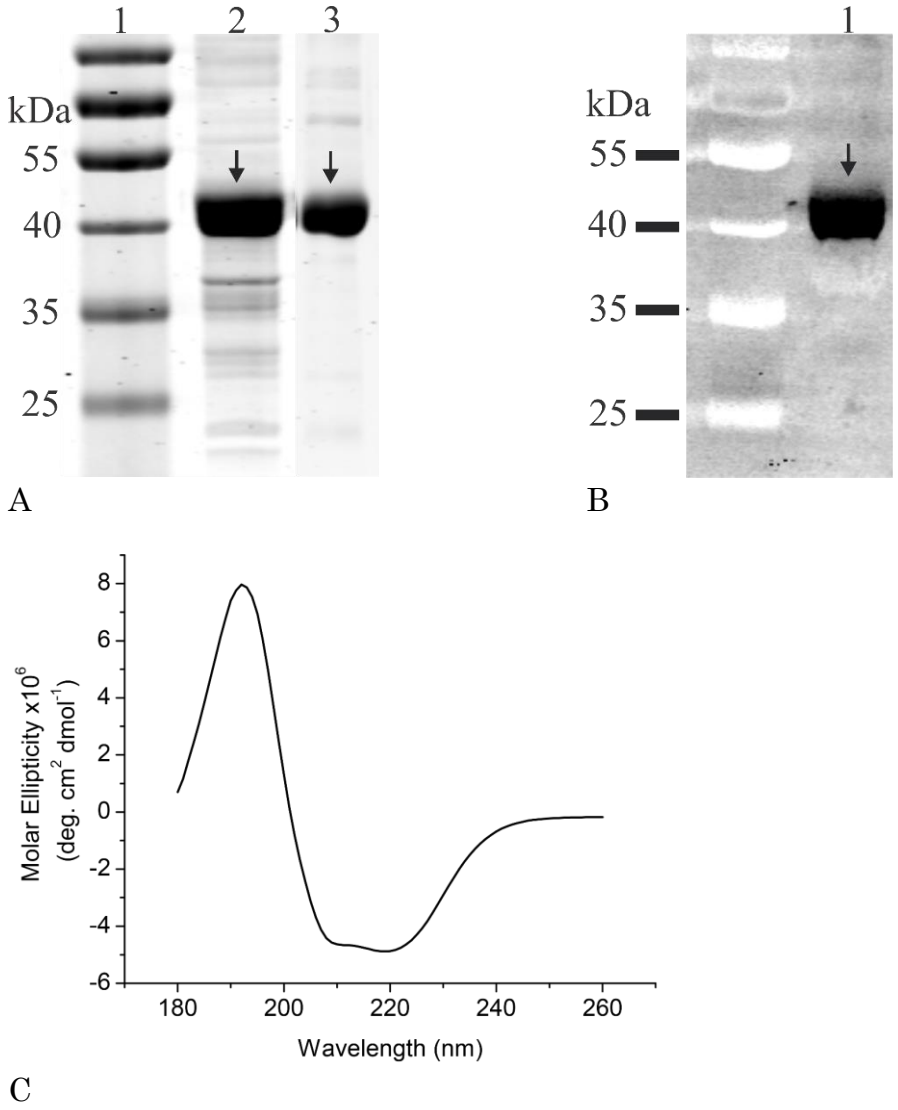
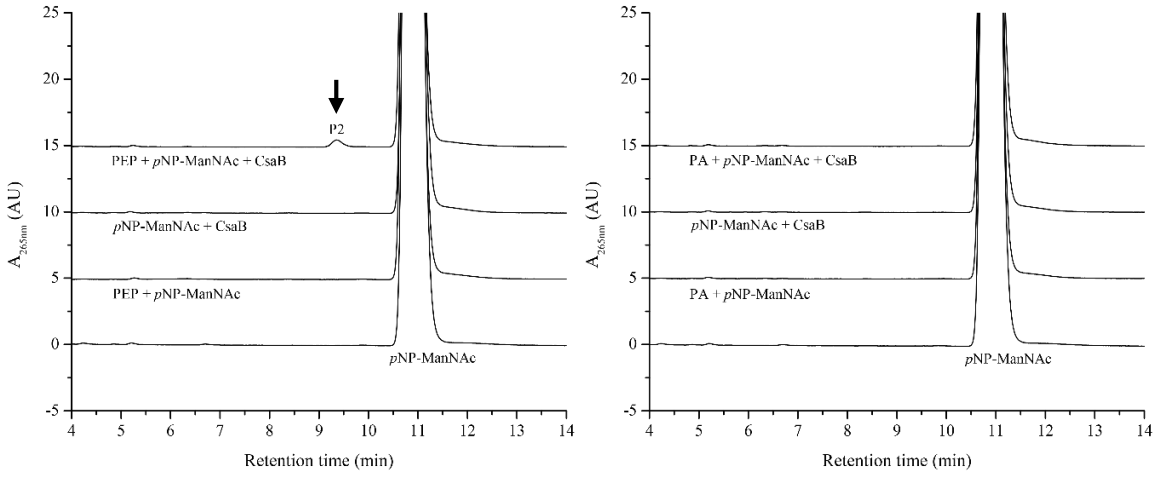


Figure 24. *P. alvei* CCM 2051^T recombinant CsaB overexpression. (A) SDS-PAGE (12%) analysis of recombinant, His6-tagged CsaB (~42 kDa) expressed in *E. coli* BL 21 after CBB staining. Lane 1, protein molecular mass ladder; Lane 2, rCsaB after induction of expression; Lane 3, rCsaB after purification by metal-affinity chromatography. (B) Western immunoblot analysis with fluorescence detection of rCsaB using an anti-His6 tag antibody. Lane 1, cell crude extract after induction of expression. The values on the left indicate the apparent molecular masses in kDa. The arrows indicate the recombinant CsaB protein. (C) Far-UV circular dichroism (CD) spectrum of purified rCsaB.

4.18 Biochemical characterization of *P. alvei* CCM 2051^T recombinant CsaB

Information about bacterial polysaccharide pyruvyl transferases is very limited regarding substrates and enzymatic mechanisms. The only study on the enzymatic properties of a pyruvyl transferase has been performed so far with the enzyme Pvg1p from the fission yeast *S. pombe* (Yoritsune *et al.* 2013). In that case, Pvg1p was shown to transfer a pyruvyl group from PEP to *p*NP-β-galactose, producing a 4,6-Pyr-β-D-galactose, confirming Pvg1p pyruvyl transferase activity *in vitro*. Recently, Higuchi and co-workers solved the crystal structure of a recombinant truncated variant of Pvg1p at 2.5 Å resolution (Higuchi *et al.* 2016). The structure of Pvg1p revealed two α/β/α domains at the N- and C-terminus, resembling the fold of GT-B glycosyltransferases, and similar to the predicted fold of *P. alvei* CCM 2051^T CsaB (**Figure 23**).

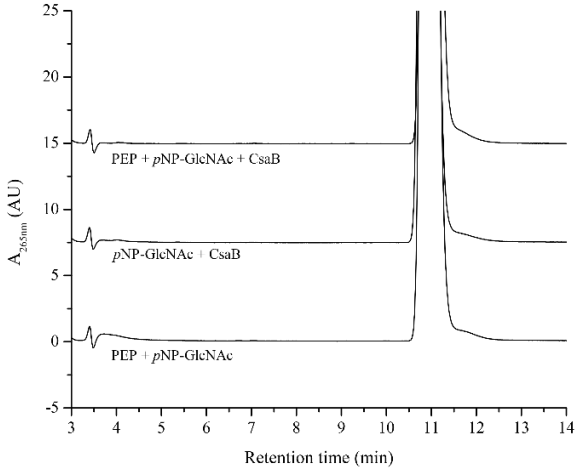
Therefore PEP and *p*NP-β-ManNAc were chosen as donor and acceptor substrates respectively, since PEP has a high-energetic phosphate bond and β-ManNAc is the pyruvylated residue in *P. alvei* CCM 2051^T SCWP (Schäffer *et al.* 2000b, Schäffer and Messner 2005). Purified, recombinant CsaB was incubated with the substrates at 37°C for one hour or overnight. The enzyme was removed from the reactions mixtures and samples were then analyzed by RP-HPLC. A peak for *p*NP-β-ManNAc was observed at 10.8 minutes (**Figure 25A**). After incubation with PEP and rCsaB, a new, small peak was observed at 9.4 minutes. The new peak was not observed when PEP or rCsaB were excluded from the reaction mixtures. The new peak was neither observed when pyruvate nor *p*NP-β-GlcNAc were used as donor or acceptor substrates, respectively (**Figure 25B, C**). The incubation time or the presence of a divalent cation (Mg^{+2}) did not have an effect on substrate conversion, since the size of the new peak did not increase (**Figure 25 D, E**). On the other hand, rCsaB concentration seemed to have a positive effect on product formation, as higher concentrations of rCsaB produced a new peak with greater intensity (**Figure 25F**).


P2
PEP + pNP-ManNAc + CsaB
pNP-ManNAc + CsaB
PEP + pNP-ManNAc
pNP-ManNAc
A265nm (AU)
Retention time (min)

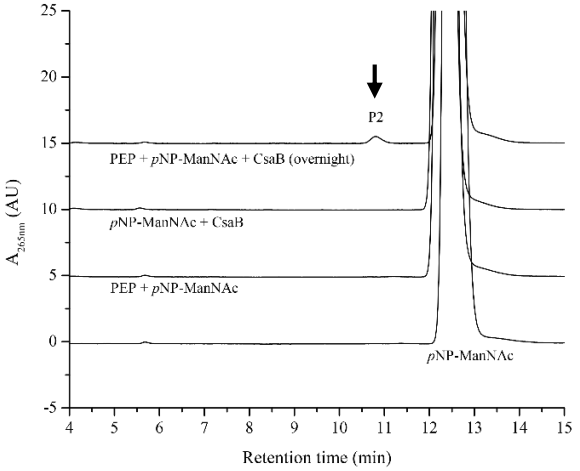
A

PA + pNP-ManNAc + CsaB
pNP-ManNAc + CsaB
PA + pNP-ManNAc
pNP-ManNAc
A265nm (AU)
Retention time (min)

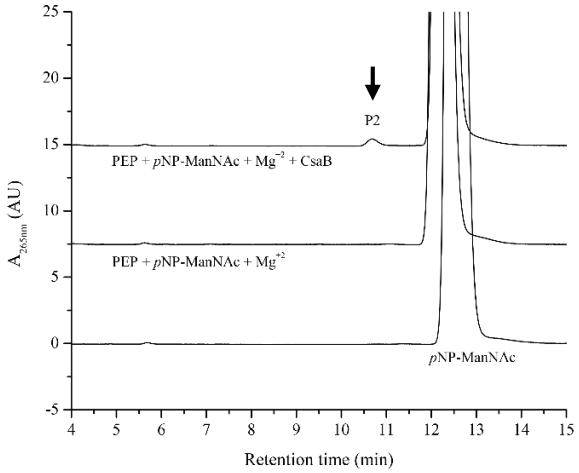
B


PEP + pNP-GlcNAc + CsaB
pNP-GlcNAc + CsaB
PEP + pNP-GlcNAc
A265nm (AU)
Retention time (min)

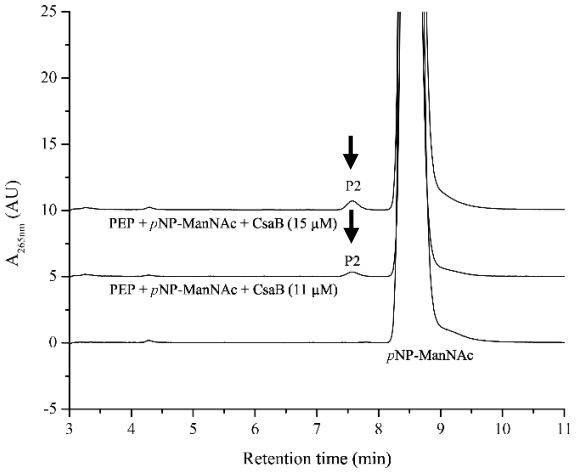
C


P2
PEP + pNP-ManNAc + CsaB (overnight)
pNP-ManNAc + CsaB
PEP + pNP-ManNAc
pNP-ManNAc
A265nm (AU)
Retention time (min)

D


P2
PEP + pNP-ManNAc + Mg+2 + CsaB
PEP + pNP-ManNAc + Mg+2
pNP-ManNAc
A265nm (AU)
Retention time (min)

E


P2
PEP + pNP-ManNAc + CsaB (15 µM)
P2
PEP + pNP-ManNAc + CsaB (11 µM)
pNP-ManNAc
A265nm (AU)
Retention time (min)

F

Figure 25. HPLC profile of *P. alvei* CCM 2051^T rCsaB reactions. Analyses of the reaction products upon 1 hour of incubation with rCsaB (A) with pNP-β-ManNAc and PEP, and (B) with pNP-β-ManNAc and PA as substrates. (C) Reaction profile of rCsaB incubated with PEP and pNP-β-GlcNAc as substrates. A new peak (P2) was observed solely when rCsaB, PEP and pNP-β-ManNAc were present in the reaction mixture. (D) Analyses of the effect of overnight incubation (D) and presence of Mg^{+2} in a final concentration of 5 mM (E) on the reaction yield of rCsaB incubated with PEP and pNP-β-ManNAc. (F) A new peak (P2) with greater intensity could be observed when the rCsaB concentration was increased around 35%.

These results suggest that the new peak may be a product of the rCsaB pyruvyltransferase activity, however pyruvylation of *p*NP-β-ManNAc could not be confirmed by NMR analyses or other analytical technique given to low, if at all, yield of potential product, which might be due to low activity of rCsaB under the chosen conditions.

On the other hand, the enzymatic activity of rCsaB could not be determined by the green malachite assay. The amount of phosphate released from PEP due to rCsaB activity could not be estimated with this assay, since the control samples showed a high degree of free phosphate compared to the enzymatic reactions. The high amount of free phosphate after treatment of the samples with the phosphate detection reagents is probably due to acid-hydrolysis of PEP, since the green malachite reagent (Pi ColorLock™ Gold reagent, Innova Biosciences) contains 5 M HCl. In fact, enol phosphate compounds (as PEP) are high energy compounds, which are very labile to acidic conditions. Therefore, the acid-hydrolysis of PEP may have hampered, if at all, the release of phosphate due to rCsaB activity, especially at PEP concentrations higher than 100 μM (**Figure 26**).

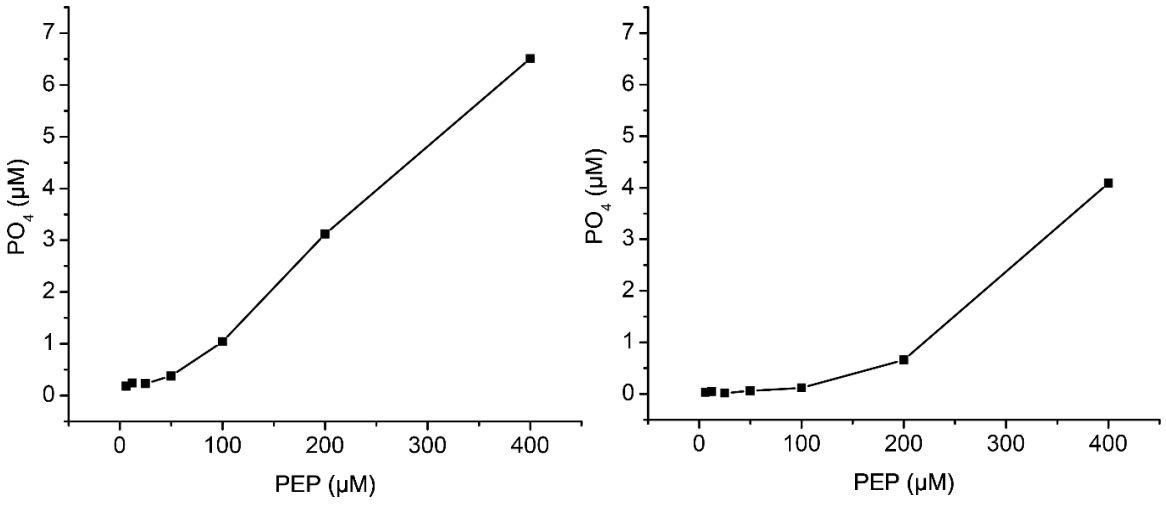
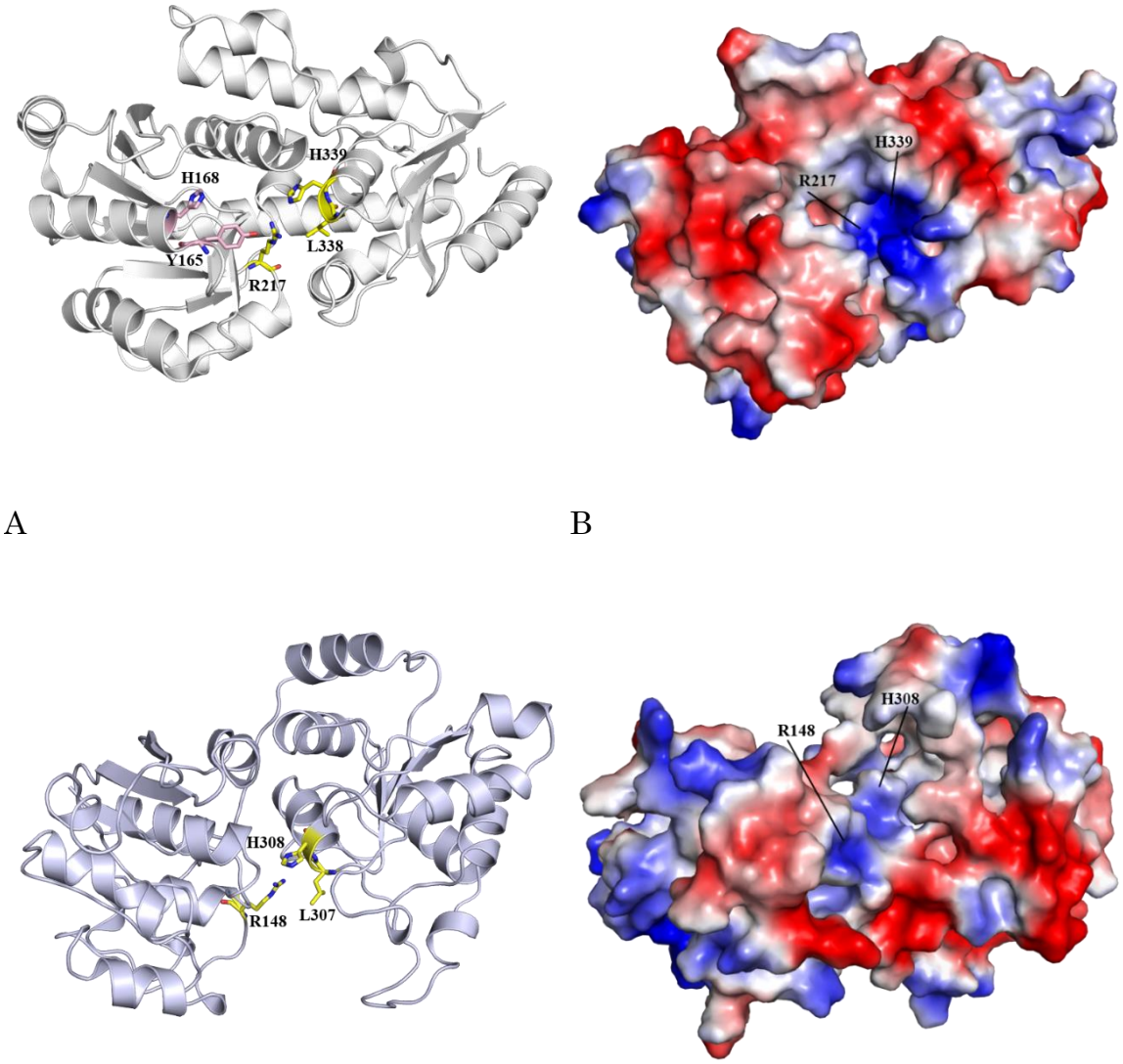


Figure 26. Detection of free inorganic phosphate by the malachite green assay in enzyme reactions and control samples. Determination of released phosphate from PEP at 595 nm upon incubation with (A) or without (B) *P. alvei* CCM 2051^{T} recombinant CsaB at 37°C.

4.19 Comparison between the crystal structure of *Saccharomyces pombe* Pvg1p and the predicted structure of *P. alvei* CCM 2051^{T} CsaB

Even though there is no homology between the primary sequence of CsaB and Pvg1p, the predicted fold of *P. alvei* CCM 2051^{T} CsaB is similar to the structure of Pvg1p. Both structures possess the type-B glycosyltransferase fold. Furthermore, Higuchi and co-workers generated a molecular model of complex Pvg1p with PEP as donor substrate and lactose as an acceptor substrate, and were able to identify the potential residues involved in ligand binding (Higuchi *et al.* 2016). Residues Arg^{217}, Leu^{338} and His^{339} are involved in binding PEP, whereas Tyr^{165} and His^{168} bind the acceptor ligand lactose (**Figure 27A**). Moreover, a charge surface representation of Pvg1p clearly shows a positively charged region located inside the cleft between the N- and C-terminal domains (**Figure 27B**). This region is formed by residues Arg^{217}, Leu^{338} and His^{339} and is suitable to bind a negatively charged substrate such as PEP.



A B

C D

Figure 27. Comparison of the crystal structure of Pvg1p and predicted structure of CsaB. Structure of Pvg1p is shown as a ribbon diagram (A) and as an electrostatic surface model (B) (Higuchi *et al.* 2016). The predicted structure of *P. alvei* CCM 2051^{T} CsaB is shown as a ribbon (C) and as an electrostatic surface model (B). Residues involved in binding are shown as sticks and in yellow for donor substrate (PEP) and in pink for acceptor substrate (β-galactose). In the electrostatic surface models, positive and negative charges are depicted in blue and red, respectively. Residues inside the positively charged pocket between the N- and C-termini are highlighted.

A positively charged region is also found between the N- and C-terminal domains of CsaB. Remarkably, this region is formed by residues equivalent to

those found in Pvg1p; in CsaB residues Arg^{148}, Leu^{307} and His^{308} are involved, which may bind the substrate PEP (**Figure 27C, D**). It is noteworthy to mention that these three residues, which are found in the motifs SVRD and MRLH, respectively, are highly conserved in all CsaB homologues from different Firmicutes (**Figure 21**). Therefore, these three residues are potential candidates as catalytic residues for the pyruvylation activity of CsaB, and thus for prospective mutational studies.

On the other hand, equivalent residues for binding of the acceptor substrate were not identified in CsaB, probably due to the difference of targeted acceptor substrates between CsaB and Pvg1p, which are β-ManNAc and β-Gal, respectively.

5 Discussion

5.1 Three-dimensional crystallization attempts of full-size SpaA in complex with purified, native *P. alvei* CCM 2051^T SCWP

Hastie and Briton were the first to report that the SCWP serves as anchoring molecule for binding of S-layer proteins to the peptidoglycan-containing layer for *Lysinibacillus sphaericus* (Hastie and Brinton 1979). Later on, extraction and chemical characterization experiments of the PG-containing sacculi from *G. stearothermophilus* PV72/p6 showed that a secondary cell wall polysaccharide composed of GlcNAc and Glc, and not the peptidoglycan or a classical SCWP (e.g. teichoic acid) is the binding site for the S-layer protein in this Gram-positive bacterium (Sára *et al.* 1996). Afterwards, Ries and co-workers demonstrated that in *G. stearothermophilus* PV72/p2, an oxygen-induced strain, possesses a SCWP composed of GlcNAc and ManNAc, which also serves as binding site for the S-layer protein SbsB (Ries *et al.* 1997).

Now, it is widely recognized that the N-terminus of S layer proteins containing typical SLH domains selectively recognizes the SCWP in Gram-positive bacteria, (Messner *et al.* 2009). Although, in some cases S-layer proteins are also able to recognize the PG layer devoid of SCWP, like SbsB from *G. stearothermophilus* PV72/p2 (Sára *et al.* 1998b) and SpaA from *P. alvei* CCVM 2051^T (Janesch *et al.* 2013b). In addition, to an anchoring function, SCWPs also influence the refolding and the self-assembling process of S-layer proteins *in vitro*. The SCWP of *G. stearothermophilus* PV72/p2 could inhibit the formation of self-assembly products and kept the extracted S-layer protein in a water-soluble state (Sára *et al.* 1998a).

A similar effect was observed for the S-layer protein of *A. thermoaerophilus*, which was retained in a water-soluble state after extraction from its natural source as a consequence of its interaction with the SCWP (Steindl *et al.* 2002).

However, upon dissociation of the SCWP from the S-layer protein, the self-assembly process was fully reestablished.

As was shown in the present study, purified, native *P. alvei* CCM 2051^T SCWP also influences the water-solubility of recombinant SpaA self-assembly products. A significant water-solubility effect on rSpaA self-assembling products could be achieved with SpaA:SCWP (w/w) ratios between 1:5 and 1:10 as shown in **Table 3**; already a 1:1 (w/w) ratio influences the solubility of the rSpaA self-assembly products, yielding solubility above 50%. On the other hand, no significant solubility effects were observed below these ratios (**Table 3**). Thus, solubilization was attempted using an rSpaA-SCWP 1:1 (w/w) ratio to obtain three-dimensional crystals of full-size rSpaA. The rSpaA-SCWP complex crystals obtained with under this condition were about 0.05 mm long (data not shown, personal communication with Blackler RB and Dr. Evans SV), however the crystals did not diffract properly and were very difficult to reproduce. The low reproducibility in these crystallization attempts is probably a consequence of both the heterogeneity of the SCWP as obtained from its natural source and the high SCWP:rSpaA molar ratio required to solubilize the S-layer protein. It is likely that the SCWP interacts in a non-ordered manner and has a non-specific coating effect of the protein. The SCWP must bind to freely accessible binding sites of SpaA self-assembly products, meaning that it cannot interact with SpaA monomers or disintegrate the SpaA oligomers. This non-specific coating effect would result in loosely packed regions within a disordered SCWP-SpaA complex, producing non-diffractive crystals. Therefore, the subsequent strategy was to create a truncated version of the S-layer SpaA that does not follow the self-assembly route, but still achieves a stable interaction with the SCWP.

5.2 Truncated variant $SpaA_{SLH}$ did not produce self-assembly products

A truncated variant of the *P. alvei* CCM 2051^T S-layer protein SpaA that does no longer self-assemble was created, since all attempts to crystallize a full-size

SpaA-SCWP complex did neither produce diffractive nor reproducible crystals. The truncated variant, termed $SpaA_{SLH}$, is a polypeptide comprising residue Val^{21} to residue Gly^{193}, and is deprived of the SpaA self-assembly domain (**Figure 9A**). $SpaA_{SLH}$ was expressed in *E. coli* BL21 as a soluble protein and could be recovered from the cytoplasmic fraction with excellent purity. Moreover, $SpaA_{SLH}$ folded correctly and showed the expected α-helical content (**Figure 9B, C**), which is also in agreement with the secondary structure of the truncated variant of *B. anthracis* S-layer protein Sap (Kern *et al.* 2011), which served as a reference protein in our studies.

5.3 Structure of $SpaA_{SLH}$ and location of the TRAE conserved motifs

Display of S-layer proteins through the interaction of their SLH domains with the corresponding SCWPs is one of the conserved mechanisms for protein cell surface display that Gram-positive bacteria have developed in the course of evolution. Interestingly, SLH domains share low sequence homology but maintain a number of conserved residues, *exempli gratia* the TRAE and GIIxG motifs and Trp and Asp residues (**Figure 15**) (Engelhardt and Peters 1998, Zona and Janeček 2005). Among all these residues, until now the TRAE motif has attracted most attention in several studies to elucidate its function in SCWP binding (May *et al.* 2006, Kern *et al.* 2011, Janesch *et al.* 2013b). The *P. alvei* CCM 2051^T SpaA S-layer protein has one TRAE motif and two naturally mutated variants, the TVEE and TRAQ motifs, at the N-terminus.

$SpaA_{SLH}$ possesses a similar overall fold to the three-pronged spindle trimer described for the truncated variant Sap_{SLH} of *B. anthracis* (Kern *et al.* 2011), where each SLH domain participates with one core helix to create a three-helix bundle core, and a second helix rotated approximately 90° from the bundle core to form the three prongs (**Figure 10**). Interestingly, this "pseudo-trimer" fold seems to be a structurally conserved carbohydrate-binding module in nature to anchor proteins on the surface of Gram-positive bacteria, as surface proteins of

C. difficile remarkably share a similar triangular trimer fold (**Figure 4F**) (Usenik *et al.* 2017).

The highly conserved TRAE, TVEE and TRAQ motifs of SpaA are located at the N-termini of the core helices, adjacent to the grooves formed by the three prongs, named in this work G1, G2 and G3 (**Figure 10A** and **11A**). The motifs contribute unequally to the architecture of the three grooves, as the side chain of Arg^{61} of TRAE motif (SLH_1) and Arg^{177} of TRAQ motif (SLH_3) thread under the connecting loops of the adjacent SLH domains into G2 and G1, respectively. In the case of the TVEE motif (SLH_2), the shorter side chain of Val^{125} does not protrude into the neighboring G3. The final residue of each motif, Glu^{63}, Glu^{127} and Gln^{179} lines the groove of its originating SLH domain. Additional residues contribute to the architecture of each groove, in that way G1 and G2 are the most similar grooves, and therefore are expected to serve as binding grooves of SpaA protein to its SCWP ligand (**Figure 11A-D**).

This assumption coincides with previous mutagenesis studies of full-size SpaA, which showed that mutation of the TRAE, TVEE and TRAQ motifs to TAAA provoked a decrease in SCWP binding to 63%, 12% and 50%, respectively, compared to wild-type SpaA (Janesch *et al.* 2013b). The TRAE motif of SLH_1 contributes with Arg^{61} to G2 and Glu^{63} to G1, therefore its mutation affects both grooves, consistent with the largest reduction in binding described by the previous study. The TVEE motif of SLH_2 lacks the conserved arginine residue, and instead threads Val^{125} to G3, but contributes with Glu^{127} to G2, which forms a salt bridge with Arg^{61}. Mutation of TVEE to TAAA has a minor effect in SCWP binding, suggesting that a structural role of Glu^{127} in G2 is not critical. The TRAQ motif of SLH_3 projects Arg^{177} into G1 in a similar manner as Arg^{61} in G2, and Gln^{179} of TRAQ lies in G3. Therefore, the contribution of the TRAQ motif to only one of the putative binding grooves explains the lesser effect of the TRAQ/TAAA mutation in comparison to the TRAE/TAAA mutation in SCWP binding (Janesch *et al.* 2013b).

5.4 The conserved TRAE motif contributes to SLH three-pronged folding

Attempts to overexpress $SpaA_{SLH}/TAAA_{12}$ (where the TRAE in SLH_1 and TVEE in SLH_2 are both mutated to TAAA) and $SpaA_{SLH}/TAAA_{13}$ (where the TRAE in SLH_1 and TRAQ in SLH_3 are both mutated to TAAA) mutants in *E. coli* were performed. Sufficient amounts of both proteins were overexpressed, however, SDS-PAGE analysis revealed that both proteins accumulate in the insoluble fraction (**Figure 20**). This finding is in contrast to a previous study by Janesch and co-workers which produced these mutants of full-size SpaA (Janesch *et al.* 2013b). In that case, the full-size of SpaA may have preserved the solubility of the double or even the triple mutant. Similarly, significantly reduced solubility has also been reported for *T. thermosulfurigenes* XynA TAAA mutants compared to wild-type XynA. However, in that case, the mutant proteins were recovered in sufficient yield to perform peptidoglycan binding assays (May *et al.* 2006).

The reduced solubility of $SpaA_{SLH}$ as a result of the mutation of residues of the conserved TRAE motifs implies that in addition of being essential for binding SCWP, these residues are also important for the SLH domain structure. Thus, simultaneous mutation of the TRAE motifs may affect the proper folding of the SLH three-pronged trimer. The presence of oligomeric SLH domains may also guarantee adequate steric requirements and folding for S-layer protein binding to the SCWP.

5.5 Synthetic *P. alvei* CCM 2051^T SCWP ligands bind to $SpaA_{SLH}$

The co-crystal structure of $SpaA_{SLH}$ with the synthetic monosaccharide 4,6-Pyr-β-D-ManNAc-*O*-Me shows ligand bound in groove G2 in a narrow pocket formed by conserved residues from all three SLH domains: Arg^{61} (TRAE motif) of SLH_1; Met^{107}, Gln^{108}, Gly^{109} (second glycine of GIIxG motif), Lys^{110}, Glu^{127} (TVEE motif) and Lys^{130} of SLH_2; and Trp^{151} of SLH_3 (**Figure 13A**). The pyruvate moiety binds deeply in G2 through salt bridge interactions to Arg^{61} and Lys^{130} and hydrogen bonds to the main-chain N of Gln^{108} and Gly^{109}. The nitrogen

atom of ManNAc hydrogen bonds to the main-chain O of Gly[109], while the hydrophobic face of the ring stacks against Trp[151].

Residues 44-54 of the connecting loop between the two helices of SLH_1 adjacent to groove G1 displayed multiple conformations in unliganded and liganded co-structures of $SpaA_{SLH}$. Thus, the effect of this disorder on potential ligand binding in G1 was investigated by co-crystallizing the $SpaA_{SLH}$ with the monosaccharide ligand in different space groups. Space groups $P2_12_12_1$, P1 and C2 showed different conformations of residues 44-54. A similar ligand binding in G2 is observed in all cases with the exception of the C2 structure, which displays a sulfate ion coordinated to Lys[130] and ligand in G2 (**Figure 13A**). Only the C2 structure shows an electron density of a partially bound ligand in pocket G1 of molecule A in the AU, whereas G1 of molecule B is occluded by molecule A. This weakly bound 4,6-Pyr-β-D-ManNAc-*O*-Me ligand shows similar interactions to the bound ligand in groove G2 (**Figure 13B**). The pyruvate moiety binds to Arg[177] through salt bridge interactions and hydrogen bonds to the main-chain N of Ser[45] and Gly[46], similarly to Arg[61], Gln[108] and Gly[109] in pocket G2. However, in this case Trp[93] does not form stacking interactions with the ManNAc ring.

On the other hand, isothermal titration calorimetry (ITC) was used to investigate the binding behavior of the synthetic SCWP ligands to $SpaA_{SLH}$ in more similar conditions to the biological context. The ITC studies displayed a 1:1 binding stoichiometry of $SpaA_{SLH}$ to 4,6-Pyr-β-D-ManNAc-*O*-Me with a strong affinity for the ligand, and a K_d of 29 nM (**Table 5**). The observed discrepancy in the stoichiometry with the weakly bound monosaccharide in groove G1 is most likely an artifact of crystal packing interactions that do not take place in solution.

A similar effect occurred with the synthetic disaccharide β-D-GlcNAc-(1→3)-4,6-Pyr-β-D-ManNAc-*O*-Me, which does not show binding to $SpaA_{SLH}$ in solution (**Figure 12B**) but was observed bound in the co-crystal structures. The

disaccharide ligand bound in groove G2, with the 4,6-Pyr-β-D-ManNAc-*O*-Me moiety interacting in a similar manner to the monosaccharide ligand. In contrast, the GlcNAc moiety displayed fragmented electron density, suggesting two different conformations in binding in each protein molecule in the AU (Supplementary information **Fig. S6A-D**), neither of which forms additional hydrogen bonds with the protein (**Figure 14A, B**). Moreover, binding of the disaccharide produced disorder in loop residues 139-152 besides G2 of molecule B (out of the two molecules in the AU), whereas these residues are ordered in all other structures of unliganded and liganded $SpaA_{SLH}$. This loop contains Trp^{151}, which usually forms stacking interactions with the ManNAc ring, thus preventing the formation of these stabilizing interactions. Reiteratively, the crystal packing forces most likely enhance the binding of the disaccharide ligand to the protein in the crystalline context.

5.6 The S-layer protein SpaA of *P. alvei* CCM 2051^T recognizes the terminal non-reducing end of its native SCWP

The structures of complex $SpaA_{SLH}$ with the monosaccharide and disaccharide ligands together with the ITC analyses showed that the 4,6-Pyr-β-D-ManNAc-*O*-Me moiety may serve as the major binding determinant for surface anchoring of the S-layer protein SpaA. The synthetic disaccharide used in this study, which possesses the 4,6-Pyr-β-D-ManNAc-*O*-Me moiety at the reducing end, represents an internal disaccharide subunit of *P. alvei* CCM 2051^T SCWP rather than the native one. In the co-crystal structure of $SpaA_{SLH}$ with the synthetic disaccharide, no additional hydrogen bonds are observed for the GlcNAc moiety, but instead residues 139-152 besides binding pocket G2 are disordered. In addition, binding of this disaccharide to $SpaA_{SLH}$ was not observed in solution (**Figure 12B**). Together these results infer that SpaA may specifically recognize the 4,6-Pyr-β-D-ManNAc-*O*-Me moiety at the distal end rather than an internal moiety of the SCWP as binding epitope.

5.7 The conserved glycine-back bond flip

SLH domains possess the highly conserved GIIxG motif in addition to the TRAE motif (**Figure 15**), a motif which has not been investigated so far. The second glycine is the most conserved residue of the SLH domain profile, but no function has been assigned to this motif in SCWP binding so far (Engelhardt and Peters 1998, Zona and Janeček 2005). Upon ligand binding in groove G2 of $SpaA_{SLH}$ the φ (phi) angle of Gly^{109} and the ψ (psi) angle of Gln^{108} each rotated approximately 180° (**Table 6**). The main-chain O of Gln^{108} is replaced in space by the main-chain N of Gly^{109}, which now forms a hydrogen bond to the pyruvyl carboxyl oxygen of 4,6-Pyr-β-D-ManNAc-*O*-Me moiety. A similar flip also occurs in pocket G1 upon ligand binding, but in this case for residues Gly^{46} and Ser^{45} (**Figure 16**). These observations together with the high conservation of this glycine residue implies that the main chain dihedral angle turn upon ligand binding is a conserved mechanism. A glycine residue is required at this position due to its flexibility, as a different residue would be disallowed for the φ-ψ angle rotation.

To investigate this hypothesis, the $SpaA_{SLH}$/G109A mutant was created, and its binding to 4,6-Pyr-β-D-ManNAc-*O*-Me was then analyzed by ITC. Surprisingly, far from being inactive, the $SpaA_{SLH}$/G109A mutant displayed a 1:1 binding stoichiometry to the synthetic monosaccharide with an affinity (K_d) of 226 nM (**Table 5**). Furthermore, the co-crystal structure of $SpaA_{SLH}$/G109A mutant with 4,6-Pyr-β-D-ManNAc-*O*-Me revealed an unexpected and remarkable structural change. The mutation of the conserved Gly^{109} to alanine certainly inactivates the binding pocket G2. However, in response, the protein meaningfully alters its conformation in order to bind deeply the ligand into the G1 pocket, where the rotation of the φ-ψ angle between Ser^{45} and Gly^{46} is observed (**Figure 16B, 18A** to **C**; **Table 6**). This co-crystal structure further emphasizes the importance of a glycine-back bond flip in SCWP binding.

The double mutant SpaA$_{SLH}$/G46A/G109A conclusively demonstrated the signifycance of the glycine-back bond flip for SCWP binding, as ITC analysis with this mutant does not reveal binding to 4,6-Pyr-β-D-ManNAc-*O*-Me (**Figure 17B**).

Still, a co-crystal structure of SpaA$_{SLH}$/G46A/G109A mutant with bound 4,6-Pyr-β-D-ManNAc-*O*-Me in groove G2 was attained, even when crystallized with a deficient ligand to protein molar ratio, indicating once more that the affinity for the synthetic monosaccharide ligand is significantly increased in the crystalline context in comparison to that in solution. A similar scenario was observed with the crystal structure in space group C2 of complex SpaA$_{SLH}$ with 4,6-Pyr-β-D-ManNAc-*O*-Me. While the crystalline context showed a 2:1 binding, the ITC analyses display unequivocally a 1:1 binding to monosaccharide, and with the observation of a complex SpaA$_{SLH}$ with the disaccharide despite no binding to that ligand in solution.

Nevertheless, in the co-crystal structure of the double mutant with 4,6-Pyr-β-D-ManNAc-*O*-Me the φ angle of Ala^{109} and the ψ angle of Gln^{108} did not rotate upon ligand binding as in the wild-type SpaA$_{SLH}$ protein. As a consequence, the main-chain O of Gln^{108} is oriented towards the pyruvyl carboxyl moiety of 4,6-Pyr-β-D-ManNAc-*O*-Me, producing a slight displacement of the bound ligand in G2 (**Figure 19B**).

5.8 *Paenibacillus alvei* CCM 2051^{T} SpaA$_{SLH}$ has two mutually-exclusive SCWP binding sites

The single mutant SpaA$_{SLH}$/G109A shows a binding stoichiometry of 1:1 to 4,6-Pyr-β-D-ManNAc-*O*-Me with an affinity (K_d) of 226 nM, 7.8 fold weaker compared to the 29 nM K_d of SpaA$_{SLH}$ for the same ligand in G2, that now can be attributed to the bound ligand in G1 (**Figure 13A**, **18B**; **Table 5**). These data jointly indicate that both grooves G1 and G2 can bind *P. alvei* CCM 2051^{T} SCWP with high affinity, but not simultaneously. Therefore, SpaA$_{SLH}$ most

likely reveals a unique form of negative cooperativity in ligand binding, where an important structural change modulates two mutually-exclusive binding sites for the same ligand.

The observed 7.8 fold difference in K_d between the two grooves, with nearly identical interactions made to the ligand, can be attributed to an energetic penalty of the structural change correlated to binding in G1, which in any case is a moderate penalty given to the magnitude of that change. This suggests that SpaA, and likely other proteins bearing SLH domains; have evolved by a mutually-exclusive SCWP binding strategy that is energetically feasible.

In the $SpaA_{SLH}$ protein, Trp^{151} stacks against the hydrophobic face of 4,6-Pyr-β-D-ManNAc-*O*-Me bound in the primary pocket G2, which emphasizes its function through its conservation (**Figure 13A**). Meanwhile, Trp^{93} in the secondary pocket G1 of $SpaA_{SLH}$ is located too distant to form stacking interactions with bound ligand (**Figure 13B**), however, upon the inactivation of the G2 pocket by the $Gly^{109}Ala$ mutation, a significant structural change occurred that narrows pocket G1 and relocates Trp^{93} to form stacking interactions with bound ligand (**Figure 18B**).

The conservation of Gly^{46}, Trp^{93} and Arg^{177} in G1 strongly indicates that $SpaA_{SLH}$ certainly uses G1 as binding groove. Therefore, the presence of highly conserved tryptophan residues in addition to the highly conserved glycine and arginine residues in both binding pockets G1 and G2 of $SpaA_{SLH}$ together with the moderate energy barrier of mutually-exclusive binding, implies an important biological role in surface protein anchoring. For instance, relocation of bound S-layer proteins among different SCWP strands could attenuate the strain of maintaining the lattice symmetry and size of the unit cell of the S-layer during cell growth or division, or could allow mobility of bound SLH-proteins on the cell surface. Electron microscopy of freeze-etched cell preparations has revealed S-layer lattice faults and dislocations in local regions of the S-layer lattice, which are apparently needed for a complete coverage of a

curved cell surface (Pum *et al.* 1991). These lattice faults introduce local stress that is mildly attenuated by forming invaginations on the cell surface, but may also be alleviated by the relocation of bound SLH-proteins through the switchable binding observed in this work for $SpaA_{SLH}$. Lattice faults are also hypothesized as sites for incorporation of new S-layer units or as initiation points for the process of cell division, which would also benefit from relocation in the anchoring of neighboring morphological units (Sleytr and Messner 1989).

5.9 SpaA *O*-glycosylation sites are contiguous to SCWP binding grooves

At least two *O*-glycosylation sites are known for *P. alvei* CCM 2051^T S-layer protein SpaA, namely Tyr^{47} and Tyr^{155}. SpaA is glycosylated with 23 repeating units (on average) of the glycan [→3)-β-D-Gal*p*-(1→)-[α-D-Glc*p*-(1→6)]→4)-β-D-Man*p*NAc-(1→] that are linked by the adaptor -[GroA-2→OPO2→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 the S-layer protein (Altman *et al.* 1991, Messner *et al.* 1995, Zarschler *et al.* 2010b). The crystal structures of $SpaA_{SLH}$ revealed that the tyrosine residues lie adjacent to the binding pockets G1 and G2 (**Figure 10A**), therefore it is likely that their glycosylation could influence SCWP binding. It is noteworthy to mention that Tyr^{47} lies within the connecting loop between the two helices of SLH_1, which was observed in multiple conformations, thus its glycosylation may stabilize this loop and affects G1 structure. Remarkably, groove G3, which lacks a known glycosylation site, has revealed no SCWP binding so far.

5.10 Insights into the structure and function of *P. alvei* CCM 2051^T CsaB

It is now widely recognized that non-covalent binding of SLH domains to pyruvylated secondary cell wall polysaccharides of the bacterial cell wall is an evolutionarily conserved mechanism for protein surface anchoring in Gram-positive bacteria (Cava *et al.* 2004, Messner *et al.* 2009). In a previous study, it was demonstrated that the product of the gene *csaB* is responsible for transfer

of pyruvate to a peptidoglycan-linked polysaccharide and that this pyruvylated polysaccharide is the ligand for *B. anthracis* SLH-domain carrying proteins (Mesnage *et al.* 2000). In *B. anthracis*, the *csaB* gene is part of the cell surface anchoring AB (*csaAB*) operon and lies next to the *B. anthracis* genes *sap* and *eag*, whose S-layer proteins possess three consecutive SLH domains. Similar operons have also been identified in *L. sphaericus* CCM 2177 (Pleschberger *et al.* 2013), *T. thermophilus* (Cava *et al.* 2004) and *P. alvei* CCM 2051^T (Zarschler *et al.* 2010a). On the other hand, no real function has been attributed yet to the product of gene *csaA*. However, because of the high hydrophobicity of CsaA, which has 14 putative transmembrane helical segments, it is suggested to function as an oligosaccharide transporter by flipping lipid-linked oligosaccharides across the cytoplasmic membrane (Mesnage *et al.* 2000). The putative protein encoded by *ORF1* from *P. alvei* CCM 2051^T SCWP biosynthesis locus (**Figure 7**), is also highly hydrophobic and possesses 9 transmembrane segments. Thus, the protein encoded by *ORF1* may be involved in the biosynthesis of polysaccharides in *P. alvei* CCM 2051^T.

Further studies revealed that the SCWPs of *L. sphaericus* CCM 2177 and *P. alvei* CCM 2051^T consist of ManNAc and GlcNAc in a 1:1 molar ratio where each second β-D-ManNAc is modified at position O-4 and O-6 by a ketal pyruvate in *L. sphaericus* CCM 2177 (Ilk *et al.* 1999) and each β-D-ManNAc residue in the case of *P. alvei* CCM 2051^T (Schäffer *et al.* 2000b). On the contrary, in the *B. anthracis* SCWP, which was released from the peptidoglycan by treatment with 48% HF acid, only one pyruvate substituent was identified at O-4 and O-6 of a terminal β-D-ManNAc residue (Forsberg *et al.* 2012). The *B. anthracis* SCWP terminates with the epitope 4,6-Pyr-β-D-ManNAc-(1→4)-3-*O*-acetyl-β-D-GlcNAc-(1→6)-α-D-Glc-NH_2-(1→.

Paenibacillus alvei CCM 2051^T CsaB shares significant homology at the protein sequence level with other CsaB orthologs from *L. sphaericus*, *B. anthracis, B. cereus* and other firmicutes (**Figure 21**), as well as similar properties such as hydrophobicity and molecular mass. Furthermore, the

BLAST and the conserved domain search indicated that the *P. alvei* CCM 2051^T CsaB protein belongs to the polysaccharide pyruvyl transferase family (PF04230.11) (**Figure 22**). Interestingly, the BLAST prediction also revealed that the pyruvyl transferase domain is classified as a glycosyltransferase GT-B type fold. This is confirmed by the fold prediction analysis, which shows that *P. alvei* CCM 2051^T CsaB models are all composed of two globular domains, connected by a flexible linker, where each domain consists of a single parallel β-sheet flanked by several α-helices, resembling the Rossmann-fold that is frequently found in several nucleotide binding proteins (Rossmann *et al.* 1974, Lesk 1995), including glycosyltransferases type-B fold. Each model was constructed with a sequence protein coverage above 90% and a confidence value of 99% based on the solved crystal structure of different glycosyltransferases. It is noteworthy to mention that the crystal structure of the pyruvyl transferase Pvg1p from the fission yeast *S. pombe*, which transfers a ketal pyruvate at positions O-4 and O-6 of a β-D-galactose residues and is the only crystallized pyruvyl transferase thus far, exhibits an overall fold similar to the type-B fold, that is adopted by some glycosyltransferases, including sialyltransferases (**Figure 27A**) (Higuchi *et al.* 2016). These findings imply that glycosyltransferases type-B and pyruvyl transferases may have evolved from common ancestor enzymes.

Paenibacillus alvei CCM 2051^T CsaB was successfully produced recombinantly in *E. coli* cells and purified from the soluble fraction of cell crude extracts, implying that rCsaB was overexpressed as soluble protein. This in agreement with the fact that the *P. alvei* CCM 2051^T CsaB lacks a signal peptide and, with the cell distribution found for *L. sphaericus* CCM 2177 CsaB, which is described as a cytoplasmic protein (Pleschberger *et al.* 2013). After purification of *P. alvei* rCsaB, the secondary structure of purified protein was determined. The far-UV CD analysis showed that rCsaB contains α-helical and β-sheets structures (**Figure 24C**), which is in accordance with the secondary structure prediction performed by fold recognition.

In the present study, we tried to evidence that recombinant *P. alvei* CCM 2051^{T} CsaB is able to transfer a pyruvyl moiety from PEP to *p*NP-β-ManNAc. Pyruvylation of hexoses is not restricted to prokaryotes, in a former study it was demonstrated *in vitro* that *S. pombe* Pvg1p is able to transfer a pyruvate ketal from PEP to *p*NP-β-Gal, producing a 4,6-Pyr-β-D-galactose (Yoritsune *et al.* 2013). In this fission yeast, pyruvylation of *N*-linked glycans at terminal galactose residues is important for cell-cell recognition events in sexual and non-sexual agglutination of the yeast (Yoritsune *et al.* 2013).

The RP-HPLC analyses of the reaction mixtures of rCsaB incubated with PEP and *p*NP-β-ManNAc revealed the presence of a new peak with a shorter retention time than the *p*NP-β-ManNAc substrate. This result is agreement with the expected retention time of pyruvylated *p*NP-β-ManNAc, as this modification would convert the substrate into a more polar molecule with less affinity for the hydrophobic column used in this analysis. In addition, the putative *P. alvei* CCM 2051^{T} CsaB pyruvyltransferase displayed specificity for its substrates, while rCsaB recognized *p*NP-β-ManNAc and PEP, it did not produce an additional peak when incubated with *p*NP-β-GlcNAc or PA, suggesting that the β-linked ManAc is an essential structure for recognition by CsaB (**Figure 25 A-F**). However, the identity of this additional peak could not be determined by NMR spectroscopy due to the low yield of material. Even longer incubation times or the addition of co-factors (Mg^{+2}) did not improve the product formation. Likewise, the pyruvyl transferase activity of rCsaB could not be determined through the phosphate release method due to high background signals, which arise from acid hydrolysis of PEP by the reagent used to detect free phosphate.

The scarce transferase activity display of rCsaB can be attributed to the absence of a known acceptor substrate. Now, it is recognized that the *csaB* gene, together with the *tagO* and *tagA* genes comprise the SCWP biosynthesis gene locus of *P. alvei* CCM 2051^{T}, which is located upstream of the S-layer gene spaA (**Figure 7**) (Zarschler *et al.* 2010a). The *tagO* and *tagA* orthologs are required

for the synthesis of the ManNAc-GlcNAc linkage unit of teichoic acids (Zhang *et al.* 2006, Brown *et al.* 2013). *Paenibacillus alvei* CCM 2051^T SCWP also has the ManNAc-GlcNAc disaccharide in its structure, but as a repeating unit. Therefore, it is very likely that CsaB requires the ManNAc-GlcNAc lipid-linked disaccharide as an acceptor substrate for pyruvylation, although it is not known at which stage of polymer biosynthesis the transfer of pyruvate to ManNAc residues takes place in the cell.

The crystal structure of Pvg1p revealed that the residues Arg^{217}, Leu^{338} and His^{339} involved in binding of PEP are located in a positively charged region within the cleft between the N- and C-terminus (Higuchi *et al.* 2016) (**Figure 27A, B**). Remarkably, the predicted three-dimensional structure of *P. alvei* CCM 2051^T CsaB displays equivalent positively charged residues in the same region (Figure 27C, D). Furthermore, these similar residues, Arg^{148}, Leu^{307} and His^{308}, are within two highly conserved motifs of the pyruvyl transferase CsaB orthologs from different *Bacillaceae*, the xVRD and xMRLH motifs, respectively (**Figure 21**), suggesting their involvement in the transfer of pyruvate to ManNAc residues in the biosynthesis of the SCWP in *P. alvei* CCM 2051^T and other *Bacillaceae*. Moreover, these residues are potential candidates to elucidate the pyruvyl transfer mechanism of *P. alvei* CCM 2051^T. On the other hand, no equivalent residues possibly involved in binding of the acceptors substrate could be assigned in *P. alvei* CCM 2051^T CsaB when compared to Pvg1p.

This may likely reside in the different substrates used by both enzymes, whereas Pvg1p pyruvylates terminal β-D-galactose residues, CsaB presumably pyruvylates each β-D-ManNAc residue of *P. alvei* CCM 2051^T secondary cell wall polysaccharide (Schäffer *et al.* 2000b).

6 Conclusions and perspectives

In the course of this thesis, two significant features of the S-layer protein anchoring system of the model Gram-positive bacterium *Paenibacillus alvei* CCM 2051^T were investigated.

The three-dimensional structure of a truncated variant of the S-layer protein SpaA of *P. alvei* CCM 2051^T comprising its three surface layer homology (SLH) domains was elucidated, revealing that SpaA$_{SLH}$ possesses a three-pronged spindle overall fold similar to the structure described for the SLH domains of *B. anthracis* surface array protein Sap (Kern *et al.* 2011). This triangular fold seems to be a highly conserved structure of carbohydrate-binding modules in Gram-positive bacteria, as a similar overall fold only recently has been as well described for the surface proteins of *C. difficile* (Usenik *et al.* 2017). The three consecutive SLH domains of SpaA possess a highly conserved TRAE motif and two naturally mutated variants TVEE and TRAQ motifs that are critical to anchor the S-layer protein onto the cell wall of *P. alvei* CCM 2051^T (Janesch *et al.* 2013b). In addition to the crystal structure of SpaA$_{SLH}$, the binding mechanism underlying SpaA and its peptidoglycan-linked ligand, a pyruvylated SCWP, was determined at a biophysical and an atomic level.

The crystal structures presented in this work together with the binding analyses of SpaA$_{SLH}$ and the mutants SpaA$_{SLH}$/G109A and SpaA$_{SLH}$/G46A/G109A with synthetic SCWP fragments demonstrate the functional contribution of various conserved SLH domains residues, particularly Gly^{46}, Gly^{109}, Arg^{61}, Arg^{177}, Trp^{93} and Trp^{151}, which are all distributed along the three SLH domains and contribute to generate the two binding grooves of SpaA$_{SLH}$. SpaA$_{SLH}$ seems to be specific for the distal Pyr-β-D-ManNAc moiety of the *P. alvei* CCM 2051^T SCWP, since the presence of a non-reducing end GlcNAc (which represents an alternative *P. alvei* CCM 2051^T SCWP repeating unit) does not produce additional hydrogen bonds,

destabilizes the primary binding groove G2 and abolishes binding in solution (ITC experiments).

Furthermore, the role of the most conserved residues within the SLH domains and until now completely unexplored residues, namely Gly^{46} and Gly^{109}, was determined. The flexibility of Gly^{46} and Gly^{109} permits a peptide bond flip upon ligand binding in the two binding sites of $SpaA_{SLH}$ that is most likely critical for SCWP binding. The co-crystal structures of $SpaA_{SLH}$ variants together with the ITC analyses reveal that $SpaA_{SLH}$ possesses two mutually-exclusive binding sites that may represent a novel binding mechanism within Gram-positive bacteria, where either G1 or G2 can bind 4,6-Pyr-β-D-ManNAc-*O*-Me with high affinity, but not at the same time. This was evidenced by the $SpaA_{SLH}$/G109A mutant, of which mutation Gly^{109} to alanine prevents the peptide bond flip and abolishes binding of the synthetic monosaccharide SCWP in pocket G2, but stunningly permits binding in pocket G1 through a significant structural change.

The conservation of the aforementioned glycine, arginine and tryptophan residues in both binding grooves emphatically demonstrates that *P. alvei* CCM 2051^T utilizes both sites with a mutually-exclusive binding mechanism, which may have a biological significance allowing mobility of bound SLH-proteins on the cell surface to relieve S-layer strain during cell growth and division (Sleytr and Messner 1989).

Secondly, the mechanism of *P. alvei* CCM 2051^T SCWP biosynthesis was explored during this work, in particular the transfer of pyruvate to the SCWP. Pyruvylation confers a net negative charge to the cell surface of prokaryotes and eukaryotes, which has a significant biological consequence, such as in protein surface anchoring in bacteria (Mesnage *et al.* 2000, Cava *et al.* 2004) or in cell-cell recognition in eukaryotes (Yoritsune *et al.* 2013). The anchoring of SLH domain-carrying proteins on the cell surface through pyruvylated SCWP is an ancestral conserved mechanism in Gram-positive bacteria (Cava *et al.*

2004). *P. alvei* CCM 2051^T produces SLH-proteins and possesses the genes necessary to synthesizes pyruvylated SCWP (Zarschler *et al.* 2010a).

The SCWP biosynthesis locus, in *P. alvei* CCM 2051^T located upstream of the S-layer structural gene *spaA*, contains the genes *tagO*, *tagA* and *csaB*. TagO and TagA are expected to build up the SCWP disaccharide unit ManNAc-GlcNAc similar to the teichoic acid linkage-unit, whereas CsaB is predicted to transfer pyruvate to the ManNAc residues. In the course of this work, it was demonstrated that *P. alvei* CCM 2051^T CsaB is a cytoplasmic protein that can be heterologously produced as a soluble protein. The bioinformatics analyses at protein sequence and structural level strongly indicate that CsaB may be involved in the cell wall metabolism in *P. alvei* CCM 2051^T. CsaB shares high sequence homology to other CsaB orthologs from different Gram-positive bacteria, and remarkably its predicted three-dimensional structure displays high similarity to the glycosyltransferase type-B fold solved structure of the yeast pyruvyl transferase Pvg1p (Higuchi *et al.* 2016). Both proteins possess comparable residues with similar spatial orientation; that are involved in PEP binding. Interestingly, these residues are highly conserved among various CsaB orthologs in *Bacillaceae*, making them ideal candidates for prospective mutational studies to understand the substrate specificity of CsaB on a structural basis.

Even though the pyruvyl transferase activity of *P. alvei* CCM 2051^T CsaB could not be conclusively determined, the biochemical characterization shows that CsaB specifically recognizes the ManNAc moiety as acceptor substrate for the transfer of pyruvate within the SCWP ManNAc-GlcNAc unit. The observed low CsaB activity might be due to the inavailability of the correct acceptor molecule, the ManNAc-GlcNAc lipid-linked component. Mimicking the synthesis of *P. alvei* CCM 2051^T SCWP *in vitro* by using the TagO, TagA and CsaB proteins may clarify, if transfer of pyruvate occurs before or after polymerization of the ManNAc-GlcNAc unit.

Finally, since pyruvylation and sialylation of oligosaccharides provides negative charges, it is conceivable that both modifications may endow similar functional effects, making pyruvylation a promising alternative to engineer novel glycopeptides with therapeutic applications.

7 References

Acosta F, De Pedro MA and Berenguer J. 2012. Homogeneous incorporation of secondary cell wall polysaccharides to the cell wall of *Thermus thermophilus* HB27. Extremophiles 16: 485-495.

Adachi S. 1965. Thin-layer chromatography of carbohydrates in the presence of bisulfite. J Chromatogr A 17: 295-299.

Adams PD, Afonine PV, Bunkoczi G, Chen VB, Davis IW, Echols N, Headd JJ, Hung LW, Kapral GJ, Grosse-Kunstleve RW, Mccoy AJ, Moriarty NW, Oeffner R, Read RJ, Richardson DC, Richardson JS, Terwilliger TC and Zwart PH. 2010. PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr D Biol Crystallogr 66: 213-221.

Albers SV and Meyer BH. 2011. The archaeal cell envelope. Nat Rev Microbiol 9: 414-426.

Altman E, Brisson JR, Messner P and Sleytr UB. 1990. Chemical characterization of the regularly arranged surface layer glycoprotein of *Clostridium thermosaccharolyticum* D120-70. Eur J Biochem 188: 73-82.

Altman E, Brisson JR, Messner P and Sleytr UB. 1991. Structure of the glycan chain from the surface layer glycoprotein of *Bacillus alvei* CCM 2051. Biochem Cell Bio 69: 72-78.

Altman E, Schäffer C, Brisson JR and Messner P. 1996. Isolation and characterization of an amino sugar-rich glycopeptide from the surface layer glycoprotein of *Thermoanaerobacterium thermosaccharolyticum* E207-71. Carbohydr Res 295: 245-253.

Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W and Lipman DJ. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25: 3389-3402.

Amano K, Hazama S, Araki Y and Ito E. 1977. Isolation and characterization of structural components of *Bacillus cereus* AHU 1356 cell walls. Eur J Biochem 75: 513-522.

Antikainen J, Anton L, Sillanpää J and Korhonen TK. 2002. Domains in the S-layer protein CbsA of *Lactobacillus crispatus* involved in adherence to collagens, laminin and lipoteichoic acids and in self-assembly. Mol Microbiol 46: 381-394.

Anzengruber J, Pabst M, Neumann L, Sekot G, Heinl S, Grabherr R, Altmann F, Messner P and Schäffer C. 2014. Protein *O*-glucosylation in *Lactobacillus buchneri*. Glycoconj J 31: 117-131.

Araki Y and Ito E. 1989. Linkage units in cell walls of Gram-positive bacteria. Crit Rev Microbiol 17: 121-135.

Ash C, Priest FG and Collins MD. 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.

Baranova E, Fronzes R, Garcia-Pino A, Van Gerven N, Papapostolou D, Péhau-Arnaudet G, Pardon E, Steyaert J, Howorka S and Remaut H. 2012. SbsB structure and lattice reconstruction unveil Ca^{2+} triggered S-layer assembly. Nature 487: 119-122.

Baumeister W and Engelhardt H 1987. Three-dimensional structure of bacterial surface layers. *In*: Harris JR and Horne RW (eds), Electron Microscopy of Proteins, vol. 6. Membraneous Structure. Academic Press, London, pp. 109-154.

Bayer EA, Belaich JP, Shoham Y and Lamed R. 2004. The cellulosomes: multienzyme machines for degradation of plant cell wall polysaccharides. Annu Rev Microbiol 58: 521-554.

Beveridge TJ. 1994. Bacterial S-layers. Curr Opin Struct Biol 4: 204-212.

Bock K, Schuster-Kolbe J, Altman E, Allmaier G, Stahl B, Christian R, Sleytr UB and Messner P. 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. J Biol Chem 269: 7137-7144.

Boot HJ and Pouwels PH. 1996. Expression, secretion and antigenic variation of bacterial S-layer proteins. Mol Microbiol 21: 1117-1123.

Bradford M. 1976. A rapid and sensitive method for the quantitation of microgram quantities utilizing the principle of protein dye binding. Anal Biochem 72: 248-254.

Brechtel E and Bahl H. 1999. In *Thermoanaerobacterium thermosulfurigenes* EM1 S-layer homology domains do not attach to peptidoglycan. J Bacteriol 181: 5017-5023.

Brechtel E, Matuschek M, Hellberg A, Egelseer EM, Schmid R and Bahl H. 1999. Cell wall of *Thermoanaerobacterium thermosulfurigenes* EM1: isolation of its components and attachment of the xylanase XynA. Arch Microbiol 171: 159-165.

Breitwieser A, Iturri J, Toca-Herrera JL, Sleytr UB and Pum D. 2017. In vitro characterization of the two-stage non-classical reassembly pathway of S-layers. Int J Mol Sci 18: 400.

Brown S, Santa Maria Jr. PJ and Walker S. 2013. Wall teichoic acids of Gram-positive bacteria. Annu Rev Microbiol 67: 313-336.

Cava F, De Pedro MA, Schwarz H, Henne A and Berenguer J. 2004. Binding to pyruvylated compounds as an ancestral mechanism to anchor the outer envelope in primitive bacteria. Mol Microbiol 52: 677-690.

Cejka Z and Baumeister W. 1987. Three-dimensional structure of the surface protein of *Clostridium thermosaccharolyticum*. FEMS Microbiol Lett 44: 13-18.

Chauvaux S, Matuschek M and Beguin P. 1999. Distinct affinity of binding sites for S-layer homologous domains in *Clostridium thermocellum* and *Bacillus anthracis* cell envelopes. J Bacteriol 181: 2455-2458.

Cheng HR and Jiang N. 2006. Extremely rapid extraction of DNA from bacteria and yeasts. Biotechnol Lett 28: 55-59.

Choudhury B, Leoff C, Saile E, Wilkins P, Quinn CP, Kannenberg EL and Carlson RW. 2006. The structure of the major cell wall polysaccharide of *Bacillus anthracis* is species-specific. J Biol Chem 281: 27932-27941.

Chung S, Shin SH, Bertozzi CR and De Yoreo JJ. 2010. Self-catalyzed growth of S layers via an amorphous-to-crystalline transition limited by folding kinetics. Proc Natl Acad Sci U S A 107: 16536-16541.

Cohen I, Ron IG and Ben-Jacob E. 2000. From branching to nebula patterning during colonial development of the *Paenibacillus alvei* bacteria. Physica A 286: 321-336.

Crooks GE, Hon G, Chandonia J-M and Brenner SE. 2004. WebLogo: a sequence logo generator. Genome Res 14: 1188-1190.

Delcour J, Ferain T, Deghorain M, Palumbo E and Hols P 1999. The biosynthesis and functionality of the cell wall of lactic acid bacteria. *In*: Konings W, Kuipers OP and Huis in't Veld JHJ (eds), Lactic Acid Bacteria: Genetics, Metabolism and Applications. Springer-Science+Business Media, Dordrecht, The Nederlands, pp. 159-184.

Desvaux M, Dumas E, Chafsey I and Hébraud M. 2006. Protein cell surface display in Gram-positive bacteria: from single protein to macromolecular protein structure. FEMS Microbiol Lett 256: 1-15.

Djukic M, Becker D, Poehlein A, Voget S and Daniel R. 2012. Genome sequence of *Paenibacillus alvei* DSM 29, a secondary invader during European foulbrood outbreaks. J Bacteriol 194: 6365-6365.

Ebner A, Kienberger F, Huber C, Kamruzzahan ASM, Pastushenko VP, Tang JL, Kada G, Gruber HJ, Sleytr UB, Sára M and Hinterdorfer P. 2006. Atomic-force-microscopy imaging and molecular-recognition-force

microscopy of recrystallized heterotetramers comprising an S-layer-streptavidin fusion protein. Chembiochem 7: 588-591.

Egelseer EM, Leitner K, Jarosch M, Hotzy C, Zayni S, Sleytr UB and Sára M. 1998. The S-layer proteins of two *Bacillus stearothermophilus* wild-type strains are bound via their N-terminal region to a secondary cell wall polymer of identical chemical composition. J Bacteriol 180: 1488-1495.

Eichler J and Adams MWW. 2005. Posttranslational protein modification in *Archaea*. Microbiol Mol Biol Rev 69: 393-425.

Eichler J and Maupin-Furlow J. 2013. Post-translation modification in Archaea: lessons from *Haloferax volcanii* and other haloarchaea. FEMS Microbiol Rev 37: 583-606.

Emsley P, Lohkamp B, Scott WG and Cowtan K. 2010. Features and development of *Coot*. Acta Crystallogr D Biol Crystallogr 66: 486-501.

Endl J, Seidl P, Fiedler F and Schleifer K. 1984. Determination of cell wall teichoic acid structure of staphylococci by rapid chemical and serological screening methods. Arch Microbiol 137: 272-280.

Engelhardt H and Peters J. 1998. Structural research on surface layers: A focus on stability, surface layer homology domains, and surface layer-cell wall interactions. J Struct Biol 124: 276-302.

Evans SV. 1993. *SETOR*: Hardware-lighted three-dimensional solid model representations of macromolecules. J Mol Graph 11: 134-138.

Fagan RP, Albesa-Jove D, Qazi O, Svergun DI, Brown KA and Fairweather NF. 2009. Structural insights into the molecular organization of the S-layer from *Clostridium difficile*. Mol Microbiol 71: 1308-1322.

Fagan RP and Fairweather NF. 2011. *Clostridium difficile* has two parallel and essential Sec secretion systems. J Biol Chem 286: 27483-27493.

Ferner-Ortner J, Mader C, Ilk N, Sleytr UB and Egelseer EM. 2007. High-affinity interaction between the S-layer protein SbsC and the secondary cell wall polymer of *Geobacillus stearothermophilus* ATCC 12980 determined by surface plasmon resonance technology. J Bacteriol 189: 7154-7158.

Finn RD, Bateman A, Clements J, Coggill P, Eberhardt RY, Eddy SR, Heger A, Hetherington K, Holm L, Mistry J, Sonnhammer ELL, Tate J and Punta M. 2014. Pfam: the protein families database. Nucleic Acids Res 42: D222-D230.

Ford MJ, Nomellini JF and Smit J. 2007. S-layer anchoring and localization of an S-layer-associated protease in *Caulobacter crescentus*. J Bacteriol 189: 2226-2237.

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

Forsberg LS, Abshire TG, Friedlander A, Quinn CP, Kannenberg EL and Carlson RW. 2012. Localization and structural analysis of a conserved pyruvylated epitope in *Bacillus anthracis* secondary cell wall polysaccharides and characterization of the galactose-deficient wall polysaccharide from avirulent *B. anthracis* CDC 684. Glycobiology 22: 1103-1117.

Forsberg LS, Choudhury B, Leoff C, Marston CK, Hoffmaster AR, Saile E, Quinn CP, Kannenberg EL and Carlson RW. 2011. Secondary cell wall polysaccharides from *Bacillus cereus* strains G9241, 03BB87 and 03BB102 causing fatal pneumonia share similar glycosyl structures with the polysaccharides from *Bacillus anthracis*. Glycobiology 21: 934-948.

Forsgren E. 2010. European foulbrood in honey bees. J Invertebr Pathol 103: S5-S9.

Fünfhaus A, Ashiralieva A, Borriss R and Genersch E. 2009. Use of suppression subtractive hybridization to identify genetic differences between differentially virulent genotypes of *Paenibacillus larvae*, the etiological agent of American foulbrood of honeybees. Environ Microbiol Rep 1: 240-250.

Fünfhaus A and Genersch E. 2012. Proteome analysis of *Paenibacillus larvae* reveals the existence of a putative S-layer protein. Environ Microbiol Rep 4: 194-202.

Ganeshapillai J, Vinogradov E, Rousseau J, Weese JS and Monteiro MA. 2008. *Clostridium difficile* cell-surface polysaccharides composed of pentaglycosyl and hexaglycosyl phosphate repeating units. Carbohydr Res 343: 703-710.

Garcia-Gonzalez E, Müller S, Ensle P, Süssmuth RD and Genersch E. 2014a. Elucidation of sevadicin, a novel non-ribosomal peptide secondary metabolite produced by the honey bee pathogenic bacterium *Paenibacillus larvae*. Environ Microbiol 16: 1297-1309.

Garcia-Gonzalez E, Müller S, Hertlein G, Heid N, Süssmuth RD and Genersch E. 2014b. Biological effects of paenilamicin, a secondary metabolite antibiotic produced by the honey bee pathogenic bacterium *Paenibacillus larvae*. Microbiologyopen 3: 642-656.

Genersch E. 2010. American foulbrood in honeybees and its causative agent, *Paenibacillus larvae*. J Invertebr Pathol 103 Suppl 1: S10-19.

Genersch E, Forsgren E, Pentikäinen J, Ashiralieva A, Rauch S, Kilwinski J and Fries I. 2006. Reclassification of *Paenibacillus larvae* subsp. pulvifaciens and *Paenibacillus larvae* subsp. larvae as *Paenibacillus*

larvae without subspecies differentiation. Int J Syst Evol Microbiol 56: 501-511.

Ghuysen J-M and Strominger JL. 1963a. Structure of the cell wall of *Staphylococcus aureus*, strain Copenhagen. I. Preparation of fragments by enzymatic hydrolysis. Biochemistry 2: 1110-1119.

Ghuysen J-M and Strominger JL. 1963b. Structure of the cell wall of *Staphylococcus aureus*, strain Copenhagen. II. Separation and structure of disaccharides. Biochemistry 2: 1119-1125.

Ginsberg C, Zhang YH, Yuan YQ and Walker S. 2006. *In vitro* reconstitution of two essential steps in wall teichoic acid biosynthesis. ACS Chem Biol 1: 25-28.

Godzik A 2005. Fold recognition methods. *In*: Bourne PE and Weissig H (eds), Structural Bioinformatics, vol. 44. John Wiley & Sons, Hoboken, NJ, USA, pp. 525-546.

Györvary ES, Stein O, Pum D and Sleytr UB. 2003. Self-assembly and recrystallization of bacterial S-layer proteins at silicon supports imaged in real time by atomic force microscopy. J Microsc 212: 300-306.

Haft DH, Payne SH and Selengut JD. 2012. Archaeosortases and exosortases are widely distributed systems linking membrane transit with posttranslational modification. J Bacteriol 194: 36-48.

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

Hastie AT and Brinton CC. 1979. Specific interaction of the tetragonally arrayed protein layer of *Bacillus sphaericus* with its peptidoglycan sacculus. J Bacteriol 138: 1010-1021.

Higuchi Y, Yoshinaga S, Yoritsune K-I, Tateno H, Hirabayashi J, Nakakita S-I, Kanekiyo M, Kakuta Y and Takegawa K. 2016. A rationally engineered yeast pyruvyltransferase Pvg1p introduces sialylation-like properties in neo-human-type complex oligosaccharide. Sci Rep 6: 26349.

Hoch JA and Demoss RD. 1965. Physiological effects of a constitutive tryptophanase in *Bacillus alvei*. J Bacteriol 90: 604-610.

Horejs C, Pum D, Sleytr UB, Peterlik H, Jungbauer A and Tscheliessnig R. 2010. Surface layer protein characterization by small angle x-ray scattering and a fractal mean force concept: From protein structure to nanodisk assemblies. J Chem Phys 133: 175102.

Horejs C, Pum D, Sleytr UB and Tscheliessnig R. 2008. Structure prediction of an S-layer protein by the mean force method. J Chem Phys 128: 065106.

Houwink A. 1953. A macromolecular mono-layer in the cell wall of *Spirillum* spec. Biochim Biophys Acta 10: 360-366.

Houwink A and Le Poole J. 1952. Eine Struktur in der Zellmembran einer Bakterie. Physikalische Verhandlungen 3: 98.

Hovmöller S, Sjögren A and Wang DN. 1988. The structure of crystalline bacterial surface-layers. Prog Biophys Mol Biol 51: 131-163.

Howorka S, Sára M, Wang YJ, Kuen B, Sleytr UB, Lubitz W and Bayley H. 2000. Surface-accessible residues in the monomeric and assembled forms of a bacterial surface layer protein. J Biol Chem 275: 37876-37886.

Huber C, Ilk N, Rünzler D, Egelseer EM, Weigert S, Sleytr UB and Sára M. 2005. The three S-layer-like homology motifs of the S-layer protein SbpA of *Bacillus sphaericus* CCM 2177 are not sufficient for binding to the pyruvylated secondary cell wall polymer. Mol Microbiol 55: 197-205.

Ilk N, Kosma P, Puchberger M, Egelseer EM, Mayer HF, Sleytr UB and Sára M. 1999. Structural and functional analyses of the secondary cell wall polymer of *Bacillus sphaericus* CCM 2177 that serves as an S-layer-specific anchor. J Bacteriol 181: 7643-7646.

Janesch B, Koerdt A, Messner P and Schäffer C. 2013a. The S-layer homology domain-containing protein SlhA from *Paenibacillus alvei* CCM 2051^T is important for swarming and biofilm formation. PLoS One 8: e76566.

Janesch B, Messner P and Schäffer C. 2013b. Are the surface layer homology domains essential for cell surface display and glycosylation of the S-layer protein from *Paenibacillus alvei* CCM 2051^T? J Bacteriol 195: 565-575.

Janesch B, Schirmeister F, Maresch D, Altmann F, Messner P, Kolarich D and Schäffer C. 2016. Flagellin glycosylation in *Paenibacillus alvei* CCM 2051^T. Glycobiology 26: 74-87.

Jarosch M, Egelseer EM, Huber C, Moll D, Mattanovich D, Sleytr UB and Sára M. 2001. Analysis of the structure-function relationship of the S-layer protein SbsC of *Bacillus stearothermophilus* ATCC 12980 by producing truncated forms. Microbiology 147: 1353-1363.

Jarosch M, Egelseer EM, Mattanovich D, Sleytr UB and Sára M. 2000. The S-layer gene *sbsC* of *Bacillus stearothermophilus* ATCC 12980: molecular characterization and heterologous expression in *Escherichia coli*. Microbiology 146 273-281.

Kandiba L, Guan Z and Eichler J. 2013. Lipid modification gives rise to two distinct *Haloferax volcanii* S-layer glycoprotein populations. Biochim Biophys Acta-Biomembranes 1828: 938-943.

Kelley LA, Mezulis S, Yates CM, Wass MN and Sternberg MJE. 2015. The *Phyre2* web portal for protein modeling, prediction and analysis. Nat Protoc 10: 845-858.

Kern J, Wilton R, Zhang RG, Binkowski TA, Joachimiak A and Schneewind O. 2011. Structure of surface layer homology (SLH) domains from *Bacillus anthracis* surface array protein. J Biol Chem 286: 26042-26049.

Kim Y-G, Lee J-H, Cho MH and Lee J. 2011. Indole and 3-indolylacetonitrile inhibit spore maturation in *Paenibacillus alvei*. BMC Microbiol 11: 119-129.

Kinns H, Badelt-Lichtblau H, Egelseer EM, Sleytr UB and Howorka S. 2010. Identifying assembly-inhibiting and assembly-tolerant sites in the SbsB S-layer protein from *Geobacillus stearothermophilus*. J Mol Biol 395: 742-753.

Kinns H and Howorka S. 2008. The surface location of individual residues in a bacterial S-layer protein. J Mol Biol 377: 589-604.

Kojima N, Araki Y and Ito E. 1985. Structure of the linkage units between ribitol teichoic acids and peptidoglycan. J Bacteriol 161: 299-306.

Konrad Z and Eichler J. 2002. Lipid modification of proteins in Archaea: attachment of a mevalonic acid-based lipid moiety to the surface-layer glycoprotein of *Haloferax volcanii* follows protein translocation. Biochem J 366: 959-964.

Korkmaz N, Ostermann K and Rödel G. 2011. Calcium dependent formation of tubular assemblies by recombinant S-layer proteins *in vivo* and *in vitro*. Nanotechnology 22: 095601.

Labischinski H and Maidhof H 1994. Bacterial peptidoglycan: overview and evolving concepts. *In*: Ghuysen JM, Hakenbeck R (eds), New Comprehensive Biochemistry, vol. 27. Elsevier, Amsterdam, pp. 23-38.

Laemmli UK. 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature 227: 680-685.

Lang W, Glassey K and Archibald A. 1982. Influence of phosphate supply on teichoic acid and teichuronic acid content of *Bacillus subtilis* cell walls. J Bacteriol 151: 367-375.

Lanzetta PA, Alvarez LJ, Reinach PS and Candia OA. 1979. An improved assay for nanomole amounts of inorganic phosphate. Anal Biochem 100: 95-97.

Lechner J and Wieland F. 1989. Structure and biosynthesis of prokaryotic glycoproteins. Annu Rev Biochem 58: 173-194.

Lemaire M, Miras I, Gounon P and Béguin P. 1998. Identification of a region responsible for binding to the cell wall within the S-layer protein of *Clostridium thermocellum*. Microbiology 144: 211-217.

Leoff C, Choudhury B, Saile E, Quinn CP, Carlson RW and Kannenberg EL. 2008. Structural elucidation of the nonclassical secondary cell wall polysaccharide from *Bacillus cereus* ATCC 10987. Comparison with the polysaccharides from *Bacillus anthracis* and *B. Cereus* type strain ATCC 14579 reveals both unique and common structural features. J Biol Chem 283: 29812-29821.

Lesk AM. 1995. NAD-binding domains of dehydrogenases. Curr Opin Struct Biol 5: 775-783.

Lopez AE, Moreno-Flores S, Pum D, Sleytr UB and Toca-Herrera JL. 2010. Surface dependence of protein nanocrystal formation. Small 6: 396-403.

Lupas A, Engelhardt H, Peters J, Santarius U, Volker S and Baumeister W. 1994. Domain structure of the *Acetogenium kivui* surface layer revealed by electron crystallography and sequence analysis. J Bacteriol 176: 1224-1233.

Mader C, Huber C, Moll D, Sleytr UB and Sára M. 2004. Interaction of the crystalline bacterial cell surface layer protein SbsB and the secondary cell wall polymer of *Geobacillus stearothermophilus* PV72 assessed by real-time surface plasmon resonance biosensor technology. J Bacteriol 186: 1758-1768.

Marchler-Bauer A, Derbyshire MK, Gonzales NR, Lu SN, Chitsaz F, Geer LY, Geer RC, He J, Gwadz M, Hurwitz DI, Lanczycki CJ, Lu F, Marchler GH, Song JS, Thanki N, Wang ZX, Yamashita RA, Zhang DC, Zheng CJ and Bryant SH. 2015. CDD: NCBI's conserved domain database. Nucleic Acids Res 43: D222-D226.

Marchler-Bauer A, Lu SN, Anderson JB, Chitsaz F, Derbyshire MK, Deweese-Scott C, Fong JH, Geer LY, Geer RC, Gonzales NR, Gwadz M, Hurwitz DI, Jackson JD, Ke ZX, Lanczycki CJ, Lu F, Marchler GH, Mullokandov M, Omelchenko MV, Robertson CL, Song JS, Thanki N, Yamashita RA, Zhang DC, Zhang NG, Zheng CJ and Bryant SH. 2011. CDD: a Conserved Domain Database for the functional annotation of proteins. Nucleic Acids Res 39: D225-D229.

May A, Pusztahelyi T, Hoffmann N, Fischer RJ and Bahl H. 2006. Mutagenesis of conserved charged amino acids in SLH domains of *Thermoanaerobacterium thermosulfurigenes* EM1 affects attachment to cell wall sacculi. Arch Microbiol 185: 263-269.

McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC and Read RJ. 2007. *Phaser* crystallographic software. J Appl Crystallogr 40: 658-674.

Mescher MF and Strominger JL. 1976. Purification and characterization of a prokaryotic glycoprotein from cell-envelope of *Halobacterium salinarium*. J Biol Chem 251: 2005-2014.

Mesnage S, Fontaine T, Mignot T, Delepierre M, Mock M and Fouet A. 2000. Bacterial SLH domain proteins are non-covalently anchored to the cell surface via a conserved mechanism involving wall polysaccharide pyruvylation. EMBO J 19: 4473-4484.

Mesnage S, Tosi-Couture E, Mock M and Fouet A. 1999. The S-layer homology domain as a means for anchoring heterologous proteins on the cell surface of *Bacillus anthracis*. J Appl Microbiol 87: 256-260.

Messner P. 1997. Bacterial glycoproteins. Glycoconj J 14: 3-11.

Messner P, Allmaier G, Schäffer C, Wugeditsch T, Lortal S, König H, Niemetz R and Dorner M. 1997. Biochemistry of S-layers. FEMS Microbiol Rev 20: 25-46.

Messner P, Christian R, Neuninger C and Schulz G. 1995. Similarity of "core" structures in 2 different glycans of tyrosine-linked eubacterial S-layer glycoproteins. J Bacteriol 177: 2188-2193.

Messner P, Egelseer EM, Sleytr UB and Schäffer C 2009. Surface layer glycoproteins and secondary cell wall polymers. *In*: Moran AP, Brennan PJ, Holst O, and von Itzstein M (eds), Microbial Glycobiology: Structures, Relevance and Applications. Academic Press-Elsevier, San Diego, CA, USA, pp. 109-128

Messner P, Pum D, Sára M, Stetter KO and Sleytr UB. 1986. Ultrastructure of the cell-envelope of the archaebacteria *Thermoproteus tenax* and *Thermoproteus neutrophilus*. J Bacteriol 166: 1046-1054.

Messner P, Schäffer C and Kosma P. 2013. Bacterial cell-envelope glycoconjugates. Adv Carbohydr Chem Biochem 69: 209-272.

Messner P and Sleytr UB. 1991a. Bacterial surface layer glycoproteins. Glycobiology 1: 545-551.

Messner P and Sleytr UB 1991b. The use of freeze-etching and freeze-drying to evaluate crystalline cell surface layers (S-layers). *In*: Mozes N, Handley PS, Busscher HJ and Rouxhet PG (eds), Microbial Cell Surface Analysis. Structural and Physicochemical Methods. VHC Publishers, New York, pp. 109-125.

Messner P and Sleytr UB. 1992. Crystalline bacterial cell-surface layers. Adv Microb Physiol 33: 213-275.

Messner P, Sleytr UB, Christian R, Schulz G and Unger FM. 1987. Isolation and structure determination of a diacetamidodideoxyuronic acid-containing glycan chain from the S-layer glycoprotein of *Bacillus stearothermophilus* NRS 2004/3a. Carbohydr Res 168: 211-218.

Messner P, Steiner K, Zarschler K and Schäffer C. 2008. S-layer nanoglycobiology of bacteria. Carbohydr Res 343: 1934-1951.

Messner P, Wellan M, Kubelka W and Sleytr UB. 1993. Reversible cross-linking of crystalline bacterial surface-layer glycoproteins through their glycan chains. Appl Microbiol Biotechnol 40: 7-11.

Moreno-Flores S, Kasry A, Butt HJ, Vavilala C, Schmittel M, Pum D, Sleytr UB and Toca-Herrera JL. 2008. From native to non-native two-dimensional protein lattices through underlying hydrophilic/ hydrophobic nanoprotrusions. Angew Chem Int Ed 47: 4707-4710.

Motomizu S and Li Z-H. 2005. Trace and ultratrace analysis methods for the determination of phosphorus by flow-injection techniques. Talanta 66: 332-340.

Müller DJ, Baumeister W and Engel A. 1996. Conformational change of the hexagonally packed intermediate layer of *Deinococcus radiodurans* monitored by atomic force microscopy. J Bacteriol 178: 3025-3030.

Müller DJ, Baumeister W and Engel A. 1999. Controlled unzipping of a bacterial surface layer with atomic force microscopy. Proc Natl Acad Sci U S A 96: 13170-13174.

Müller S, Garcia-Gonzalez E, Genersch E and Süssmuth RD. 2015. Involvement of secondary metabolites in the pathogenesis of the American foulbrood of honey bees caused by *Paenibacillus larvae*. Nat Prod Rep 32: 765-778.

Müller S, Garcia-Gonzalez E, Mainz A, Hertlein G, Heid NC, Mosker E, Van Den Elst H, Overkleeft HS, Genersch E and Süssmuth RD. 2014. Paenilamicin: structure and biosynthesis of a hybrid nonribosomal peptide/polyketide antibiotic from the bee pathogen *Paenibacillus larvae*. Angew Chem Int Ed 53: 10821-10825.

Murshudov GN, Skubak P, Lebedev AA, Pannu NS, Steiner RA, Nicholls RA, Winn MD, Long F and Vagin AA. 2011. *REFMAC5* for the refinement of macromolecular crystal structures. Acta Crystallogr D Biol Crystallogr 67: 355-367.

Naumova IB, Shashkov AS, Tul'skaya EM, Streshinskaya GM, Kozlova YI, Potekhina NV, Evtushenko LI and Stackebrandt E. 2001. Cell wall teichoic acids: structural diversity, species specificity in the genus *Nocardiopsis*, and chemotaxonomic perspective. FEMS Microbiol Rev 25: 269-283.

Navarre WW and Schneewind O. 1999. Surface proteins of Gram-positive bacteria and mechanisms of their targeting to the cell wall envelope. Microbiol Mol Biol Rev 63: 174-229.

Norville JE, Kelly DF, Knight TF, Belcher AM and Walz T. 2007. 7Å projection map of the S-layer protein SbpA obtained with trehalose-embedded monolayer crystals. J Struct Biol 160: 313-323.

Novotny R, Schäffer C, Strauss J and Messner P. 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.

Olabarria G, Carrascosa JL, De Pedro MA and Berenguer J. 1996. A conserved motif in S-layer proteins is involved in peptidoglycan binding in *Thermus thermophilus*. J Bacteriol 178: 4765-4772.

Otwinowski Z and Minor W. 1997. Processing of X-ray diffraction data collected in oscillation mode. Methods Enzymol 276: 307-326.

Pavkov-Keller T, Howorka S and Keller W. 2011. The structure of bacterial S-layer proteins. Prog Mol Biol Transl Sci 103: 73-130.

Pavkov T, Egelseer EM, Tesarz M, Svergun DI, Sleytr UB and Keller W. 2008. The structure and binding behavior of the bacterial cell surface layer protein SbsC. Structure 16: 1226-1237.

Pavkov T, Oberer M, Egelseer EM, Sára M, Sleytr UB and Keller W. 2003. Crystallization and preliminary structure determination of the C-terminal truncated domain of the S-layer protein SbsC. Acta Crystallogr D Biol Crystallogr 59: 1466-1468.

Petersen BO, Sára M, Mader C, Mayer HF, Sleytr UB, Pabst M, Puchberger M, Krause E, Hofinger A, Duus JØ and Kosma P. 2008. Structural characterization of the acid-degraded secondary cell wall polymer of *Geobacillus stearothermophilus* PV72/p2 Carbohydr Res 343: 1346-1358.

Pleschberger M, Hildner F, Rünzler D, Gelbmann N, Mayer HF, Sleytr UB and Egelseer EM. 2013. Identification of a novel gene cluster in the upstream region of the S-layer gene *sbpA* involved in cell wall metabolism of *Lysinibacillus sphaericus* CCM 2177 and characterization of the recombinantly produced autolysin and pyruvyl transferase. Arch Microbiol 195: 323-337.

Poppinga L, Janesch B, Fünfhaus A, Sekot G, Garcia-Gonzalez E, Hertlein G, Hedtke K, Schäffer C and Genersch E. 2012. Identification and functional analysis of the S-layer protein SplA of *Paenibacillus larvae*, the causative agent of American foulbrood of honey bees. PLoS Pathog 8: e1002716.

Posch G, Pabst M, Brecker L, Altmann F, Messner P and Schäffer C. 2011. Characterization and scope of S-layer protein *O*-glycosylation in *Tannerella forsythia*. J Biol Chem 286: 38714-38724.

Pum D, Messner P and Sleytr UB. 1991. Role of the S-layer in morphogenesis and cell-division of the archaebacterium *Methanocorpusculum sinense*. J Bacteriol 173: 6865-6873.

Pum D and Sleytr UB. 1994. Large-scale reconstitution of crystalline bacterial surface layer proteins at the air-water interface and on lipids. Thin Solid Films 244: 882-886.

Pum D and Sleytr UB. 1995. Anisotropic crystal growth of the S-layer of *Bacillus sphaericus* CCM 2177 at the air/water interface. Colloids Surf A Physicochem Eng Asp 102: 99-104.

Pum D, Toca-Herrera JL and Sleytr UB. 2013. S-layer protein self-assembly. Int J Mol Sci 14: 2484-2501.

Reboli AC, Bryan CS and Farrar WE. 1989. Bacteremia and infection of a hip prosthesis caused by *Bacillus alvei*. J Clin Microbiol 27: 1395-1396.

Reichmann NT and Gründling A. 2011. Location, synthesis and function of glycolipids and polyglycerolphosphate lipoteichoic acid in Gram-positive bacteria of the phylum Firmicutes. FEMS Microbiol Lett 319: 97-105.

Reid CW, Vinogradov E, Li J, Jarrell HC, Logan SM and Brisson J-R. 2012. Structural characterization of surface glycans from *Clostridium difficile*. Carbohydr Res 354: 65-73.

Ries W, Hotzy C, Schocher I, Sleytr UB and Sára M. 1997. Evidence that the N-terminal part of the S-layer protein from *Bacillus stearothermophilus* PV72/p2 recognizes a secondary cell wall polymer. J Bacteriol 179: 3892-3898.

Ristl R, Janesch B, Anzengruber J, Forsthuber A, Blaha J, Messner P and Schäffer C. 2012. Description of a putative oligosaccharyl: S-layer protein transferase from the tyrosine *O*-glycosylation system of *Paenibacillus alvei* CCM 2051^T. Adv Microbiol 2: 537-546.

Roberts IS. 1996. The biochemistry and genetics of capsular polysaccharide production in bacteria. Annu Rev Microbiol 50: 285-315.

Rossmann MG, Moras D and Olsen KW. 1974. Chemical and biological evolution of a nucleotide-binding protein. Nature 250: 194-199.

Rünzler D, Huber C, Moll D, Köhler G and Sára M. 2004. Biophysical characterization of the entire bacterial surface layer protein SbsB and its two distinct functional domains. J Biol Chem 279: 5207-5215.

Sambrook J, Fritsch EF and Maniatis T. 1989. Molecular cloning: A Laboratory Manual, 2nd Ed, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, USA.

Sára M. 2001. Conserved anchoring mechanisms between crystalline cell surface S-layer proteins and secondary cell wall polymers in Gram-positive bacteria? Trends Microbiol 9: 47-48.

Sára M, Dekitsch C, Mayer HF, Egelseer EM and Sleytr UB. 1998a. Influence of the secondary cell wall polymer on the reassembly, recrystallization,

and stability properties of the S-layer protein from *Bacillus stearothermophilus* PV72/p2. J Bacteriol 180: 4146-4153.

Sára M, Egelseer EM, Dekitsch C and Sleytr UB. 1998b. Identification of two binding domains, one for peptidoglycan and another for a secondary cell wall polymer, on the N-terminal part of the S-layer protein SbsB from *Bacillus stearothermophilus* PV72/p2. J Bacteriol 180: 6780-6783.

Sára M, Kuen B, Mayer HF, Mandl F, Schuster KC and Sleytr UB. 1996. Dynamics in oxygen-induced changes in S-layer protein synthesis from *Bacillus stearothermophilus* PV72 and the S-layer-deficient variant T5 in continuous culture and studies of the cell wall composition. J Bacteriol 178: 2108-2117.

Sára M, Küpcü S and Sleytr UB. 1989. Localization of the carbohydrate residue of the S-layer glycoprotein from *Clostridium thermohydrosulfuricum* L111-69. Arch Microbiol 151: 416-420.

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

Sára M and Sleytr UB. 1987a. Molecular sieving through S layers of *Bacillus stearothermophilu*s strains. J Bacteriol 169: 4092-4098.

Sára M and Sleytr UB. 1987b. Production and characteristics of ultrafiltration membranes with uniform pores from two-dimensional arrays of proteins. J Membr Sci 33: 27-49.

Sára M and Sleytr UB. 1996. Crystalline bacterial cell surface layers (S-layers): from cell structure to biomimetics. Prog Biophys Mol Biol 65: 83-111.

Sára M and Sleytr UB. 2000. S-layer proteins. J Bacteriol 182: 859-868.

Schäffer C, Dietrich K, Unger B, Scheberl A, Rainey FA, Kählig H and Messner P. 2000a. A novel type of carbohydrate-protein linkage region in the tyrosine-bound S-layer glycan of *Thermoanaerobacterium thermo-saccharolyticum* D120-70. Eur J Biochem 267: 5482-5492.

Schäffer C, Kählig H, Christian R, Schulz G, Zayni S and Messner P. 1999a. The diacetamidodideoxyuronic-acid-containing glycan chain of *Bacillus stearothermophilus* NRS 2004/3a represents the secondary cell-wall polymer of wild-type *B. stearothermophilus* strains. Microbiology 145: 1575-1583.

Schäffer C and Messner P. 2004. Surface-layer glycoproteins: an example for the diversity of bacterial glycosylation with promising impacts on nanobiotechnology. Glycobiology 14: 31R-42R.

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

Schäffer C and Messner P. 2017. Emerging facets of prokaryotic glycosylation. FEMS Microbiol Rev 41: 49-91.

Schäffer C, Müller N, Christian R, Graninger M, Wugeditsch T, Scheberl A and Messner P. 1999b. Complete glycan structure of the S-layer glycoprotein of *Aneurinibacillus thermoaerophilus* GS4-97. Glycobiology 9: 407-414.

Schäffer C, Müller N, Mandal PK, Christian R, Zayni S and Messner P. 2000b. A pyrophosphate bridge links the pyruvate-containing secondary cell wall polymer of *Paenibacillus alvei* CCM 2051 to muramic acid. Glycoconj J 17: 681-690.

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

Scott JR and Barnett TC. 2006. Surface proteins of Gram-positive bacteria and how they get there. Annu Rev Microbiol 60: 397-423.

Sekot G, Posch G, Messner P, Matejka M, Rausch-Fan X, Andrukhov O and Schäffer C. 2011. Potential of the *Tannerella forsythia* S-layer to delay the immune response. J Dent Res 90: 109-114.

Sekot G, Schuster D, Messner P, Pum D, Peterlik H and Schäffer C. 2013. Small-angle x-ray scattering for imaging of surface layers on intact bacteria in the native environment. J Bacteriol 195: 2408-2414.

Shashkov AS, Kosmachevskaya LN, Streshinskaya GM, Evtushenko LI, Bueva OV, Denisenko VA, Naumova IB and Stackebrandt E. 2002. A polymer with a backbone of 3-deoxy-D-glycero-D-galacto-non-2-ulopyranosonic acid, a teichuronic acid, and a β-glucosylated ribitol teichoic acid in the cell wall of plant pathogenic *Streptomyces sp*. VKM Ac-2124. Eur J Biochem 269: 6020-6025.

Shin SH, Chung S, Sanii B, Comolli LR, Bertozzi CR and De Yoreo JJ. 2012. Direct observation of kinetic traps associated with structural transformations leading to multiple pathways of S-layer assembly. Proc Natl Acad Sci U S A 109: 12968-12973.

Sigrist CJA, De Castro E, Cerutti L, Cuche BA, Hulo N, Bridge A, Bougueleret L and Xenarios I. 2013. New and continuing developments at *PROSITE*. Nucleic Acids Res 41: D344-D347.

Sleytr UB. 1975. Heterologous reattachment of regular arrays of glycoproteins on bacterial surfaces. Nature 257: 400-402.

Sleytr UB. 1976. Self-assembly of the hexagonally and tetragonally arranged subunits of bacterial surface layers and their reattachment to cell walls. J Ultrastruct Res 55: 360-377.

Sleytr UB. 1978. Regular arrays of macromolecules on bacterial cell walls: structure, chemistry, assembly, and function. Int Rev Cytol 53: 1-64.

Sleytr UB. 1997. I. Basic and applied S-layer research: an overview. FEMS Microbiol Rev 20: 5-12.

Sleytr UB and Beveridge TJ. 1999. Bacterial S-layers. Trends Microbiol 7: 253-260.

Sleytr UB and Glauert AM. 1975. Analysis of regular arrays of subunits on bacterial surfaces; evidence for a dynamic process of assembly. J Ultrastruct Res 50: 103-116.

Sleytr UB, Huber C, Ilk N, Pum D, Schuster B and Egelseer EM. 2007. S-layers as a tool kit for nanobiotechnological applications. FEMS Microbiol Lett 267: 131-144.

Sleytr UB and Messner P. 1983. Crystalline surface layers on bacteria. Annu Rev Microbiol 37: 311-339.

Sleytr UB and Messner P 1989. Self-assembly of crystalline bacterial cell surface layers (S-layers), *In*: Plattner H. (ed), Electron Microscopy of Subcellular Dynamics. CRC Press, Boca Raton, FL, USA, pp. 13-31.

Sleytr UB, Messner P and Pum D. 1988. Analysis of crystalline bacterial surface-layers by freeze-etching, metal shadowing, negative staining and ultra-thin sectioning. Method Microbiol 20: 29-60.

Sleytr UB, Messner P, Pum D and Sára M. 1993. Crystalline bacterial cell surface layers. Mol Microbiol 10: 911-916.

Sleytr UB, Messner P, Pum D and Sára M. 1999. Crystalline bacterial cell surface layers (S layers): from supramolecular cell structure to biomimetics and nanotechnology. Angew Chem Int Ed 38: 1034-1054.

Sleytr UB and Sára M. 1986. Ultrafiltration membranes with uniform pores from crystalline bacterial-cell envelope layers. Appl Microbiol Biotechnol 25: 83-90.

Sleytr UB and Sára M. 1997. Bacterial and archaeal S-layer proteins: structure-function relationships and their biotechnological applications. Trends Biotechnol 15: 20-26.

Sleytr UB, Sára M, Pum D, Schuster B, Messner P and Schäffer C 2002. Self-assembling protein systems: microbial S-layers. *In*: Steinbüchel A and Fahnestock SR (eds), Biopolymers, vol. 7. Polyamides and Complex Proteinaceous Matrices I. Wiley-VCH, Weinheim, pp. 285-338.

Sleytr UB, Schuster B, Egelseer EM and Pum D. 2014. S-layers: principles and applications. FEMS Microbiol Rev 38: 823-864.

Sleytr UB and Thorne KJI. 1976. Chemical characterization of regularly arranged surface layers of *Clostridium thermosaccharolyticum* and *Clostridium thermohydrosulfuricum*. J Bacteriol 126: 377-383.

Smit E, Oling F, Demel R, Martinez B and Pouwels PH. 2001. The S-layer protein of *Lactobacillus acidophilus* ATCC 4356: Identification and characterisation of domains responsible for S-protein assembly and cell wall binding. J Mol Biol 305: 245-257.

Smit E and Pouwels PH. 2002. One repeat of the cell wall binding domain is sufficient for anchoring the *Lactobacillus acidophilus* surface layer protein. J Bacteriol 184: 4617-4619.

Soldo B, Lazarevic V, Pagni M and Karamata D. 1999. Teichuronic acid operon of *Bacillus subtilis* 168. Mol Microbiol 31: 795-805.

Steindl C, Schäffer C, Smrečki V, Messner P and Müller N. 2005. The secondary cell wall polymer of *Geobacillus tepidamans* GS5-97T: structure of different glycoforms. Carbohydr Res 340: 2290-2296.

Steindl C, Schäffer C, Wugeditsch T, Graninger M, Matecko I, Müller N and Messner P. 2002. The first biantennary bacterial secondary cell wall polymer and its influence on S-layer glycoprotein assembly. Biochem J 368: 483-494.

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

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

Strominger JL and Ghuysen J-M. 1967. Mechanisms of enzymatic bacteriaolysis. Cell walls of bacteri are solubilized by action of either specific carbohydrases or specific peptidases. Science 156: 213-221.

Sumihiro H and Matsushima Y. 1972. Structural studies on a glucose-containing polysaccharide obtained from cell walls of *Micrococcus lysodeikticus*. J Biochem 72: 1117-1128.

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

Szabo Z and Pohlschroder M. 2012. Diversity and subcellular distribution of archaeal secreted proteins. Front Microbiol 3: 207.

Tang J, Ebner A, Huber C, Ilk N, Zhu R, Pastushenko V, Sára M and Hinterdorfer P. 2007. High resolution atomic force microscopy imaging and single molecule force microscopy studies of S-layer-strep-tag I and S-layer-strep-tag II proteins. Biophys J Suppl S: 513A.

Teixeira LM, Strickland A, Mark SS, Bergkvist M, Sierra-Sastre Y and Batt CA. 2010. Entropically driven self-assembly of *Lysinibacillus sphaericus* S-layer proteins analyzed under various environmental conditions. Macromol Biosci 10: 147-155.

Thornley MJ, Glauert AM and Sleytr UB. 1974. Structure and assembly of bacterial surface-layers composed of regular arrays of subunits. Philos Trans R Soc Lond B Biol Sci 268: 147-153.

Usenik A, Renko M, Mihelič M, Lindič N, Borišek J, Perdih A, Pretnar G, Müller U and Turk D. 2017. The CWB2 cell wall-anchoring module is revealed by the crystal structures of the *Clostridium difficile* cell wall proteins Cwp8 and Cwp6. Structure 25: 514-521.

Varki A, Cummings RD, Aebi M, Packer NH, Seeberger PH, Esko JD, Stanley P, Hart G, Darvill A, Kinoshita T, Prestegard JJ, Schnaar RL, Freeze HH, Marth JD, Bertozzi CR, Etzler ME, Frank M, Vliegenthart JFG, Lutteke T, Perez S, Bolton E, Rudd P, Paulson J, Kanehisa M, Toukach P, Aoki-Kinoshita KF, Dell A, Narimatsu H, York W, Taniguchi N and Kornfeld S. 2015. Essentials of glycobiology [Internet]. 3rd edition. Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press. Appendix 1B, Symbol Nomenclature for Glycans (SNFG). Glycobiology 25: 1323-1324.

Walker SG, Karunaratne DN, Ravenscroft N and Smit J. 1994. Characterization of mutants of *Caulobacter crescentus* defective in surface attachment of the paracrystalline surface layer. J Bacteriol 176: 6312-6323.

Ward JB. 1981. Teichoic and teichuronic acids: biosynthesis, assembly, and location. Microbiol Rev 45: 211-243.

Weidenmaier C and Peschel A. 2008. Teichoic acids and related cell-wall glycopolymers in Gram-positive physiology and host interactions. Nat Rev Microbiol 6: 276-287.

Weygand M, Kjaer K, Howes PB, Wetzer B, Pum D, Sleytr UB and Lösche M. 2002. Structural reorganization of phospholipid headgroups upon recrystallization of an S-layer lattice. J Phys Chem B 106: 5793-5799.

Weygand M, Schalke M, Howes PB, Kjaer K, Friedmann J, Wetzer B, Pum D, Sleytr UB and Lösche M. 2000. Coupling of protein sheet crystals (S-layers) to phospholipid monolayers. J Mater Chem 10: 141-148.

Weygand M, Wetzer B, Pum D, Sleytr UB, Cuvillier N, Kjaer K, Howes PB and Lösche M. 1999. Bacterial S-layer protein coupling to lipids: X-ray reflectivity and grazing incidence diffraction studies. Biophys J 76: 458-468.

Willing SE, Candela T, Shaw HA, Seager Z, Mesnage S, Fagan RP and Fairweather NF. 2015. *Clostridium difficile* surface proteins are anchored to the cell wall using CWB2 motifs that recognise the anionic polymer PSII. Mol Microbiol 96: 596-608.

Winn MD, Ballard CC, Cowtan KD, Dodson EJ, Emsley P, Evans PR, Keegan RM, Krissinel EB, Leslie AGW, Mccoy A, Mcnicholas SJ, Murshudov GN, Pannu NS, Potterton EA, Powell HR, Read RJ, Vagin A and Wilson KS. 2011. Overview of the *CCP4* suite and current developments. Acta Crystallogr D Biol Crystallogr 67: 235-242.

Yoritsune K-I, Matsuzawa T, Ohashi T and Takegawa K. 2013. The fission yeast Pvg1p has galactose-specific pyruvyltransferase activity. FEBS Lett 587: 917-921.

Yue D, Nordhoff M, Wieler LH and Genersch E. 2008. Fluorescence *in situ* hybridization (FISH) analysis of the interactions between honeybee larvae and *Paenibacillus larvae*, the causative agent of American foulbrood of honeybees (*Apis mellifera*). Environ Microbiol 10: 1612-1620.

Zarschler K, Janesch B, Kainz B, Ristl R, Messner P and Schäffer C. 2010a. Cell surface display of chimeric glycoproteins via the S-layer of *Paenibacillus alvei*. Carbohydr Res 345: 1422-1431.

Zarschler K, Janesch B, Pabst M, Altmann F, Messner P and Schäffer C. 2010b. Protein tyrosine *O*-glycosylation–A rather unexplored prokaryotic glycosylation system. Glycobiology 20: 787-798.

Zarschler K, Janesch B, Zayni S, Schäffer C and 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. Appl Environ Microbiol 75: 3077-3085.

Zhang Y-H, Ginsberg C, Yuan Y and Walker S. 2006. Acceptor substrate selectivity and kinetic mechanism of *Bacillus subtilis* TagA. Biochemistry 45: 10895-10904.

Zona R and Janeček Š. 2005. Relationships between SLH motifs from different glycoside hydrolase families. Biologia (Bratisl) 60: 115-121.

8 Supplementary information

Supplementary methods

General synthetic methods

All purchased chemicals were used without further purification unless stated otherwise. Solvents (CH_2Cl_2, DMF[1], pyridine) were dried over activated 4 Å molecular sieves. Dry MeOH (secco solv) was purchased from Merck. Compound **1** and donor **9** were prepared according to literature (Murphy *et al.* 2003, Ellervik and Magnusson 1996). Concentration of organic solutions was performed under reduced pressure < 40 °C. Optical rotations were measured with a Perkin-Elmer 243 B Polarimeter. $[\alpha]_D^{20}$ values are given in units of 10^{-1}deg cm^2g^{-1}. Thin layer chromatography was performed on Merck precoated plates (Vienna, Austria): generally on 5 x 10 cm, layer thickness 0.25 mm, Silica Gel $60F_{254}$; alternatively on HP-TLC plates with 2.5 cm concentration zone (Merck, Vienna, Austria). Spots were detected by dipping reagent (anisaldehyde-H_2SO_4). For column chromatography silica gel (0.040 – 0.063 mm) was used. HP-column chromatography was performed on pre-packed columns (YMC-Pack SIL-06, 0.005 mm, 250x10 mm and 250x20 mm). Size exclusion chromatography was performed on Bio-Gel® P-2 Gel extra fine < 45 µm (wet) (1 x 30 cm). NMR spectra were recorded with a Bruker Avance III 600 instrument (600.22 MHz for ^1H, 150.93 MHz for ^{13}C) using standard Bruker NMR software. ^1H NMR spectra were referenced to 7.26 ($CDCl_3$) and 0.00 (D_2O, external calibration to 2,2-dimethyl-2-silapentane-5-sulfonic acid) ppm. ^{13}C NMR spectra were referenced to 77.00 ($CDCl_3$) and 67.40 (D_2O, external calibration to 1,4-dioxane) ppm. ESI-MS data were obtained on a Waters Micromass Q-TOF Ultima Global instrument (Waters, Elstree, Herts, UK).

Synthesis of mono- and disaccharide ligands

The synthesis of the ligands **8** and **12** is shown in **Figure S1**. The monosaccharide ligand **8** was prepared from known methyl glucoside **1** (Murphy *et al*. 2003). Inversion of configuration at position 2 and introduction of the azide group was achieved by reaction of an intermediate 2-*O*-triflate and subsequent treatment with sodium azide in DMF in 84% yield. Staudinger reduction of the azide with triphenylphosphine afforded the 2-deoxy-2-amino-mannopyranoside **3**. Chromatographic removal of the phosphine-derived products, however, could not be fully accomplished. Hence polymer-bound triphenylphosphine was used. Repeated hydrolysis of the resin was needed in order to fully recover product **3**. *N*-Acetylation of **3** giving **4** and subsequent hydrolysis of the benzylidene group with trifluoroacetic acid proceeded smoothly to afford the 3-*O*-benzyl glycoside **5**. Introduction of the pyruvic acid ketal was achieved by reaction of **5** with methyl pyruvate in the presence of TMSO-triflate (Schüle and Ziegler, 1996) which gave the *S*-configured pyruvyl derivative **6** (53%) and the crystalline debenzylated product **7** (33%). The *S*-configuration of the stereogenic center of the pyruvic acetal was assigned on the basis of the characteristic ^{1}H and ^{13}C NMR chemical shifts (Jansson *et al*. 1993). Hydrogenation of **6** with 10% Pd-C in methanol gave compound **7** which was treated with aqueous NaOH and subjected to purification on BioGel-P2 to give the monosaccharide ligand **8** as sodium salt in 91% yield.

Compound **7** served as glycosyl acceptor in a TMSO-triflate promoted coupling with the *N*-Troc-protected trichloroacetimidate donor **9** (Ellervik and Magnusson, 1993) in MeCN to afford the β-(1→3)-linked disaccharide derivative **10** in modest yield (32%) with additional byproduct formation and with recovery of unreacted acceptor **7** (19%). The low reactivity of pyruvyl-substituted glycosyl acceptors has also been observed in the literature (Schüle and Ziegler, 1996). Troc-removal by reaction with Zn-dust was followed by *N*-acetylation to produce disaccharide **11** in 82% yield. Despite several attempts by HPLC-separations, product **11** could not be fully purified and still

contained appr. 7% of an unknown constituent. Zemplén-de-*O*-acetylation and alkaline hydrolysis afforded the target disaccharide ligand **12** in 87% yield. NMR spectra of ligands **8** and **12** are shown in **Figures S2-S5**.

*Methyl 2-azido-3-*O*-benzyl-4,6-*O*-benzylidene-2-deoxy-β-D-*manno*-pyranoside (2)*

Compound **1** (298 mg, 0.80 mmol) was co-evaporated twice with toluene before dry DCM (5.14 ml) and dry pyridine (2.57 ml) were added. The solution was stirred at 0°C under Ar and triflic anhydride (0.326 ml, 1.936 mmol) was added. After 45 minutes additional triflic anhydride (0.050 ml, 0.297 mmol) was injected into the solution as TLC showed incomplete conversion. The reaction was stopped after 75 min. The mixture was diluted with DCM, washed with saturated aq $NaHCO_3$, dried ($MgSO_4$) and concentrated. The residue was dissolved in anhydrous DMF (8.57 ml). Sodium azide (260.1 mg, 4.00 mmol) was added and the reaction mixture was stirred for 21 hours at 70 °C. The off-white solid was removed by filtration of the reaction mixture over Celite. The filtrate was concentrated and the residue was suspended in DCM. The suspension was filtered once more over Celite and the yellow filtrate was concentrated. Purification of the crude product by column chromatography (toluene-EtOAc 6:1) gave the azido-derivative **2** as light yellow syrup. Yield: 269 mg (84%); R_f 0.61 (toluene-EtOAc 6:1); ^{1}H-NMR (600 MHz, $CDCl_3$): δ 7.50-7.48 and 7.41-7.25 (m, 10H, 1 x CHPh and 1 x OCH_2Ph), 5.60 (s, 1H, CHPh), 4.89 and 4.75 (2d, 2H, ^2J 12.4 Hz, OCH_2Ph), 4.46 (d, 1H, $^3J_{1,2}$ 1.4 Hz, H-1), 4.32 (dd, 1H, $^3J_{5,6a}$ 5.0 Hz, $^2J_{6a,6b}$ 10.5 Hz, H-6a), 4.04 (app. t, 1H, $^3J_{3,4} = {}^3J_{4,5}$ 9.5 Hz, H-4), 3.98 (dd, 1H, $^3J_{2,3}$ 3.7 Hz, H-2), 3.88 (app. t, 1H, H-6b), 3.75 (dd, 1H, H-3), 3.54 (s, 3H, OMe), 3.46 (ddd, 1H, $^3J_{5,6b}$ 10.0 Hz, H-5) ppm. The chemical shifts agree with published data (Augé *et al.* 1980).

Methyl 2-acetamido-3-O-benzyl-4,6-O-benzylidene-2-deoxy-β-D-manno-pyranoside (4)

Compound **2** (787.0 mg, 1.980 mmol) and polymer-bound PPh_3 (1.6 mmol/g resin, 100-200 mesh, 3.094 g, 4.951 mmol) were treated with dry DCM (26.0 ml). The mixture was stirred at room temperature under Ar for 18 h. The polymer was filtered and washed several times with EtOAc, DCM and MeOH. Concentration of the filtrate gave a low amount of the intermediate amine **3** (50 mg). The polymer resin was then suspended in CH_3CN (20.0 ml), H_2O (600 µl, 33.3 mmol) was added and the mixture was stirred for 30 min, which gave 20 mg of **3**. The polymer was then further treated with THF (20.0 ml) and H_2O (360 µl, 20.0 mmol) for 1 h at 55 °C and for 15 h at room temperature. The polymer was washed with EtOAc several times and concentration of the combined filtrates gave the amine **3** (550 mg). Hydrolysis of the polymer was continued under the same conditions and stirring for 2 hours at 55 °C, which provided an additional amount of the **3** (120 mg). An aliquot of **3** (650 mg, 1.75 mmol) was dissolved in pyridine (2.0 ml) and the solution was treated with acetic anhydride (1.3 ml, 13.8 mmol) for 105 min at room temperature under Ar. MeOH was added at 0 °C, the solution was concentrated in vacuo and co-evaporated once with toluene. A second portion of the intermediate **3** (122 mg, 0.328 mmol) was acetylated with pyridine (1.5 ml) and Ac_2O (0.5 ml, 5.29 mmol) for 30 min and processed as described. The combined acetylated crude products were purified by MPLC (EtOAc-toluene 3:1 →EtOAc, flow rate: 60 to 50 ml/min) which afforded **4** as a yellow syrup (615 mg, 75%); R_f 0.42 (EtOAc, HPTLC); $[\alpha]_D^{20}$ - 62.9 (*c* 1.0, $CHCl_3$). ^{1}H-NMR (600 MHz, $CDCl_3$): δ 7.50-7.48 and 7.41-7.29 (m, 10H, 1 x CHPh and 1 x OCH_2Ph), 5.76 (d, 1H, $^3J_{NH,2}$ 9.6 Hz, NH), 5.59 (s, 1H, CHPh), 4.88 (ddd, 1H, $^3J_{1,2}$ 1.9, $^3J_{2,3}$ 4.1, $^3J_{NH,2}$ 9.6 Hz, H-2), 4.81 and 4.66 (2d, 2H, ^2J 12.2 Hz, OCH_2Ph), 4.53 (d, 1H, H-1), 4.35 (dd, 1H, $^3J_{5,6a}$ 4.9, $^2J_{6a,6b}$ 10.4 Hz, H-6a), 3.80 (app. t, 1H, $^3J_{5,6b} = {}^2J_{6a,6b}$ 10.3 Hz, H-6b), 3.79-3.75 (m, 2H, H-3 and H-4), 3.50 (s, 3H, OMe), 3.46 (dt, 1H, $^3J_{4,5} = {}^3J_{5,6b}$ 9.7 Hz, H-5), 2.08 (s, 3H, CH_3CO) ppm. ^{13}C-NMR (150 MHz, $CDCl_3$): δ 170.83 (s,

CH_3CO), 137.93 and 137.35 (2s, 1 x CHPh and 1 x OCH_2Ph), 129.16-126.18 (10 d, 5 x CHPh and 5 x OCH_2Ph), 101.76 (d, CHPh), 101.48 (d, C-1), 79.06 and 75.63 (2d, C-3 and C-4), 71.80 (t, O*C*H_2Ph), 68.92 (t, C-6), 67.16 (d, C-5), 57.15 (q, OMe), 50.03 (q, CH_3CO) ppm. HR-MS: $[M+Na]^+$ *m/z* calcd: 436.1731; found: 436.1715

Methyl 2-acetamido-3-*O*-benzyl-2-deoxy-β-D-*manno*-pyranoside (5)

Compound **4** (605.9 mg, 1.465 mmol) was dissolved in dry DCM (20 ml). Trifluoroacetic acid (12.1 ml, 157.32 mmol) was added to the solution at 0 °C under Ar and the solution was stirred for 1 h. The solution was concentrated and the resulting brown residue was co-evaporated twice with toluene. Purification by column chromatography (EtOAc-toluene 3:1 → EtOAc-EtOH 4:1) furnished unreacted **4** (173 mg, 28.5%) followed by **5** as syrup (296 mg, 62%); R_f 0.33 (EtOAc-EtOH 9:1, HPTLC);$[\alpha]_D^{23}$ – 44.6 (c 0.75, MeOH). ^{1}H-NMR (600 MHz, MeOH-d_4): δ 7.40-7.38, 7.32-7.29 and 7.27-7.24 (m, 5H, OCH_2Ph), 4.82 and 4.48 (2d, 2H, ^2J 11.1 Hz, OCH_2Ph), 4.74 (dd, 1H, $^3J_{1,2}$ 1.5, $^3J_{2,3}$ 4.2 Hz, H-2), 4.52 (d, 1H, H-1), 3.86 (app. d, 2H, $^3J_{5,6a} = {}^3J_{5,6b}$ 3.4 Hz, H-6a and H-6b), 3.61 (app. t, 1H, $^3J_{3,4} = {}^3J_{4,5}$ 9.7 Hz, H-4), 3.75 (dd, 1H, H-3), 3.48 (s, 3H, OMe), 3.26 (dt, 1H, $^3J_{4,5}$ 9.8 Hz, H-5), 2.01 (s, 3H, CH_3CO) ppm. ^{13}C-NMR (150 MHz, MeOH-d_4): 173.87 (s, CH_3CO), 139.64 (s, OCH_2Ph), 129.22 and 128.60 (5 d, OCH_2Ph), 102.02 (d, C-1), 81.60 (d, C-3), 78.18 (d, C-5), 72.20 (t, OCH_2Ph), 67.04 (d, C-4), 61.95 (t, C-6), 57.00 (q, OMe), 50.88 (d, C-2), 22.60 (q, CH_3CO) ppm. HR-MS: $[M+H]^+$ *m/z* calcd: 326.1598; found: 326.1599.

Methyl 2-acetamido-3-*O*-benzyl-2-deoxy-4,6-*O*-[1-(methoxycarbonyl) ethylidene]-β-D-*manno*-pyranoside (6)

Methyl pyruvate (190 µl, 2.108 mmol) and TMSO-triflate (382 µl, 2.108 mmol) were added in succession to a stirred solution of compound **5** (298.2 mg, 0.917 mmol) in CH_3CN (3 ml) at room temperature. Stirring was continued for 105 min when TLC showed formation of undesired side products. The solution was concentrated, diluted with EtOAc, neutralized with Et_3N and concentrated.

The crude mixture was purified by column chromatography (EtOAc-toluene 2.5:1 → EtOAc-EtOH 9:1) which afforded **6** (198.7 mg, 53%) followed by debenzyl derivative **7** (97 mg, 33%). For preparation of an analytically pure sample, aliquots of compound **6** were purified by column chromatography (EtOAc-toluene 5:1) and by HPLC (EtOAc: toluene 6:1, column: YMC 250x20, flow rate: 15ml/min), respectively; R_f 0.74 (EtOAc-EtOH 9:1, HPTLC); $[\alpha]_D^{23}$ – 35.0 (*c* 0.98, $CHCl_3$).^{1}H-NMR (600 MHz, $CDCl_3$): δ 7.41-7.39, 7.35-7.33 and 7.29-7.27 (m, 5H, OCH_2Ph), 5.76 (d, 1H, $^3J_{NH,2}$ 9.5 Hz, NH), 4.77 (app. s, 2H, OCH_2Ph), 4.73 (dd, 1H, $^3J_{1,2}$ 2.0, $^3J_{2,3}$ 5.5 Hz, H-2), 4.47 (d, 1H, H-1), 4.08 (dd, 1H, $^3J_{5,6a}$ 5.0, $^2J_{6a,6b}$ 10.7 Hz, H-6a), 3.84 (s, 3H, CO_2Me), 3.72 (app. t, 1H, $^3J_{4,5}$ 9.2 Hz, H-4), 3.75 (app. t, 1H, $^3J_{5,6b}$ 10.7 Hz, H-6b), 3.74 (dd, 1H, $^3J_{3,4}$ 8.5 Hz, H-3), 3.42 (s, 3H, OMe), 3.38 (m, 1H, H-5), 2.03 (s, 3H, CH_3CO), 1.57 (s, 3H, Me) ppm. ^{13}C-NMR (150 MHz, $CDCl_3$): δ 170.49 and 170.42 (2s*, CH_3CO), 170.13 (s, $MeCCO_2Me$), 138.04 (s, OCH_2Ph), 100.82 (d, C-1), 99.15 (s, $MeCCO_2Me$), 75.81 (d, C-4), 74.52 (d, C-3), 71.44 (t, OCH_2Ph), 65.94 (d, C-5), 65.54 (t, C-6), 56.62 (q, OMe), 52.76 (q, $MeCCO_2Me$), 49.36 and 49.28 (2d*, C-2), 25.61 (q, $MeCCO_2Me$), 23.47 and 23.42 (2q*, CH_3CO) ppm. * Signal duplication was observed. HR-MS: $[M+Na]^+$ *m/z* calcd: 432.1629; found: 432.1627.

*Methyl 2-acetamido-2-deoxy-4,6-***O***-[1-(methoxycarbonyl)ethylidene]-β-D-***manno***-pyranoside (7)***

A solution of **6** (87.0 mg, 0.212 mmol) in dry MeOH was hydrogenated at atmospheric pressure in the presence of 33.6 mg 10% Pd-C for 5.5 h at room temperature. The suspension was filtered over Celite and the filtrate was concentrated. The product **7** was partly crystallized in a mixture of *n*-hexane and EtOAc. The remaining mother liquor was concentrated and purified by column chromatography (EtOAc-MeOH 95:5). Combined yield for **7**: 44.7 mg (66%). Colorless crystals, m.p. 192-195 °C; *Rf* 0.29 (EtOAc-EtOH 9:1, HPTLC), $[\alpha]_D^{20}$ – 49.9 (*c* 0.62, MeOH). ^{1}H-NMR (600 MHz, MeOH-d4): δ 4.58 (d, 1H, $^3J_{1,2}$ 1.9 Hz, H-1), 4.54 (dd, 1H, $^3J_{2,3}$ 4.7 Hz, H-2), 3.99 (dd, 1H, $^3J_{5,6a}$ 5.0 Hz, $^2J_{6a}$,6b 10.5 Hz, H-6a), 3.83 (dd, 1H, $^3J_{3,4}$ 9.7 Hz, H-3), 3.83 (s, 3H, CO_2Me), 3.78 (app.

t, 1H, $^{3}J_{5,6b} = {}^{2}J_{6a,6b}$ 10.5 Hz, H-6b), 3.55 (app. t, 1H, $^{3}J_{3,4} = {}^{3}J_{4,5}$ 9.7 Hz, H-4), 3.44 (s, 3H, OMe), 3.31 (m, 1H, H-5), 2.01 (s, 3H, CH_3CO), 1.48 (s, 3H, Me) ppm.^{13}C-NMR (150 MHz, MeOH-d_4): δ 174.77 (s, CH_3CO), 172.10 (s, $MeCCO_2Me$), 102.60 (d, C-1), 100.70 (s, MeCCO2Me), 76.31 (d, C-4), 71.06 (d, C-3), 68.29 (d, C-5), 65.93 (t, C-6), 57.17 (q, OMe), 54.72 (q, $MeCCO_2Me$), 53.17 (d, C-2), 25.92 (q, $MeCCO_2Me$), 22.64 (q, CH_3CO) ppm. HR-MS: $[M+H]^+$ *m/z* calcd: 320.1340; found: 320.1340.

Methyl 2-acetamido-2-deoxy-4,6-*O*-[1-carboxyethylidene]-β-D-manno-pyranoside sodium salt (8)

A solution of 0.2 M NaOH (1.5 ml) was added to compound **7** (6.2 mg, 0.0194 mmol). The mixture was stirred at room temperature for 3 h 15 min. Dowex® AG1X cation-exchange resin (H^+-form) (Sigma-Aldrich) was added until pH 7.5. The suspension was filtered and the filtrate was lyophilized. Product **8** was desalted by size exclusion chromatography using a BioGel P2-column (5% aq EtOH) and the product fraction was lyophilized. Yield: 5.8 mg of **8** as a colorless solid (91%); R_f 0.94 ($CHCl_3$-MeOH-H_2O 10:10:3); $[\alpha]_D^{20}$ – 34.2 (*c* 0.51, H_2O). ^{1}H-NMR (600 MHz, D_2O): δ 4.71 (d, 1H, $^{3}J_{1,2}$ 1.9 Hz, H-1), 4.51 (dd, 1H, $^{3}J_{2,3}$ 4.6 Hz, H-2), 4.01 (dd, 1H, $^{3}J_{5,6a}$ 5.0, $^{2}J_{6a,6b}$ 10.6 Hz, H-6a), 3.93 (dd, 1H, $^{3}J_{3,4}$ 10.1 Hz, H-3), 3.70 (app. t, 1H, $^{3}J_{5,6b} = {}^{2}J_{6a,6b}$ 10.6 Hz, H-6b), 3.58 (app. t, 1H, $^{3}J_{4,5}$ 9.9 Hz, H-4), 3.45 (s, 3H, OMe), 3.40 (ddd, 1H, H-5), 2.02 (s, 3H, CH_3CO), 1.43 (s, 3H, Me) ppm. ^{13}C-NMR (150 MHz, D_2O): δ 176.36 and 176.27 (2s, $MeCCO_2Na$ and CH_3CO), 102.65 (s, $MeCCO_2Na$), 101.81 (d, C-1), 72.82 (d, C-4), 70.08 (d, C-3), 67.67 (d, C-5), 64.88 (t, C-6), 58.02 (q, OMe), 54.01 (d, C-2), 25.49 (q, $MeCCO_2Na$), 22.80 (q, CH_3CO) ppm. HR-MS: $[M+H]^+$ *m/z* calcd: 328.1003; found: 328.1008.

Methyl [3,4,5-tri-*O*-acetyl-2-deoxy-2-(2,2,2-trichloroethoxycarbonylamino)-β-D-gluco-pyranosyl]-(1→3)-2-acetamido-2-deoxy-4,6-*O*-[1 (methoxycarbonyl) ethylidene]-β-D-manno-pyranoside (10)

A solution of predried donor **9** (151.1 mg, 0.242 mmol) in dry MeCN (6 ml) was added to acceptor **7** (38.6 mg, 0.121 mmol) and powdered acid-washed 4 Å molecular sieves, and the suspension was stirred under Ar at room temperature for 30 min. An aliquot of a 0.242 M TMS-triflate stock solution (100 µl, 0.0242 mmol) was added and stirring was continued for 1 h 45 min. Another aliquot of TMSO-triflate (100 µl, 0.0242 mmol) was added. TLC showed two products and some side products due to hydrolysis. The reaction was stopped after 2 h 50 min by adding a few drops of Et_3N. The suspension was filtered over Celite and the filtrate was concentrated. Purification of the residue by column chromatography (EtOAc-toluene 6:1→ to EtOAc: EtOH 4:1) afforded crude **10** as a colorless syrup (30.2 mg, 32%), an unidentified by-product (9.5 mg) and unreacted acceptor **7** (7.3 mg, 19%). Final purification of **10** was achieved by HPLC (EtOAc→ toluene 2:1, column: YMC 250x10, flow rate: 5 ml/min). R_f = 0.43 (EtOAc→: toluene 6:1); $[\alpha]_D^{20}$ – 38.2 (*c* 0.9, $CHCl_3$). ^{1}H-NMR (600 MHz, $CDCl_3$): δ 6.02 (d, 1H, $^3J_{NH,2'}$ 8.8 Hz, NH'), 5.87 (d, 1H, $^3J_{NH,2}$ 8.2 Hz, $NHCOCH_3$), 5.23 (app. t, 1H, $^3J_{2',3'} = {}^3J_{3',4'}$ 9.8 Hz, H-3'), 5.12 (app. t, 1H, $^3J_{4',5'}$ 9.6 Hz, H-4'), 5.00 (d, 1H, $^3J_{1',2'}$ 8.4 Hz, H-1'), 4.85 and 4.59 (2d, 2H, ^2J = 12.1 Hz, Cl_3CCH_2), 4.64 (m, 1H, H-2), 4.50 (d, 1H, $^3J_{1,2}$ 1.9 Hz, H-1), 4.16 (dd, 1H, $^3J_{5',6a'}$ 4.3, $^2J_{6a',6b'}$ 12.3 Hz, H-6a'), 4.21 (dd, 1H, $^3J_{5',6b'}$ 2.8 Hz, H-6b'), 4.08 (dd, 1H, $^3J_{5,6a}$ 5.0, $^2J_{6a,6b}$ 10.7 Hz, H-6a), 4.05 (dd, 1H, $^3J_{2,3}$ 4.7, $^3J_{3,4}$ 9.4 Hz, H-3), 3.87 (dt, 1H, H-2'), 3.80 (s, 3H, CO_2Me), 3.74 (app. t, 1H, $^3J_{5,6b}$ 10.7 Hz, H-6b), 3.73 (ddd, 1H, $^3J_{4',5'}$ 10.0 Hz, H-5'), 3.54 (app. t, 1H, $^3J_{4,5}$ 9.6 Hz, H-4), 3.46 (s, 3H, OMe), 3.38 (dt, 1H, H-5), 2.09-1.98 (4 s, 12H, 1 x $NHCOCH_3$ and 3 x $OCOCH_3$), 1.56 (s, 3H, Me) ppm. ^{13}C-NMR (150 MHz, $CDCl_3$): δ 172.43-169.65 (5 s, 3 x $OCOCH_3$, 1 x $NHCOCH_3$ and 1 x $NHCOOCH_2CCl_3$), 170.23 (s, $MeCCO_2Me$), 154.69 (s, CH_2CCl_3), 100.90 (d, C-1), 99.55 (s, $MeCCO_2Me$), 96.40 (d, C-1'), 74.56 (t, CH_2CCl_3), 73.70 (d, C-3), 73.37 (d, C-3'), 73.15 (d, C-4), 72.22 (d, C-5'), 69.09 (d, C-4'), 67.10 (d, C-5), 65.25 (t, C-6), 62.32 (t, C-6'), 57.12 (q, OMe), 55.91 (d, C-2'), 52.79 (q, $MeCCO_2Me$), 49.66 (d, C-2), 25.44 (q, $MeCCO_2Me$), 23.72-20.78 (4q, 3 x $OCOCH_3$ and 1 x $NHCOCH_3$) ppm. HR-MS: $[M+H]^+$ *m/z* calcd: 781.1387; found: 781.1371.

Methyl (2-acetamido-3,4,5-tri*-O-*acetyl-2-deoxy-β-D*-gluco-*pyranosyl)-(1→3)-2-acet-amido-2-deoxy-4,6*-O-*[1-(methoxycarbonyl)ethylidene]-β-D*-manno-*pyranoside (11)

Compound **10** (18.1 mg, 0.023 mmol) was dissolved in glacial acetic acid (2.0 ml) and Zn-powder (10 µm, 75.67 mg, 1.157 mmol) was added. The reaction mixture was stirred at room temperature under Ar for 18 h. As conversion was not completed, another portion of Zn-powder was added and stirring was continued for 1 h. The suspension was filtered over Celite and the solids were washed several times with glacial acetic acid. The combined filtrate was concentrated and co-evaporated three times with toluene. The off-white residue was dissolved in pyridine (1.2 ml) and acetic anhydride (600 µl, 6.347 mmol) was added and the solution was stirred for 105 min at room temperature. The reaction was quenched with MeOH (300 µl), solvents were evaporated and the residue was co-evaporated twice with toluene. Purification by column chromatography (EtOAc-MeOH 9.5:0.5 → EtOAc-MeOH 9:1) gave **11** containing a trace impurity (~7%) as a colorless amorphous solid (12.4 mg, 82%); R_f 0.35 (EtOAc-MeOH 9:1, HPTLC); $[\alpha]_D^{20}$ – 53.0 (*c* 0.92, MeOH). ^{1}H-NMR (600 MHz, MeOH-d_4): δ 5.18 (app. t, 1H, $^3J_{2',3'} = {}^3J_{3',4'}$ 9.7 Hz, H-3'), 5.05 (app. t, 1H, $^3J_{4',5'}$ 9.6 Hz, H-4'), 4.82 (d, 1H, H-1'), 4.69 (dd, 1H, $^3J_{1,2}$ 1.6, $^3J_{2,3}$ 4.4 Hz, H-2), 4.56 (d, 1H, H-1), 4.33 (dd, 1H, $^3J_{5',6a'}$ 4.4, $^2J_{6a',6b'}$ 12.2 Hz, H-6a'), 4.18 (dd, 1H, $^3J_{5',6b'}$ 2.7 Hz, H-6b'), 4.08 (dd, 1H, $^3J_{3,4}$ 10.0 Hz, H-3), 3.99 (dd, 1H, $^3J_{5,6a}$ 5.0, $^2J_{6a,6b}$ 10.5 Hz, H-6a), 3.96 (dd, 1H, $^3J_{1',2'}$ 8.2 Hz, H-2'), 3.82 (m, 1H, H-5'), 3.81 (s, 3H, CO_2Me), 3.80 (app. t, 1H, $^3J_{5,6b}$ 10.6 Hz, H-6b), 3.66 (app. t, 1H, $^3J_{4,5}$ 9.8 Hz, H-4), 3.44 (s, 3H, OMe), 3.35 (dt, 1H, H-5), 2.06-1.98 (5 s, 15H, 2 x $NHCOCH_3$ and 3 x $OCOCH_3$), 1.48 (s, 3H, Me) ppm; ^{13}C-NMR (150 MHz, MeOH-d_4): δ 174.45-171.30 (5 s, 3 x $OCOCH_3$ and 2 x $NHCOCH_3$), 171.91 (s, $MeCCO_2Me$), 102.50 (d, C-1), 100.74 (s, $MeCCO_2Me$), 98.32 (d, C-1'), 75.30 (d, C-3), 74.93 (d, C-3'), 74.18 (d, C-4), 73.08 (d, C-5'), 70.24 (d, C-4'), 68.59 (d, C-5), 65.86 (t, C-6), 63.36 (t, C-6'), 57.23 (q, OMe), 55.23 (d, C-2'), 53.10 (q, $MeCCO_2Me$), 51.28 (d, C-2), 25.87 (q, $MeCCO_2Me$), 23.05-20.60 (5 q, 3 x

$OCOCH_3$ and 2 x $NHCOCH_3$) ppm. HR-MS: $[M+Na]^+$ *m/z* calcd: 671.2270; found: 671.2266.

Methyl (2-acetamido-2-deoxy-β-D-*gluco*-pyranosyl)-(1→3)-2-acetamido-2-deoxy-4,6-O-(1-carboxyethylidene)-β-D-*manno*-pyranoside sodium salt (12)

0.1 M NaOMe-solution (1.0 ml) was added to a solution of compound **11** (4.5 mg, 0.007 mmol) in dry MeOH (2.0 ml) and stirred at room temperature for 80 min. The pH was adjusted to 7.0 by addition of Dowex H^+ ion exchange resin, the resin was filtered off and the filtrate was concentrated in vacuo. The residue was treated with 0.2 M NaOH-solution (1.0 ml) at room temperature for 3 h. The pH was adjusted to 7.5 by addition of Dowex H^+ ion exchange resin. The suspension was filtered and the filtrate was lyophilized. Purification by size exclusion chromatography on a BioGel-P2-column gave **12** as colorless amorphous solid. Yield: 3.2 mg (87%, containing ~7% impurity); R_f 0.32 (EtOAc-MeOH 1:1, HPTLC); $[\alpha]_D^{23}$ – 56.0 (*c* 0.37, H_2O). ^{1}H-NMR (600 MHz, D_2O): δ 4.63 (m, 2H, H-1 and H-2), 4.59 (d, 1H, $^3J_{1',2'}$ 8.5 Hz, H-1'), 4.17 (dd, 1H, $^3J_{2,3}$ 4.8, $^3J_{3,4}$ 10.4 Hz, H-3), 3.99 (dd, 1H, $^3J_{5,6a}$ 5.0, $^2J_{6a,6b}$ 10.7 Hz, H-6a), 3.90 (dd, 1H, $^3J_{5',6a'}$ 2.2, $^2J_{6a',6b'}$ 12.4 Hz, H-6a'), 3.72 (m, 1H, H-4), 3.71 (m, 1H, H-6b), 3.70 (m, 1H, H-2'), 3.69 (m, 1H, H-6b'), 3.50 (dd, 1H, $^3J_{2',3'}$ 10.3, $^3J_{3',4'}$ 8.8 Hz, H-3'), 3.42 (m, 1H, H-5'), 3.43 (s, 3H, OMe), 3.39 (m, 1H, H-5), 3.38 (m, 1H, H-4'), 2.01 and 1.99 (2 s, 6H, 2 x NHCOCH3), 1.43 (s, 3H, Me) ppm. 13C-NMR (150 MHz, D2O): δ 176.09 (s, $MeCCO_2Na$), 175.56 and 175.12 (2 s, 2 x $NHCOCH_3$), 102.83 (s, $MeCCO_2Na$), 102.09 (d, C-1), 98.87 (d, C-1'), 76.78 (d, C-5'), 74.87 (d, C-3'), 74.79 (d, C-3), 73.03 (d, C-4), 70.63 (d, C-4'), 67.82 (d, C-5), 64.81 (t, C-6), 61.61 (t, C-6'), 58.06 (q, OMe), 56.14 (d, C-2'), 51.05 (d, C-2), 25.41 (q, $MeCCO_2Na$), 23.25 and 22.86 (2q, 2 x $NHCOCH_3$) ppm. HR-MS: $[M+H]^+$ *m/z* calcd: 531.1797; found: 531.1796.

Abbreviations used are: Ac, acetyl; Ar, argon, Bn, benzyl; D_2O, deuterium oxide, DCM, dichloromethane; DMF, *N,N*-dimethylformamide; EtOAc, ethyl acetate, H_2, hydrogen; *J*, coupling constant; MeCN, acetonitrile, MPLC, medium pressure liquid chromatography; NaOMe, sodium methoxide; Pd, palladium; Ph, phenyl; Pyr, pyruvyl; TFA, trifluoroacetic acid; THF, tetrahydrofuran; TLC, thin layer chromatography; TMSOTf, trimethylsilyl trifluoromethanesulfonate; Troc; trichloroethoxycarbonyl.

Supplementary information references

Augé J, David S, Guibe L and Jugie G. 1980. A nuclear-quadrupole resonance study of halogen substituted methyl pyranosides-dependence of the p-orbital population of an halogen at position 2 upon the anomeric configuration and a possible correlation with chemical-reactivity. Noveau J Chim J Chem 4: 481–486.

Ellervik U and Magnusson G. 1996. Glycosylation with N-Troc-protected glycosyl donors. Carbohydr Res 280: 251–260.

Jansson P-E, Lindberg J and Widhalm G. 1993. Syntheses and NMR studies of pyruvic acid 4,6-acetals of some methyl hexopyranosides. Acta Chem Scand 47: 711–715.

Murphy PV, O'Brien JL, Gorey-Feret LJ and Smith AB. 2003. Synthesis of novel HIV-1 protease inhibitors based on carbohydrate scaffolds. Tetrahedron 59: 2259–2271.

Pokorny B and Kosma P. 2015. Synthesis of *Chlamydia* lipopolysaccharide haptens through the use of α-specific 3-iodo-Kdo fluoride donors. Chem Eur J 21: 305-313.

Schüle G and Ziegler T. 1996. Efficient convergent block synthesis of a pyruvated tetrasaccharide 5-aminopentyl glycoside related to *Streptococcus pneumoniae* type 27. Tetrahedron 52: 2925–2936.

Supplementary Figures

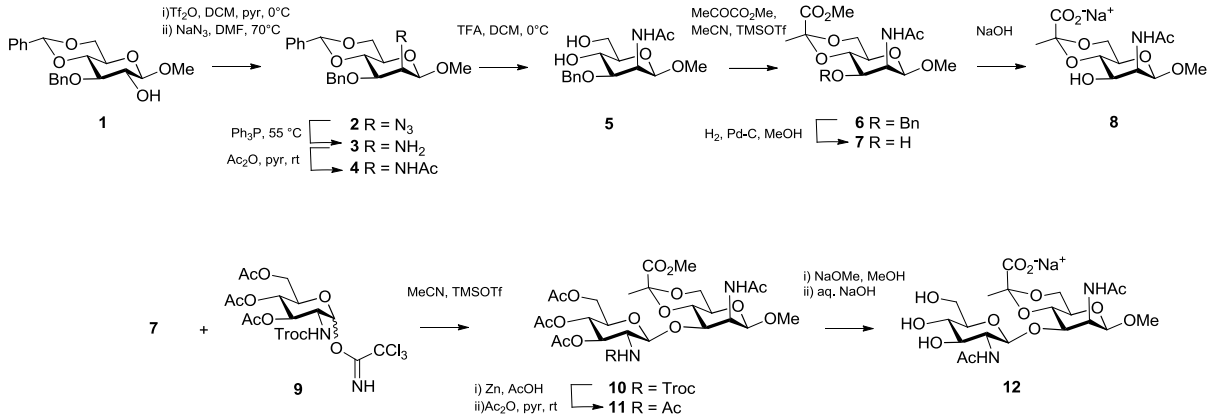


Figure S1. Reaction scheme for the synthesis of 4,6-*O*-pyr-β-D-Man*p*NAc-*O*-Me (**8**) and β-D-Glc*p*NAc-(1→3)-4,6-*O*-pyr-β-D-Man*p*NAc-*O*-Me (**12**)

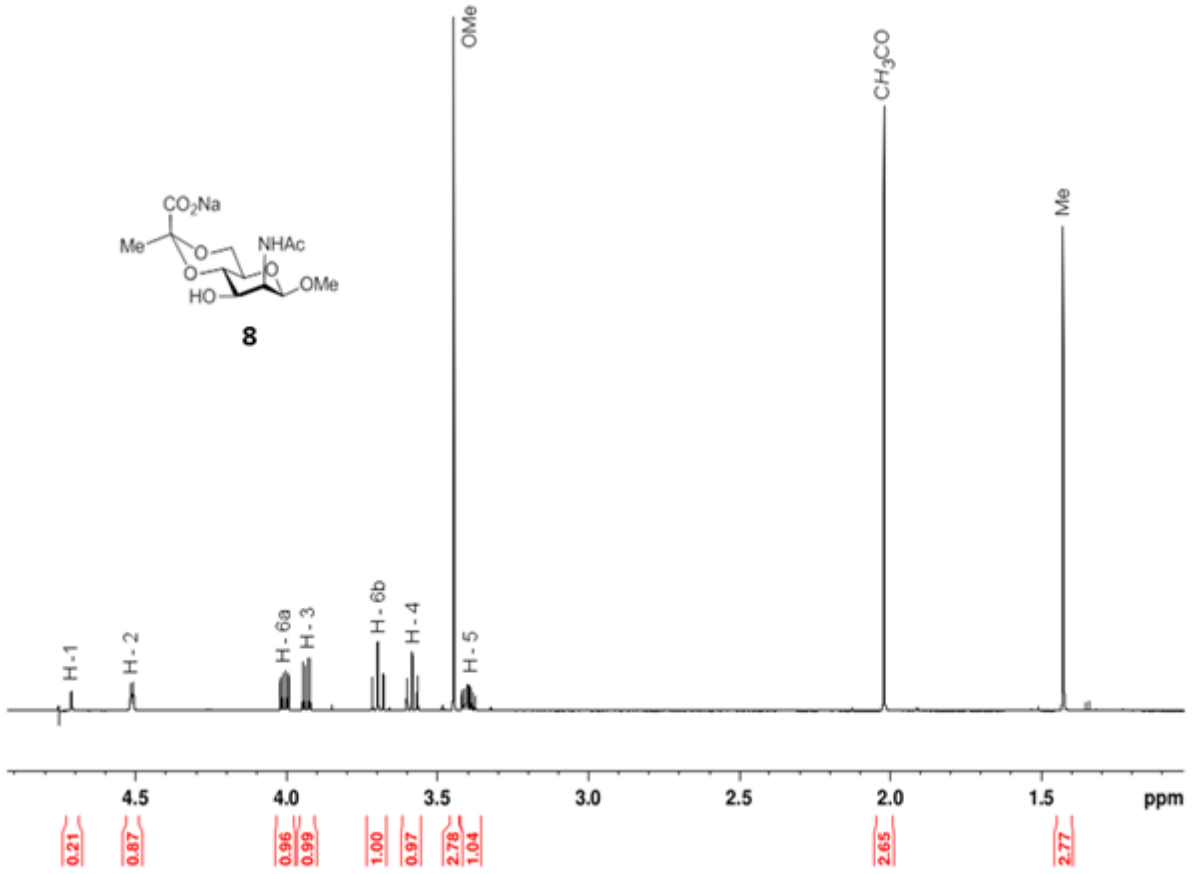


Figure S2. ^{1}H NMR spectrum (600 MHz) of 4,6-*O*-pyr-β-D-Man*p*NAc-*O*-Me (**8**) in D_2O

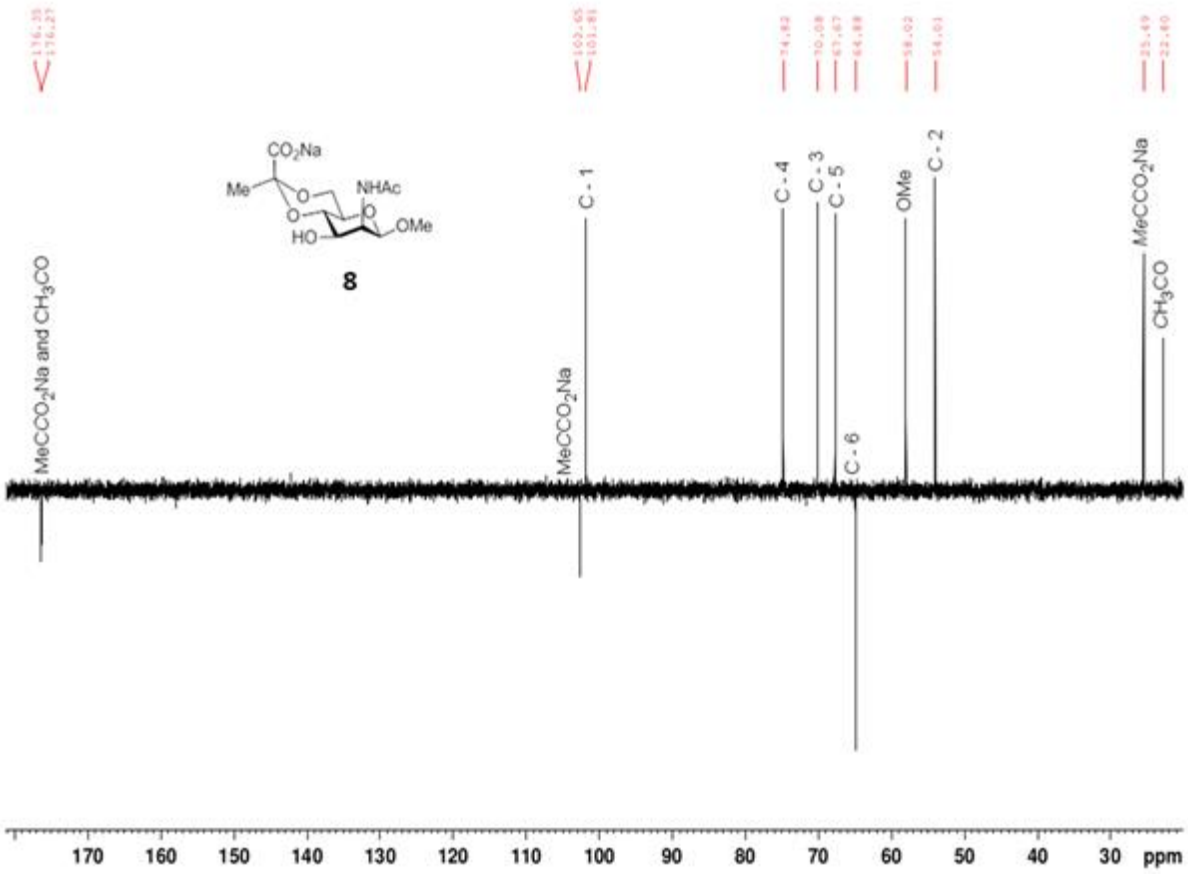


Figure S3. ^{13}C NMR spectrum (150 MHz) of 4,6-*O*-pyr-β-D-Man*p*NAc-*O*-Me (**8**) in D_2O

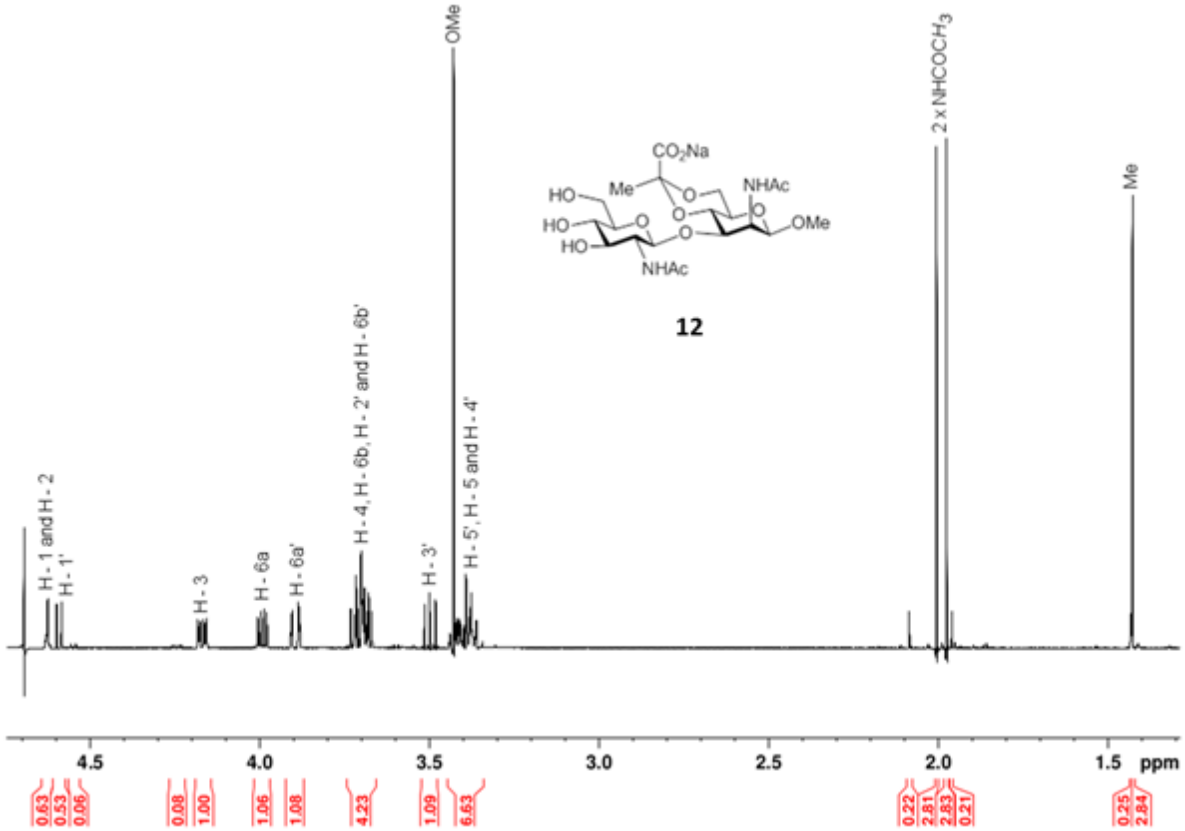


Figure S4. ^{1}H NMR spectrum (600 MHz) of β-D-Glc*p*NAc-(1→3)-4,6-*O*-pyr-β-D-Man*p*NAc-*O*-Me (**12**) in D_2O.

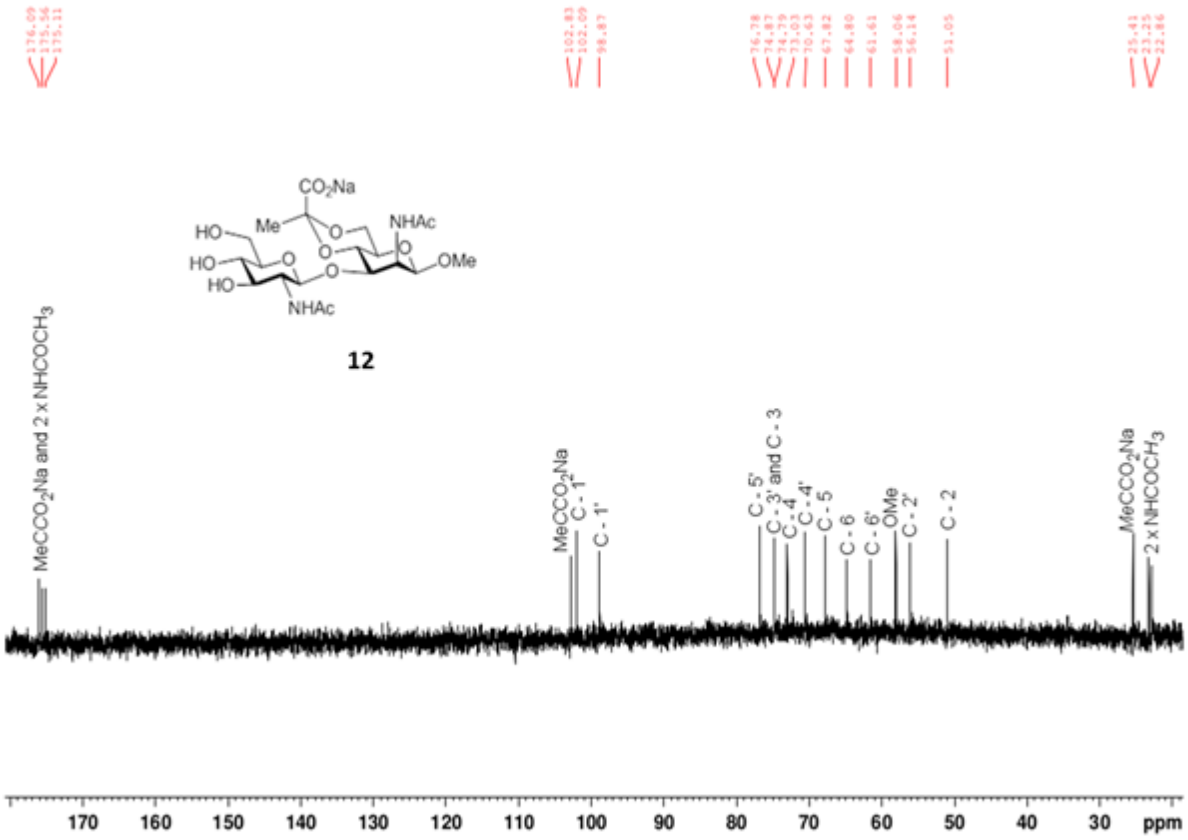
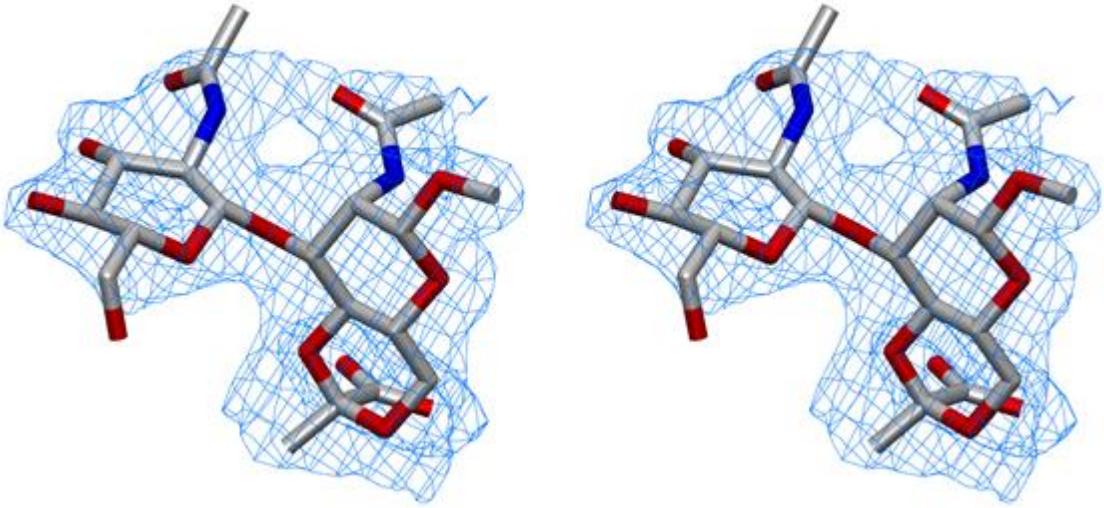
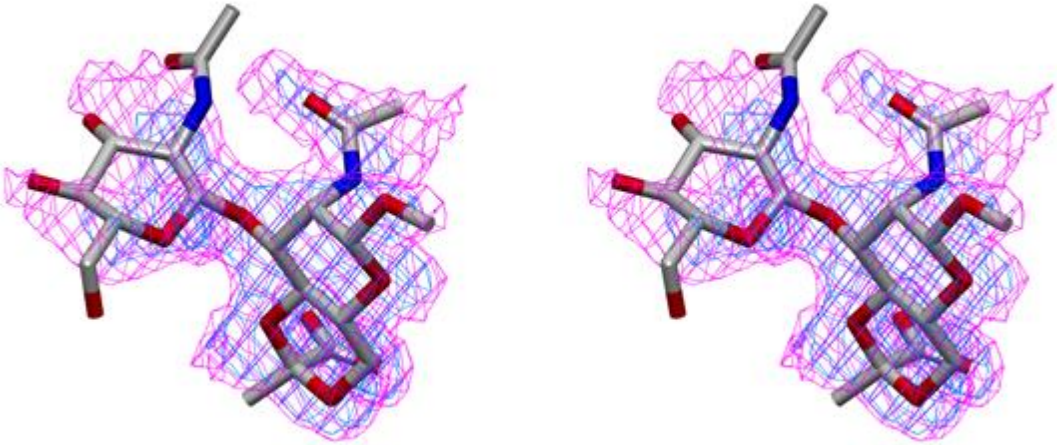


Figure S5. ^{13}C NMR spectrum (600 MHz) of β-D-Glc*p*NAc-(1→3)-4,6-*O*-pyr-β-D-Man*p*NAc-*O*-Me (12) in D_2O.

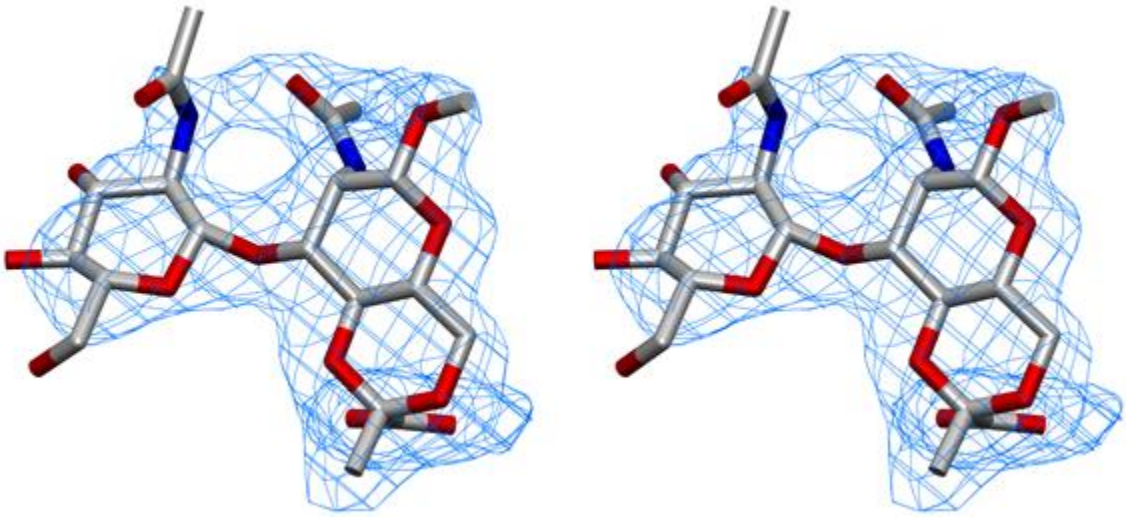
A



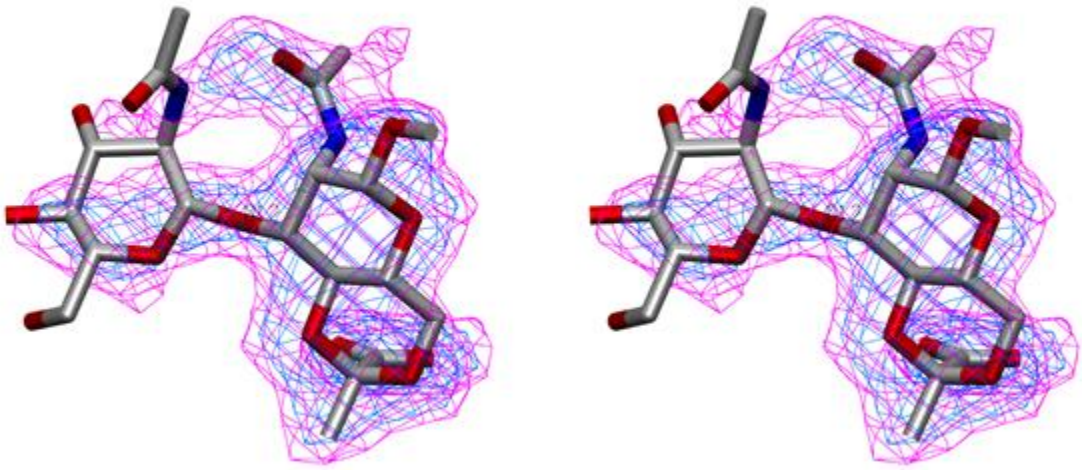
B



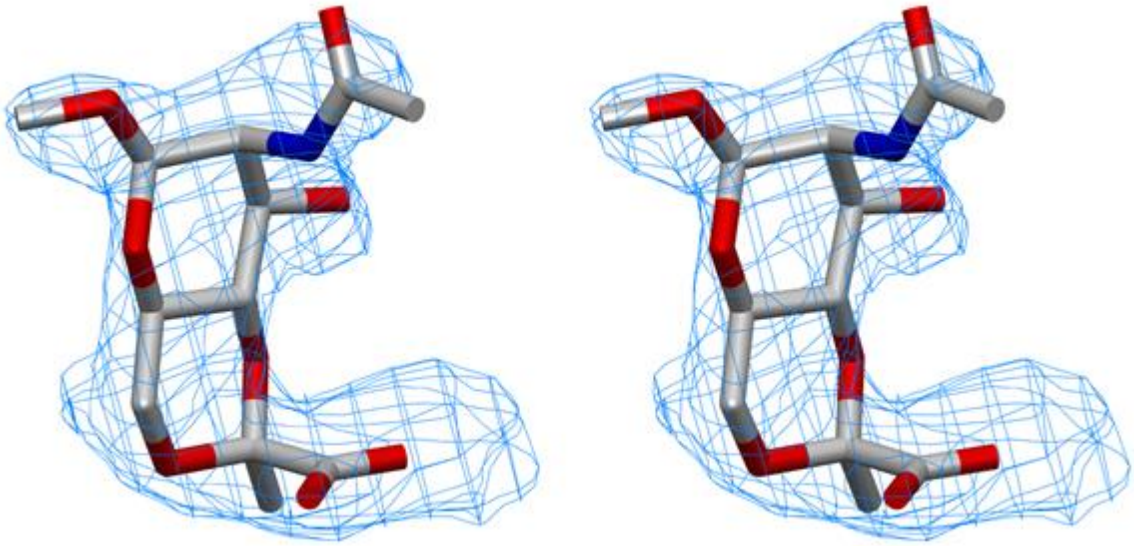
C



D



E



F

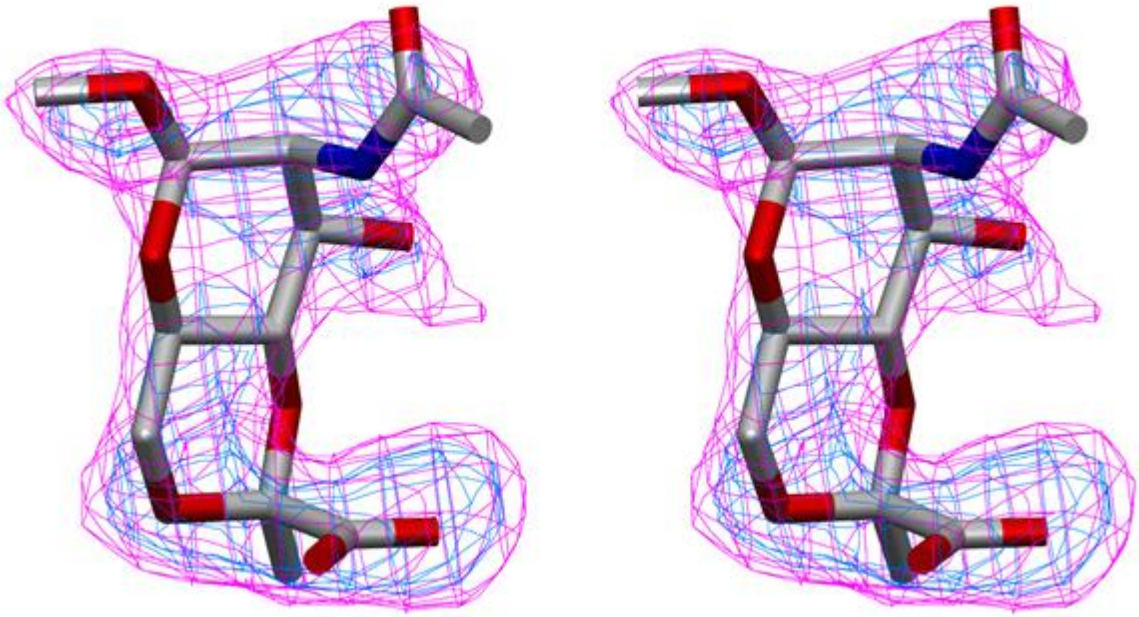


Figure S6. Electron density maps for β-D-GlcNAc-(1→3)-4,6-Pyr-β-D-ManNAc-*O*-Me bound to $SpaA_{SLH}$ (A) 2Fo-Fc map contoured to 1σ at molecule A binding site. (B) Omit Fo-Fc map contoured to 2σ (Magenta) and 3σ (Blue) at molecule A binding site. (C) 2Fo-Fc map at molecule B binding site. (D) Omit Fo-Fc map at molecule B binding site. Electron density for 4,6-Pyr-β-D-ManNAc-*O*-Me bound in the G1 pocket of the C2 space group structure is shown in (E) 2Fo-Fc map and (F) omit Fo-Fc map.

9 Appendix

List of publications

- Structural basis of cell wall anchoring by SLH domains in *P. alvei*. Ryan J. Blackler, **Arturo López-Guzmán**, Gudrun Martinz, Susannah M. L. Gagnon, Omid Haji-Ghassemi, Paul Kosma, Paul Messner, Christina Schäffer and Stephen V. Evans. (Manuscript in preparation)

Conference contributions

Oral presentations

Presenting author is underlined.

- **Crystal structure of *Paenibacillus alvei* S-layer protein SpaA SLH domains in complex with synthetic secondary cell wall polymer ligands, reveals a basis for differential binding by variant SLH motifs.** López-Guzmán A., Blackler R., Martinz G., Kosma P., Messner P., Schäffer C., Evans S. [19th Austrian Carbohydrate workshop, Graz, Austria 12.02.2015]

- **Strategy for *in vivo* or *in vitro* surface display of functional epitopes.** López-Guzmán A. [Key note Lecture at COST CM 1102/IBCarb Spring Training School, Synthesis for nano-and glyco-sciences, Bangor, United Kingdom, 9.-11.04.2015]

- **Synthesis and binding interactions of pyruvylated *N*-acetyl-mannosamine ligands to the S-layer homology domains of *Paenibacillus alvei*.** Martinz G., Blackler R., López-Guzmán A., Messner P., Schäffer C., Evans S. V., Kosma P. [The 18th European carbohydrate symposium, Moscow, Russia, 2.-6.08.2015]

Poster presentations

Presenting author is underlined.

- **Crystal structure of Surface Layer Homology domains from *Paenibacillus alvei* S-layer protein SpaA provides insight to secondary cell wall polymer recognition.** Blackler R. J., López-Guzmán A., Messner P., Schäffer C., Kosma P. and Stephen V. Evans. [Joint Meeting of the Society of Glycobiology and the Japanese Society for Carbohydrate Research, Honolulu, Hawaii, USA, 16.-19.11.2014]

- **Functional biochemical analysis and characterization of selected components of the surface layer linkage system of *Paenibacillus alvei* CCM 2051^T**. López-Guzmán A., Blackler R. J., Evans S., Kosma P., Schäffer C., Messner P. [FASEB SRC Microbial Glycobiology, Itasca, Illinois, USA, 8.-13.06.2014]

- **Crystallization experiments of selected components from the S-layer glycosylation pathway of *Paenibacillus alvei* CCM 2051^T.** López-Guzmán A., Anzengruber J., Janesch B., Ristl R., Schäffer C., Messner P. (2011). [21st International Symosium on Glycoconjugates, Vienna, Austria, 21.-26.08.2011]

Crystal structure of *Paenibacillus alvei* S-layer protein SpaA SLH domains in complex with synthetic secondary cell wall polymer ligands reveals a basis for differential binding by variant SLH motifs

Arturo López-Guzmán[1], Ryan Blackler[2], Paul Messner[1], Stephen Evans[2], Paul Kosma[3], Christina Schäffer[1]

[1]Department of NanoBiotechnology, *NanoGlycobiology* unit, Universität für Bodenkultur Wien, Muthgasse 11, 1190 Vienna, Austria
[2]Department of Biochemistry & Microbiology, University of Victoria, Victoria, BC, V8W 3P6 Canada
[3]Department of Chemistry, Universität für Bodenkultur Wien, Muthgasse 18, 1190 Vienna, Austria

E-mail: Arturo.lopez-guzman@boku.ac.at

Paenibacillus alvei CCM 2051^T is a Gram-positive, mesophilic bacterium that is completely covered with a 2D crystalline S-layer protein lattice that is *O*-glycosylated at distinct tyrosine residues (1). The S-layer protein SpaA is non-covalently linked to the bacterial cell wall via a "non-classical" secondary cell wall polymer (SCWP). This polymer is composed of repeating units, as follows: →3)-[4,6-*O*-(1-carboxyethylidene)]-β-D-Man*p*NAc-(1→4)-β-D-Glc*p*NAc-(1→, and is covalently linked to carbon 6 of β-D-*N*-acetylmuramic acid residues of the peptidoglycan layer (2).

The SpaA protein contains three consecutive repeats of surface layer homology (SLH) domains at the N-terminus, which are important for anchoring the S-layer protein to the peptidoglycan (PG) matrix (3). Furthermore, it has been shown that conserved four-amino acid motifs within the SLH-domains play a critical role for the interaction between SLH domains and pyruvylated SCWPs (4). In addition, the *csaB* gene, which encodes the pyruvyl transferase CsaB, is a pre-requisite for SpaA assembly and S-layer attachment to the cell surface of *P. alvei* (5). *P. alvei* SpaA protein possess a TRAE motif in domain SLH 1 and two variants, TVEE and TRAQ in the SLH 2 and SLH 3 domains respectively. The SLH domains are rich in α-helical structures, where each SLH domain contributes to a three-helix bundle core with three-helical perpendicular prongs. Each prong forms a groove, where the TRAE motif and its variants are contained, hence generating a potential binding pocket.

In order to study the binding features in molecular detail, mono- and disaccharide synthetic SCWP ligands were generated from methyl β-D-glucoside. Crystal structures of a recombinant truncated $SpaA_{SLH}$ form complexed with the mono and disaccharide ligand, respectively, revealed binding to groove 2. Complementary ITC measurements evidenced that the binding affinity of the ligands is a consequence of the intricate equilibrium of hydrogen bonding, hydrophobic interactions and entropic restrictions to form complexes of $SpaA_{SLH}$ and synthetic SCWPs.

This work was supported by the Austrian Science Fund, FWF projects P22791-B11 (to PM) and P27374-B22 (to CS).

1. Altman, E., Brisson, J. R., Messner, P., and Sleytr, U. B. (1991) *Biochem Cell Biol* **69**, 72-78
2. Schäffer, C., Müller, N., Mandal, P. K., Christian, R., Zayni, S., and Messner, P. (2000) *Glycoconjugate J* **17**, 681-690
3. Janesch, B., Messner, P., and Schäffer, C. (2013) *J. Bacteriol***195**(3), 565-575**.**
4. Kern, J., Wilton, R., Zhang, R., Binkowski, T. A., Joachimiak, A., and Schneewind, O. (2011) *J Biol Chem* **286**, 26042-26049
5. Zarschler, K., Janesch, B., Kainz, B., Ristl, R., Messner, P., and Schäffer, C. (2010) *Carbohydr Res* **345**, 1422-1431

Crystallization Experiments of Selected Components from the S-Layer Glycosylation Pathway of *Paenibacillus alvei* CCM 2051^T

Arturo López-Guzmán, Julia Anzengruber, Bettina Janesch, Robin Ristl, Christina Schäffer, Paul Messner

Department of Nanobiotechnology, Vienna Institute of Bio Technology, University of Natural Resources and Life Sciences, Mutgasse 11, 1190 Vienna, Austria.

Paenibacillus alvei CCM 2051^T is a Gram-positive, mesophilic bacterium that is completely covered with a glycosylated surface (S-) layer protein. The O-glycosidically linked S-layer glycans are polymers of [-3)-β-D-Gal-(1[α-D-Glc-(1-6)]-4)-β-D-ManNAc-(1-] repeating units that are linked via the adaptor GroA-2-OPO2-4-β-D-ManNAc-(1-4)]-3)-[α-L-Rha-(1-3)]n=3 β-D-Gal-(1- to specific tyrosine residues of the S-layer protein. The S-layer protein itself is non-covalently linked by a secondary cell wall polymer (SCWP) with the structure 3)-[4,6-*O*-(1-carboxyethyldiene)]-β-D-ManNAc-(1-4)-β-D-GlcNAc-(1- to carbon 6 of *N*-acetylmuramic acid residues of the peptidoglycan. The genes involved in the glycosylation process are clustered in distinct, closely spaced *slg* (S-layer glycosylation) gene loci comprising approximately 24 kb of known sequence.

Usually S-layer proteins are insoluble in aqueous environment and have the inherent ability to rapidly crystallize in large two-dimensional lattices, either on intact cells or *in vitro*. Therefore, three-dimensional crystallization experiments of the complete S-layer proteins have been unsuccessful so far. Recently it has been demonstrated that SCWPs are able to render isolated S-layer proteins soluble in aqueous systems. Since the S-layer protein-SCWP complex of *P. alvei* CCM 2051^T obtained after purification of the S-layer protein from native host cells has shown to be water soluble, X-ray crystallography will be applied for characterization of the type of interaction between S-layer protein and SCWP and, thus, the attachment of the S-layer protein to the cell wall.

[1] Zarschler, K., Janesch, B., Zayni, S., Schäffer, C. and Messner, P., *Applied and Environmental Microbiology*, **2009**, 75, 3077-3085.
[2] Schäffer, C., Müller, N., Mandal, K. P., Christian, R., Zayni, S. and Messner P., *Glycoconjugate Journal*, **2000**, 17, 681-690.
[3] Norville, J. E., Deborah, F. K., Knight, T. F., Belcher, A. M. and Walz, T., *Biotechnology Journal* **2011**, 6, 1-5.

Crystallization Experiments of Selected Components From the S-Layer Glycosylation Pathway of *Paenibacillus alvei* CCM 2051^T

Arturo López-Guzmán, Julia Anzengruber, Bettina Janesch, Robin Ristl, Christina Schäffer, Paul Messner

Department of Nanobiotechnology, Vienna Institute of Biotechnology, University of Natural Resources and Life Sciences, Muthgasse 11, 1190 Vienna, Austria.

Introduction

Paenibacillus alvei CCM 2051^T is a Gram-positive, mesophilic bacterium, which is entirely covered with a glycosylated surface layer (S-layer) protein [1]. The S-layer protein (SpaA) of *P. alvei* CCM 2051^T is post-translationally modified with a tyrosine linked *O*-glycans [1]. It is non-covalently linked to the bacterial cell wall through an auxiliary molecule, named secondary cell wall polymer (SCWP). The SCWP is a polymer of D-ManNAc and D-GlcNAc repeats, pyruvylated at the distal end and at the proximal end covalently linked to the peptidoglycan via a phosphodiester bond (**Fig. 1**) [2]. Isolated S-layer proteins are quite insoluble in aqueous environments, due to their instrinsic capacity to form 2-D crystalline arrays either on intact cells or *in vitro* at the air water-interface and on solid supports. Thus, this attribute prevents these glycoproteins from crystallyzing to 3-D structures. It has been demonstrated that the interaction of SCWP with S-layer proteins renders the protein soluble in an aqueous milieu. Therefore, in future experiments spectroscopic and calorimetric characterization of the type of interaction involving these molecules will be performed along with 3-D crystallization, in order to eventually elucidate the tertiary structure of an S-layer protein in complex with SCWP.

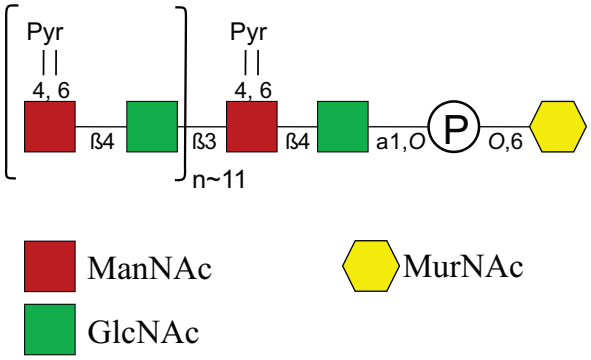


Figure 1. Structure of the cell wall anchor of the S-layer protein SpaA of *Paenibacillus alvei* CCM 2051^T. ManNAc (*N*-acetylmannosamine), GlcNAc (*N*-acetylglucosamine), MurNAc (*N*-acetylmuramic acid), Pyr (pyruvic acid).

Results

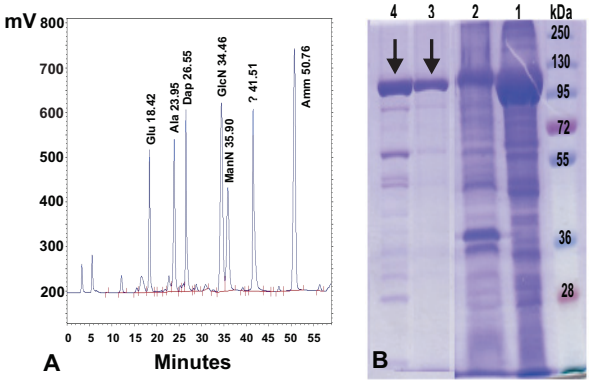


Figure 2. HPLC separation with pulsed amperometric detection analysis for amino sugars and amino acids of the SCWP (**A**). SDS-PAGE analysis of the expression and purification of recombinant SpaA-His^6 (arrow) (**B**). Lane 1: cell lysate, Lane 2: cell pellet, Lane 3: 100 mM imidazole, Lane 4: 400 mM imidazole.

The amino acid and the amino sugar composition of the SCWP after size exclusion chromatography is shown in **Figure 2A**. The chromatogram shows the presence of *N*-acetylglucosamine and *N*-acetylmannosamine, which are the main components of this anchor molecule. Besides, the presence of some amino acids, mainly glutamic acid and lysine can be observed. These amino acids are constituents of the linkage peptide of *P. alvei* CCM 2051^T peptidoglycan. The co-presence of the latter amino acids together with the amino sugars is a conclusive probe that SCWP and peptidoglycan are covalently linked [2], since during the purification procedure there is no step to cleave this linkage. It is also important to mention that the integrity of the polymer is intact. However, additional analyses are required in future work.

To produce recombinant SpaA, the *spaA* gene was amplified by PCR, cloned into the pET28a vector, and expressed in *E. coli* BL21(D3) strain [3]. The SDS-PAGE analysis showed that rSpaA-His$_6$ partioned between the soluble fraction and inclusion bodies (**Fig. 2B**, lane 1 and 2, respectively). After His-tag affinity chromatography, the recombinant protein was enriched and is available in almost pure form (**Fig. 2B**, lane 3).

Conclusions & Outlook

- Purification of the SCWP was almost completed, and most important, its molecular integrity was kept after the purification procedure.
- Cleavage of the linking phosphodiester bond is planned to remove the remaining portion of the peptidoglycan, attached to SCWP.
- Determination of the binding forces involved between the SCWP and the S-layer protein SpaA as well as the type of interaction between native SCWP and recombinant SpaA protein will be analyzed in future.
- To determine whether the S-layer intermolecular interactions are stronger than the interactions for binding to the bacterial surface.
- This data will help to find the best conditions for 3-D crystallization of the S-layer SpaA-SCWP complex of *P. alvei* CCM

The work is supported by Austrian Science Fund, projects P20745-B11 and P22791-B11 (to PM) and P21954-B20 (to CS) and BioTop W1244 to (PM & CS).

References:
[1] Zarschler, K., *Applied and Environmental Microbiology* **2009**, 75, 10, 3077-3085.
[2] Schäffer, C., *Glycoconjugate Journal* **2000**, 17, 681-690.
[3]Norville, J. E., Biotechnology Journal **2011**, 6, 1-5.

FASEB SRC Microbial Glycobiology, June 8 - 13, 2014, Itasca, Illinois (USA)

Functional biochemical analysis and characterization of selected components of the surface layer linkage system of *Paenibacillus alvei*.

Arturo Lopez-Guzman[1], Ryan Blackler[2], Stephen Evans[2], Paul Kosma[3], Christina Schäffer[1], Paul Messner[1]

[1]Department of NanoBiotechnology, *NanoGlycobiology* unit, Universität für Bodenkultur Wien, Muthgasse 11, 1190 Vienna, Austria
[2]Department of Biochemistry & Microbiology, University of Victoria, Victoria, BC, V8W 3P6 Canada
[3]Department of Chemistry, Universität für Bodenkultur Wien, Muthgasse 18, 1190 Vienna, Austria

E-mail: Arturo.lopez-guzman@boku.ac.at

Paenibacillus alvei CCM 2051^T is a Gram-positive, mesophilic bacterium that is completely covered with a glycosylated, 2D crystalline S-layer protein lattice (SpaA)(1). The S-layer protein SpaA is non-covalently linked to the bacterial cell wall via a "non-classical" secondary cell wall polymer (SCWP). This polymer is composed of repeating units, as follows: →3)-[4,6-*O*-(1-carboxyethylidene)]-β-D-Man*p*NAc-(1→4)-β-D-Glc*p*NAc-(1→, and is covalently linked to carbon 6 of β-D-*N*-acetylmuramic acid residues of the peptidoglycan(2).

The SpaA protein contains three surface layer homology (SLH) domains at the N-terminus, which are important for anchoring the S-layer protein to the peptidoglycan (PG) matrix(3). Furthermore, it has been shown that the SLH domains are rich in α-helical structures, where each SLH domain contributes to a three-prong structure that is important for the recognition of SLH-domain proteins and pyruvylated SCWP(4). In addition, the *csaB* gene, which encodes the pyruvyl transferase CsaB, is a pre-requisite for SpaA assembly and complete attachment on the cell surface of *P. alvei*(5).

By a combination of state-of-the-art approaches we are attempting to unravel the structure-function relationship of the SCWPs of *P. alvei* as a cell wall anchoring epitope for the glycosylated SpaA protein.

This work was supported by the Austria Science Fund, project P22791-B11 (to PMS).

1. Altman, E., Brisson, J. R., Messner, P., and Sleytr, U. B. (1991) *Biochem Cell Biol* **69**, 72-78
2. Schäffer, C., Müller, N., Mandal, P. K., Christian, R., Zayni, S., and Messner, P. (2000) *Glycoconjugate J* **17**, 681-690
3. Janesch, B., Messner, P., and Schäffer, C. (2013) *J. Bacteriol* **195** (3), 565-575**.**
4. Kern, J., Wilton, R., Zhang, R., Binkowski, T. A., Joachimiak, A., and Schneewind, O. (2011) *J Biol Chem* **286**, 26042-26049
5. Zarschler, K., Janesch, B., Kainz, B., Ristl, R., Messner, P., and Schäffer, C. (2010) *Carbohydr Res* **345**, 1422-1431

Functional Biochemical Analysis and Characterization of Selected Components of the Surface Layer Linkage System of *Paenibacillus alvei*

Arturo Lopez-Guzman[1], Ryan Blackler[2], Stephen Evans[2], Paul Kosma[3], Christina Schäffer[1], Paul Messner[1]

[1]Department of NanoBiotechnology, NanoGlycobiology unit, Universität für Bodenkultur Wien, Muthgasse 11, 1190 Vienna, Austria, [2]Department of Biochemistry & Microbiology, University of Victoria, Victoria, BC, V8W 3P6 Canada, [3]Department of Chemistry, Universität für Bodenkultur Wien, Muthgasse 18, 1190 Vienna, Austria.

Introduction

Paenibacillus alvei CCM 2051^T is a Gram-positive, mesophilic bacterium, which is entirely covered with a glycosylated 2D crystalline S-layer protein lattice (SpaA) which is post-translationally modified with a tyrosine linked *O*-glycan [1]. The S-layer is non-covalently linked to the bacterial cell wall through a non-classical secondary cell polymer (SCWP). The SCWP is an oligosaccharide composed of *N*-acetylglucosamine (GlcNAc) and *N*-acetylmannosamine (ManNAc), where the ManNAc residues are modified by the addition of a pyruvyl group **(Fig. 1A)**. Surface layer homology domains (SLH) have been shown to mediate binding of exocellular proteins to the cell surface of Gram-Positive bacteria [3]. In addition, the *csaB* gene, which encodes the pyruvyltransferase enzyme CsaB, is an important prerequisite for SpaA assembly and complete attachment to the cell surface of *P. alvei* [5] **(Fig. 1B)**.

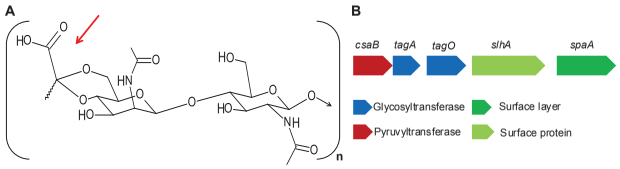


Figure 1. Chemical structure of the repeating oligosaccharide unit, →3)-[4,6-*O*-(1-carboxyethylidene)]-β-D-Man*p*NAc-(1→4)-β-D-Glc*p*NAc-(1→, of the SCWP (A). Genetic localization of the *csaB* gene in the SCWP biosynthesis locus of *P. alvei* (B).

SpaA contains three SLH domains at the N-terminus **(Fig. 2A)**. In addition, it has been found that there is a conserved motif in SLH domain-bearing proteins, known as the TRAE motif, which is directly involved in recognition of peptidoglycan-SCWP by the S-layer **(Fig. 2B)**. By a combination of state-of-the-art approaches we are attempting to unravel the structure-function relationship of the SCWPs of *P. alvei* as a cell wall anchoring epitope for the glycosylated SpaA protein.

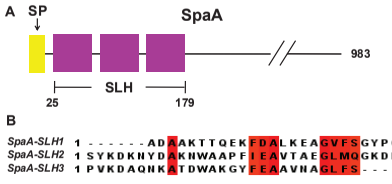


Figure 2. Primary structure of SpaA protein showing the signal peptide and the position of the SLH domains (A). Alignment of the three SLH domains. Shown (in red) the conserved amino acids and the TRAE motifs (B).

projects: P22791-B11, P21954-B20, P24305-B20

Results

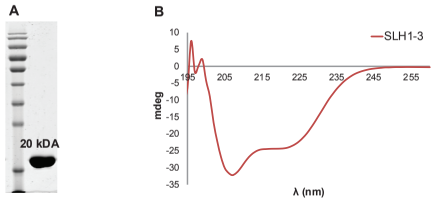


Figure 3. Expression, purification and structural analysis of SpaA-SLH of *P. alvei*. Recombinant SpaA-SLH protein analyzed by SDS-PAGE (A). Circular dichroism (CD) spectrum of purified SpaA-SLH (B).

SpaA-SLH was recombinantly expressed and was purified by affinity and size exclusion chromatography **(Fig. 3A)**. Secondary structure analysis of truncated SpaA was performed by Far-UV CD spectroscopy **(Fig. 3B)**. The predicted model of truncated SpaA shows that SpaA-SLH may adopt the shape of a three-prong spindle, where each SLH domain contains the conserved TRAE motif in the central helices **(Fig. 4A)**. TRAE motifs contain charged residues that are in close proximity and located towards the peptidoglycan of the bacterial cell wall **(Fig. 4B)**.

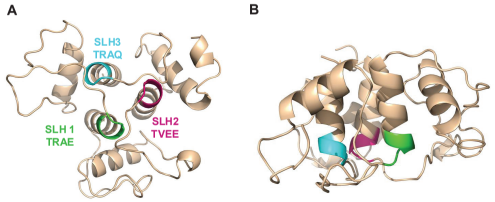


Figure 4. Modeling of the SLH domain of *P. alvei* S-layer protein SpaA. The location of the TRAE motifs is shown in each SLH domain (A). Side view of SpaA-SLH domains showing the helical core (B).

CsaB was recombinantly expressed and purified by nickel affinity chromatography **(Fig. 5A)**. Correct folding and secondary structure analysis of rCsaB was confirmed by Far-UV CD spectroscopy **(Fig. 5B)**. The predicted model of the pyruvyltransferase shows that CsaB has a similar structure to glycosyltransferases GT-B type fold, consisting of two domains with the Rossmann fold structure of GT's **(Fig. 5C)**. The putative catalytic site is shown (in red) enclosed in a groove between these two domains.

Results

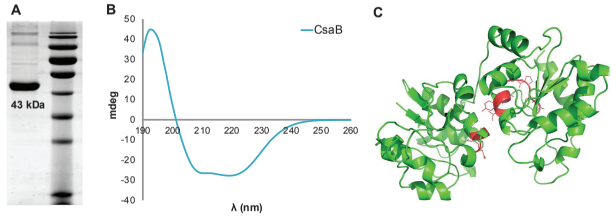


Figure 5. Expression, purification and structural analysis of recombinant pyruvyltransferase CsaB of *P. alvei*. CsaB enzyme analyzed by SDS-PAGE (A). Circular dichroism (CD) spectrum of purified CsaB (B). Modeling of CsaB showing the Rossmann fold (C).

We assume that CsaB uses phosphoenolpyruvic acid (PEP) as donor molecule for the transfer of pyruvyl groups to ManNAc residues. As acceptor substrate *p*-nitrophenyl-β-D-ManNAc (pNP-ManNAc) was used. The substrate pNP-ManNAc was detected at 12 min **(Fig. 6A)**. After incubation with rCsaB a new peak was detected at 11 min **(Fig. 6A)**. This new peak was not detected when PEP or rCsaB were excluded from the reaction mixture. This result suggests that the new peak might be the pyruvylated pNP-ManNAc product **(Fig. 6B)**.

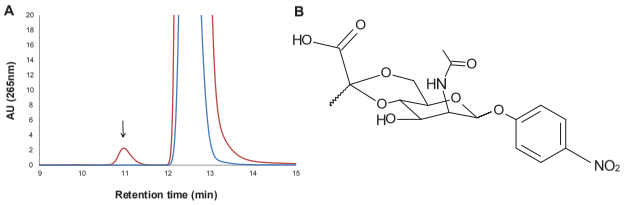


Figure 6. HPLC analysis of the enzymatic reaction carried out by CsaB. The arrow indicates the product (A). Proposed pyruvylation reaction product, *p*-nitrophenyl-4,6-*O*-β-D-ManNAc (B).

Conclusions

Positively and negatively charged residues (TRAE motifs) are highly conserved in SLH domains, which is a clear indication of the crucial role of electrostatic interactions for the recognition of the bacterial cell wall by S-layer proteins. The second important condition is the pyruvylation of the SCWP. The *csaB* gene is highly conserved in S-layer carrying Gram-positive bacteria that have SLH domain proteins. The enzymatic assays performed so far have demonstrated that rCsaB retains activity *in* vitro.

References:
[1] Altman, E., Brisson, J. R., Messner, P., and Sleytr, U. B. (1991) *Biochem Cell Biol* **69**, 72-78
[2] Schäffer, C., Müller, N., Mandal, P. K., Christian, R., Zayni, S., and Messner, P. (2000) *Glycoconjugate J* **17**, 681-690
[3] Janesch, B., Messner, P., and Schäffer, C. (2013) *J. Bacteriol* **195** (3), 565-575.
[4] Kern, J., Wilton, R., Zhang, R., Binkowski, T. A., Joachimiak, A., and Schneewind, O. (2011) *J Biol Chem* **286**, 26042-26049
[5] Zarschler, K., Janesch, B., Kainz, B., Ristl, R., Messner, P., and Schäffer, C. (2010) *Carbohydr Res* **345**, 1422-1431

http://www.nano.boku.ac.at/en/bimat/ arturo.lopez-guzman@boku.ac.at

MSc
Arturo López-Guzmán

Born 02 November 1980, Mexico City, Mexico

Education

Dec. 2010 - present
Doctoral studies of Food Chemistry and Biotechnology, University of Natural Resources and Life Sciences (BOKU), Vienna, Austria
Thesis topic:
"Functional biochemical analysis and crystallization of essential components from the surface anchoring system of *Paenibacillus alvei* CCM 2051^{T}"

Sep. 2008 - Aug. 2010
Master of Science in Biotechnology, Lund University, Lund, Sweden
Thesis topic: "*De novo* design of protein biomaterials"

Jan. 2001 - Jun. 2007
Bachelor of Science in Pharmaceutical Chemistry and Biology, National Autonomous University of Mexico (UNAM), Mexico City, Mexico

Professional experience

Dec. 2010 - present
Researcher in Glycobiology of prokaryotic microorganisms, Department of Nanobiotechnology, University of Natural Resources and Life Sciences (BOKU), Vienna, Austria

Teaching experience: Supervision of Master and Bachelor Theses

Jun. 2007 - Aug. 2008
Microbiology Quality Control Analyst, Schering-Plough S.A. de C.V., Mexico City, Mexico

Jul. 2005 - Dec. 2005
Research internship at the Faculty of Chemistry, National Autonomous University of Mexico (UNAM)

Publications

Ryan J. Blackler, **Arturo López-Guzmán**, Gudrun Martinz, Susannah M. L. Gagnon, Omid Haji-Ghassemi, Paul Kosma, Paul Messner, Christina Schäffer and Stephen V. Evans (2017). "Structural basis of cell wall anchoring by SLH domains in *P. alvei*". (Manuscript in preparation)

Skills and expertise

Field of expertise	Protein biochemistry, Microbiology, Glycobiology, Molecular biology, Structural biology, Nanobiotechnology, Analytical chemistry
Techniques	Chromatography (SEC, IEC, IMAC, HPLC, TLC), western blotting, PCR, molecular cloning, site-directed mutagenesis, vis-uv spectroscopy, fluorescence spectroscopy, circular dichroism (CD) spectroscopy, NMR spectroscopy, isothermal titration calorimetry (ITC), differential scanning calorimetry (DSC), mass spectrometry, X-ray crystallography, fluorescence microscopy
Computer literacy	MS office (Word, PowerPoint, Excel), CorelDRAW, Adobe Photoshop CC, Sci-Ed Clone Manager, PyMOL Molecular Graphics System, Sci-Ed Symyx Draw, Jalview, OriginPro
Languages	Spanish (native), English (fluent), German (good), French (intermediate)

Conference contributions

Oral presentation	**Key note Lecture at COST CM 1102/IBCarb Spring Training School, Synthesis for nano-and glyco-sciences,** *Bangor, UK, 2015*.
Oral presentation	**19th Austrian Carbohydrate workshop,** *Graz, Austria, 2015*.
Poster presentation	**FASEB Science Research Conference Microbial Glycobiology,** *Itasca, Illinois, USA, 2014*.
Poster presentation	**21st International Symposium on Glycoconjugates,** *Vienna, Austria, 2011*.

Personal interests

Literature, cinematography, playing football, photography, dancing, traveling