Advancing understanding of secondary cell wall polymer binding and synthesis in S-layers of Gram-Positive bacteria

by

Max Stefan Genuist Legg
BSc, University of Victoria, 2016

A Dissertation Submitted in Partial Fulfillment
of the Requirements for the Degree of

Doctor of Philosophy

in the Department of Biochemistry and Microbiology

Advancing understanding of secondary cell wall polymer binding and synthesis in S-layers of Gram-Positive bacteria

by

Max Stefan Genuist Legg
BSc, University of Victoria, 2016

Supervisory Committee

Dr. Stephen V. Evans **Supervisor**	(Department of Biochemistry and Microbiology)
Dr. Alisdair Boraston **Departmental Member**	(Department of Biochemistry and Microbiology)
Dr. Caroline Cameron **Departmental Member**	(Department of Biochemistry and Microbiology)
Dr. John Taylor **Outside Member**	(Department of Biology)

Abstract

Supervisory Committee

Dr. Stephen V. Evans (Department of Biochemistry and Microbiology)
Supervisor

Dr. Alisdair Boraston (Department of Biochemistry and Microbiology)
Departmental Member

Dr. Caroline Cameron (Department of Biochemistry and Microbiology)
Departmental Member

Dr. John Taylor (Department of Biology)
Outside Member

Self-assembling protein surface layers (S-layers) are ubiquitous prokaryotic cell-surface structures involved in structural maintenance, nutrient diffusion, host adhesion, virulence, and many additional processes, which makes them appealing targets for therapeutics and biotechnological applications, including live vaccines, liposome drug delivery and biosensors. Unlocking this potential requires expanding our understanding of S-layer properties, especially the details of surface-attachment.

S-layers of Gram-positive bacteria often are attached through the interaction of specialized S-layer homology (SLH) domain trimers with peptidoglycan-linked secondary cell wall polymers (SCWPs). Characterization of this interaction in the Gram-positive model organism *Paenibacillus alvei* CCM 2051^{T} reveals that, remarkably, binding-site switches can occur between two distinct SLH-domain SCWP receptor-site grooves in the S-layer protein SpaA, possibly as part of a mechanism to alleviate strain in the S-layer. To date, however, analysis of this novel mechanism has been limited to the terminal SCWP monosaccharide and the internal SCWP repeat disaccharide ligand analogues, leaving open the role of subsequent SCWP sugar residues in binding, as well as

whether the two receptor sites are also suited to accommodate longer SCWP ligands that better approximate the biological target at the surface of *P. alvei*.

To address this, the objective of this work aims to uncover and characterize the details of the SpaA SLH-domain ($SpaA_{SLH}$) SCWP-interaction by determining the co-crystal structures of $SpaA_{SLH}$, and single ($SpaA_{SLH}$/G109A) and the corresponding double ($SpaA_{SLH}$/G46A/G109A) mutants in complex with synthetic terminal disaccharide and trisaccharide analogues of the *P. alvei* CCM 2051^{T} SCWP target. These structural characterizations have been supplemented with disaccharide and trisaccharide binding data, which was obtained through thermodynamic ITC analyses carried out by collaborators.

The co-crystal structures of *P. alvei* $SpaA_{SLH}$ with synthetic, terminal SCWP disaccharide and trisaccharide analogues, together with previously published monosaccharide-bound $SpaA_{SLH}$ structures, reveal that while the SLH trimer accommodates longer biologically relevant SCWP ligands within both its primary (G2) and secondary (G1) binding sites, the terminal pyruvylated ManNAc moiety serves as the nearly-exclusive SCWP anchoring point. Binding is accompanied by displacement of a flexible loop adjacent to the receptor site that enhances the complementarity between protein and ligand, including electrostatic complementarity with the terminal pyruvate moiety. Remarkably, binding of the pyruvylated monosaccharide SCWP fragment alone is sufficient to cause rearrangement of the receptor binding sites in a manner necessary to accommodate longer SCWP fragments. The observation of multiple conformations for longer oligosaccharides bound to the protein, together with the demonstrated functionality of two of the three SCWP receptor binding sites, reveals how the $SpaA_{SLH}$-SCWP interaction has evolved to accommodate longer SCWP ligands and alleviate the strain inherent to bacterial S-layer adhesion during growth and division.

In addition, to further clarify the steps involved in SCWP biosynthesis, we present a crystal structure of the unliganded UDP-GlcNAc 2-epimerase enzyme MnaA, which catalyzes the interconversion of UDP-GlcNAc into UDP-ManNAc—an essential building block of the *P. alvei* SCWP target. The *P. alvei* MnaA epimerase adopts a GT-B fold that is consistent with the architecture of previously published structures of other bacterial non-hydrolyzing UDP-GlcNAc 2-epimerase enzymes for which substrate binding is observed in the cleft located between the two domains. Characterization of this structure, coupled with an analysis of the sequence of the MnaA protein, reveals the presence of conserved residues that define the catalytic and allosteric sites in homologous enzymes from different organisms. These residues are positioned to accommodate substrate within the MnaA binding cleft in much the same manner as the published enzyme homologues, suggesting that allosteric regulation as a mechanism for enzyme regulation is conserved in *P. alvei* MnaA.

These investigations are part of a greater effort toward understanding SLH domain-mediated SCWP-interactions in Gram-positive organisms, and provide insight into the structure and putative function of this SCWP biosynthetic enzyme. By understanding these processes, this knowledge may contribute to providing a platform for the rational design of Gram-positive inhibitors. Such inhibitors could selectively target, for example, the bacterial S-layer SCWP-binding interaction, or perhaps the essential biosynthetic enzymes involved in producing the exclusive targets that these S-layer proteins recognize and bind, and would thus represent a new class of antimicrobial therapeutics.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

Bis-tris	2-[Bis(2-hydroxyethyl)amino]-2-(hydroxymethyl)propane-1,3-diol
Cα	Alpha carbon
COOT	Crystallographic Object-Oriented Toolkit
CCP4	Collaborative Crystallography Project 4
Delta G	Gibbs free energy
G1	Groove 1 of $SpaA_{SLH}$
G2	Groove 2 of $SpaA_{SLH}$
G3	Groove 3 of $SpaA_{SLH}$
GlcNAc	N-acetylglucosamine
GT	Glycosyltransferase
GT-B	Glycosyltransferase fold B
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
IPTG	Isopropyl-β-D-thiogalactopyranosid
ITC	Isothermal titration calorimetry
K_D	Dissociation constant
ManNAc	N-acetylmannosamine
MPD	2-Methyl-2,4-pentanediol
NMR	Nuclear magnetic resonance
PEG	Polyethylene glycol
PDB	Protein data bank
RMSD	Root-mean-square deviation

Sap_{SLH}	SLH domains of *Bacillus anthracis* S-layer protein Sap
SCWP	Secondary cell wall polymer
S-layer	Surface layer
SLH1	First SLH domain of the protein sequence
SLH2	Second SLH domain of the protein sequence
SLH3	Third SLH domain of the protein sequence
SLH domain	S-layer homology domain
$SpaA_{SLH}$	SLH domains of *P. alvei* S-layer protein SpaA
$SpaA_{SLH}$/G109A	$SpaA_{SLH}$ single mutant
$SpaA_{SLH}$/G46A/G109A	$SpaA_{SLH}$ double mutant
TCEP	tris(2-carboxyethyl)phosphine
Tris	tris(hydroxymethyl)aminomethane
UPD	Uridine diphosphate

Acknowledgments

I am grateful to my supervisor, Prof. Stephen V. Evans. His encouragement and support throughout the course of my PhD have been instrumental in completing this research.

I also wish to thank my committee members, Profs. Alisdair Boraston, Caroline Cameron and John Taylor for their input and guidance over the past 5 years.

I owe special thanks to my collaborators Fiona Hager-Mair, Cordula Stefanović, Prof. Christina Schäffer and Prof. Paul Kosma, who's scientific contributions have made this work possible.

My time in the Evans lab has been enriched by amazing coworkers, with many of whom I have become close friends. It has been a pleasure getting to know all of you over the years.

Finally, to my loved ones. You fill my life with warmth and joy. Thank you.

Dedication

To my family.

Chapter 1: Introduction

Portions of this chapter are adapted or reproduced with permission (see Appendix A) from:

Legg, M.L., Hager-Mair, F.F., Krauter, S., Gagnon, S.M.L., Lòpez-Guzmán, A., Lim, C., Blaukopf, M., Kosma, P., C. Schäffer and S. V. Evans (2022). "The S-layer homology domains of *Paenibacillus alvei* surface protein SpaA bind to cell wall polysaccharide through the terminal monosaccharide residue." Journal of Biological Chemistry **298**(4): 101745.

1.1. Importance of S-layers

1.1.1. A history of S-layer discovery

Life on Earth falls into three domains, *Bacteria*, *Archaea*, and *Eukaryota*, based on the biological classification system Carl Woese *et al*. developed and published in 1990 (Woese and Fox 1977, Woese, Kandler *et al.,* 1990). "Prokaryote" is an even broader designation, encompassing both bacteria and archaea, which are unicellular organisms that lack a true nucleus. Typically, prokaryotes also possess a supramolecular and layered "wall" structure outside the cell membrane (Sapp 2005). This outer prokaryotic wall functions, in part, to maintain cell shape by resisting turgor pressure from the protoplast (Beveridge 1981, Beveridge 1988). The outer wall and cytoplasmic membrane together form the cell envelope, which regulates molecular exchange between the cell and its local environment (Beveridge 1981, Beveridge 1988). Throughout evolution, prokaryotes have accumulated a multiplicity of features that facilitate survival in complex and highly competitive ecosystems, including a dynamic and diverse set of cell envelope structures that are particularly notable at the outermost boundary layers of the cell. One of the most common cell envelope structures is the surface layer (S-layer)—a two-dimensional proteinaceous lattice-like array that encases the entire cell surface (Sleytr 1976, Sleytr, Schuster *et al.,* 2014).

Discovery of S-layers — In the early 1950s, Adriën Loo Houwink and Jan Bart Le Poole discovered the prokaryotic S-layer through electron microscopy experiments, where they observed a distinct "macromolecular mono-layer" present on cell fragments from the bacterium *Spirillum serpens* (Houwink 1953). Soon thereafter, Houwink noticed a similarly patterned array on the surface of the archaeal organism *Halobacterium salinarum* (Houwink 1956). This was the first documented evidence suggesting that S-layers were a conserved feature of the prokaryotic cell envelope.

However, neither Houwink nor his contemporaries recognized this at the time (Sleytr and Messner 1988).

Ubiquity of S-layers in Archaea & Bacteria — Since these initial discoveries, researchers have identified S-layers in most species of *Archaea* as well as in hundreds of different species of walled *Bacteria* from almost every taxonomic group (Messner and Sleytr 1992, Sleytr, Messner *et al.*, 1996, Sleytr, Messner *et al.*, 1999, Sleytr, Sára *et al.*, 2002, Claus, Akca *et al.*, 2005, König, Claus *et al.*, 2010, Sleytr, Egelseer *et al.*, 2010, Albers and Meyer 2011, Hynonen and Palva 2013). Despite growing interest in S-layers, no recognized term existed to describe these structures as a unified molecular class until the mid-1970s. Scientists had referred instead to S-layers disjointedly using various terms, including: "paracrystalline arrays," "regular-structured layers," "planar crystalline layers," and "surface layers" (Sleytr 2016). Uwe B. Sleytr coined the term "S-layer" in his 1976 description of S-layer reattachment in clostridia (Sleytr 1976). The scientific community adopted Sleytr's terminology after reaching a consensus at the First International Workshop on Crystalline Bacterial Cell Surface Layers, held in Vienna (Austria) in 1984.

Spontaneous S-layer Assembly — Beyond his taxonomic contributions (Sleytr 1976), Sleytr conducted a remarkable experiment that confirmed a growing body of evidence suggesting that S-layers could spontaneously assemble with or without cell wall supports (Brinton, McNary *et al.*, 1969, Buckmire and Murray 1970, Buckmire and Murray 1973, Glaubert and Sleytr 1975, Sleytr 1975). Sleytr's work demonstrated that the S-layer proteins from one organism could reassemble onto the surface of another organism, and that the proteins adopted the *same lattice symmetry* in each case (Sleytr 1976). As part of this research Sleytr also observed that the same reassembly phenomenon could occur in the absence of *any* cell wall support. This revealed another

extraordinary S-layer property: the dependence of S-layer lattice symmetry on protein identity is independent of underlying cell wall attachment (Sleytr 1976).

S-layers as model systems — S-layers are ubiquitous and modular in their assembly. As one of the simplest biological protein membrane structures to evolve in unicellular life—usually composed of a single protein or glycoprotein species (Sleytr 1975, Sleytr and Messner 1983)—they offer a useful model system for studying the synthesis, secretion, attachment, and assembly processes of prokaryotic extracellular proteins. Further, researchers have found therapeutic and biotechnological uses for S-layers (Kirk, Gebhart *et al.,* 2017, Fioravanti, Van Hauwermeiren *et al.,* 2019, Oatley, Kirk *et al.,* 2020, Ravi and Fioravanti 2021), and at present there remains a robust scientific interest in uncovering the molecular details of S-layer structure, function, and attachment, in order to explore this potential.

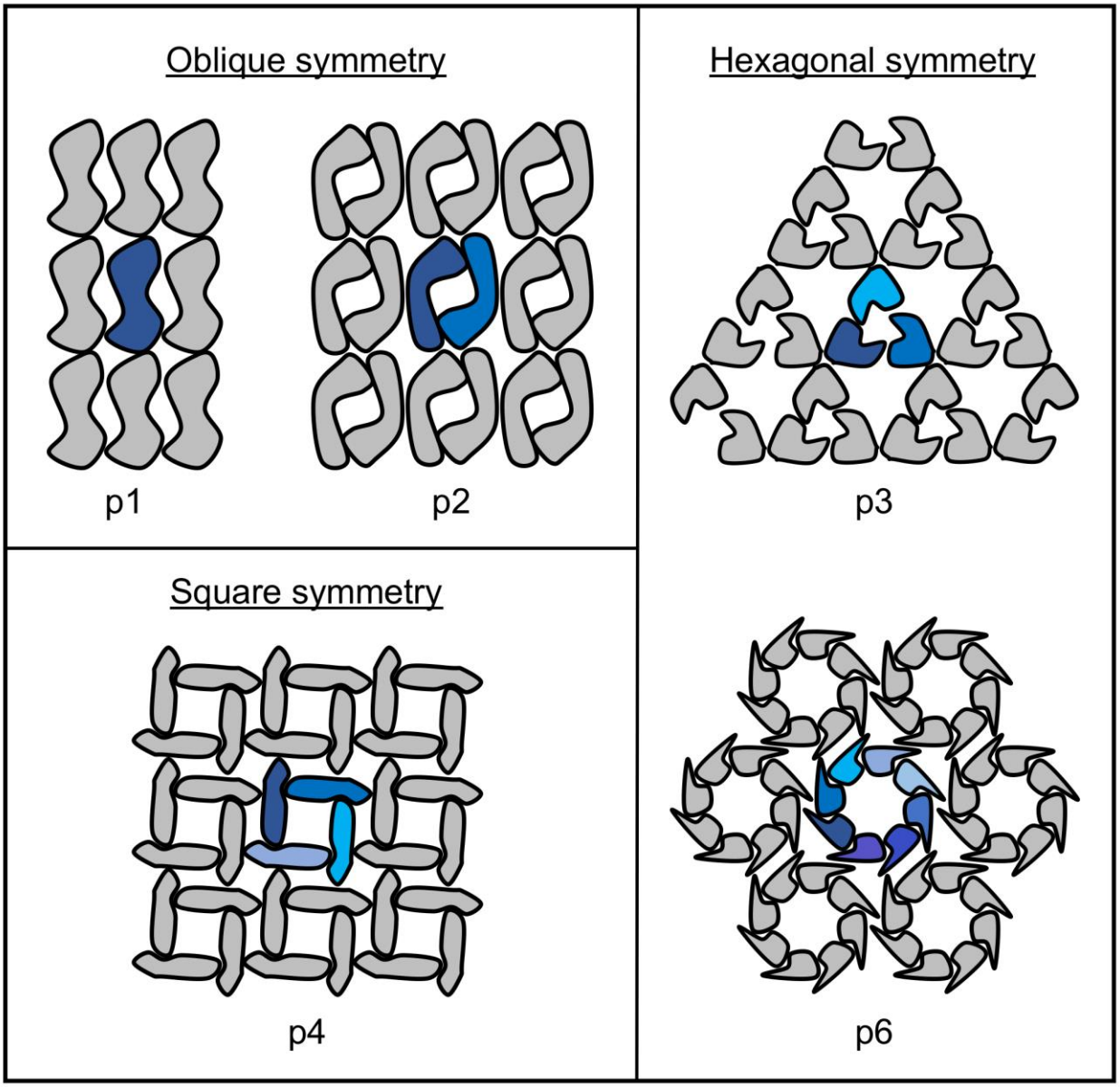
1.1.2. S-layer location & structure

As "the simplest type of biological membrane developed during evolution of prokaryotic cells" (Sleytr, Messner *et al.,* 1996), S-layers have attracted substantial scientific attention and have been studied in a wide variety of *Bacteria* and *Archaea* (Sleytr, Schuster *et al.,* 2014). Knowledge of S-layer structure and function, however, remains incomplete owing to a number factors: (i) S-layers are absent in the classical model organisms *Escherichia coli* and *Bacillus subtilis*, which has delayed their introduction into the scientific "main-stream"; (ii) S-layers are prone to spontaneous self-assembly, which hampers characterization of their structural and biochemical features; and (iii) S-layers exhibit low amino acid sequence identity, which makes sequence-based identifications of these proteins a challenge (Ravi and Fioravanti 2021).

Given that S-layers feature prominently at the outermost bounds of the cell, scientists have turned to electron and atomic force microscopy techniques to visualize these structures. Imaging reveals that S-layers assemble into regularized monomolecular arrays that envelope the entire cell surface and are typically present throughout all stages of prokaryotic growth and division (Sleytr, Schuster *et al.,* 2014). In some cases, two distinct S-layer lattices can sometimes be observed superimposed at the surface of certain organisms, though this phenomenon is a somewhat rare occurrence and has only been recorded a handful of times (Watson and Remsen 1970, Beveridge and Murray 1974, Sleytr, Schuster *et al.,* 2014, Suhr, Lederer *et al.,* 2016).

Despite the enormous sequence diversity of S-layer proteins (Bharat, von Kugelgen *et al.,* 2021), S-layer lattices are observed to adopt a limited number of symmetrical arrangements (**Figure 1**). The central morphological unit in these arrangements generally consists of one protein (or glycoprotein) species that is repeated throughout the S-layer structure (Sleytr, Messner *et al.,* 1999). This unit is formed from either one, two, three, four, or six protein subunits, which together exhibit either oblique (plane groups *p*1 or *p*2), square (plane group *p*4), or hexagonal (plane groups *p*3 or *p*6) lattice symmetry (Sleytr, Messner *et al.,* 1999) (**Figure 1**). In *Archaea*, the predominant (read: *nearly-exclusive*) form of lattice symmetry arrangement is hexagonal (Sleytr, Schuster *et al.,* 2014). By comparison, bacterial S-layers do not show a particular preference for hexagonal symmetries, and instead show a phylogenetically uncorrelated distribution of all five known lattice symmetries (Sleytr, Schuster *et al.,* 2014). Notably, certain bacteria, or their different strains, display lattice symmetries of alternate morphology for different S-layer proteins or S-layer protein species of different molecular weight (Fujimoto, Takade *et al.,* 1991, Sara and Sleytr 1994, Egelseer, Danhorn *et al.,* 2001, Schaffer, Wugeditsch *et al.,* 2002).

Figure 1. *Schematic of the different lattice symmetries possible in prokaryotic S-layers*



S-layers arrange into regular protein arrays that exhibit either oblique (*p*1, *p*2), square (*p*4), or hexagonal symmetry (*p*3, *p*6). The drawings shown above are based on the models proposed by Sleytr *et al.* (1999). The number of identical protein morphological units are highlighted in alternate shades of blue and/or purple.

Sequence analyses show that S-layer proteins vary in size between 40 and 200 kDa, and typically fold into multi-domain structures, some of which contain conserved secondary structural elements (Sleytr, Schuster *et al.,* 2014). Taken together, these properties (*i.e.*, protein shape and size), as well as the amino acid composition of the protein, influence center-to-center spacing (4-35 nm) between morphological units within the S-layer lattice (Sleytr, Schuster *et al.,* 2014). Spacing between these units creates a highly porous (30–70% porosity) and regularized structure with pores that are uniform in size (ranging between 2 and 8 nm) and that, in some cases, adopt more than one distinct type of morphology (Sleytr, Sára *et al.,* 2002, Albers and Meyer 2011). These features substantiate the proposal that S-layers can function as precise molecular sieves, whereby pores provide sharp cutoff levels for prokaryotic cells to restrict the free passage of molecules exceeding the molecular weight exclusion limits imposed by these structures (Sleytr, Sára *et al.*, 2002, Albers and Meyer 2011).

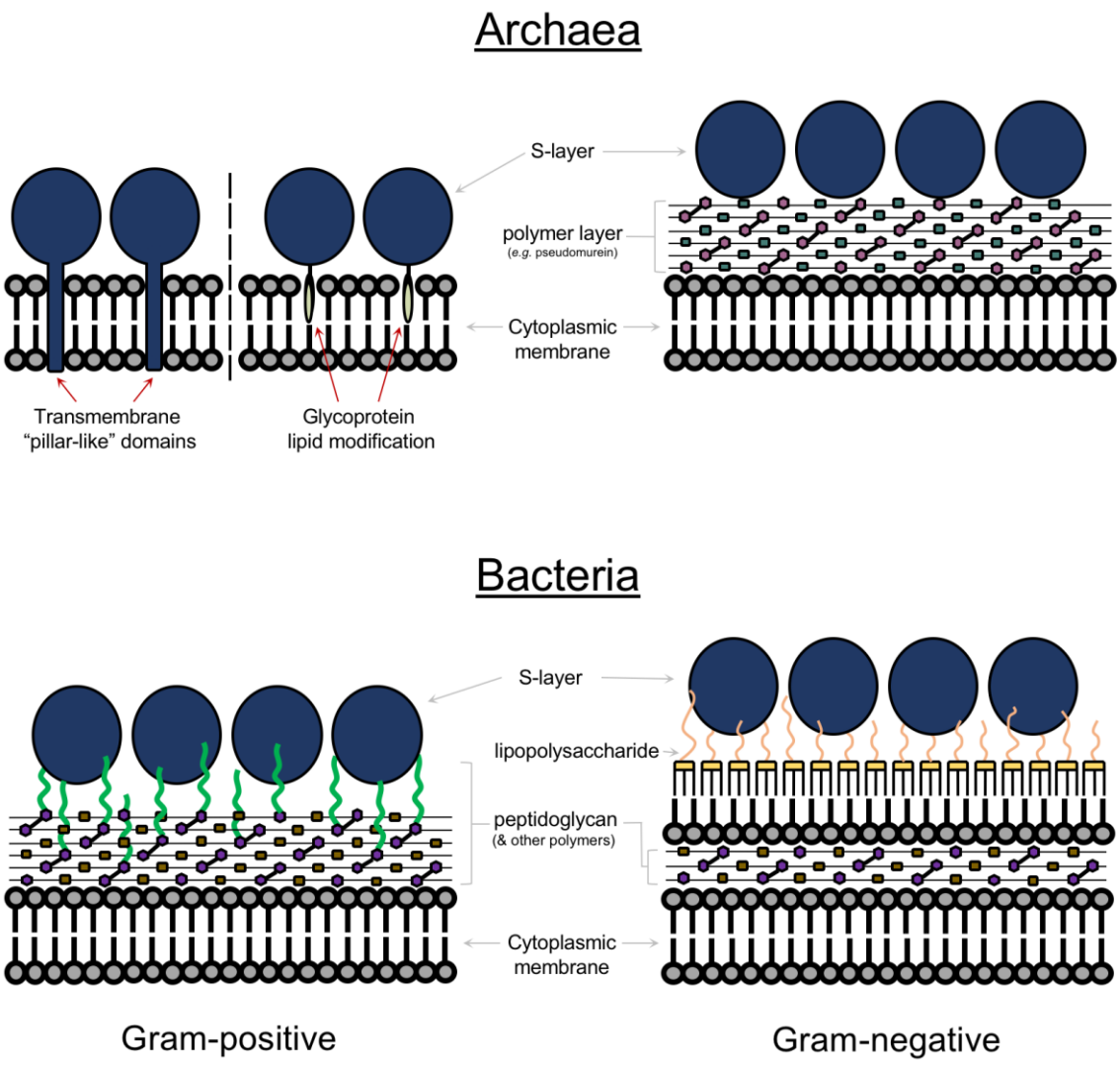
S-layer structure and organization is distinct among Archaea, Gram-negative bacteria, and Gram-positive bacteria. In Archaea, S-layers are often the only cell wall component that exists beyond the plasma membrane (Sleytr, Messner *et al.,* 1999), suggesting that beyond their filtration abilities S-layers are also critical determinants of cell shape and likely play roles in the archaeal mechanism of cell division. Due to the dearth of other structured extracellular support elements, archaeal S-layer proteins must therefore often anchor directly to the plasma membrane *via* either a hydrophobic trans-membrane "pillar-like" domain that embeds into the lipid membrane, or else *via* a glycoprotein lipid modification that inserts directly into the lipid membrane (Albers and Meyer 2011, Rodrigues-Oliveira, Belmok *et al.,* 2017) (**Figure 2**). Less frequently, in cases where *Archaea* do possess a rigid layer atop the plasma membrane (*e.g.*, pseudomurein), the S-layer

proteins will instead interact with this cell-wall feature rather than anchoring directly to the underlying lipid membrane (Sleytr, Schuster *et al.*, 2014) (**Figure 2**).

In Gram-negative bacteria, S-layer proteins generally associated with the lipopolysaccharide layer coating the outer regions of the plasma membrane (Sleytr, Schuster *et al.*, 2014). No unified S-layer-anchoring motif has yet been described for these proteins (Bharat, von Kugelgen *et al.*, 2021), however Gram-negative S-layer attachment occurs *via* non-covalent interactions, which typically include either ionic, carbohydrate-carbohydrate, protein-carbohydrate, or protein-protein contacts (Sleytr, Schuster *et al.*, 2014) (**Figure 2**).

S-layer anchoring for Gram-positive bacteria is more clearly defined than for Gram-negative bacteria. Contacts between Gram-positive S-layer proteins and the bacterial cell wall frequently involve non-covalent interactions with either the peptidoglycan layer or, as has been demonstrated more recently (Kern, Wilton *et al.*, 2011, Blackler, Lopez-Guzman *et al.*, 2018, Sychantha, Chapman *et al.*, 2018), with secondary cell wall polymer (SCWP) glycan structures that are covalently linked to the underlying peptidoglycan layer and that branch out from the cell surface (**Figure 2**). In either case, S-layer protein cell-wall anchoring is mediated *via* the binding action of specialized S-layer protein anchoring domains (Sleytr, Schuster *et al.*, 2014).

Figure 2*. Illustration of the major classes of prokaryotic cell envelope containing S-layers*



Archaeal S-layers (*top*) are composed of protein subunits that anchor to the cell surface *via* either hydrophobic trans-membrane domains or a lipid-modified glycoprotein subunit (*top, left*). In certain archaea that possess a rigid wall layer (*e.g.*, pseudomurein in methanogenic organisms) the S-layer will typically anchor to this structure instead of the plasma membrane (*top, right*). S-layers in Gram-positive bacteria (*bottom, left*) are bound to polymers that are linked to the rigid peptidoglycan layer. In Gram-negative bacteria (*bottom, right*), S-layers associate with the lipopolysaccharide of the outer membrane.

1.1.3. S-layers and bacterial pathogenesis

Beyond contributing to cell-maintenance functions in a more general sense (Sleytr, Schuster *et al.,* 2014), S-layers for several Gram-positive and Gram-negative bacterial pathogens play potentially significant roles in virulence through a variety of mechanisms (Carl and Dasch 1989, Kawai, Akatsuka *et al.,* 1998, Mignot, Denis *et al.,* 2001, Thompson 2002, Shimotahira, Oogai *et al.,* 2013, Rasmussen-Ivey, Figueras *et al.,* 2016, Ravi and Fioravanti 2021).

In *Tannerella forsythia*, a Gram-negative organism contributing to the "red complex" group of bacteria that cause severe and chronic periodontitis, S-layer function is associated with adhesion and invasion during infection while also operating to suppress the host proinflammatory cytokine response (Shimotahira, Oogai *et al.,* 2013). In yet other Gram-negative pathogens, S-layers promote virulence through alternate methods where, for example in *Aeromonas hydrophila* (Rasmussen-Ivey, Figueras *et al.*, 2016) and a number of *Campylobacter* species (Thompson 2002), S-layers are implicated in immune evasion strategies. *C. fetus*, which is an organism that mainly threatens people at higher risk of infection like the elderly and immunocompromised (Wagenaar, van Bergen *et al.,* 2014), has evolved a DNA inversion mechanism that generates antigenic variability in the S-layer that permits the organism to evade the host antibody response (Thompson 2002).

The best-characterized S-layer-carrying human pathogens to date, however, are the Gram-positive organisms *Bacillus anthracis* and *Clostridioides difficile*, although these pathogens are most known for reasons extending well beyond the realm of S-layer research.

B. anthracis is the etiological agent of anthrax and is considered a CDC Tier 1 biological agent that is infamous for its application as a bioweapon during the anthrax attacks of 2001 (Jernigan, Stephens *et al.,* 2001, Goel 2015). The cell surface in *B. anthracis* is encased by one of

two mutually-exclusive S-layer proteins, which form either the Sap or EA1 S-layer (Fioravanti, Van Hauwermeiren *et al.,* 2019). These S-layer structures exhibit distinct two-dimensional lattice morphologies, where switching between these two arrays occurs regularly throughout the life cycle of the cell and also during the course of infection (Ravi and Fioravanti 2021). Switching in this manner serves to temporarily change the phenotype (and function) of the cell-wall, which can limit imminent action of host immunity against this pathogen (Baillie, Hebdon *et al.*, 2003). However, while switching between Sap and EA1 immunogenic proteins makes sense as an immune evasion strategy during host infection (Baillie, Hebdon *et al.,* 2003), its purpose during other non-infectious stages of bacterial growth remains unclear, particularly given that S-layer remodeling is an energetically expensive phenomenon. (Ravi and Fioravanti 2021). Although the precise mechanism of the S-layer contribution toward *B. anthracis* virulence has not been established, recent studies have demonstrated that disrupting S-layer integrity shows promise as a mechanism for therapeutic intervention of anthrax, and this application may extend to other S-layer-carrying pathogens (Fioravanti, Van Hauwermeiren *et al.,* 2019).

C. difficile is an emergent threat in modern-day healthcare settings, where an increasing number of nosocomial infections are derived from antibiotic-resistant strains of this organism (Kirk, Gebhart *et al.,* 2017, Peng, Addisu *et al.,* 2017, Ravi and Fioravanti 2021). The S-layer in *C. difficile* is a rare example where the lattice is comprised of a repeat assembly of heterodimers that include one low-molecular-weight (LMW) and one high-molecular-weight (HMW) variant of the S-layer protein SlpA that form a tightly associated non-covalent complex (Calabi, Ward *et al.,* 2001). The LMW SlpA protein exhibits a high degree of antigenic variability between strains and is likely present at the outermost region of the cell surface (Calabi, Ward *et al.,* 2001, Willing, Candela *et al.,* 2015) where it is the component of the heterodimeric complex that is directly

involved in the assembly of the S-layer itself. The HMW SlpA protein is the component of the heterodimeric complex that is implicated in cell-wall anchoring (Calabi, Ward *et al.,* 2001, Willing, Candela *et al.,* 2015). Although the SlpA S-layer protein in *C. difficile* is associated with pathogenicity, for example, it is required for adherence to gastrointestinal tissue (Calabi, Calabi *et al.,* 2002), further biochemical characterizations are needed to elucidate if there are additional specific functions of this protein (Ravi and Fioravanti 2021).

Beyond these well-characterized examples of Gram-positive pathogenic organisms, other members of Firmicutes also form S-layer structures that are linked to virulence and pathogenicity (Sleytr and Messner 1983, Fagan and Fairweather 2014). The genus *Paenibacillus,* for example, includes mesophilic, endospore-forming, Gram-positive microorganisms with implications in agriculture, medicine, industry, and apiculture, that contain S-layers.

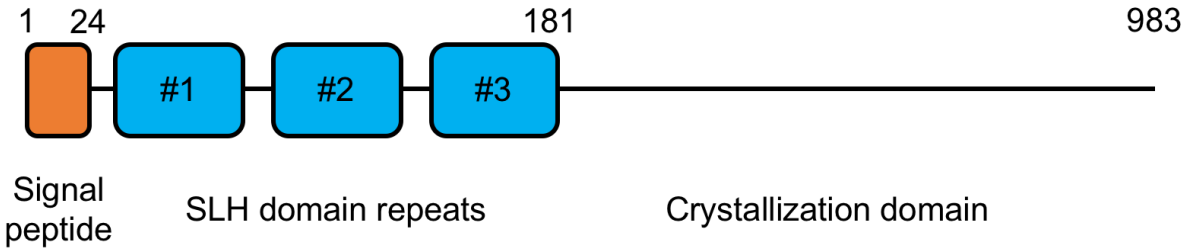
Paenibacillus larvae is the etiological agent of American foulbrood disease and a primary pathogen of honeybees (Poppinga, Janesch *et al.,* 2012). *P. larvae* only affects honeybees that are in the larval stage of development (Genersch, Forsgren *et al.,* 2006). Infection occurs through oral uptake of spore-contaminated larval food. These spores will settle in the midgut of the honeybee, where the vegetative form of the bacteria is able to proliferate and penetrate the epithelium in order to invade the haemocoel and eventually kill the host (Yue, Nordhoff *et al.,* 2008, Funfhaus and Genersch 2012). Proteomic and bioinformatic analyses revealed the presence of an S-layer protein, SlpA, in *P. larvae* (Funfhaus and Genersch 2012, Poppinga, Janesch *et al.,* 2012), which plays an important role as a virulence mediator for cell adhesion during the early stages of honeybee infection while also contributing to the determination of cell morphology (Poppinga, Janesch *et al.,* 2012).

P. alvei is another example of S-layer-carrying *Paenibacillus* species. This microorganism was reclassified from the *Bacillaceae* family into the *Paenibacillus* genus in 1994 (Ash, Priest *et al.,* 1993) and is a motile spore-forming bacterium that can strenuously swarm on semi-solid surfaces (Djukic, Becker *et al.*, 2012, Janesch, Koerdt *et al.,* 2013) and is known as a secondary invader in European foulbrood disease of honeybees (Forsgren 2010, Djukic, Becker *et al.,* 2012, Grossar, Kilchenmann *et al.*, 2020). Recently, Janesch *et al.* showed that *Paenibacillus alvei* CCM 2051^T possesses polar flagella that embed into the S-layer and that the flagellin protein Hag exerts a critical influence on the bacterium's biofilm life-style (Janesch, Koerdt *et al.,* 2013). *P. alvei* CCM 2051^T is enveloped by a proteinaceous S-layer composed of a single glycosylated protein species, consisting usually of either the SpaA or SlhA proteins, which arrange into a lattice with oblique symmetry (Zarschler, Janesch *et al.,* 2010). The presence of SlhA in *P. alvei* is vital for cell function, where disruption of its ability to attach to the cell wall impairs biofilm formation and also causes a loss of swarming motility (Janesch, Messner *et al.,* 2013). The SpaA protein, meanwhile, has been the target of recent investigations by our group and others (Zarschler, Janesch *et al.,* 2010, Blackler, Lopez-Guzman *et al.,* 2018) as it shares a set of conserved anchoring motifs in common with the SlhA protein of *P. alvei* and in other Gram-positive S-layers. Specifically, the *spaA* gene encodes a 983 amino acid protein with an N-terminal 24-amino acid signal peptide that is followed by three consecutive SLH domain repeats that correspond to the SLH domain superfamily known to be implicated in S-layer protein cell-wall anchoring. The remainder of the protein forms a crystallization domain that is involved in S-layer self-assembly and the formation of the two-dimensional protein lattice (**Figure 3**) (Zarschler, Janesch *et al.,* 2010).

Given that S-layers play significant roles in sustaining virulence and pathogenicity in a wide range of bacteria, surface-exposure, and strain-specificity offer promising targets for

diagnostic and therapeutic purposes. Studies aimed at characterizing these structures are therefore paramount toward exploiting this potential.

Figure 3. *Schematic representation of the domain architecture of* P. alvei *S-layer protein SpaA*



SpaA is comprised of an N-terminal signal peptide (residues 1-24), three SLH anchoring motifs (residues 25-181), and a crystallization domain that is implicated in S-layer self-assembly (residues 182-983).

1.2. Secondary cell wall polymers

1.2.1. SCWPs in Gram-positive bacteria

SCWP functions and location — The cell envelope in Gram-positive bacteria is decorated by a variety of accessory polysaccharides where a significant portion of these molecules attach covalently to the underlying rigid peptidoglycan exoskeleton *via* phosphodiester bonds, and extend out from the cell surface (Schaffer and Messner 2005, Silhavy, Kahne *et al.,* 2010). These polymers, most of which are anionic, are classified into three distinct groups on the basis of their structural characteristics as either (i) teichoic acids (Archibald, Baddiley *et al.,* 1968), (ii) teichuronic acids (Ward 1981), or (iii) other polysaccharides of neutral or acidic character that

cannot be classified into either of the former two groupings (Araki and Ito 1989, Naumova and Shashkov 1997). As a class, these compounds contribute to cell wall function largely in a secondary capacity (*e.g.*, by binding cations, by binding protons to acidify the cell wall during growth and division, by maintaining the peptidoglycan sacculus in an expanded state *via* charge repulsion, by forming a physical barrier to prevent metabolite and nutrients diffusion (Sara and Sleytr 2000)), and thus have been aptly termed "secondary" cell wall polymers (SCWPs) to reflect their biological roles (Schaffer and Messner 2005).

Classical SCWPs — Teichoic and teichuronic acids (now dubbed as the "classical" SCWPs (Schaffer and Messner 2005) are among the best characterized SCWPs based on analyses of their structure and function (Messner, Schaffer *et al.*, 2013). Their net negative charge derives predominantly from phosphate groups (as is the case in teichoic acids) or from carboxyl groups (as is the case in teichuronic acids, which are phosphate-free polymers that are synthesized in greater quantity when phosphate levels in the cell are low enough to halt teichoic acid biosynthesis) (Ellwood and Tempest 1969, Lang, Glassey *et al.*, 1982, Lahooti and Harwood 1999, Bhavsar, Erdman *et al.*, 2004). Acidic side chains containing glycerol, phosphate, sulphate, or organic acids (such as pyruvic acid) also contribute to the overall anionic character of these polymer (Schaffer and Messner 2005), to which several roles have been ascribed. These functions include (i) maintaining divalent cation homeostasis in the extracellular matrix in order to buffer the cell membrane against environmental cation fluxes, (ii) obtaining and sequestering metal ions that are required to facilitate metal-ion-dependent biochemical processes, (iii) binding and regulating the attachment of autolysins, some of which are required to break down peptidoglycan and enable turnover of the cell wall, and (iv) provision of phosphate as a natural reserve for various cell functions (Schaffer and Messner 2005). As these polymers account for 10–60 % (by weight) of

the bacterial cell wall, their biosynthesis necessitates a large expenditure of cellular resources and energy (Schaffer and Messner 2005), indicating that these compounds are vital in sustaining normal cell function (Schaffer and Messner 2005, Brown, Santa Maria *et al.,* 2013).

Non-classical SCWPs — Beyond this accumulated wealth of information concerning the features of teichoic and teichuronic acids from the cell walls of Gram-positive bacteria, there is an additional class of cell-wall polymer that is often present in these organisms. These "alternate" polymers were discovered in a variety of bacterial sources as by-products that co-isolated during the purification of S-layer glycoproteins (Messner, Sleytr *et al.*, 1987, Altman, Brisson *et al.*, 1990, Altman, Schaffer *et al.*, 1996, Schaffer, Kahlig *et al.,* 1999, Schaffer, Muller *et al.,* 2000, Steindl, Schaffer *et al.,* 2002, Schaffer and Messner 2004). With the release of information around that time on the multi-glycosylated state of S-layer proteins, early interpretations suggested that these glycan co-isolates corresponded to a second set of low-abundance glycan chains derived from the pool of S-layer glycoprotein glycoconjugates (Schaffer and Messner 2005). However, with the advent of improved purification and separation methods these glycan chains were instead revealed to represent a separate class of SCWP compound (*i.e.*, the "non-classical" SCWPs) that is a substantial constituent (7–15%) of the peptidoglycan layer (Schaffer and Messner 2005).

1.2.2. Non-classical SCWPs in Bacillaceae

Since the early days of cell wall research, researchers have identified peptidoglycan-linked compounds as common features underlying a variety of essential cell-maintenance functions (Schaffer and Messner 2005), and primary among these compounds are the "classical" SCWPs teichoic and teichuronic acid polymers. Schäffer and Messner (2005) have also defined a separate class of SCWP, the "non-classical" SCWPs in *Bacillaceae*, in order to further categorize this new

pool of SCWP compounds that match neither the biochemical nor structural definitions of the other class. Non-classical SCWPs have since been characterized as mediators for the non-covalent attachment of S-layer (glyco)proteins to the bacterial cell wall (Mesnage, Fontaine *et al.,* 2000, Sara 2001, Cava, de Pedro *et al.,* 2004).

In the *Bacillaceae* family, non-classical SCWPs constitute a large part of the peptidoglycan matrix, suggesting their presence is critical to normal cell function (Schaffer and Messner 2005). Typically just one type of non-classical SCWP is displayed on the surface of a given microorganism, meaning these compounds are homogeneous and distinctly strain-specific features (Schaffer and Messner 2005). Investigations by nuclear magnetic resonance (NMR) spectroscopy reveal SCWPs are linked to approximately 20–25% of *N*-acetylmuramic acid residues in the peptidoglycan layer *via* either pyrophosphate bridges or phosphodiester bonds to the C-6 atom of the muramyl peptidoglycan residues (Schaffer, Kahlig *et al.,* 1999). This direct attachment to peptidoglycan corroborates a major proposed function of non-classical SCWPs, where this class of compound serves predominantly as *in vivo* cell-surface anchors for S-layer (glyco)proteins in Gram-positive bacteria (Sara 2001).

Although somewhat varied in their structure, non-classical SCWPs share a number of common features: (i) they are linear or branched heteropolysaccharides and can be anionic or neutral in character, (ii) they may possess non-carbohydrate substituents such as pyruvate, acetate, or phosphate, and (iii) they vary in mass, usually ranging between 4000–6000 Da (Schaffer and Messner 2005). On the basis of these structural characteristics Schaffer and Messner (2005) created three further subgroupings to distinguish between non-classical SCWPs within S-layer carrying *Bacillaceae*, as follows:

- ***Group I* non-classical SCWPs** are comprised of polymers that most frequently contain the disaccharide backbone repeat [→3)-β-D-Man*p*NAc-(1→4)-β-D-Glc*p*NAc-(1→] (**Figure 4, *top left***). Although the ManNAc-GlcNAc disaccharide resembles the linkage unit of certain teichoic acids (Araki and Ito 1989), in non-classical SCWPs this disaccharide unit is repeated several times in order to form the entire polymer chain. In these SCWPs, ManNAc residues are often modified by pyruvic acid substitutions or attachment of a D-ribofuranose. The pyruvate modification confers an overall anionic character to the SCWP, as is the case for the SCWP observed in *P. alvei* CCM 2051^{T} (Schaffer, Muller *et al.*, 2000), and *Lysinibacillus sphaericus* CCM 2177 (Ilk, Kosma *et al.*, 1999). While certain strains of *Bacillus anthracis* and *Bacillus cereus* also possess pyruvylated SCWPs (Mesnage, Fontaine *et al.*, 2000), the branched polymers from these organisms have not yet been assigned to a further subgrouping of non-classical SCWP.
- ***Group II* non-classical SCWPs** consist of anionic polymers that contain the tetrasaccharide backbone structure [→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→] that is repeated on average six times throughout the polymer (**Figure 4, *top right***). A polymer from this group was the first SCWP-peptidoglycan complex to be fully elucidated following discovery in *Geobacillus stearothermophilus* NRS 2004/3a (Messner, Sleytr *et al.*, 1987, Schaffer, Kahlig *et al.*, 1999). In some cases, the 2,3-diacetamido-2,3-dideoxymannuronic acid moiety of the ManNAc residues is modified by amidation, as is the case in *G. tepidamans* GS5-97^{T}, which turns the anionic character of the glycan into a neutral one (Schaffer and Messner 2005).
- ***Group III* non-classical SCWPs** contain the same linkage-region teichoic disaccharide unit (as is seen in Group I), however this unit is extended even further in group III polymers and

includes branched repeats that can often contain neutral or amino sugars (**Figure 4, *bottom right***). An example is the charge-neutral SCWP isolated from *Aneurinibacillus thermoaerophilus* DSM 10155, which adopts a biantennary structure unique to bacteria. This biantennary structure resembles the glycans of a eukaryotic glycoprotein rather than the polysaccharides of a prokaryotic organisms (Schaffer and Messner 2005).

Overall, the variations in SCWP glycan length and structure could serve to accommodate specific *in vivo* functions, where the repetition of backbone units may improve the flexibility of these polymers in order to benefit S-layer (glyco)protein anchoring and, in these cases, longer and (sometimes) branched glycan chains would be better suited to provide more spatial flexibility to the proteins that accumulate and anchor to the cell surface during S-layer lattice expansion.

***Figure 4**. Structures of non-classical SCWP of S-layered Gram-positive bacteria*

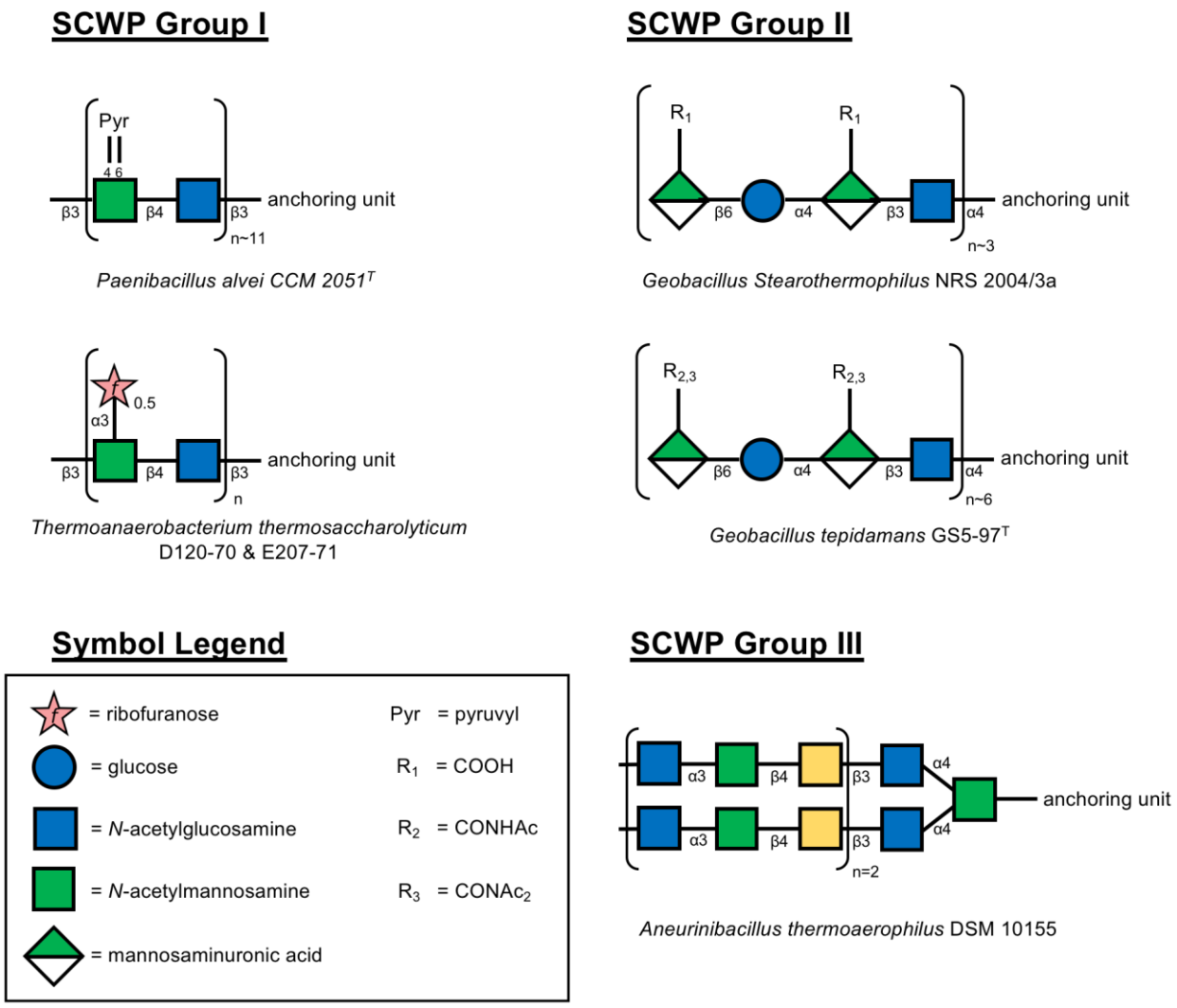


Figure adapted from (Schaffer and Messner 2005)

1.3. S-layer cell-wall anchoring

1.3.1. Anchoring targets in Gram-positive bacteria

Since the first cloning and sequencing reports on S-layer were published in the mid-eighties, thousands of S-layer genes from the domains *Archaea* and *Bacteria* have been sequenced and cloned (Sleytr, Schuster *et al.,* 2014). With this accumulation of genetic information, screening for putative sequence identities in S-layer proteins is now a possibility and provides useful insight into their putative functions. Although S-layer proteins generally exhibit low sequence identity, secondary structure predictions and analyses reveal conserved organizational elements within these sequences (Sleytr, Schuster *et al.,* 2014).

Generally, S-layer proteins adopt a bipartite arrangement with a lattice forming domain and an anchoring segment. The anchoring region, most often located at the protein N-terminus, is responsible for targeting and anchoring to the cell surface by binding to specific cell-surface structures. The self-assembly region, which is usually the larger of the two regions and located at the C-terminus, is responsible for 2-D structural arrangement between subunits of the S-layer lattice, with many of these interactions occurring in a manner that is metal-ion-dependent (Kessel, Wildhaber *et al.,* 1988, Baranova, Fronzes *et al.*, 2012, Sleytr, Schuster *et al.*, 2014, von Kugelgen, Tang *et al.,* 2020, Bharat, von Kugelgen *et al.,* 2021). Although the lattice-forming elements of S-layer proteins have remarkably divergent sequences, the anchoring domains exhibit distinctly less variation. Often, similar anchoring motifs are present in S-layer proteins from species with otherwise divergent S-layer protein sequences.

One example of this is from Gram-positive bacteria, where many S-layer proteins contain three tandem repeats of the cell-wall-binding protein modules (Kern, Wilton *et al.,* 2011, Usenik, Renko *et al.,* 2017, Blackler, Lopez-Guzman *et al.,* 2018). Although the exact position of the wall-

targeting region in S-layer proteins may vary among bacterial species, in *Bacillaceae* these domains are uniformly located at the protein N-terminus and bind specifically to SCWP targets covalently linked to the peptidoglycan layer of the cell (Sleytr, Schuster *et al.,* 2014).

Amino acid sequence analyses reveal that these specialized SCWP-binding domains are comprised of recurring structural motifs. Each motif consists of approximately 55 amino acid residues, of which 10–15 are conserved, and appear mostly as triplicate tandem repeats at the N-terminus of the S-layer protein (Sleytr, Schuster *et al.,* 2014). These so-called "S-layer homology" (SLH) domains are defined by their ability to bind specific SCWP ligands, which are comprised nearly exclusively of group I non-classical SCWPs that carry pyruvic acid modifications as part of their polymer structure (Ries, Hotzy *et al.,* 1997, Lemaire, Miras *et al.,* 1998, Chauvaux, Matuschek *et al.,* 1999, Ilk, Kosma *et al.,* 1999, Mesnage, Fontaine *et al.,* 2000, Cava, de Pedro *et al.,* 2004, Mader, Huber *et al.,* 2004, Runzler, Huber *et al.,* 2004, Huber, Ilk *et al.,* 2005).

Beyond roles in S-layer protein anchoring, SLH domains are also present in cellulosomes (Bayer, Belaich *et al.,* 2004) and other surface-associated enzymes (Liu, Gherardini *et al.,* 1996), indicating that SLH-mediated anchoring as a mechanism is not restricted uniquely to S-layer protein interactions. These motifs therefore represent a conserved form of surface attachment that is widely utilized to anchor cell-associated proteins, either permanently or transiently, to the Gram-positive bacterial cell surface. Further, in contrast with the previous assumption that peptidoglycan serves as the exclusive S-layer binding site (Lupas, Engelhardt *et al.,* 1994), it is now clear that secondary cell wall polymers are important structures required for anchoring of many proteins in addition to S-layer proteins, such as cell-associated exoenzymes, and exoproteins in Gram-positive organisms (Ries, Hotzy *et al.,* 1997, Lemaire, Miras *et al.,* 1998, Chauvaux, Matuschek *et al.,* 1999, Ilk, Kosma *et al.,* 1999, Mesnage, Fontaine *et al.,* 2000).

It should be noted, however, that not all bacterial S-layer proteins possess SLH domains, and other cell wall binding mechanisms have been proposed for those organisms that lack SLH domains. For instance, in the Gram-positive bacterium *Aneurinibacillus thermoaerophilus* (formerly *Bacillus thermoaerophilus*) cell wall anchoring is through a biantennary SCWP compound with an overall neutral charge (**Figure 4, *bottom right***) that is recognized by a region of basic residues from the S-layer protein (Steindl, Schaffer *et al.*, 2002). In *G. stearothermophilus* PV72/p6 and ATC 12980, the S-layer proteins SbsA and SbsC possess N-terminal regions with highly conserved tyrosine, lysine, and arginine residues that form charged-residue interactions with the negatively charged mannuronic acid-containing SCWPs. In a similar fashion, positively charged regions from the C-terminus of the S-layer proteins S_A and CbsA of *Lactobacillus acidophilus* ATCC4356 and *Lactobacillus cripsatus* JCM 5810, respectively, recognize teichoic and lipoteichoic acids as cell wall anchoring targets (Antikainen, Anton *et al.*, 2002, Smit and Pouwels 2002). Finally, in *C. difficile*, the SlpA S-layer protein, the Cwp66 adhesin, and the Cwp2, Cwp6, and Cwp8 proteins each contain three tandem repeats of the cell wall binding 2 (CWB2) domain that recognize and bind a negatively charged peptidoglycan-bound glycosyl phosphate polymer containing hexaglycosyl phosphate repeating units (Willing, Candela *et al.*, 2015).

1.3.2. Pyruvylation as an essential modification for SLH-anchoring

The generation of knock-out mutants in *Bacillus anthracis* and *Thermus thermophilus* (Mesnage, Fontaine *et al.*, 2000, Cava, de Pedro *et al.*, 2004) reveal that pyruvate modification of SCWP targets is necessary for SLH-mediated S-layer attachment. Specifically, the deletion of the gene locus encoding CsaB—a pyruvyltransferase enzyme responsible for the addition of pyruvate

to SCWP compounds—is sufficient to disrupt S-layer attachment (Mesnage, Fontaine *et al.*, 2000, Cava, de Pedro *et al.*, 2004). This phenomenon was confirmed by surface plasmon resonance studies, where the S-layer protein SbsB of *Geobacillus stearothermophilus* PV72/p2 showed a reduced ability to bind depyruvylated SCWP when compared to binding of the native SCWP ligand (Mader, Huber *et al.*, 2004). These results demonstrate a strong correlation between the presence of pyruvylated non-classical SCWPs, the CsaB pyruvyltransferase enzyme, and the existence of SLH motifs in Gram-positive organisms with S-layer proteins.

The importance of pyruvate modification of SCWPs was further substantiated in *B. anthracis*, where Δ*csaB* mutant strains expressed S-layer proteins that were unable to associate with the bacterial cell surface (Mesnage, Fontaine *et al.*, 2000). Pyruvate modifications of non-classical SCWP compounds has since been documented in several bacterial species (Mesnage, Fontaine *et al.*, 2000), indicating that pyruvylation, or perhaps more generally, the negative charge of these SCWP compounds constitutes a widespread motif that is necessarily conserved to permit SLH-mediated S-layer anchoring (Sleytr, Schuster *et al.*, 2014).

In fact, in organisms that possess an S-layer but no associated SLH-anchoring motifs, no pyruvylation of SCWPs is observed, often with the corresponding SCWP structure displaying an overall net-neutral charge (Schaffer and Messner 2005), and, additional binding mechanisms are undoubtedly required to mediate S-layer protein interactions for these Gram-positive organisms (Mesnage, Fontaine *et al.*, 2000, Sara 2001, Cava, de Pedro *et al.*, 2004, Janesch, Messner *et al.*, 2013).

Presumably, this diversity of SCWP structures and the associated variety of S-layer anchoring motifs is responsible for creating microenvironments that favour certain organisms under a range of conditions (Sleytr, Schuster *et al.*, 2014). Characterization of both the

biosynthesis and structure of several nonclassical SCWPs, as well as the details of the SCWP-SLH domain interaction, will be necessary to broaden our understanding of mechanisms that underlie S-layer (glyco)protein tethering to the Gram-positive cell surface.

1.3.3. SLH domain structure and characterizations

Recently, our group (Blackler, Lopez-Guzman *et al.,* 2018) and others (Janesch, Messner *et al.,* 2013) have characterized the role of the SLH trimer from the glycosylated S-layer protein SpaA of *Panibacillus alvei* CCM 2051^{T}. *In vivo* binding assays show that the presence of the SpaA SLH trimer ($SpaA_{SLH}$) is sufficient to display foreign proteins at the cell surface of *P. alvei* (Janesch, Messner *et al.,* 2013). In an interesting observation from the same publication, *in vitro* binding experiments show that recombinant SpaA can bind to native peptidoglycan-containing cell wall sacculi of *P. alvei* CCM 2051^{T} that either contains or is deprived of SCWP (although binding to the SCWP-free peptidoglycan structure is greatly reduced relative to that of the SCWP-containing structure) (Janesch, Messner *et al.,* 2013). Together, these results demonstrate that the SLH domains of SpaA can function *via* a dual-recognition binding mechanism where the protein can anchor either to SCWP or to peptidoglycan (Janesch, Messner *et al.,* 2013).

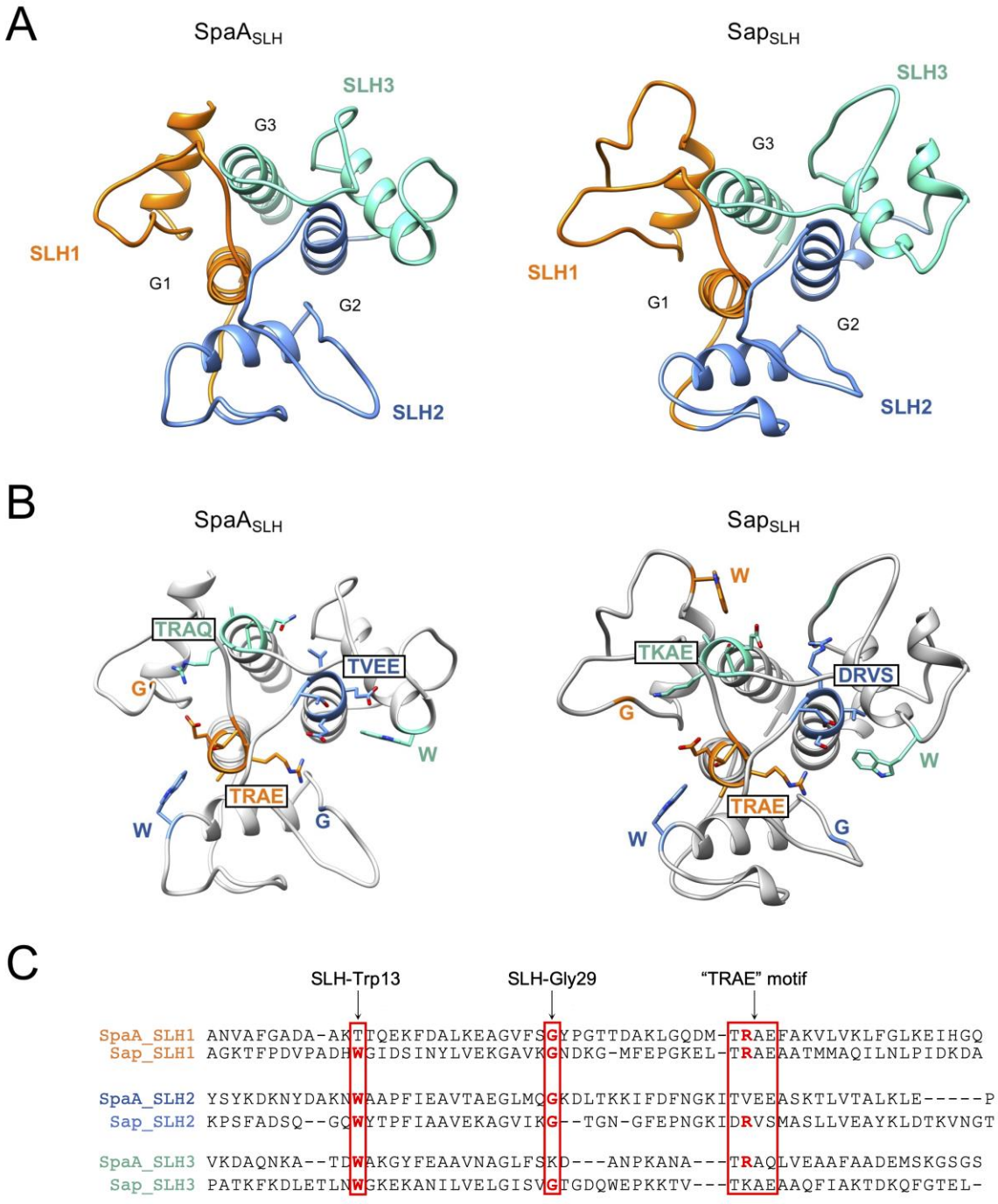
More recently, our group investigated the structure of the *P. alvei* SpaA SLH-trimer using X-ray crystallographic and thermodynamic binding methods. These studies showed that the SLH trimer of *P. alvei* SpaA adopts a three-pronged symmetrical trefoil arrangement that correlates well with the published structure of the SLH trimer from the *B. anthracis* S-layer protein Sap (Sap_{SLH}) (Kern, Wilton *et al.,* 2011) (**Figure 5**). Although these two SLH domain trimers exhibit rather low overall amino acid sequence identity, each SLH domain contains a variation of the 'TRAE' amino acid sequence-motif, as well as a copy of the highly conserved glycine and

tryptophan residues SLH-Gly29 and SLH-Trp13, respectively (**Figure 5**), which are together important features required for the binding to SCWP (Kern, Wilton *et al.,* 2011, Janesch, Messner *et al.,* 2013, Blackler, Lopez-Guzman *et al.,* 2018). Overall, the crystal structures of $SpaA_{SLH}$ (of *P. alvei*) and Sap_{SLH} (of *B. anthracis*) demonstrate that the SLH trimers are arranged to form three distinct grooves interspersed between "lobe-like" regions of the protein structure. In the unliganded state, the conserved TRAE motif and SLH-Gly29 residue (Kern, Wilton *et al.,* 2011, Blackler, Lopez-Guzman *et al.,* 2018), which are located near the base of each of the binding pockets, are exposed to solvent.

As part of these structural characterizations, Blackler *et al.* (2018) demonstrated that the protein structure of the $SpaA_{SLH}$ is quite adaptable and undergoes a remarkable conformational rearrangement to accommodate a terminal monosaccharide SCWP ligand within at least two of the three observed binding grooves (*i.e.*, in a mutually-exclusive manner in either G2 or G1). This observation offers a mechanism for this SLH domain and, conceivably, of those in S-layer proteins of other Gram-positive organisms, to relieve S-layer strain that is inherent to processes of cellular division and growth (Blackler, Lopez-Guzman *et al.,* 2018).

Beyond *P. alvei* and *B. anthracis*, SLH domain binding has also been identified in the cell wall of other organisms, such as *Thermus thermophilus*, where there is a strong interaction between the SLH domain of the S-layer protein and the pyruvylated component of the SCWP at the cell surface (Cava, de Pedro *et al.,* 2004), as well as in *G. stearothermophilus* PV72/p2, where the SLH domain of the SbsB S-layer protein was determined to be exclusively responsible for binding to SCWP (Mader, Huber *et al.,* 2004).

Figure 5. *SLH trimers from SpaA (*P. alvei*) and Sap (*B. anthracis*) S-layer proteins adopt a trefoil fold*



(**A**) The SLH1 (orange), SLH2 (blue), and SLH3 (cyan) domains from SpaA (*P. alvei*, PDB: 6CWC) and Sap (*B. anthracis*, PDB: 3PYW) arrange into a trimeric structure with three grooves, labeled as: G1, G2, and G3. (**B**) Conserved residues generally important to binding are shown within the trimeric SpaA and Sap SLH structures using single-letter amino acid codes. (**C**) Structure-based amino acid sequence alignment (using the pairwise structure alignment tool available at https://www.rcsb.org) between the SLH1, SLH2 and SLH3 domains of the SpaA and Sap proteins shows that despite having the same overall protein fold, relatively few amino acid residues are conserved among SLH domains.

1.4. $SpaA_{SLH}$ binding and structure elucidation

The crystal structures of *P. alvei* CCM 2051^T $SpaA_{SLH}$ show that the protein trimer contains a conserved TRAE motif at the N-termini of each core helix (**Figure 5**) (Blackler, Lopez-Guzman *et al.*, 2018). The exact TRAE motif is present in the SLH1 domain, while the SLH2 and SLH3 domains contain TVEE and TRAQ sequence variants, respectively. Interestingly, the residues of each motif do not all contribute to the same groove. The TRAE and TRAQ motifs of SLH1 and SLH2, respectively, each contribute to two of the three binding groves *via* an arginine residue, which protrudes under the connecting loops of polypeptide adjoining the SLH domain α-helices and into the neighboring groove (**Figure 5**), while the remaining TVEE sequence (of SLH2) has a valine residue in this place. The final residue of each motif (Glu63, Glu127, Gln179) lines the groove adjacent to its parent helix (**Figure 5**). Therefore, TRAE of SLH 1 contributes residues to G1 and G2, while TVEE of SLH2 contributes residues to G2 and G3, and TRAQ of SLH3 contributes residues to G1 and G3 (**Figure 5**).

In vitro cell binding assays show that these motifs do not contribute equally to binding, where mutated recombinant SpaA variants bind to *P. alvei* CCM 2051^T cell wall peptidoglycan with reduced affinities (Janesch, Messner *et al.*, 2013). Relative to the wild-type SpaA protein, TAAA mutagenesis of SLH1, 2, and 3 reduces cell wall binding by 37%, 88%, and 51%, respectively (Janesch, Messner *et al.*, 2013), suggesting that conservation of arginine in this motif is important to SCWP binding. This is validated in the published monosaccharide-bound $SpaA_{SLH}$ co-crystal structures, where the positively-charged side chain of the TRAE arginine residue is able to accommodate the negatively-charged pyruvate moiety of the terminal SCWP ligand at the base of the binding sites in either G2 (*via* the SLH1 TRAE arginine residue) or G1 (*via* the SLH3 TRAQ arginine residue) (Blackler, Lopez-Guzman *et al.*, 2018).

Further inspection of these crystal structures revealed additional conserved residues are also important for SCWP monosaccharide binding (Blackler, Lopez-Guzman *et al.,* 2018). For example, each SLH domain contains a highly conserved GIIxG motif located in the loop between its α-helices (Blackler, Lopez-Guzman *et al.,* 2018). Remarkably, monosaccharide-binding in G2 of wild-type $SpaA_{SLH}$, induces a "backbone flip", where the phi angle of SLH-Gly29 (Gly109 of $SpaA_{SLH}$) and the psi angle of SLH-x28 (the last two residues in the GIIxG motif) each change by ~180 ° in order to accommodate the ligand's pyruvyl carboxyl group (Blackler, Lopez-Guzman *et al.,* 2018). The backbone flip relies on SLH-Gly29's inherent main chain flexibility, which explains the near-universal conservation of this residue.

Blackler *et al.* (2018) probed the significance of this conserved glycine by mutating SLH2 residue Gly109 (corresponding to SLH-Gly29 of G2) to alanine in order to reduce the conformational flexibility of the residue at this position. The resultant $SpaA_{SLH}$/G109A mutant bound SCWP monosaccharide with reduced affinity relative to wild-type $SpaA_{SLH}$ (an ~8-fold reduction as measured by ITC); however, and quite remarkably, the monosaccharide-bound $SpaA_{SLH}$/G109A co-crystal structure undergoes significant conformational rearrangement relative to the wild-type unliganded and monosaccharide-bound structures in order to accommodate the monosaccharide ligand in G1 instead of the deactivated G2. This ability to switch between two sites was postulated as a potential mechanism wherein SpaA is able to utilize these two binding grooves to alleviate S-layer strain inherent to processes of cellular division and growth (Blackler, Lopez-Guzman *et al.,* 2018).

Blackler *et al.* (2018) further examined the effects of a double deactivating pair of mutations ($SpaA_{SLH}$/G46A/G109A) intended to impede binding in G2 (*via* the G109A mutation) and G1 (*via* the G46A mutation). In a somewhat surprising observation, the

$SpaA_{SLH}$/G46A/G109A construct showed no detectable binding toward the SCWP monosaccharide *via* ITC, and yet the monosaccharide was observed bound within the deactivated G2 site of the protein. However, the co-crystal structure itself reveals that no backbone flip occurs in G2 when glycine is replaced with alanine, which explains the lack of detectable binding measured by ITC and suggests that the observation of monosaccharide in G2 was an artefact of crystallization. This result also emphasizes the apparent lack of binding by G3 and that, even without a backbone flip, G2 exhibits significant structural complementarity toward the terminal SCWP monosaccharide.

While certainly illuminating, $SpaA_{SLH}$ structural studies had to that point been limited to the terminal monosaccharide 4,6-Pyr-β-D-ManNAcOMe and the internal repeat disaccharide β-D-GlcNAc-(1→3)-4,6-Pyr-β-D-ManNAcOMe SCWP ligand analogues. In order to advance understanding of the SCWP-binding process and define the specific epitopes critical for binding, a full study into the binding of terminal disaccharide (or even longer ligands) was deemed necessary, and is thus the main topic of the work reported within this dissertation (see *Summary of Objectives* in Chapter 1.7, and all content presented in Chapter 2).

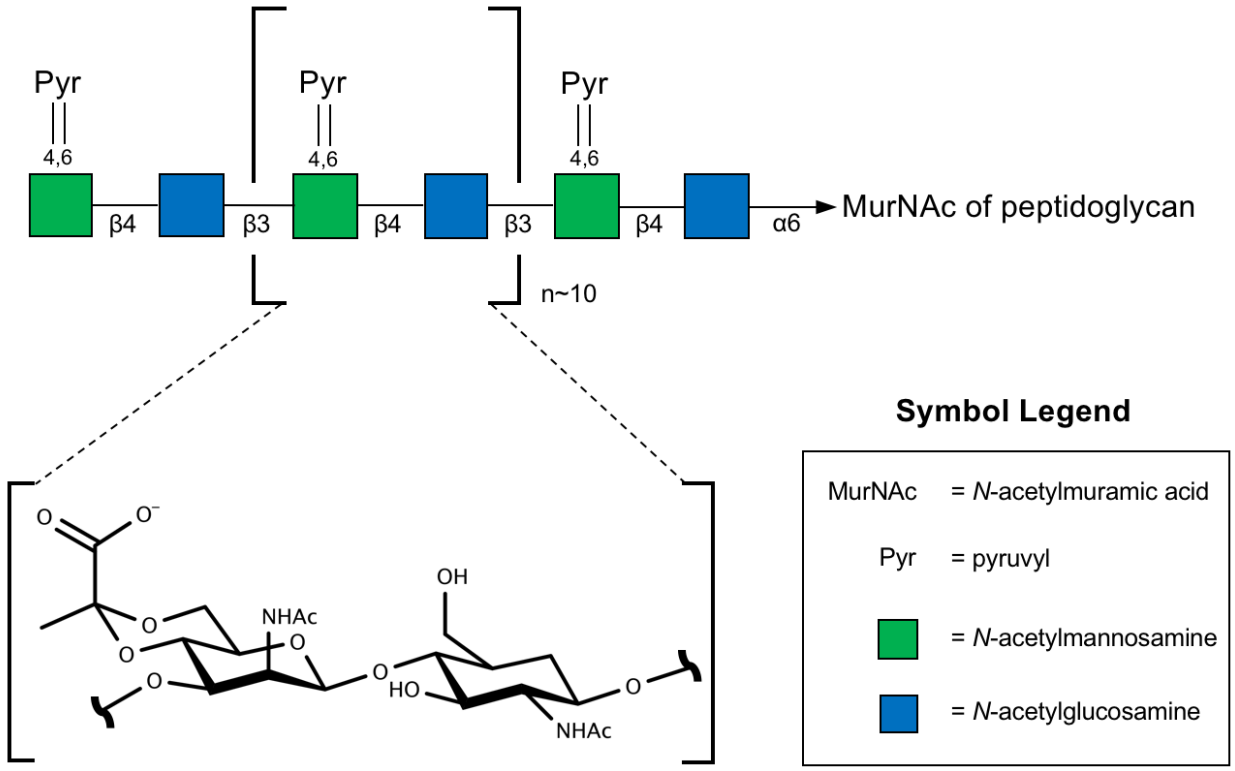
1.5. Biosynthesis of SCWPs in **P. alvei** *CCM 2051^T*

Non-classical SCWPs are species- and strain-specific polysaccharides carrying a neutral or negative net charge that are often covalently linked to the muramic acid residues of bacterial peptidoglycan by a phosphodiester bond (Schaffer, Muller *et al.*, 2000, Schaffer and Messner 2005). In *P. alvei* CCM 2051^T, SCWPs consist of the pyruvylated GlcNAc-ManNAc disaccharide repeat with a backbone structure of [→4)-β-D-GlcNAc-(1→3)-4,6-Pyr-β-D-ManNAc-(1→]

(**Figure 6**) (Schaffer, Muller *et al.,* 2000). The CsaB pyruvyltransferase enzyme catalyzes ketal pyruvylation of each ManNAc residues in the polymer chain (Hager, Lopez-Guzman *et al.,* 2018). This modification imparts negative charge throughout the entire SCWP structure and supports assembly of the S-layer envelope of *P. alvei*, which maintains bacteria cell wall integrity *via* SLH domain-mediated interactions with S-layer protein SpaA (Janesch, Messner *et al.,* 2013) and SlhA (Janesch, Koerdt *et al.,* 2013).

Within the *P. alvei* genome, the *spaA* and *slhA* S-layer genes are located in close proximity to an SCWP biosynthetic gene locus that contains five ORFs: *orf1* (PAV_RS07430), *csaB* (PAV_RS07425), *tagA* (PAV_RS07420), *tagO* (PAV_RS07415), and *orf7* (PAV_RS07395) (Zarschler, Janesch *et al.,* 2010, Hager, Lopez-Guzman *et al.,* 2018). Two of these ORFs, *tagA* and *tagO*, encode proteins that share 32 – 42% sequence identity with GTs from different Bacillaceae with proven roles in polyribitol wall teichoic acid biosynthesis (Ginsberg, Zhang *et al.,* 2006, Zarschler, Janesch *et al.,* 2010). TagO, the first enzyme in this pathway, transfers GlcNAc-phosphate from donor UDP-GlcNAc to the undecaprenylphosphate carrier lipid acceptor, which is embedded in the cytoplasmic membrane (**Figure 7**). To extend the sugar chain, the GT TagA catalyzes ManNAc transfer from UDP-ManNAc to the terminal GlcNAc, producing the disaccharide substrate that corresponds to the repeat backbone unit of the *P. alvei* SCWP (Schaffer, Muller *et al.,* 2000) (**Figure 6**). As a final step, the pyruvyltransferase CsaB transfers a pyruvate to the ManNAc-GlcNAc acceptor (Hager, Lopez-Guzman *et al.,* 2018) (**Figure 7**). CsaB gene deletion in Gram-positives severely reduces protein-cell wall interactions and pathogenicity (Mesnage, Fontaine *et al.,* 2000, Kern, Wilton *et al.,* 2011), emphasizing the importance of pyruvate and, presumably, its associated negative charge.

Figure 6. *Schematic of* P. alvei *CCM 2051^T SCWP*

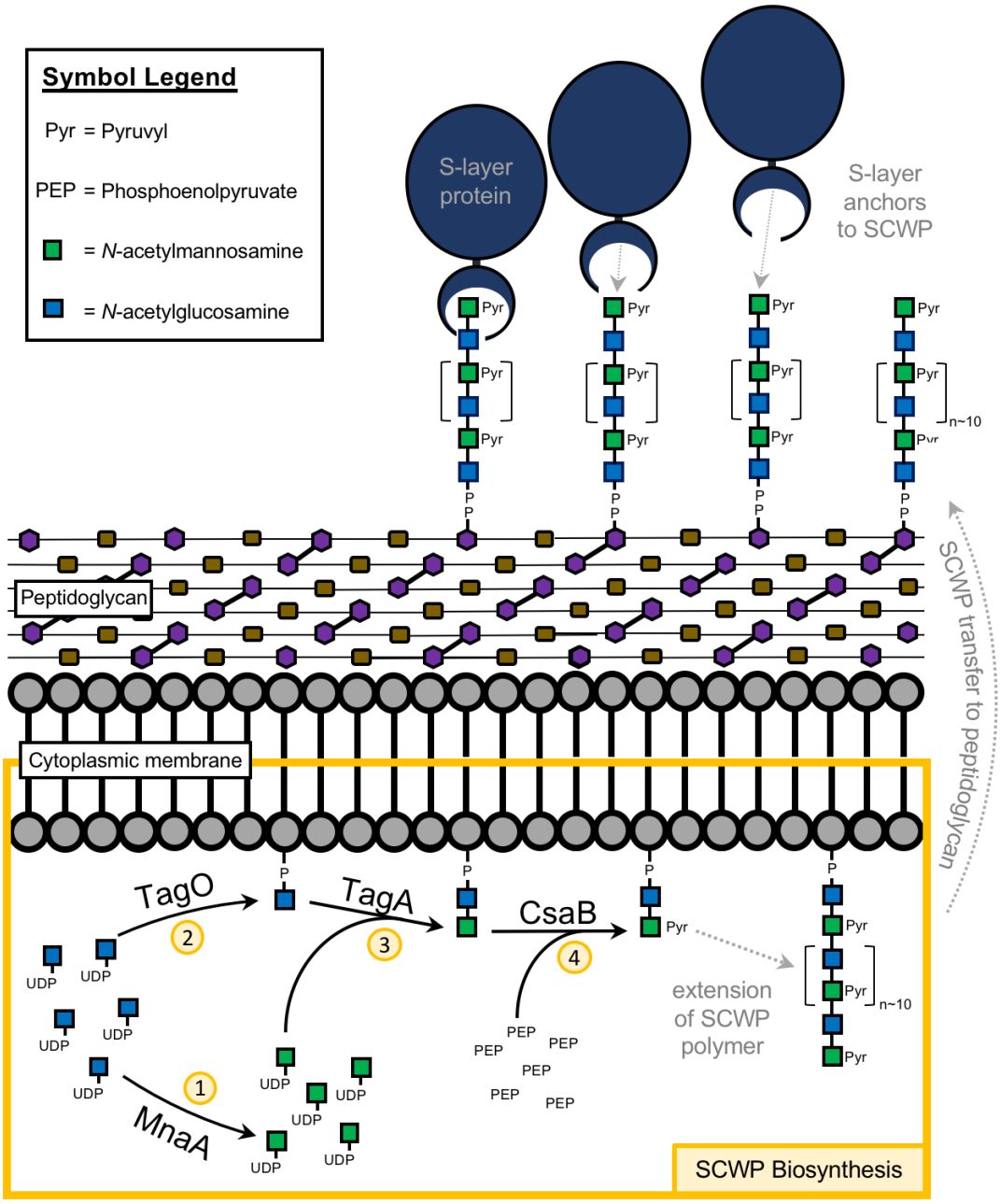


The *P. alvei* CCM 2051^T SCWP consists of ~eleven (~11) →3)-4,6-Pyr-β-D-Man*p*NAc-(1→4)-β-D-Glc*p*NAc-(1→ disaccharide repeats. A detailed chemical diagram of one repeat (*left*) is shown to illustrate the stereochemistry of the SCWP.

Obviously, ManNAc is a key residue of this well-characterized *P. alvei* SCWP disaccharide repeating unit and is the substrate of TagA, the second enzyme in the proposed biosynthetic pathway. *P. alvei* MnaA, of the UDP-GlcNAc 2-epimerase protein family (PF02350) (El-Gebali, Mistry *et al.*, 2019), catalyzes the reversible conversion of UDP-GlcNAc to UDP-ManNAc (**Figure 8**) (Velloso, Bhaskaran *et al.*, 2008, Hager, Lopez-Guzman *et al.*, 2018), and this reaction generates the UDP-ManNAc substrate that is subsequently required for TagA to incorporate ManNAc into the *P. alvei* SCWP (**Figure 7**).

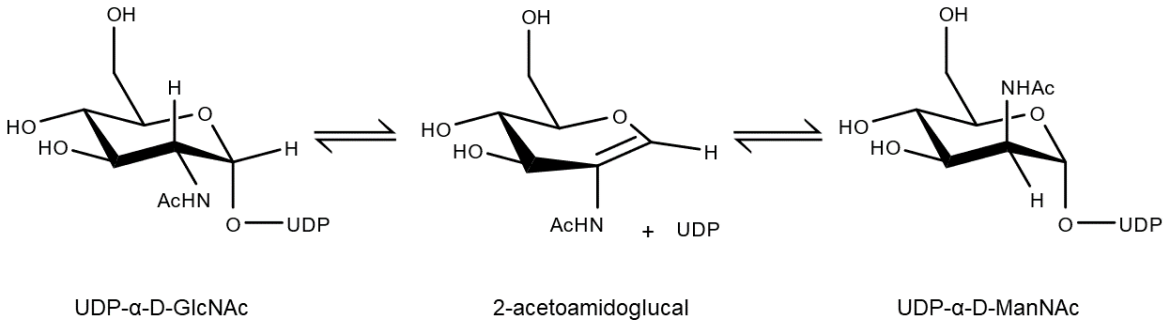
To confirm the biosynthetic roles of these enzymes, Hager *et al.* (2018) demonstrated that a lipid-linked *P. alvei* SCWP disaccharide backbone repeat could be produced *via* a "one-pot" reaction where recombinant versions of the TagA UDP-ManNAc transferase, the CsaB pyruvyltransferase, and the MnaA UDP-GlcNAc 2-epimerase enzymes, were combined together with 11-phenoxyundecyl-diphosphoryl-α-GlcNAc acceptor, and phosphoenolpyruvate and UDP-GlcNAc donor molecules. In a separate experiment, the authors of this work also showed that the *csaB*, *tagA*, and *tagO* genes are linked as one transcriptional unit, which makes their concerted action a likely scenario (Hager, Lopez-Guzman *et al.*, 2018). While the *mnaA* gene (PAV_RS07610) is located elsewhere on the genome, the UDP-GlcNAc 2-epimerase MnaA was included in this biosynthetic *in vitro* study since, *in vivo*, UDP-GlcNAc epimerization is a necessary prerequisite for provision of the UDP-ManNAc donor substrate to TagA, and only a single copy of each of these genes was identified in the *P. alvei* genome, suggesting they are important enzyme components of this biosynthetic pathway as no other enzymes have yet been identified that could perform the same essential functions.

Figure 7. *Proposed SCWP biosynthetic pathway in* P. alvei *CCM 2051*^{T}



Schematic representation of the multi-enzyme biosynthetic pathway proposed for the production of a lipid-linked SCWP repeat in *P. alvei* CCM 2051^{T} (Hager, Lopez-Guzman *et al.*, 2018). The non-hydrolyzing UDP-GlcNAc 2-epimerase MnaA (1) catalyzes the reversible epimerization of UDP-α-D-GlcNAc to UDP-α-D-ManNAc in order to provide UDP-α-D-ManNAc as a donor substrate for incorporation into the SCWP biopolymer. The TagO glycosyltransferase (2) transfers GlcNAc-phosphate from a donor UDP-GlcNAc molecule to the undecaprenylphosphate carrier lipid acceptor, which is embedded in the cytoplasmic membrane. TagA (3) catalyzes ManNAc transfer from UDP-ManNAc to the terminal GlcNAc, producing the disaccharide substrate that corresponds to the repeat backbone unit of the *P. alvei* SCWP (Schaffer, Muller *et al.*, 2000). The pyruvyltransferase CsaB (4) transfers a pyruvate to the ManNAc-GlcNAc acceptor.

Figure 8. *Epimerization reaction catalyzed by MnaA UDP-GlcNAc 2-epimerase*



The non-hydrolysing UDP-GlcNAc 2-epimerase MnaA catalyzes the reversible epimerization of UDP-α-D-GlcNAc to UDP-α-D-ManNAc at the C2 position, which occurs *via* the intermediates 2-acetoamidoglucal and UDP. UDP-ManNAc generated by MnaA is used in SCWP synthesis in *P. alvei* CCM 2051^{T}.

1.6. Structural knowledge of enzymes in the P. alvei *SCWP biosynthetic pathway*

To date no crystal structure of any SCWP biosynthetic enzyme from *P. alvei* has been determined, and thus all structural insights into the enzymatic components of this pathway are derived from the protein structures of related enzymes in other organisms.

Insight into MnaA structure — The MnaA epimerase of *P. alvei* is most closely related (67.2 % amino acid sequence identity) to the bacterial non-hydrolyzing UDP-GlcNAc 2-epimerase enzyme of *Bacillus anthracis* for which a crystal structure has been determined (Velloso, Bhaskaran *et al.*, 2008). This enzyme adopts a GT-B fold that is characterized by two tightly-associated α/β/α Rossman-like domains, one at the protein N-terminus and one at the protein C-terminus, which are attached by a short linker region to form a substrate binding cleft between the two lobes. The substrate-bound co-crystal structure from *B. anthracis* reveals that the epimerase possesses both a catalytic site (with electron density that corresponds to only UDP of the UDP-GlcNAc substrate bound at this site) and an allosteric site (with electron density that corresponds to the entire UDP-

GlcNAc effector molecule bound at this site) situated within the binding cleft, both of which are in close proximity to one another (Velloso, Bhaskaran *et al.,* 2008). While no substrate-bound UDP-GlcNAc 2-epimerase structure has yet shown the full set of contacts that occur between the effector GlcNAc residue in the allosteric site and substrate GlcNAc residue in the catalytic site, the structure of *B. anthracis* MnaA and other UDP-GlcNAc 2-epimerases do reveal that residues involved in binding (at both sites) are highly conserved among other non-hydrolyzing bacterial UDP-GlcNAc 2-epimerase enzymes. Therefore, if a crystal structure of *P. alvei* MnaA with the substrate or a substrate analog bound at the catalytic site could be generated, the residues involved in catalysis could be determined. This enzyme is the subject of the second research aim presented within this dissertation (see *Summary of Objectives* in Chapter 1.7, and all content presented in Chapter 3).

Current TagA structures — TagA, which is an metal-independent inverting GT-B fold glycosyltransferase that catalyzes ManNAc transfer from donor UDP-ManNAc to a lipid-linked acceptor, is classified as a GT26 enzyme in the Carbohydrate-Active enZyme database (CAZy). The Clubb lab from the University of California, Los Angeles has recently published the only known structures for any enzyme of this class (Kattke, Gosschalk *et al.,* 2019, Martinez, Mahoney *et al.,* 2021), where they show that the GT26 domain of TagA from *Thermoanaerobacter italicus* adopts a unique fold that is distinct from other glycosyltranferases (*e.g.*, GT-A, GT-B, GT-C, and GT-D) and is followed by a C-terminal tail that targets the enzyme to the cell membrane. Functional studies from the same group reveal that, in addition to stabilizing the enzyme through its association with the membrane, the C-terminal tail facilitates catalysis by encapsulating the UDP-ManNAc substrate in the enzyme active site (Kattke, Gosschalk *et al.,* 2019, Martinez, Mahoney *et al.,* 2021). However, the molecular basis of substrate recognition and catalysis remains

incompletely understood, and therefore further biochemical and structural studies are warranted to examine these enzymes as a class. This enzyme is the subject of the third research aim presented within this dissertation (see *Summary of Objectives* in Chapter 1.7, and a summary of expression constructs presented in Appendix B).

Current pyruvyltransferase structures — The only reported pyruvyltransferase structure to date is of unliganded Pvg1P from *Schizosaccharomyces pombe* (PDB: 5AXY), which has led to an *in silico* model (generated using MOE software in combination with ASEDock docking simulations) that attempts to describe Pvg1P substrate binding (Higuchi, Yoshinaga *et al.*, 2016). The Pvg1P structure forms a positively-charged cleft between its N- and C-terminal Rossman-like domains, which is predicted to facilitate binding of the negatively-charged pyruvate moiety (Higuchi, Yoshinaga *et al.*, 2016). Within this cleft, modeled ligands indicate the presence of three candidate catalytic residues, which are conserved in CsaB homologs from bacteria of the phylum Firmicutes. However, despite their conserved nature, the precise catalytic role of the residues remains uncertain and, with less than 30% sequence identity between CsaB and Pvg1P, deducing any such potential substrate binding modes for CsaB has been challenging. Further, given that Pvg1p and CsaB accommodate different sugar acceptor substrates (β-Gal versus β-ManNAc), the binding cleft geometry and the identity of key residues involved in substrate binding may differ. Therefore, to reveal overarching structural features of CsaB, including important secondary structure elements, active site architecture, as well as potential domain rearrangements that occur upon substrate binding, further structural studies of this enzyme are warranted. This enzyme is also the subject of the third research aim presented within this dissertation (see *Summary of Objectives* in Chapter 1.7, and a summary of expression constructs presented in Appendix B).

1.7. Summary of Objectives

S-layers play critical roles in cell maintenance and growth, and often function at the frontline for mediating host-pathogen interactions. Their abundance at the cell surface provides an accessible structure that makes S-layers amenable targets for diagnostic, vaccine, or therapeutic purposes. Understanding the molecular basis of S-layer cell-wall attachment is a requisite step toward exploiting this potential. Further, the biosynthesis of SCWPs warrants deeper examination as the suite of enzymes involved in this critical pathway are necessary to establish S-layer attachment. Targeting these enzymes, thereby limiting the production of SCWPs, could provide an alternate route for disruption of S-layer formation and the life cycle of these organisms. The research presented in this dissertation will address aspects of these topics within the context of the following research objectives:

(1) Using novel synthetic terminal di- and trisaccharide SCWP analogues, expand upon our current biophysical and structural understanding to elucidate more fully the cell-wall anchoring mechanisms of the *P. alvei* S-layer protein SpaA, and its mutant variants, $SpaA_{SLH}$/G109A and $SpaA_{SLH}$/G46A/G109A.

(2) Analyze the structure and sequence of the *P. alvei* MnaA UDP-GlcNAc 2-epimerase enzyme in order to characterize its putative catalytic and allosteric substrate-binding sites.

(3) The *P. alvei* CCM 2051^{T} enzymes CsaB and TagA were initially included as objectives for this research, but both ultimately proved intractable to structural studies. A summary of the experimental approaches used in an attempt to generate soluble protein and any associated crystallization studies is presented in Appendix B.

Chapter 2: Exploring the recognition of terminal disaccharide and terminal trisaccharide secondary cell wall polysaccharide fragments by *Paenibacillus alvei* SLH domains

Adapted with permission (see Appendix A) from:

Legg, M.L., Hager-Mair, F.F., Krauter, S., Gagnon, S.M.L., Lòpez-Guzmán, A., Lim, C., Blaukopf, M., Kosma, P., C. Schäffer and S. V. Evans (2022). "The S-layer homology domains of *Paenibacillus alvei* surface protein SpaA bind to cell wall polysaccharide through the terminal monosaccharide residue." Journal of Biological Chemistry **298**(4): 101745.

Contributions:

The terminal disaccharide and terminal trisaccharide SCWP analogues used in this study were synthesized by Simon Krauter, Charlie Lim, Markus Blaukopf and Paul Kosma from the Department of Chemistry, Institute of Organic Chemistry, at the Universität für Bodenkultur Wien in Vienna, Austria. The $SpaA_{SLH}$ wild-type and mutant protein constructs utilized in this study were produced by my collaborators Fiona F. Hager-Mair, Arturo Lòpez-Guzmán and Christina Schäffer from the Department of NanoBiotechnology, *NanoGlycobiology* Unit, at the Universität für Bodenkultur Wien in Vienna, Austria. Fiona F. Hager-Mair, Arturo Lòpez-Guzmán and Christina Schäffer also performed all ITC experimentation and analysis for this research. Susannah Gagnon assisted with single crystal X-ray diffraction data collection. All purifications, concentrations, crystallizations, single crystal X-ray diffraction data collections and analyses, and the development of all resulting structure-function correlations were performed by the author.

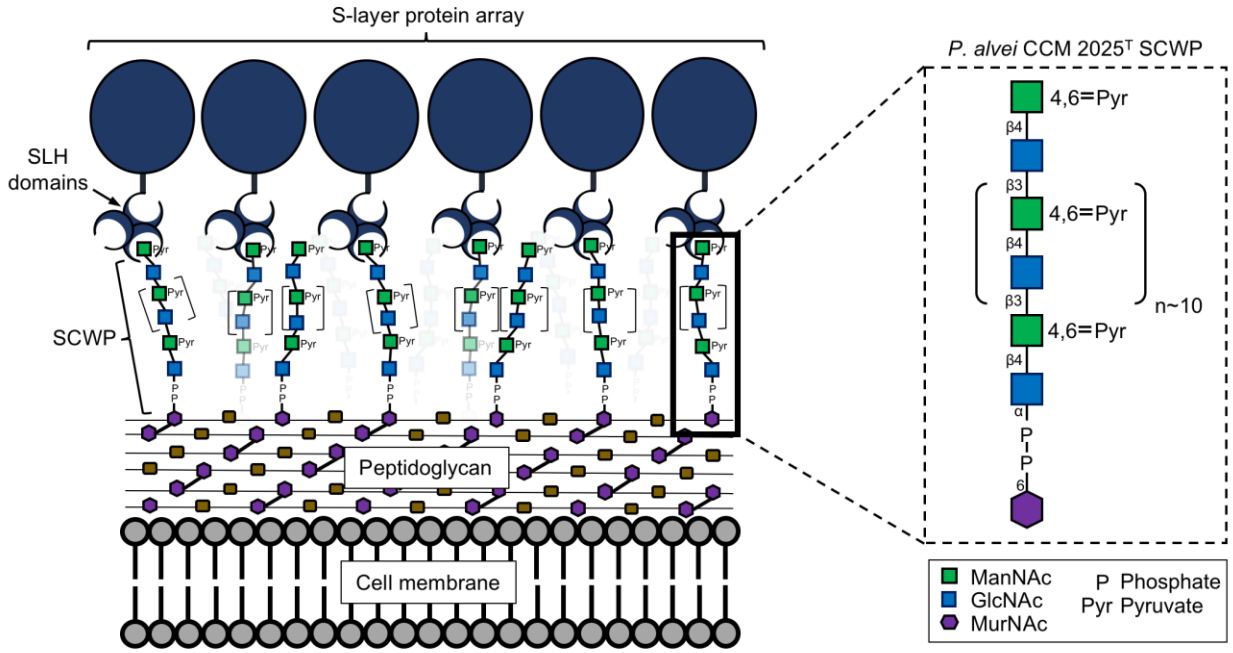
2.1. Introduction to Chapter 2

Briefly, as mentioned above, S-layers are present in archaea (Albers and Meyer 2011, Sleytr, Schuster *et al.,* 2014, Rodrigues-Oliveira, Belmok *et al.,* 2017) and Gram-negative bacteria (Lupas, Engelhardt *et al.,* 1994, Awram and Smit 2001, Fagan and Fairweather 2014), however only in Gram-positive bacteria are they bound to the cell membrane *via* evolutionarily-conserved S-layer homology (SLH) domains (Engelhardt and Peters 1998, Fagan and Fairweather 2014, Sleytr, Schuster *et al.,* 2014, Xu, Resch *et al.,* 2016, Usenik, Renko *et al.,* 2017), which occur in tandem as triplicate repeats (designated SLH1, SLH2, and SLH3) that are located immediately after the protein C-terminus or N-terminal signal peptide (Engelhardt and Peters 1998, Fagan and Fairweather 2014, Sleytr, Schuster *et al.,* 2014, Xu, Resch *et al.,* 2016, Usenik, Renko *et al.,* 2017). SLH domains bind non-covalently to anionic polysaccharides called non-classical secondary cell wall polymers (SCWPs) that are themselves covalently linked to the muramic acid residues of peptidoglycan (**Figure 9**) (Schäffer and Messner 2017).

The only two published structures of the SLH-domain trimers from the S-layer proteins SpaA (of *Paenibacillus alvei*) (Blackler, Lopez-Guzman *et al.,* 2018) and Sap (of *Bacillus anthracis*) (Kern, Wilton *et al.,* 2011, Sychantha, Chapman *et al.,* 2018) exhibit overall three-fold symmetry (**Figure 5**), where each repeat consists of ~55 residues arranged into a pair of α-helices linked by an extended loop structure (Kern, Wilton *et al.,* 2011, Blackler, Lopez-Guzman *et al.,* 2018, Sychantha, Chapman *et al.,* 2018). The C-terminal helix from each SLH repeat contributes to a three-helix core bundle, while the corresponding N-terminal helix is rotated ~90° from the three-fold axis to form three lobes and grooves (Kern, Wilton *et al.,* 2011, Blackler, Lopez-Guzman *et al.,* 2018) (**Figure 5**). For *P. alvei* SpaA, the N-terminus of each core SLH helix contributes a conserved 'TRAE' motif towards these binding grooves, where TRAE of SLH1,

TVEE of SLH2, and TRAQ of SLH3 are primarily situated within grooves 1 (G1), 2 (G2) and 3 (G3), respectively (Kern, Wilton *et al.*, 2011, Blackler, Lopez-Guzman *et al.*, 2018) (**Figure 5**). The side chains of the arginine residues from the TRAE (SLH1) and TRAQ (SLH3) motifs extend through the hydrophobic core to emerge at the base of the neighboring G2 and G1 grooves, respectively, where the positively charged guanidino group is positioned to complement the negative charge of the pyruvylated SCWP ligands.

Figure 9. *Schematic diagram of S-layer protein binding on Gram-positive bacteria*



The cell surface of many prokaryotes is coated in a surface layer (S-layer) consisting of one or more (glyco)proteins arranged in a repeating two-dimensional array. S-layer protein binding on Gram-positive bacteria is mediated by specialized S-layer homology (SLH) domains that recognize and bind peptidoglycan-linked secondary cell wall polymers (SCWPs), with the inset showing the structure of the SCWP from *P. alvei* CCM 2051^{T} (Schäffer, Müller *et al.*, 2000, Schäffer and Messner 2017, Blackler, Lopez-Guzman *et al.*, 2018)

The SCWP polymer of *P. alvei* CCM 2051^{T} is composed of eleven →3)-4,6-Pyr-β-D-Man*p*NAc-(1→4)-β-D-Glc*p*NAc-(1→ repeats (Schäffer, Müller *et al.,* 2000) (**Figure 6**). The binding of a synthetic, terminal 4,6-Pyr-β-D-ManNAcOMe SCWP monosaccharide analogue to an SLH domain has been shown to induce a remarkable structural rearrangement involving an ~180° phi angle backbone flip of a conserved glycine residue (consensus sequence SLH-Gly29) fundamental to the receptor site's ability to bind ligand, including accommodation of the negative charge of the terminal pyruvate moiety (Blackler, Lopez-Guzman *et al.*, 2018). Together with enzymatic and biochemical analyses (Mesnage, Fontaine *et al.,* 2000, Schaffer, Muller *et al.,* 2000, Cava, de Pedro *et al.,* 2004, Zarschler, Janesch *et al.,* 2010), these structural findings confirm that, for SpaA, Sap, and other Gram-positive S-layer proteins, SLH-mediated SCWP-anchoring is reliant on a negatively charged ketal-pyruvate moiety located on the terminal sugar residue of the SCWP (Blackler, Lopez-Guzman *et al.,* 2018, Sychantha, Chapman *et al.,* 2018).

P. alvei $SpaA_{SLH}$ monosaccharide-bound structures allowed the identification of separate SLH-domain binding sites in G2 and G1 (Blackler, Lopez-Guzman *et al.,* 2018), where the wild-type $SpaA_{SLH}$ protein co-crystallized with ligand bound in G2 and inactivation of binding in G2 resulted in a conformational shift that forced the mutant construct to instead utilize G1 for binding, resulting in a concomitant ten-fold decrease of the associated ITC monosaccharide binding affinity (Blackler, Lopez-Guzman *et al.,* 2018). Together, these observations show that monosaccharide SCWP binding is likely centered about G2, with G1 serving as an alternate site for SCWP binding (Blackler, Lopez-Guzman *et al.,* 2018).

Significantly, the ability of $SpaA_{SLH}$ to accommodate monosaccharide binding in G2 or G1 suggested a biological mechanism for the relief of S-layer strain encountered during processes of cellular growth and division (Blackler, Lopez-Guzman *et al.,* 2018), where the transfer of SCWP

interactions between SLH-domain binding sites could facilitate S-layer rearrangement and expansion without disrupting the overall confluence of the S-layer protein lattice. To date, however, the structural analysis of this novel binding site switching mechanism has been limited to the terminal monosaccharide 4,6-Pyr-β-D-ManNAcOMe and the internal repeat disaccharide β-D-GlcNAc-(1→3)-4,6-Pyr-β-D-ManNAcOMe SCWP ligand analogues (Blackler, Lopez-Guzman *et al.,* 2018), leaving open the role of the subsequent SCWP sugar residues in binding, as well as whether the primary G2 and secondary G1 sites can accommodate longer ligands that better approximate SCWP at the surface of *P. alvei*. To address this, here, we characterize the details of the $SpaA_{SLH}$-SCWP interaction using thermodynamic ITC analysis of disaccharide and trisaccharide binding and through the determination of the co-crystal structures of $SpaA_{SLH}$, and single ($SpaA_{SLH}$/G109A) and the corresponding double ($SpaA_{SLH}$/G46A/G109A) mutants in complex with synthetic terminal disaccharide and trisaccharide analogues of the *P. alvei* CCM 2051^{T} SCWP ligand.

2.2. Results

2.2.1. Binding parameters of SpaA$_{SLH}$ with disaccharide and trisaccharide SCWP ligands

Isothermal titration calorimetry (ITC) was used to determine the binding affinities of SpaA$_{SLH}$, SpaA$_{SLH}$/G109A and SpaA$_{SLH}$G46A/G109A toward the terminal 4,6-Pyr-β-D-ManNAc-(1→4)-β-D-GlcNAcOMe disaccharide (**Figure 10**) and the terminal 4,6-Pyr-β-D-ManNAc-(1→4)-β-D-GlcNAc-(1→3)-4,6-Pyr-β-D-ManNAcOMe trisaccharide (**Figure 10**) analogues in solution (**Table 1**). The SpaA$_{SLH}$ construct binds disaccharide with 1:1 stoichiometry and an apparent dissociation constant (K_D) of 64 nM at 20°C, while the SpaA$_{SLH}$/G109A construct binds disaccharide with a 1:1 stoichiometry and an apparent K_D of 288 nM at 20°C (**Table 1**). For each construct, the disaccharide binding affinity is slightly lower than the corresponding monosaccharide binding affinity measured for SpaA$_{SLH}$ (29 nM) and SpaA$_{SLH}$/G109A (226 nM), respectively (**Table 1**). No measurable binding was detected by ITC between the SpaA$_{SLH}$G46A/G109A double mutant and the disaccharide analogue, which corroborates the corresponding lack of measurable binding for monosaccharide with this same protein construct (**Table 1**).

Wild-type SpaA$_{SLH}$ binds trisaccharide analogue with 1:1 stoichiometry and a K_D (4.6 nM at 20°C) that is slightly lower than the K_D measured for monosaccharide (29 nM (Blackler, Lopez-Guzman *et al.*, 2018)), indicating that SpaA$_{SLH}$ has a slightly higher binding affinity towards the trisaccharide ligand than the monosaccharide ligand. The SpaA$_{SLH}$/G109A construct also binds trisaccharide analogue with 1:1 stoichiometry and a K_D of 260 nM at 20°C (**Table 1**).

Despite similar Gibbs free energies (ΔG) for mono-, di-, and trisaccharide binding, the entropic contributions for both trisaccharide and disaccharide binding are comparably stronger than for monosaccharide (**Table 1**), resulting in overall lower enthalpic contributions for these

longer ligands. These binding data support favorable hydrogen bond formation and hydrophobic interaction with di- and trisaccharide analogues, whereas the binding data for monosaccharide (Blackler, Lopez-Guzman *et al.*, 2018) supports good hydrogen bonding but with unfavorable conformational change. The $SpaA_{SLH}$G46A/G109A double mutant does not show detectable binding by ITC to the trisaccharide analogue in solution (**Table 1**).

Figure 10. *Synthetic terminal di- and trisaccharide ligands used in this study*

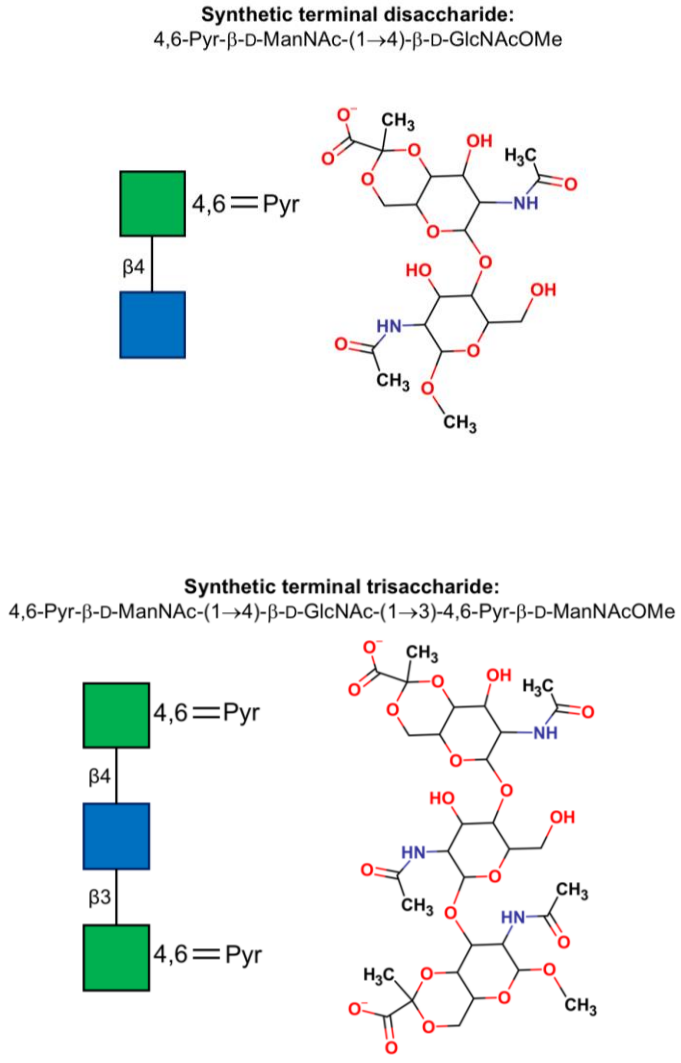


Table 1. *Binding affinities of $SpaA_{SLH}$ toward synthetic terminal SCWP oligosaccharide ligands.*

ITC analysis of $SpaA_{SLH}$ binding of the SCWP terminal disaccharide analogue (4,6-Pyr-β-D-ManNAc-(1→4)-β-D-GlcNAcOMe) and the SCWP terminal trisaccharide analogue (4,6-Pyr-β-D-ManNAc-(1→4)-β-D-GlcNAc-(1→3)-4,6-Pyr-β-D-ManNAcOMe). The published ITC data for binding to the SCWP terminal monosaccharide ligand (4,6-Pyr-β-D-ManNAcOMe) (Blackler, Lopez-Guzman et al., 2018) is included for comparison. Data collection for di- and trisaccharide ligands was carried out by Fiona F. Hager-Mair, Arturo Lòpez-Guzmán and Christina Schäffer from the Department of NanoBiotechnology, NanoGlycobiology Unit, at the Universität für Bodenkultur Wien in Vienna, Austria.

$SpaA_{SLH}$ construct	**Temperature (°C)**	**$-T\Delta S$ (kJ/mol)**	**ΔH (kJ/mol)**	**ΔG (kJ/mol)**	**Stoichiometry**	**K_A (M^{-1})**	**K_D (nM)**
Trisaccharide-binding data							
WT	20	-18.70± 5.57	-27.14 ± 3.29	-45.84 ± 2.29	1.16 ± 0.10	$2.14 \times 10^8 \pm 1.46$	4.6
G109A	20	-15.99 ± 1.00	-20.92 ± 0.57.	-36.91 ± 0.47	1.44 ± 0.25	$3.80 \times 10^6 \pm 0.68$	260
G46A/G109A	20	No binding					
Disaccharide-binding data							
WT	20	-9.95 ± 2.08	-30.36 ± 1.64	-40.31 ± 0.65	1.20 ± 0.16	$1.56 \times 10^7 \pm 0.37$	64
G109A	20	-15.74 ± 1.35	-20.95 ± 1.51	-36.70 ± 0.20	1.30 ± 0.19	$3.47 \times 10^6 \pm 0.29$	288
G46A/G109A	20	No binding					
Previously published (Blackler, Lopez-Guzman *et al.*, 2018) monosaccharide-binding data							
WT	20	45.74 ± 16.01	−87.85 ± 15.62	−42.10 ± 0.65	0.91 ± 0.04	$3.48 \times 10^{7 \pm} 0.36$	29
G109A	20	13.33 ± 7.0	−50.64 ± 7.29	−37.31 ± 0.33	0.92 ± 0.04	$4.48 \times 10^6 \pm 0.62$	226
G46A/G109A	20	No binding					

2.2.2. Data collection & structure determination

Data collection and refinement statistics for the ligand-bound structures of $SpaA_{SLH}$, $SpaA_{SLH}$/G109A, and $SpaA_{SLH}$G46A/G109A are shown in **Table 2**. All X-ray diffraction data were collected on a Dectris Pilatus 200K detector coupled to a Micromax-007 HF X-ray generator (Rigaku, USA) from crystals that were flash frozen at 100 °K using an Oxford Cryosystems Cryostream Unit (800 series). Each structure was solved *via* molecular replacement using PHASER of the CCP4 program suite (Winn, Ballard *et al.,* 2011) and either unliganded $SpaA_{SLH}$ (PDB: 6CWM) as the search model for the $SpaA_{SLH}$ and $SpaA_{SLH}$/G46A/G109A structures or monosaccharide-bound $SpaA_{SLH}$/G109A (PDB: 6CWN, with the monosaccharide ligand removed) used as a search model for the $SpaA_{SLH}$/G109A structure. Structures were finalized *via* refinement and iterative model improvement using REFMAC5 of the CCP4 program suite (Winn, Ballard *et al.,* 2011) in conjunction with Coot macromolecular model building software (Emsley, Lohkamp *et al.,* 2010).

The $SpaA_{SLH}$ crystal structure in complex with the terminal disaccharide SCWP analogue was determined to 1.70 Å resolution in space group $P2_12_12_1$ with 1 molecule in the asymmetric unit (AU). The final model yielded an average B-factor of 17.1 $Å^2$, with R_{work} and R_{free} values of 18.4 % and 21.3 %, respectively. The structure displayed excellent electron density for nearly all modeled protein residues apart from the first residue (Ala28) and last four modeled residues of the construct (residues 194-197) for which only partial main chain and side chain electron density was observed. The disaccharide analogue was present in G2, where it displayed excellent electron density with an average B-factor of 18.0 $Å^2$ (**Figure 11A**).

The crystal structure of $SpaA_{SLH}$ in complex with the terminal trisaccharide SCWP analogue was solved to a resolution of 2.06 Å in space group $P2_12_12_1$ with 1 molecule in the AU.

The final model yielded R_{work} and R_{free} values of 20.5 % and 24.0 %, respectively, with an overall B-factor of 26.2 $Å^2$. Almost all amino acid residues exhibited excellent electron density aside from the first residue (Ala28) and last two modeled residues of the construct (residues 196-197), which only displayed partial side-chain electron density. The trisaccharide ligand was present in G2, where it exhibited excellent electron density with an average B-factor of 34.6 $Å^2$ (**Figure 11B**).

The crystal structure of the $SpaA_{SLH}$/G109A single mutant construct in complex with the terminal disaccharide SCWP analogue was solved to 1.72 Å resolution in space group C2 with 2 molecules in the AU (labeled A and B). The final model yielded R_{work} and R_{free} values of 17.8 % and 19.5 %, respectively, with an overall B-factor of 22.2 $Å^2$. All modeled residues showed excellent electron density apart from the first residue in both of molecules A and B (Ala28), the last modeled residue of molecule A (residue 192), and last three modeled residues of molecule B (residues 189-191), for which only partial side-chain electron density is observed. The disaccharide analogue was present in G1 for both molecules A and B, where it displayed unambiguous electron density for the entire ligand and yielded an average B-factor value of 19.5 $Å^2$ (20.0 $Å^2$ for molecule A and 19.0 $Å^2$ for molecule B) (**Figure 11C**).

The crystal structure of the $SpaA_{SLH}$/G46A/G109A double mutant construct in complex with the terminal disaccharide SCWP analogue was determined to 1.85 Å resolution in space group $P2_12_12_1$ with 1 molecule in the AU. The model yielded an average B-factor value of 23.5 $Å^2$, with final R_{work} and R_{free} values of 18.6 % and 21.7 %, respectively. The structure displayed excellent electron density for nearly all modeled protein residues apart from the first residue (Ala28) and last two modeled residues of the construct (residues 191-192) for which only partial side-chain

electron density was observed. The modeled disaccharide analogue yielded an average B-factor of 37.1 $Å^2$ and displayed unambiguous electron density for the entire ligand in G2 (**Figure 11D**).

Table 2. *Data collection and refinement statistics for di- and trisaccharide-bound $SpaA_{SLH}$ constructs.*

Statistics are shown di- and trisaccharide-bound wild-type $SpaA_{SLH}$, as well as for disaccharide-bound single mutant ($SpaA_{SLH}$ /G109A) and disaccharide-bound double mutant ($SpaA_{SLH}$ /G46A/109A).

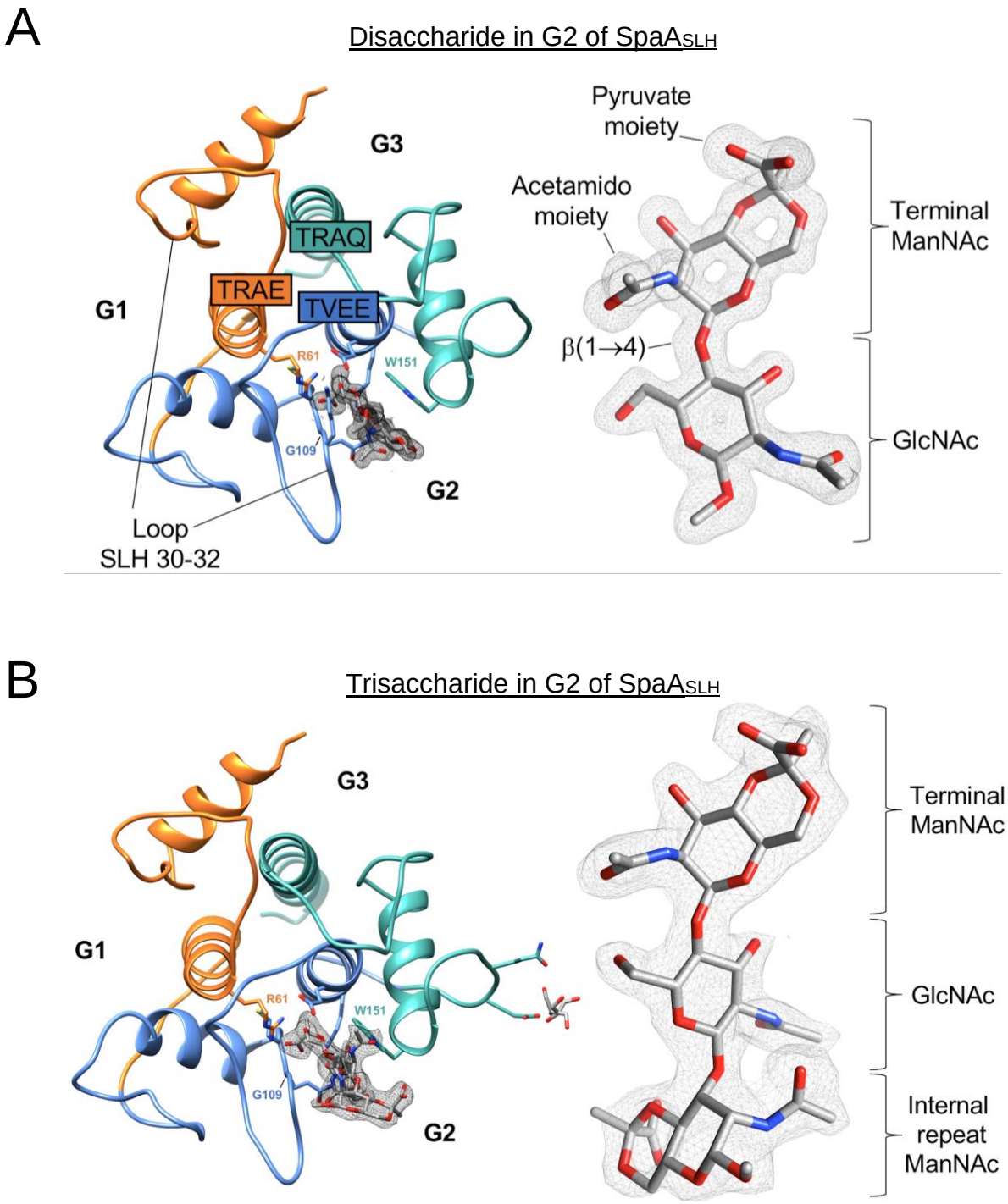
Protein	**$SpaA_{SLH}$**	**$SpaA_{SLH}$**	**$SpaA_{SLH}$/G109A**	**$SpaA_{SLH}$/G46A/G109A**
Ligand	Disaccharide[a]	Trisaccharide[b]	Disaccharide[a]	Disaccharide[a]
PDB code	7SV3	7SV4	7SV5	7SV6
Data collection				
Space group	$P2_12_12_1$	$P2_12_12_1$	C2	$P2_12_12_1$
Resolution (Å)	20.00-1.70 (1.76-1.70)	20.00-2.06 (2.13-2.06)	20.00-1.72 (1.78-1.72)	20.00-1.85 (1.92-1.85)
Cell dimensions				
a (Å)	34.65	34.69	111.85	33.30
b (Å)	67.92	67.30	45.32	65.61
c (Å)	72.91	72.75	87.49	72.29
α (°)	90	90	90	90
β (°)	90	90	129.47	90
γ (°)	90	90	90	90
Z	4	4	8	4
R_{sym}	0.145 (0.773)	0.143 (0.591)	0.071 (0.539)	0.099 (0.540)
R_{pim}	0.048 (0.369)	0.043 (0.295)	0.029 (0.301)	0.053 (0.345)
$CC_{1/2}$	(0.774)	(0.137)	(0.699)	(0.795)
I/σ(I)	8.4(2.0)	7.6 (2.3)	11.1 (1.9)	9.5 (2.1)
Completeness (%)	100.0 (99.9)	98.4 (90.1)	90.3 (82.6)	98.2 (99.3)
Redundancy	7.9 (5.2)	10.2 (3.3)	6.7 (3.6)	4.3 (3.2)
Unique reflections	19791	11130	32683	13823
Refinement				
Resolution (Å)	20.00-1.70	20.00-2.06	20.00-1.72	20.00-1.85
No. reflections	18483	10409	30968	13083
R_{work} (%)	18.4	20.5	17.8	18.6
R_{free} (%)	21.3	24.0	19.5	21.7
No. atoms				
Protein	1311	1290	2520	1281
Ligand(s)	35[a]	66[b,c]	82[a,c]	35[a]
Water	242	137	303	112
B factors (Å^2)				
Protein	15.1	25.0	21.2	22.4
Ligand	18.0	34.6	19.5	37.1
Water	27.5	32.6	31.6	31.6
Average	17.1	26.2	22.2	23.5
Ramachandran				
Favoured (%)	98.2	98.8	98.5	98.8
Allowed (%)	1.8	1.2	1.5	1.2
Disallowed (%)	0	0	0	0
r.m.s. bonds (Å)	0.0039	0.0031	0.0034	0.0040
r.m.s. angles (°)	1.25	1.23	1.21	1.23

[a] Disaccharide = 4,6-Pyr-β-D-ManNAc-(1→4)-β-D-GlcNAcOMe

[b] Trisaccharide = 4,6-Pyr-β-D-ManNAc-(1→4)-β-D-GlcNAc-(1→3)-4,6-Pyr-β-D-ManNAcOMe

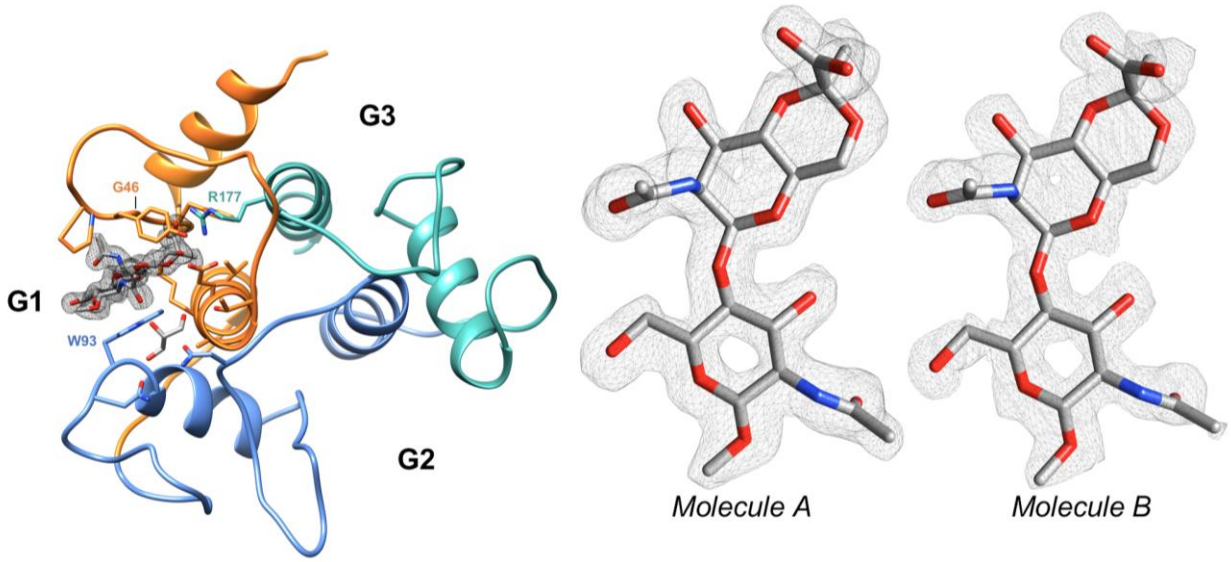
[c] Glycerol = $C_3H_8O_3$

Figure 11. *Ribbon model for the protein structure, including ligand electron density, for each $SpaA_{SLH}$ co-crystal structure in* ***Table 2***



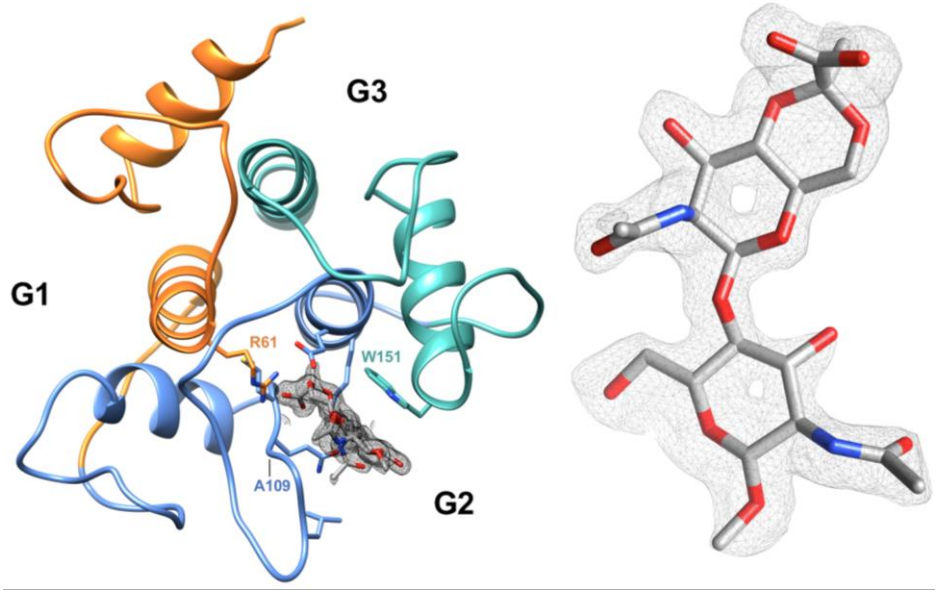
C

Disaccharide in G1 of $SpaA_{SLH}$/G109A



D

Disaccharide in G2 of $SpaA_{SLH}$/G46A/G109A



Models for (**A**) disaccharide-bound $SpaA_{SLH}$; (**B**) trisaccharide-bound $SpaA_{SLH}$; (**C**) disaccharide-bound $SpaA_{SLH}$/G109A with ligands shown from both molecules of the asymmetric unit of the co-crystal structure; and (**D**) disaccharide-bound $SpaA_{SLH}$/G46A/G109A double mutant. Grooves 1, 2, and 3 are indicated on each structure as G1, G2, and G3, respectively. SLH1 orange, SLH2 blue, and SLH3 aquamarine. Ligand atoms are colored by element with oxygen red, nitrogen blue, and carbon white. $2F_o$-F_c electron density maps for ligands are depicted as grey mesh contoured to 1σ.

2.2.3. $SpaA_{SLH}$ co-crystallization complexes with terminal di- and trisaccharide SCWP fragments

The synthesized *P. alvei* SCWP terminal disaccharide analogue (Krauter, Schäffer *et al.,* 2021) (**Figure 10**) was co-crystallized with wild-type $SpaA_{SLH}$ in order to structurally characterize $SpaA_{SLH}$ binding to a biologically relevant SCWP ligand target longer than monosaccharide.

The terminal ManNAc residue of disaccharide binds to the protein in the same manner as that of the published monosaccharide-bound $SpaA_{SLH}$ structure (Blackler, Lopez-Guzman *et al.,* 2018), where ManNAc lies in a deep and narrow positively-charged pocket comprised of residues that belong to all three SLH domains (**Figure 12A**). In this pocket, the ManNAc residue pyruvate moiety participates in a salt-bridge interaction with SLH1 residue Arg61 (corresponding to SLH-Arg43) of the TRAE motif, and forms a hydrogen bond interaction with the backbone amide of SLH2 residue Gly109 (corresponding to SLH-Gly29). The ManNAc ring also forms a stacking interaction with the side chain of SLH3 residue Trp151 (corresponding to SLH-Trp13) (**Figure 12A**).

In this structure, the β(1→4) linked GlcNAc extends outward beyond the binding pocket, as shown in **Figure 12A**, and is not observed to participate in hydrogen bond interactions with $SpaA_{SLH}$ surface residues. The GlcNAc residue does, however, maintain contacts with the protein surface through an ordered network of bridging water molecules to $SpaA_{SLH}$ surface residues Lys110, Glu127 and Trp151 (**Figure 12A**).

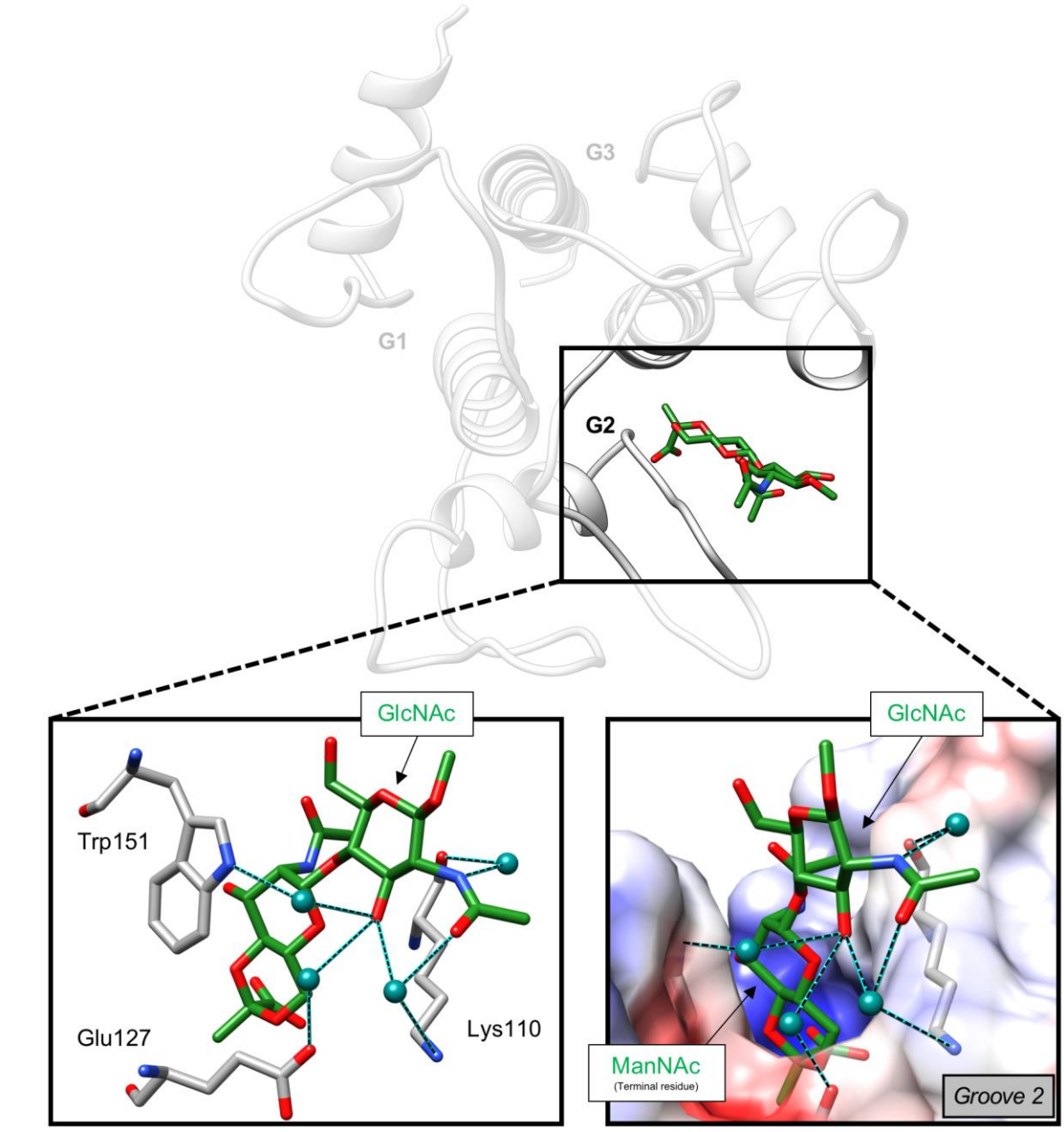
To investigate binding in G2 further, the synthesized *P. alvei* SCWP terminal trisaccharide analogue (**Figure 10**) was also co-crystallized with wild-type $SpaA_{SLH}$. The trisaccharide ligand is bound within G2 (**Figure 11B**) with the terminal ManNAc residue bound to the protein in the same manner as observed for disaccharide (above) and monosaccharide ligand-bound structures (Blackler, Lopez-Guzman *et al.,* 2018) (**Figure 12B**). As with disaccharide, the β(1→4) linked

GlcNAc extends beyond the border of the binding from G2. The extended GlcNAc residue of trisaccharide, however, adopts a rotated conformation where the acetamido moiety is displaced by 4.5 Å towards the opposite face of the binding pocket relative to disaccharide, and forms bridging water contacts with Lys110, Glu127 and Thr149 (**Figure 12B**). As part of this rotation, the trisaccharide GlcNAc O3 hydroxyl group is shifted by 2.1 Å, with respect to disaccharide, and forms a single hydrogen bond to the protein surface *via* the sidechain of Trp151 (**Figure 12B**). The trisaccharide internal-repeat ManNAc residue (the residue located furthest away from G2) forms a 3.2 Å hydrogen bond to the GlcNAc acetamido moiety as well as forming bridging water interactions to surface residues Lys110 and Asp150 (**Figure 12B**).

2.2.4. Mutant SpaA$_{SLH}$ co-crystallization complexes with terminal disaccharide SCWP ligand

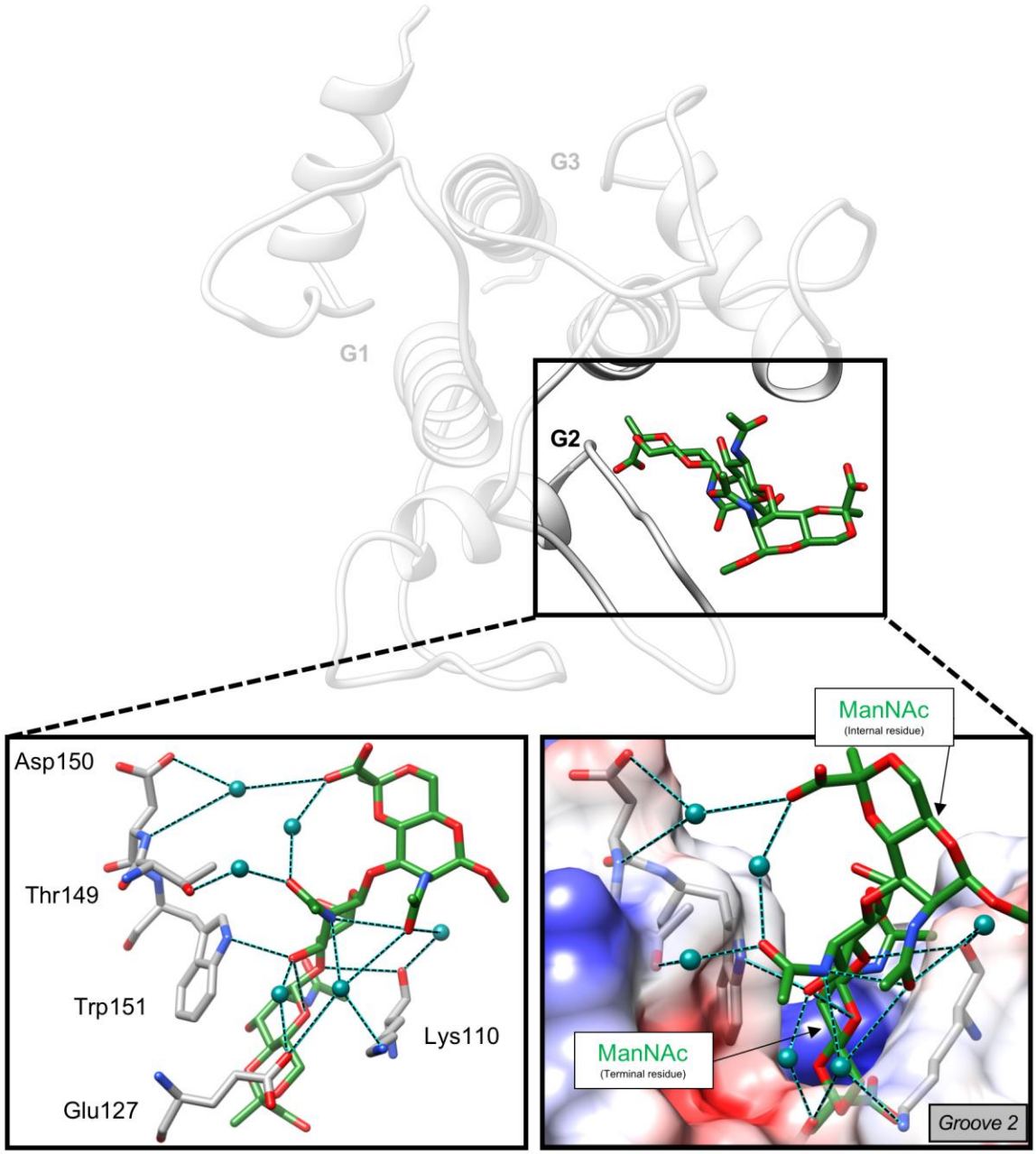
The SpaA$_{SLH}$/G109A single mutant construct, which was originally designed to destabilize ligand binding in G2 (Blackler, Lopez-Guzman *et al.*, 2018), was co-crystallized with the SCWP terminal disaccharide analogue (**Figure 11C**). This time the ligand is observed to bind in G1, where terminal ManNAc ring of the disaccharide is bound in a manner similar to that of the published monosaccharide-bound SpaA$_{SLH}$/G109A structure (Blackler, Lopez-Guzman *et al.*, 2018). The ManNAc residue sits in G1 in deep pocket formed by conserved residues Arg177 (corresponding to conserved residue SLH-Arg43 of SLH3), Trp93 (corresponding to the conserved residue SLH-Trp13 of SLH2), and Gly46 (corresponding to conserved residue SLH-Gly29 of SLH1). In contrast to binding in G2, the disaccharide's β(1→4) linkage does not extend completely past the border of the G1 pocket, and the lower portion of the GlcNAc sugar residue is partially obscured by SLH2 residues Pro48 and Trp93 (**Figure 12C**), to which the ligand forms hydrophobic contacts.

Figure 12. *$SpaA_{SLH}$ binding of terminal SCWP oligosaccharide targets*



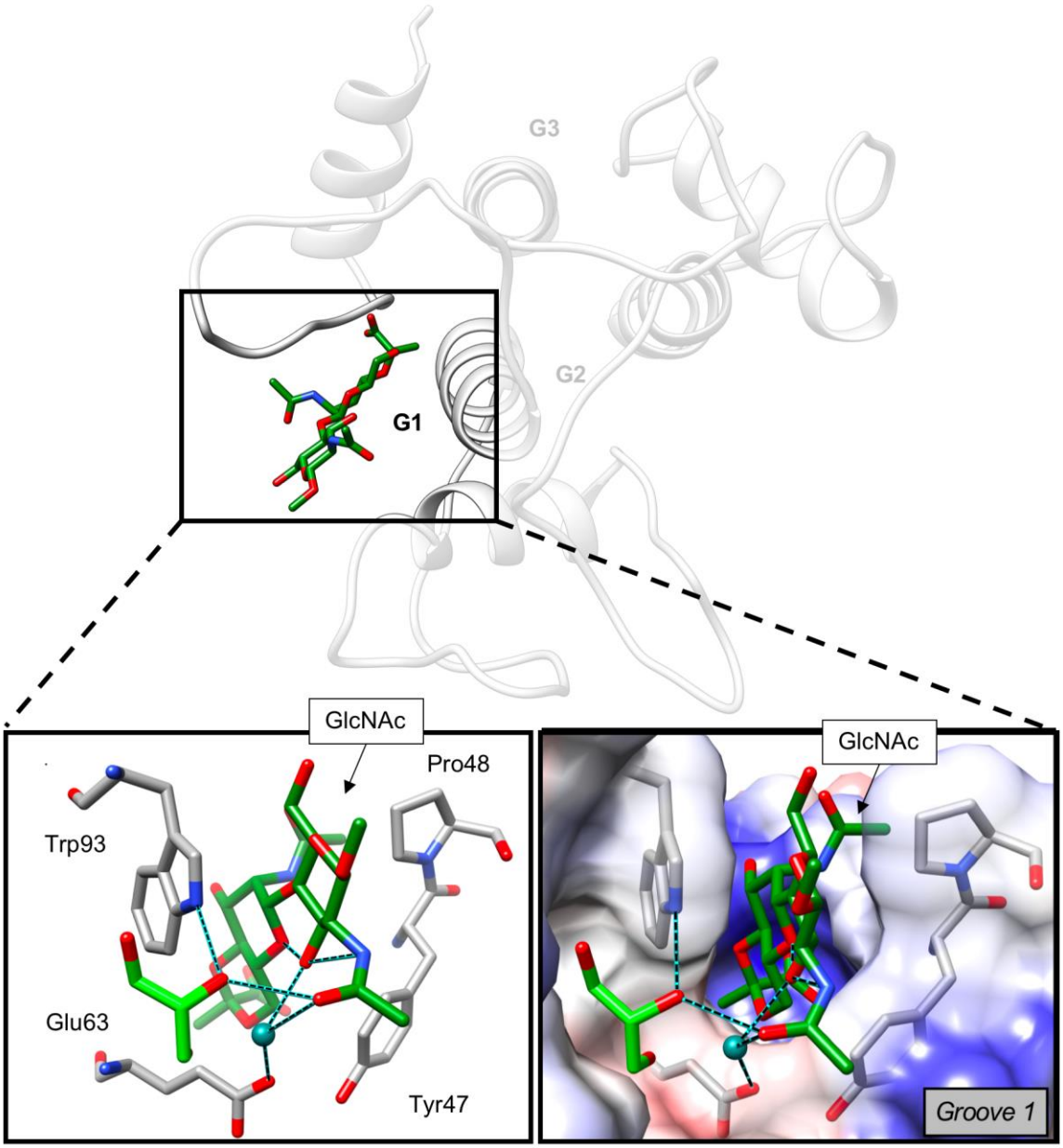
B

Trisaccharide bound in G2 of $SpaA_{SLH}$



C

Disaccharide bound in G1 of $SpaA_{SLH}$/G109A



Only the terminal monosaccharide of the di- and trisaccharide ligands forms significant direct contacts with the $SpaA_{SLH}$ protein. (**A**) The synthetic terminal disaccharide analogue binds to G2 of $SpaA_{SLH}$, with the GlcNAc residue forming no direct contact to the protein. (**B**) The synthetic terminal trisaccharide analogue binds to G2 of $SpaA_{SLH}$, with the GlcNAc residue forming a single hydrogen bond interaction with the protein. (**C**) The synthetic terminal disaccharide analogue binds to G1 of the single mutant with the terminal ManNAc again forming the only direct interactions with the protein. Water molecules are depicted as light green spheres, while atoms are colored by element with oxygen red, nitrogen blue, protein carbon white, and ligand carbon green. Charges are shown in surface representations as red (negative), blue (positive), and white (neutral).

As in G2, the GlcNAc residue does not form any direct hydrogen bonds to the protein surface, however the GlcNAc O3 hydroxyl does form a water-bridge with Glu63, while also coordinating the ManNAc O5 hydroxyl and GlcNAc *N*-acetyl nitrogen (**Figure 12C**). In addition, the acetamido substituent of the GlcNAc residue is observed to form a hydrogen bond with a nearby surface-associated glycerol molecule located close to the mouth of the G1 pocket (**Figure 12C**) in the crystal lattice.

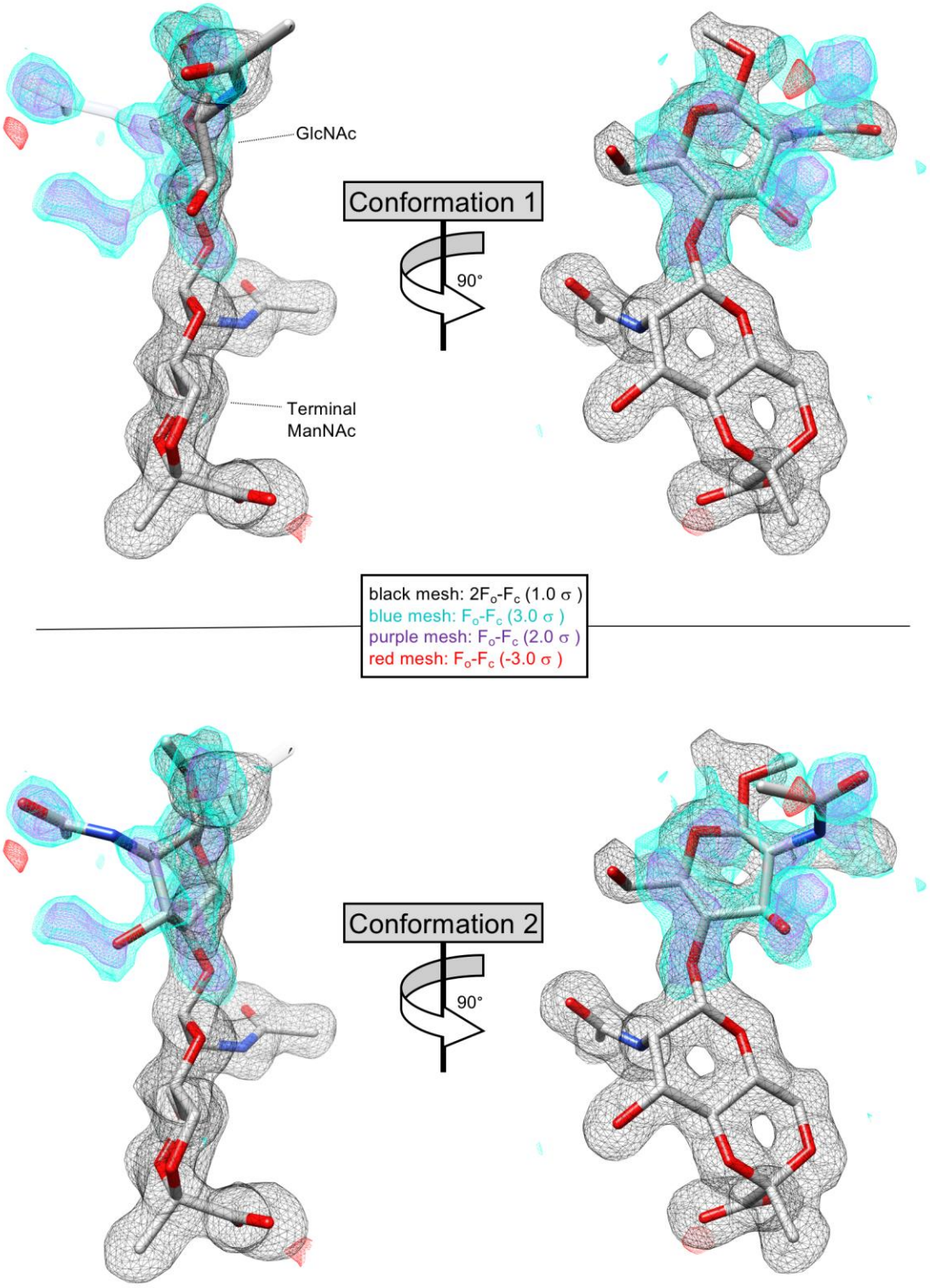
Finally, the SCWP terminal disaccharide analogue was also co-crystallized with the $SpaA_{SLH}$/G46A/G109A double mutant construct, which was originally designed to destabilize ligand binding in both G2 and G1 (Blackler, Lopez-Guzman *et al.*, 2018). Similar to the monosaccharide structure, the structure displays unambiguous electron density for the entire disaccharide ligand bound in G2, and the overall protein structure exhibits a high degree of structural homology to the disaccharide-bound $SpaA_{SLH}$ structure (reported above), with an overlap Cα rmsd of 0.44 Å.

2.2.5. Distinct GlcNAc sugar conformations observed in G2 of disaccharide-bound $SpaA_{SLH}$

In the co-crystal structure of disaccharide-bound $SpaA_{SLH}$, unexplained electron density was observed near the GlcNAc sugar moiety of the ligand bound in G2. This density can be explained by modeling the GlcNAc sugar residue of the disaccharide over two distinct conformations (**Figure 13**), with the second conformation (conformation 2) positioning the GlcNAc residue in the observed difference density. As determined *via* examination of temperature factors and electron density, the occupancy of the primary GlcNAc conformation (conformation 1) was reduced to 50% in order to yield temperature factors comparable with those of the surrounding protein and ligand atoms, which range between 10.00 and 19.00.

The GlcNAc acetamido and O3 hydroxyl groups of conformation 1 are located nearest the region of the binding pocket where residue Lys110 is situated (**Figure 12A**). In conformation 2, the acetamido and O3 hydroxyl groups are displaced by 4.5 Å and 2.1 Å, respectively, relative to conformation 1 and are positioned on the opposite side of the binding pocket where residues Trp151 and Thr149 are situated, adopting a conformation matching that of the terminal ManNAc and GlcNAc residues of the trisaccharide ligand bound in G2 of $SpaA_{SLH}$ (**Figure 12B**).

Figure 13. *Disaccharide conformations in G2 of SpaA$_{SLH}$*



The GlcNAc residue of disaccharide ligand bound to G2 of wild type SpaA$_{SLH}$ is disordered over two conformations, with the second conformation corresponding to the GlcNAc orientation observed in trisaccharide-bound SpaASLH structure. Atoms are colored by element with oxygen red, nitrogen blue, and carbon white.

2.3. Discussion

2.3.1. $SpaA_{SLH}$ binds oligosaccharide SCWP targets using both G2 & G1 sites

The wild-type and mutant SLH trimers from the *P. alvei* S-layer protein SpaA ($SpaA_{SLH}$) were previously co-crystallized in the presence of SCWP synthetic terminal monosaccharide analogue (Blackler, Lopez-Guzman *et al.,* 2018). The resultant structures showed that monosaccharide binding occurs primarily in G2 of $SpaA_{SLH}$ while G1 accommodates ligand when G2 is deactivated through mutation. These monosaccharide binding site switches demonstrate the remarkable flexibility of the SLH trimers when alternating between G2- and G1-bound states and suggest a potential mechanism for the relief of S-layer strain during cell growth and division. However, the structural characterization of the terminal monosaccharide has left an open question: what is the role of subsequent SCWP sugar residues in binding? To explore this, we report the crystal structures and binding characteristics of wild-type and mutant $SpaA_{SLH}$ variants determined in the presence of SCWP terminal di- and trisaccharide analogues (**Figure 10**).

SCWP terminal di- and trisaccharide analogues were co-crystallized with the wild-type $SpaA_{SLH}$ protein. Each structure shows the SCWP analogue bound in the G2 site with excellent electron density for the entire ligand (**Figure 11A & 11B**), once again showing that $SpaA_{SLH}$ primarily binds SCWP terminal oligosaccharide targets through interactions in G2, as was the case with terminal monosaccharide (Blackler, Lopez-Guzman *et al.,* 2018). The disaccharide ligand showed a slight decrease in binding affinity relative to monosaccharide while the trisaccharide ligand showed a slight increase in binding affinity to $SpaA_{SLH}$ (**Table 1**), but relative to the monosaccharide, both ligands had binding affinities with the same order of magnitude (nanomolar) (Blackler, Lopez-Guzman *et al.,* 2018).

The terminal SCWP disaccharide analogue co-crystallized in G1 of the $SpaA_{SLH}$/G109A protein with unambiguous electron density for the entire ligand (**Figure 11C**), indicating that the Gly109Ala mutation was sufficient to destabilize G2 binding in the crystal structure. As observed with the monosaccharide-bound structures (Blackler, Lopez-Guzman *et al.*, 2018), ligand binding in G1 forces a conformational rearrangement of the SLH trimers that enables the alternate (G1) site to accommodate oligosaccharide SCWP targets, highlighting the flexibility of this adaptable S-layer protein binding domain. Last, the terminal SCWP disaccharide analogue was co-crystallized with the $SpaA_{SLH}$/G46A/G109A double mutant. Despite the presence of inactivating mutations in both G2 (G109A mutation) and G1 (G46A mutation) binding sites, ligand is observed in G2 with unambiguous electron density (**Figure 11D**). However, neither the disaccharide nor the trisaccharide ligand showed any detectable binding affinity toward $SpaA_{SLH}$/G46A/G109A (Table 1). This suggests G2 has sufficient ligand complementarity to accommodate the terminal disaccharide fragment, even without the wild-type residue identity (mutant residue Ala109) at position 109.

2.3.2. G2 of $SpaA_{SLH}$ accommodates flexibility in oligosaccharide SCWP ligands

In both the di- and trisaccharide-bound wild-type $SpaA_{SLH}$ co-crystal structures, the terminal monosaccharide residue makes contacts to G2 while the remaining oligosaccharide residues extend into solvent to form bridging contacts to the protein surface through an ordered network of water molecules (**Figure 12A & 12B**). The disaccharide is observed in two distinct conformations (conformations 1 and 2, **Figure 13**), where the GlcNAc sugar residue is modeled in alternate pockets of corresponding electron density. The disaccharide terminal ManNAc residue, however, is observed in G2 in only one orientation, which is similar to that of previously characterized terminal monosaccharide-bound structures (Blackler, Lopez-Guzman *et al.*, 2018).

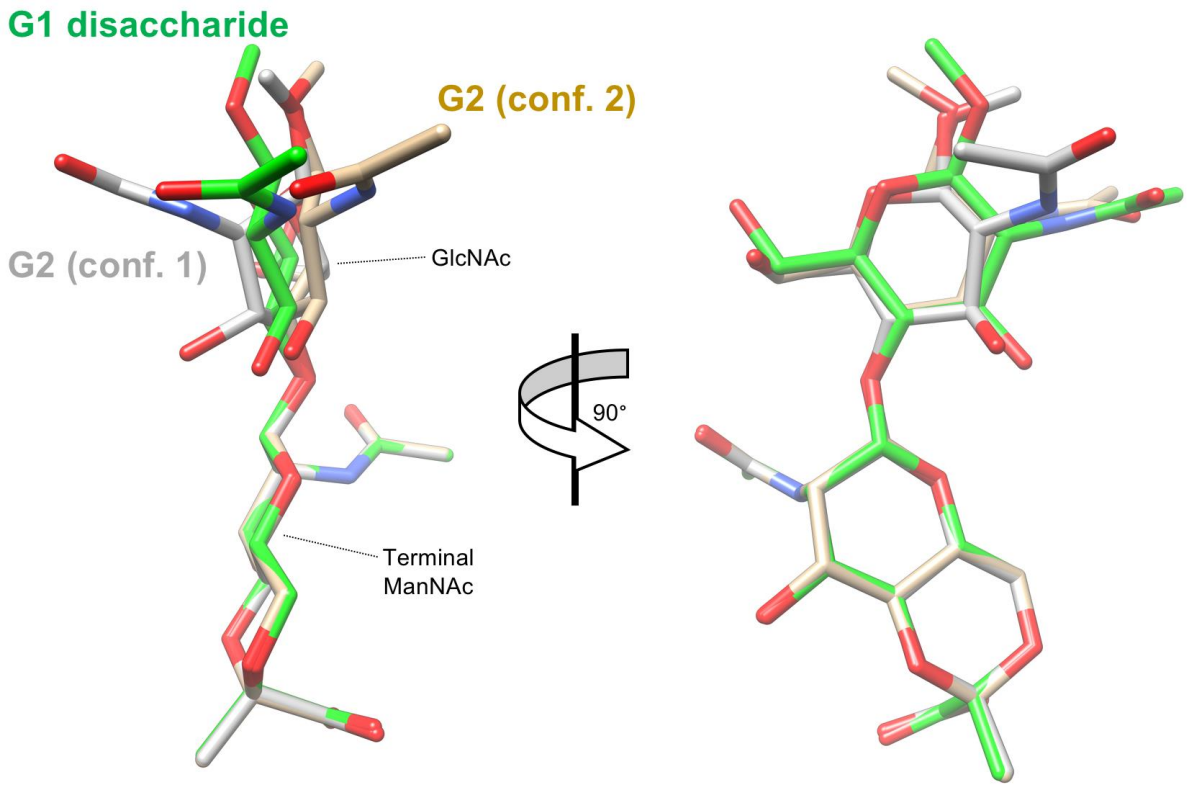
The G2-bound trisaccharide is observed in a single conformation (**Figure 12B**) corresponding to conformation 2 of the disaccharide. In this conformation, there are hydrogen bond interactions between GlcNAc-O3 and the Trp151 side chain nitrogen (3.2 Å), and between internal repeat ManNAc acetamido (3.2 Å; **Figure 12B**) and the GlcNAc acetamido nitrogen atom. Together, these conformational differences demonstrate that, alone, G2 can accommodate flexible SCWP ligands, where the G2 binding site allows the SCWP to pivot about the anchored terminal ManNAc. In the biological context, these conformational minutiae could play an important role: at the surface of *P. alvei* and other bacteria, this phenomenon would allow S-layer proteins undergoing lattice shifts or rearrangements to withstand local increases in the strain of the underlying SCWP-SLH domain interactions, thereby enabling the SCWP interaction to "flex" about the G2 site without breaking contact immediately.

The adaptability of G2 is made further apparent when comparing the structure of $SpaA_{SLH}$ bound to the SCWP terminal disaccharide (4,6-Pyr-β-D-ManNAc-(1→4)-β-D-GlcNAcOMe) with the previously published structure of $SpaA_{SLH}$ bound to the internal-repeat disaccharide (β-D-GlcNAc-(1→3)-4,6-Pyr-β-D-ManNAcOMe), where the internal ManNAc residue is seen bound to the protein in largely the same manner as the terminal ManNAc residue (Blackler, Lopez-Guzman *et al.,* 2018). The lack of corresponding measurable binding in ITC experiments with the internal-repeat SCWP indicates that this phenomenon (*i.e.*, binding the internal ManNAc residue) is unlikely to represent a common biological interaction, but it does emphasize the flexibility of G2 as an adaptable primary SCWP binding site.

2.3.3. G1 of SpaA$_{SLH}$/G109A accommodates an intermediate disaccharide conformation

In the co-crystal structure of disaccharide-bound SpaA$_{SLH}$/G109A, deactivation of G2 binding *via* G109A mutation forces the SLH trimer to rearrange in order to utilize G1 instead. The G1-liganded disaccharide is observed in an "intermediate" conformation—that is, a position between conformations 1 and 2 from G2 of wild-type SpaA$_{SLH}$ (**Figure 14**). There are some similarities between disaccharide binding in G2 of wild-type SpaA$_{SLH}$ and in G1 of the single-mutant (SpaA$_{SLH}$/G109A): only the terminal monosaccharide residue contacts the binding site, and the GlcNAc residue extends into solvent to form bridging contacts with the protein surface through an ordered network of water molecules (**Figure 12C**). This shows that the alternate G1 binding site can also accommodate the longer, more biologically-relevant SCWP terminal disaccharide ligand, and is consistent with binding site switches of terminal SCWP targets being part of a broader S-layer mechanism to relieve increased strain in the protein lattice structure.

Figure 14. *G1-bound disaccharide adopts alternate conformation from G2-bound disaccharide*



The disaccharide ligand bound to G1 of $SpaA_{SLH}$/G109A single mutant adopts an alternate conformation where the GlcNAc residue is in a position between conformations 1 and 2 of the disaccharide ligand that is bound in G2 of wild-type $SpaA_{SLH}$. Ligand atoms are colored by element with oxygen red, nitrogen blue, and carbon green for G1-bound disaccharide, grey for G2 conformation 1, and tan for G2 conformation 2.

2.3.4. SCWP terminal monosaccharide binding alone induces amino acid residue shifts required for longer ligand binding

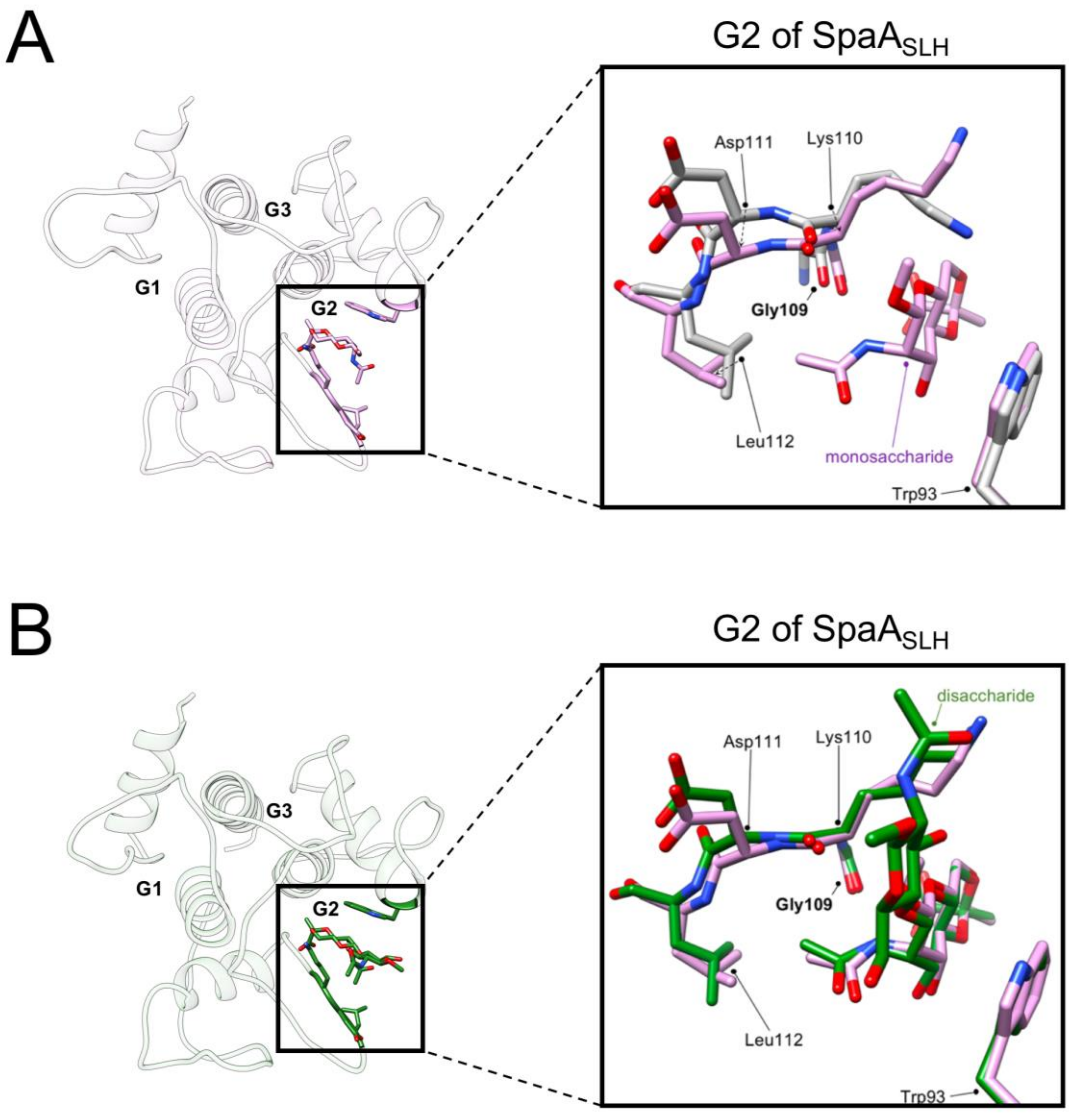
Superposition of the monosaccharide receptor sites of the unliganded wild-type $SpaA_{SLH}$ structure with those of disaccharide-bound (in G2) wild-type $SpaA_{SLH}$ and disaccharide-bound (in G1) $SpaA_{SLH}$/G109A structures reveals that displacement of amino acid residues located near the binding site is necessary for binding of SCWP fragments longer than a monosaccharide. These displacements avoid steric collision with internal oligosaccharide residues. Remarkably, the published monosaccharide-bound structures (Blackler, Lopez-Guzman *et al.,* 2018) of both wild-type $SpaA_{SLH}$ and $SpaA_{SLH}$/G109A proteins exhibit the same relative displacements observed in the structures reported here — that is to say, the binding of the SCWP terminal monosaccharide alone is sufficient to rearrange the receptor binding sites in a manner that accommodates a longer SCWP ligand.

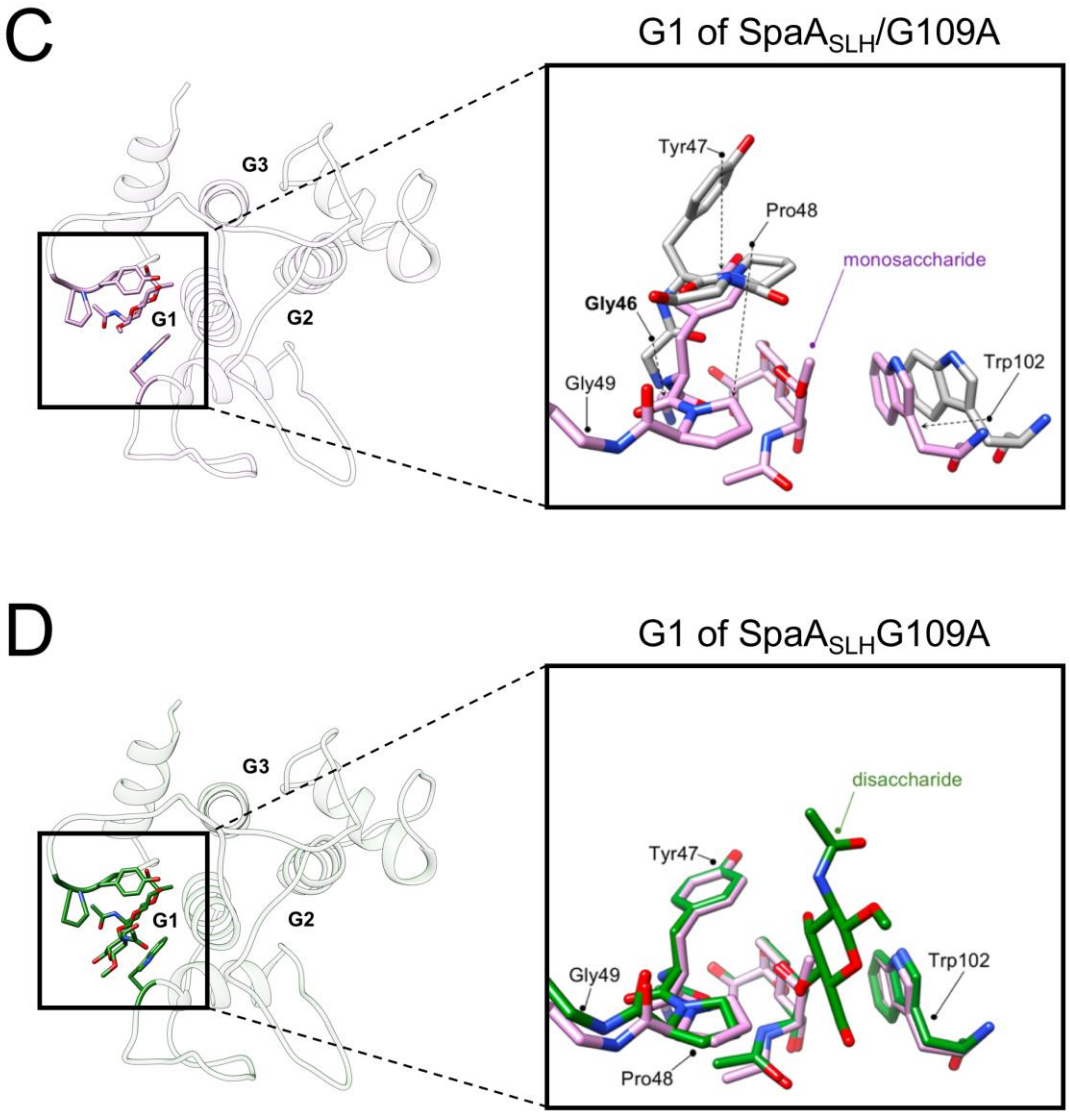
In the receptor site (G2) of disaccharide-bound wild-type $SpaA_{SLH}$, the SLH 30-31-32 (Lys110-Asp111-Leu112) α-carbon atoms are displaced by 1.0 Å, 1.1 Å, and 1.4 Å, respectively, relative to the corresponding α-carbon atoms in the unliganded structure (**Figure 15A**). These three residues, which border G2, are directly connected to conserved residue SLH-Gly29 (Gly109). Specifically, the monosaccharide acetamide moiety displaces Leu112 from the G2 opening. Together with the conserved Gly109 backbone flip induced by binding of the terminal ManNAc residue, the Lys110 main chain atoms move toward the receptor site and stretch the Lys110 side chain atoms into near-full extension in order to accommodate the SCWP disaccharide GlcNAc residue (**Figure 15B**). Interestingly, an identical set of residue shifts are observed in the monosaccharide-bound structure of $SpaA_{SLH}$. This means that the monosaccharide alone is

sufficient to induce the displacement observed in the disaccharide (and trisaccharide) ligand liganded $SpaA_{SLH}$ structures.

A somewhat more dramatic (higher magnitude displacement) but corresponding rearrangement occurs in G1 of the single mutant (**Figure 15C**). However, the displacement appears to operate through a slightly different mechanism. After compensating for the protein rearrangement required to accommodate ligand binding in G1 of the G109A mutant, the α-carbon atoms of G1 residues SLH-30-31-32 (Tyr47-Pro48-Gly49) are displaced by 4.6 Å, 6.4 Å, and 5.2 Å, respectively, relative to the SLH-30-31-32 α-carbon atoms in G1 of the unliganded structure (**Figure 15C**). Specifically, the monosaccharide acetamido moiety directly stacks against the Pro48 side chain, which displaces Pro48 from the G1 opening, providing space for internal oligosaccharide residues to occupy (**Figure 15D**) space formerly occupies by residues Tyr47 and Gly49. Once again, the same SLH-30-31-32 displacements are observed in the disaccharide-bound G109A mutant structure, indicating again that the monosaccharide unit alone induces residue displacements that will be required to accommodate longer SCWP targets.

Figure 15. *$SpaA_{SLH}$ residues SLH 30-31-32 are displaced to accommodate SCWP ligands*





SpaA$_{SLH}$ residues SLH 30-31-32 that border the monosaccharide receptor site shift in response to terminal ManNAc binding to accommodate longer, more biologically relevant SCWP ligands. (**A**) G2 of unliganded SpaA$_{SLH}$ (grey; PDB 6CWC) overlapped with G2 of monosaccharide-bound SpaA$_{SLH}$ (purple; PDB 6CWH) showing the relative displacement of these residues in G2 caused by binding of monosaccharide. (**B**) The overlap of G2 of disaccharide-bound SpaA$_{SLH}$ (green) with G2 of monosaccharide-bound SpaA$_{SLH}$ (purple; PDB 6CWH) shows similar shifts, indicating that it is the binding of the monosaccharide which is largely responsible for the shifts. (**C**) G1 of unliganded SpaA$_{SLH}$ (grey) is overlapped with G1 of monosaccharide-bound SpaA$_{SLH}$/G109A single mutant (purple; PDB 6CWN) showing the relative displacement of these residues in G1 caused by binding of the monosaccharide. (**D**) The overlap of G1 of disaccharide-bound SpaA$_{SLH}$/G109A (green) with G1 of monosaccharide-bound SpaA$_{SLH}$ (purple, PDB: 6CWN) shows similar shifts, indicating again that it is the binding of the monosaccharide which is largely responsible for the shifts. All carbon atoms are colored according to the model descriptions above, with all oxygen and nitrogen atoms colored in *red* and *blue*, respectively.

2.3.5. G1 binding affinity in SpaA$_{SLH}$/G109A is weaker than G2 binding in wild-type SpaA$_{SLH}$

The SpaA$_{SLH}$ trimer has three potential receptor sites, which can complicate binding affinity determinations of this protein. Binding assays carried out well before any structural studies show that separately mutating the three SLH domain binding motifs of SLH1 (TRAE), SLH2 (TVEE), and SLH3 (TRAQ) to TAAA reduces SpaA$_{SLH}$ cell wall binding capacity to 37%, 88%, and 50%, respectively, compared to wild-type protein (Zarschler, Janesch *et al.*, 2010, Janesch, Messner *et al.*, 2013).

Interestingly, no structure with G3 occupied by ligand is ever observed, which might seem to contradict the evidence that disruption of any of the three SLH domains affects binding. The lack of observed binding in G3 can be understood by noting that a number of key residues needed to recognize monosaccharide are missing (*i.e.*, conserved residues SLH-Trp13, SLH-Gly29, and SLH-Arg43 that present in G1 and G2 and that play important roles in binding are instead replaced by threonine, lysine, and valine residues, respectively, in G3). The observation that disruption of any of the three SLH domains affects binding can be explained by remembering that SLH1, SLH2, and SLH3, between them, contribute amino acid residues to the two binding sites known to be functional, G1 and G2.

Co-crystallization of wild-type SpaA$_{SLH}$ with an excess of the SCWP terminal monosaccharide ligand reveals occupation of the G2 site in multiple versions of this structure (Blackler, Lopez-Guzman *et al.*, 2018), which indicates that SCWP binding occurs primarily in G2 (Blackler, Lopez-Guzman *et al.*, 2018). Blackler *et al*. showed that deactivation of G2 through a G109A mutation forces a protein conformational rearrangement to accommodate monosaccharide-binding in G1, though with a significant reduction in binding affinity (K_D of 226 nM) compared to wild type (K_D of 29 nM; **Table 1**). With ligand bound, each binding site makes

a similar set of important contacts with the terminal SCWP monosaccharide through conserved residues SLH-Trp13, SLH-Gly29, and SLH-Arg43. However, the geometries of the G2 and G1 binding sites differ somewhat, including at residue positions SLH 30-31-32, and while these changes are likely contributors to the relative differences in the binding affinities of the two receptor sites a larger contributor is likely the energetic cost of the structural rearrangement.

To understand this, we need to examine the conformational changes in greater detail. In unliganded $SpaA_{SLH}$, G1 adopts a more "open" conformation (measuring 9.1 Å between Trp93 Cη2 and Gly46 O) than G2 (measuring 8.6 Å between Trp151 Cη2 and Gly109 O). When ligand binds to either G1 or G2, the corresponding receptor site width decreases uniformly, down to ~7.5 Å in both cases. This means that G1 must undergo an additional 0.5 Å displacement relative to G2 to accommodate the SCWP-interaction. The binding in G2 does not significantly affect the open conformation of G1 observed in the completely unliganded stated, whereas binding in G1 causes G2 to adopt a more open conformation than is otherwise observed in the completely unliganded state.

To accommodate ligand in G1, the G109A mutant SLH trimer undergoes a conformational change that features displacement of the SLH2 domain towards SLH1 (Blackler, Lopez-Guzman *et al.,* 2018). Part of this rearrangement includes a necessary >4.5 Å shift in G1 residues SLH 30-31-32 (**Figure 15C & 15D**), the movement of which occurs perpendicular to the SLH2 domain shift. The net result is a liganded G1 site with a strong resemblance to that of G2, with sufficient space to accommodate internal residues of oligosaccharide SCWP ligands.

Collectively, the larger groove-narrowing and polypeptide loop shifts of G1 are consistent with the observed reduction in $SpaA_{SLH}$/G109A binding affinity ((Blackler, Lopez-Guzman *et al.*, 2018) and **Table 1**) observed relative to $SpaA_{SLH}$ wild-type.

2.3.6. 'Doubly deactivated' $SpaA_{SLH}$ mutant co-crystallizes with oligosaccharide ligand in G2

The introduction of a single G109A mutation into the G2 binding site of $SpaA_{SLH}$ ($SpaA_{SLH}$/G109A) was sufficient to shift all observed binding in the crystal structure to the previously unoccupied G1 site. The simultaneous deactivation of both G1 and G2 by glycine-to-alanine mutations ($SpaA_{SLH}$/G46A/G109A) results in no measurable binding by ITC, (**Table 1**) to the terminal di- or trisaccharide ligands. However, and consistent with the monosaccharide-bound $SpaA_{SLH}$/G46A/G109A structure (Blackler, Lopez-Guzman *et al.,* 2018), the crystal structure of the double mutant soaked with an approximate five-fold excess of disaccharide shows unambiguous electron density in G2 for a fully occupied disaccharide SCWP fragment in conformation 1 (**Figure 11D**). In the presence of disaccharide ligand, the SpaA crystal structure shows displacement of double mutant residues SLH 30-31-32 relative to the unliganded wild-type $SpaA_{SLH}$, even though the Gly109Ala deactivating mutation prevents the G2 backbone flip. This highlights the importance of the backbone flip in terminal monosaccharide recognition and binding and shows that (within the crystal context) the induced SLH-30-31-32 shifts likely come at a much smaller energetic cost than the glycine backbone rearrangement, but still create sufficient space for the internal residues of oligosaccharide SCWP ligands.

2.3.7. SLH domains of Bacillus anthracis undergo similar residue shifts to those of $SpaA_{SLH}$

Currently, the crystal structures of the SLH domains from the Sap protein of *Bacillus anthracis* (Sap_{SLH}) are the only other SLH domains models available for comparison to $SpaA_{SLH}$. Superposition of the published liganded (PDB: 6BT4 (Sychantha, Chapman *et al.,* 2018)) and unliganded (PDB: 3PYW (Kern, Wilton *et al.,* 2011)) *B. anthracis* Sap_{SLH} structures reveals that G2 residues SHL 30-31-32 (Thr90-Gly91-Asn92) shift upon binding in a manner similar to that

observed for *P. alvei* $SpaA_{SLH}$. The Sap_{SLH} binding site is structurally similar to that of $SpaA_{SLH}$, and the shifting Sap_{SLH} residues, Thr90-Gly91-Asn92, are proximal to conserved residue SLH-Gly29. As in $SpaA_{SLH}$, the displacement of these residues is necessary to prevent steric collision with the internal residue of bound oligosaccharide SCWP. These observations suggest that the tri-residue shift, in addition to the conserved backbone flip mechanism, may be a common feature of the SCWP-SLH domain interaction. It remains unclear whether in Sap_{SLH} of *B. anthracis* the terminal SCWP residue alone is sufficient to induce these receptor site rearrangements as there are currently no published structures of monosaccharide-bound Sap_{SLH}. But as in *P. alvei,* the penultimate SCWP residue (GlcNAc) lacks direct contacts with the protein surface, and the terminal monosaccharide unit dominates the observed SCWP-SLH domain binding interaction.

2.4. Conclusions

The astounding ability to accommodate binding site switches between two distinct SLH domain ligand-receptor sites reveals an evolved mechanism in *Paenibacillus alvei* S-layer that allows these structures to anchor to the cell surface in a dynamic and inherently flexible manner. The studies of the SLH domain trimers from *P. alvei* S-layer protein SpaA presented in this dissertation reveal that while $SpaA_{SLH}$ accommodates longer biologically relevant SCWP ligands within both its primary (G2) and secondary (G1) binding sites, the terminal pyruvylated ManNAc moiety serves as the nearly-exclusive SCWP anchoring point. Ligand binding in both grooves is accompanied by displacement of a flexible loop adjacent to the receptor site that enhances the complementarity between protein and ligand, including electrostatic complementarity with the terminal pyruvate moiety.

Remarkably, binding of the monosaccharide SCWP fragment alone is sufficient to induce these rearrangements in a manner necessary to accommodate longer SCWP fragments. The observation of multiple conformations for longer oligosaccharides bound to the protein, together with the demonstrated functionality of two of the three SCWP receptor binding sites, reveals how the $SpaA_{SLH}$-SCWP interaction has evolved to accommodate longer SCWP ligands and alleviate sources of environmental strain such as those inherent to bacterial S-layer adhesion during growth and division.

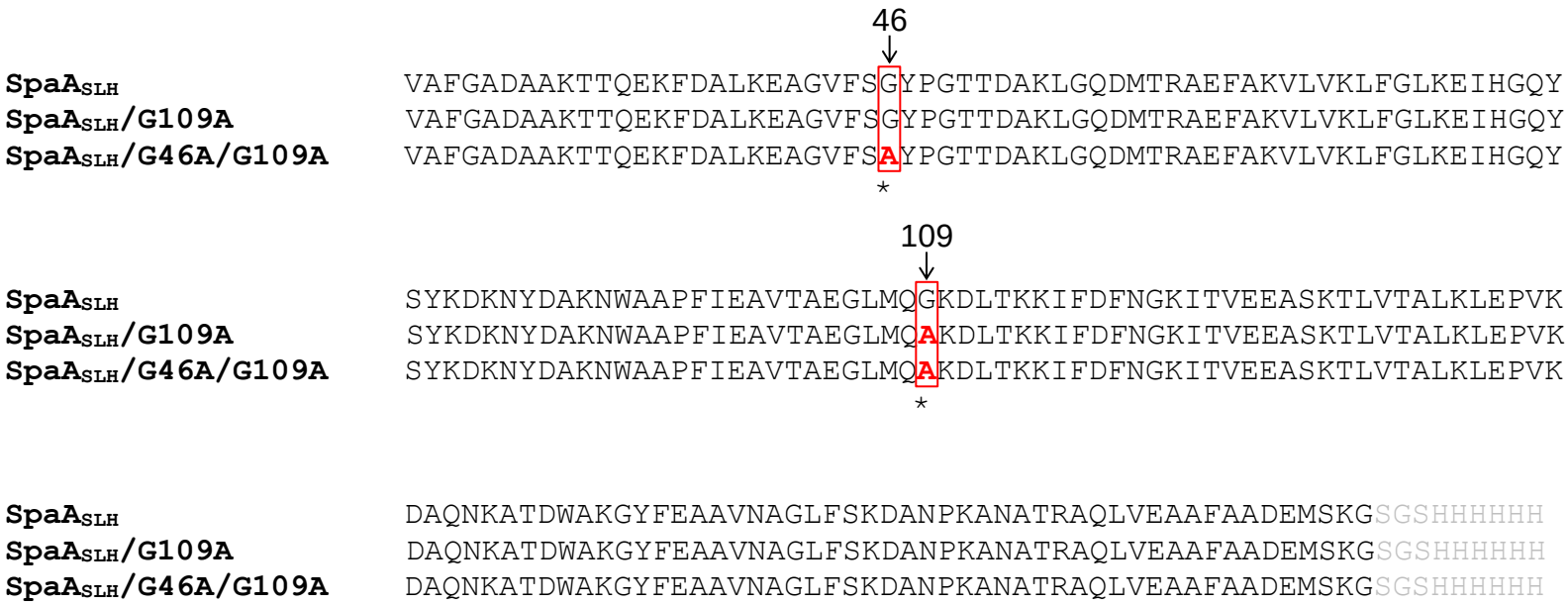
Together, these structures provide insight into the conformational dynamics and SCWP-binding mechanisms of the conserved SLH-anchoring motifs in Gram-positive organisms, which is of fundamental importance to our understanding of S-layer protein structure and function and may be combined with future experiments to provide direct practical applications in biotechnology as well as therapeutic design.

2.5. Methods

2.5.1. $SpaA_{SLH}$ expression and purification

The wild-type, single mutant (G109A) and double mutant (G46A/G109A) $SpaA_{SLH}$ protein constructs (**Figure 16**) were expressed and purified as described previously (Blackler, Lopez-Guzman *et al.*, 2018). Each variant consists of amino acids 21-193 of $SpaA_{SLH}$, with a Ser-Gly-Ser linker followed by a C-terminal His_6 tag included for purification purposes. The gene for each $SpaA_{SLH}$ protein construct was inserted into the pET22b expression vector (Novagen) and transformed into *E. coli* BL21 (DE3-Star) cells (Invitrogen) for protein overexpression in LB medium supplemented with 100 μg/mL ampicillin. All cell culture incubations occurred at 37°C with 200 rpm shaking. Protein overexpression was induced *via* addition of IPTG to a final concentration of 0.6 mM after cells reached an OD_{600} ~0.6. After induction, cells were incubated for a further 3 hours and subsequently harvested by centrifugation for 25 min at 6,500 x *g*. Cell pellets were resuspended in lysis buffer (25 mM Tris-HCl, pH 8.0, 200 mM NaCl, 5 mM imidazole) and lysed by sonication. The sonication lysate was centrifuged at 10,000 x *g* and the clarified supernatant was applied to a nickel NTA-affinity chromatography column (Qaigen) pre-equilibrated with lysis buffer. The column was washed with ten column volumes of wash buffer A (25 mM Tris-HCl, pH 8.0, 20 mM imidazole), followed by ten volumes of wash buffer B (25 mM Tris-HCl, pH 8.0, 50 mM imidazole), and eluted with elution buffer (25 mM Tris-HCl, pH 8.0, 250 mM imidazole). The eluted sample was then purified by size exclusion on a Superdex 75 16/60 column pre-equilibrated with buffer containing 20 mM HEPES, pH 7.5, and 100 mM NaCl. Fractions were analyzed for purity by SDS-PAGE on a 10% acrylamide gel that was visualized using Coomassie brilliant blue stain.

Figure 16. *Amino acid sequences of the $SpaA_{SLH}$ constructs used in this study*



Amino acid sequence alignment of the $SpaA_{SLH}$ wild-type, single mutant (G109A) and double mutant (G46A/G109A) constructs used in this study. Mutated residues at positions 46 and 109 in the protein sequence are indicated in bolded red font; the Ser-Gly-Ser linker and C-terminal His_6 tag (derived from the pET22b expression vector) are indicated in grey font.

2.5.2. $SpaA_{SLH}$ isothermal titration calorimetry

All ITC experiments were carried out by my collaborators Fiona F. Hager-Mair, Arturo Lòpez-Guzmán and Christina Schäffer, from the Department of NanoBiotechnology at the University of Natural Resources and Life Sciences in Vienna, Austria. ITC experiments were performed on a MCS titration calorimeter (Microcal, Inc., Northampton, MA) following the same protocol as previously described in (Blackler, Lopez-Guzman *et al.*, 2018) with minor modifications: the terminal disaccharide and trisaccharide SCWP analogues were used as ligands in this set of ITC experiments instead of terminal monosaccharide analogue as was done previously (Blackler, Lopez-Guzman *et al.*, 2018). All protein concentrations were determined using UV/Vis light spectroscopy and a molar extinction coefficient of 18450 $M^{-1}cm^{-1}$ at 280 nm. The concentrations of the terminal disaccharide and terminal trisaccharide SCWP ligands were determined based on dry weight. A protein-to-ligand ratio of 1:40 was used for ITC measurements, all of which were done at 20 °C in 20 mM KH_2PO_4, pH 7.0. Each experiment was performed at least in triplicate.

2.5.3. $SpaA_{SLH}$ crystallization

The purified wild-type $SpaA_{SLH}$, $SpaA_{SLH}$/G109A, and $SpaA_{SLH}$/G46A/G109A expression constructs were each concentrated using 3-kDa MWCO Amicon ultra centrifugal filter units, and were co-crystallized with either 5 mM SCWP terminal disaccharide analogue (4,6-Pyr-β-D-ManNAc-(1→4)-β-D-GlcNAcOMe) (Krauter, Schäffer *et al.*, 2021) or 1.1 mM SCWP terminal trisaccharide analogue (4,6-Pyr-β-D-ManNAc-(1→4)-β-D-GlcNAc-(1→3)-4,6-Pyr-β-D-ManNAcOMe). Crystallization screens were prepared using an Art Robbins Instruments crystal gryphon robot and Hampton 96-well Intelli plates. Screens were immediately stored at 18 °C. Co-

crystals of $SpaA_{SLH}$ (15 mg/ml) and disaccharide were obtained within one week from Hampton Index screen condition number 2 (0.1 M sodium acetate trihydrate pH 4.5, 2.0 M ammonium sulphate). Co-crystals of $SpaA_{SLH}$/G109A (12 mg/ml) and disaccharide were obtained after 1 month from the Qiagen PEG II screen condition 15 (0.1 M HEPES pH 7.5, 25% (w/v) PEG 1000). Co-crystals of $SpaA_{SLH}$/G46A/G109A (19 mg/ml) and disaccharide were obtained after one week from the Qiagen PEG II screen condition 68 (0.1 M sodium acetate, 25% (w/v) PEG 4000, 8% (w/v) isopropanol). Co-crystals of $SpaA_{SLH}$ (18 mg/ml) and trisaccharide were obtained within one week from Qiagen JCSG+ screen condition number 83 (0.2 M Magnesium Chloride, 0.1 M Bis-Tris pH 5.5, 25% PEG 3350).

2.5.4. Data collection, structure solution, and refinement

Before being flash-frozen at 100°K for data collection, each crystal was resuspended for 10-20 seconds in mother liquor containing 20% glycerol. X-ray diffraction data were collected for all structures using a Dectris Pilatus 200 K detector coupled to a Rigaku Micromax-007 HF X-ray generator. The data were scaled, averaged and integrated using HKL2000. Molecular replacement was performed to generate a structure solution *via* the use of PHASER from the CCP4 suite (Winn, Ballard *et al.*, 2011). Unliganded $SpaA_{SLH}$ (PDB: 6CWM) was used as the search model for the $SpaA_{SLH}$ and $SpaA_{SLH}$/G46A/G109A structures, and $SpaA_{SLH}$/G109A (PDB: 6CWN, with the monosaccharide ligand removed) used as a search model for the $SpaA_{SLH}$/G109A structure. After initial model improvement and structure refinement, the 4,6-Pyr-β-D-ManNAc-(1→4)-β-D-GlcNAcOMe disaccharide and 4,6-Pyr-β-D-ManNAc-(1→4)-β-D-GlcNAc-(1→3)-4,6-Pyr-β-D-ManNAcOMe trisaccharide ligands were generated using the PRODRG2 Server and incorporated into the appropriate corresponding molecular models (Schüttelkopf and van Aalten 2004). All

model building steps and refinements were performed using Coot macromolecular model building software (Emsley, Lohkamp *et al.,* 2010) and REFMAC5 from the CCP4 program suite (Winn, Ballard *et al.,* 2011).

2.5.5. Visualization and graphics

All figures that depict protein or ligand chemical structures, or electron density were created using the UCSF Chimera Extensible Molecular Modeling System, developed by the Resource for Biocomputing, Visualization, and Informatics at the University of California, San Francisco, with support from NIH P41-GM103311 (Pettersen, Goddard *et al.,* 2004).

Chapter 3: Characterization of the unliganded structure of *Paenibacillus alvei* MnaA, a UDP-GlcNAc 2-epimerase involved in secondary cell wall polymer synthesis

Contributions:

The MnaA protein construct used in this study was provided by my collaborators Cordula Stefanović and Christina Schäffer from the Department of NanoBiotechnology, *NanoGlycobiology* Unit, at the Universität für Bodenkultur Wien in Vienna, Austria. Cordula Stefanović and Nick Mateyko both contributed to the expression, purification and initial crystallization of the MnaA protein, while Nick Mateyko and the author performed the crystal optimizations and soaking experiments for this research. The author collected the data, and solved and analyzed the structure.

3.1. Introduction to Chapter 3

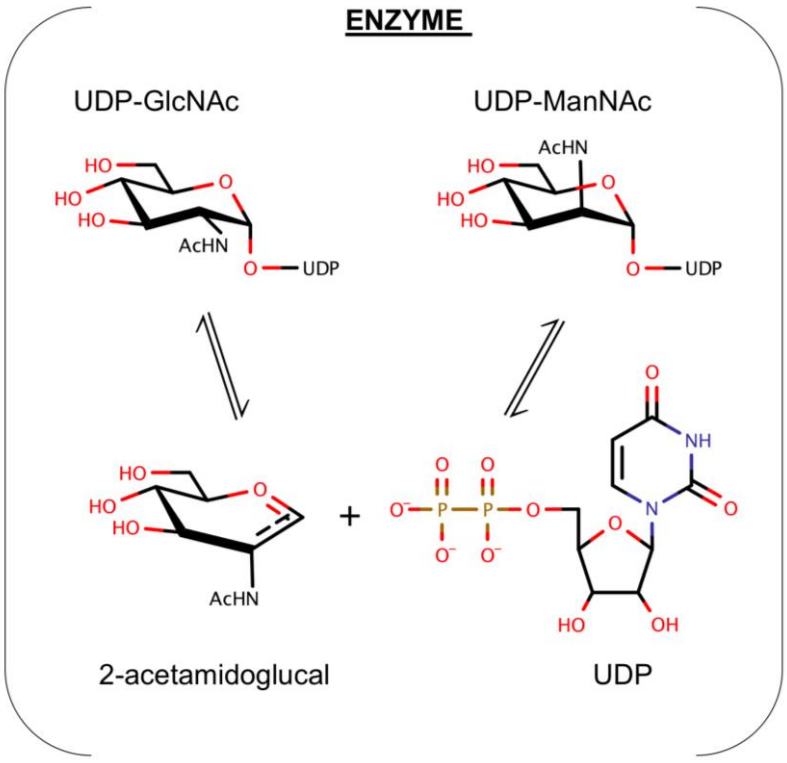
P. alvei MnaA, of the UDP-GlcNAc 2-epimerase protein family (PF02350) (El-Gebali, Mistry *et al.,* 2019), catalyzes the reversible conversion of UDP-GlcNAc to UDP-ManNAc (**Figure 17**) (Velloso, Bhaskaran *et al.,* 2008, Hager, Lopez-Guzman *et al.,* 2018). This generates the UDP-ManNAc substrate required for TagA to incorporate ManNAc into the *P. alvei* SCWP (**Figure 9**) (Hager, Lopez-Guzman *et al.,* 2018). The UDP-GlcNAc 2-epimerase protein family broadly contains two groups of enzymes: prokaryotic non-hydrolyzing UDP-GlcNAc 2-epimerases that retain the UDP group in the UDP-ManNAc product, and bifunctional hydrolyzing UDP-GlcNAc 2-epimerases involved in sialic acid synthesis in mammals that also have kinase functionality and do not retain UDP in the product (*i.e.* the bond between UDP and ManNAc is hydrolyzed) (Velloso, Bhaskaran *et al.,* 2008).

MnaA is a non-hydrolyzing UDP-GlcNAc 2-epimerase, and structures from 11 bacterial species are available from the Protein Data Bank (PDB) (Berman, Westbrook *et al.,* 2000). Both groups of enzymes have similar folds consisting of two Rossmann-like domains that form an active site in the cleft between them, which is the same structure seen in GT-B fold type glycosyltransferases (GTs), and the catalytic mechanism of both groups is believed to be similar (Chen, Huang *et al.,* 2016). Multiple pieces of chemical evidence suggest that the catalytic mechanism of UDP-GlcNAc 2-epimerases is a two-step elimination-addition mechanism with a 2-acetamidoglucal intermediate (Tanner 2002) (**Figure 17**).

The bacterial non-hydrolyzing UDP-GlcNAc 2-epimerases are allosterically activated by their substrate UDP-GlcNAc, which means that if only UDP-ManNAc is present, essentially no UDP-GlcNAc is formed. However, with the addition of only a small amount of UDP-GlcNAc, the reaction goes to equilibrium (Velloso, Bhaskaran *et al.,* 2008). Binding of UDP-GlcNAc to

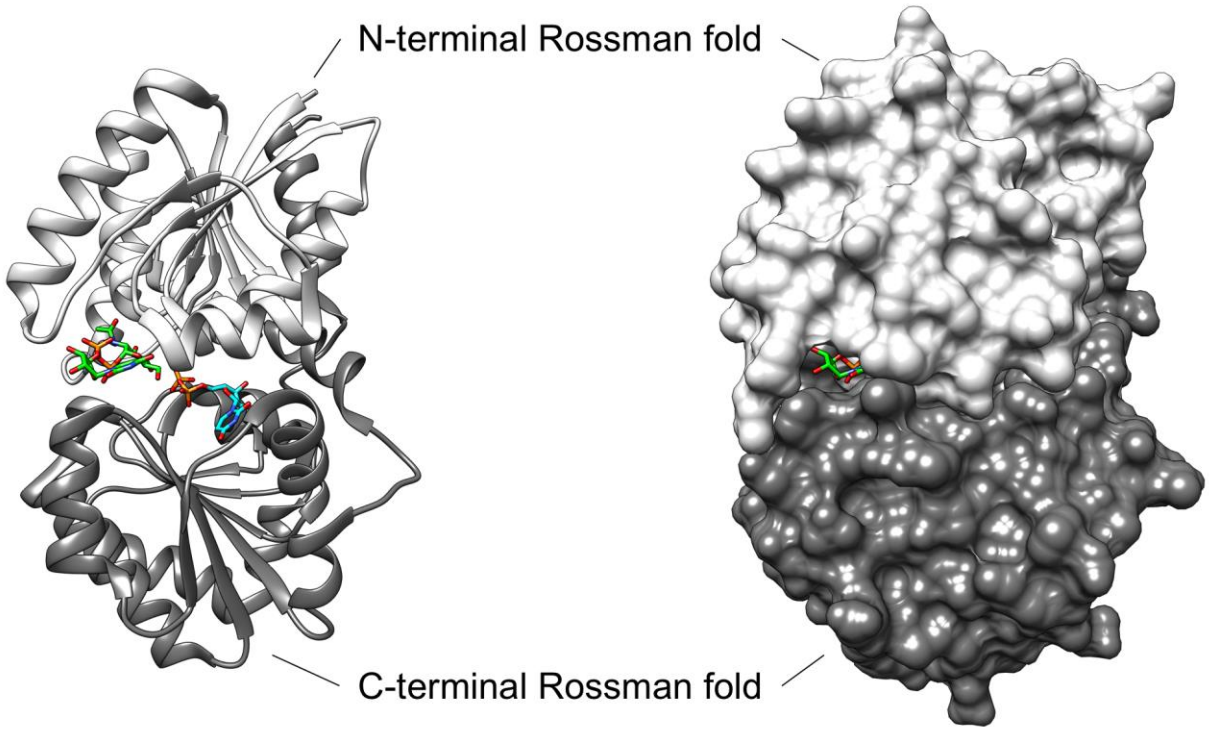
the allosteric site causes a conformational shift that closes the two Rossman-like domains onto the active site, and the UDP-GlcNAc in the allosteric site also directly interacts with the UDP-GlcNAc bound to the catalytic site and prevents access to solvent (**Figure 17**) (Velloso, Bhaskaran *et al.*, 2008). The obstruction of the active site is thought to be important because formation of the 2-acetoamidoglucal intermediate and UDP is thermodynamically favoured over both UDP-GlcNAc and UDP-ManNAc products, therefore this more stable intermediate must be trapped within the enzyme so that it does not become the major product (Tanner 2002, Velloso, Bhaskaran *et al.*, 2008). UDP-GlcNAc 2-epimerases do in fact occasionally release 2-acetoamidoglucal and UDP which can be detected after long incubation of UDP-GlcNAc with high concentrations of enzyme (Tanner 2002).

Figure 17. *Mechanism of reaction catalyzed by MnaA*



Although a number of structures of bacterial non-hydrolyzing UDP-GlcNAc 2-epimerases with bound UDP-GlcNAc have been published (Velloso, Bhaskaran *et al.*, 2008, Chen, Huang *et al.*, 2016, Mann, Muller *et al.*, 2016), in each case only UDP-GlcNAc is observed bound to the allosteric site with UDP bound at the catalytic site and the sugar absent (**Figure 18**). Comparison with the *E. coli* epimerase structure reveals that interactions with UDP in the catalytic site of the *B. anthracis* epimerase appear to be conserved (Velloso, Bhaskaran *et al.*, 2008). However, the residues responsible for proton abstraction/addition have yet to be determined unambiguously (Velloso, Bhaskaran *et al.*, 2008). Mutations in the region where the glucosamine moiety of the substrate would be positioned have identified Asp95 and Glu131 of the *E. coli* epimerase as candidate residues for proton abstraction, while Glu117 is suggested to be involved in the second step of the reaction (Velloso, Bhaskaran *et al.*, 2008). These residues are conserved in *B. anthracis* MnaA as Asp100, Glu136, and Glu122 (Velloso, Bhaskaran *et al.*, 2008), and as Asp99, Glu135, and Glu121 in *P. alvei* MnaA. Therefore, if a crystal structure of *P. alvei* MnaA epimerase with substrate or substrate analog bound at the catalytic site could be generated, a clearer picture of the structural role of these residues in catalysis could be determined.

Figure 18. *Closed conformation of UDP-GlcNAc 2-epimerase enzyme from* Bacillus anthracis



Ribbon (*left*) and surface (*right*) depictions of the UDP-GlcNAc 2-epimerase from *B. anthracis* (PDB ID: 3BEO). The enzyme adopts a closed conformation when UDP-GlcNAc (substrate with green carbon atoms) is bound at the allosteric site and UDP (substrate with blue carbon atoms) is present in the catalytic site (Velloso, Bhaskaran *et al.*, 2008).

3.2. Results

3.2.1. Data collection & structure determination

All data collection and refinement statistics for the unliganded structure of MnaA are shown in **Table 3**. The crystal structure was solved to a final resolution of 2.20 Å in space group $P12_11$ with two molecules in the AU. A Phyre2-generated search model (Kelley, Mezulis *et al.*, 2015) was used for molecular replacement in PHASER of the CCP4 program suite (Winn, Ballard *et al.*, 2011). The solution was finalized *via* iterative refinement and model improvements using REFMAC5 of the CCP4 program suite (Winn, Ballard *et al.*, 2011) and Coot macromolecular model building software (Emsley, Lohkamp *et al.*, 2010). The final structure yielded an average B-factor of 36.1 $Å^2$, with R_{work} and R_{free} values of 21.0 % and 26.2 %, respectively.

As expected, the overall structure of MnaA forms a GT-B fold, where both the N-terminal and C-terminal domains adopt distinct Rossmann folds connected by a short linker region (consisting of residues 173-200). The space between these two folds forms an interdomain cleft containing the active site (**Figure 19**).

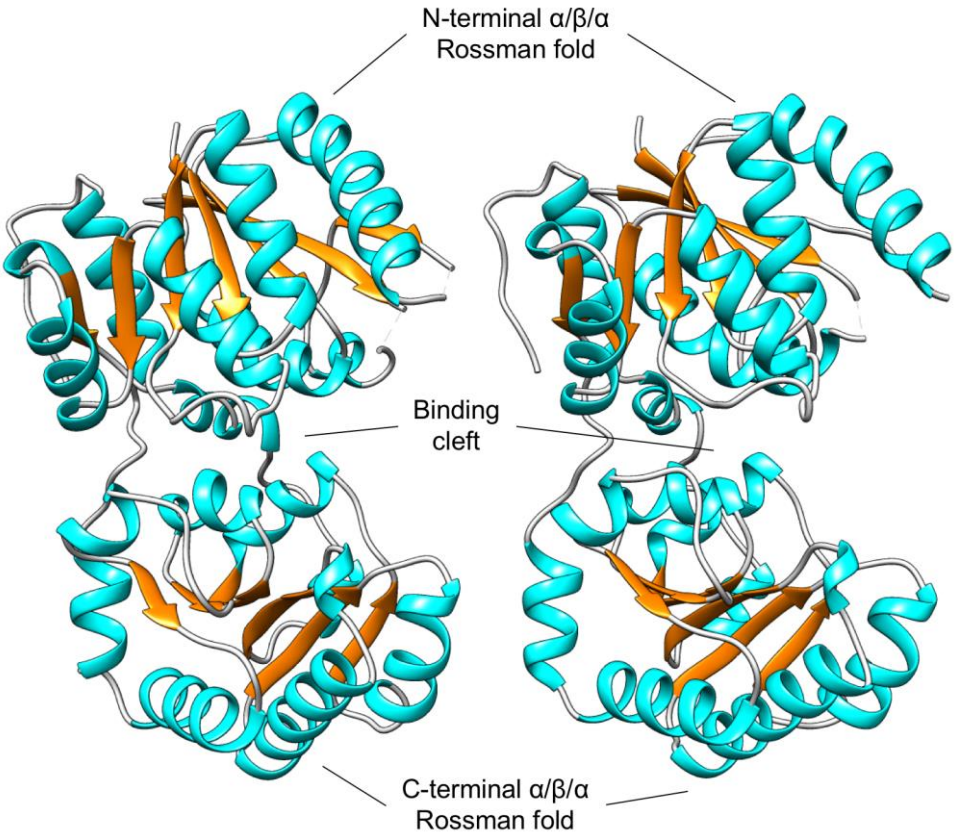
The N-terminal domain (consisting of residues 2-172) shows good electron density for almost all residues except those at positions 42-45 and 64-69 in molecule A, and those at positions 43-46 and 66-68 in molecule B, for which no corresponding electron density is observed and no residues were modeled. The N-terminal domain forms an extended β-sheet structure with 3-2-1-4-5-6-7 connectivity between the 7 β-strands (β1-β7), which are linked by 6 α-helices (α1-α6) (**Figure 19**). The unmodeled residues in molecule A and molecule B correspond to flexible loop regions that are located between the β2 and α2, and β3 and α3 secondary structural elements.

The linker region that connects the N- and C-terminal domains (residues 173-200) consists of two α-helices (α7 and α8). Residues in this region show mostly good electron density, apart

from residues 186-192 in molecule A, and residues 186-190 in molecule B, which only show partial main chain and side-chain electron density.

The C-terminal domain (consisting of residues 201-347) exhibits good electron density for all residues in molecules A and B. Similar to the N-terminal domain, the C-terminal domain forms an extended β-sheet structure (with 3-2-1-4-5-6 connectivity between the β-strands) that is comprised of 6 β-strands (β8-β13) that are linked by 5 α-helices (α9-α13). There are 3 additional α-helices (α14-α16) present after the final β-strand of the C-terminal domain, the last of which is associated to the N-terminal domain of the MnaA crystal structure (**Figure 19**).

***Figure 19**. MnaA adopts a GT-B fold*



The MnaA crystal structure was solved with two molecules (molecules A and B) in the AU (shown above). The N-terminal and C-terminal domains each adopt a distinct Rossmann fold, which is characteristic of the GT-B fold. The space between these two folds forms an interdomain cleft containing the active site.

Table 3. *Data collection and refinement statistics for MnaA*

Protein	**MnaA**
Data collection	
Space group	$P12_11$
Resolution (Å)	50.00-2.20 (2.28-2.20)
Cell dimensions	
a (Å)	45.82
b (Å)	81.76
c (Å)	107.72
α (°)	90
β (°)	97.14
γ (°)	90
Z	4
R_{sym}	0.057 (0.211)
R_{pim}	0.031 (0.133)
$C/C_{1/2}$	(0.957)
I/σ(I)	18.3 (4.4)
Completeness (%)	99.0 (91.4)
Redundancy	4.0 (3.0)
Unique reflections	39705
Refinement	
Resolution (Å)	25.41-2.20
No. reflections	37651
R_{work} (%)	21.0
R_{free} (%)	26.2
No. atoms	
Protein	5658
Ligand	14 (PEG[a])
Water	183
B factors	
Protein	36.1
Ligand	50.6
Water	36.1
Average	36.1
Ramachandran	
Favoured (%)	95.6
Allowed (%)	4.4
r.m.s. bonds (Å)	0.0078
r.m.s. angles (°)	1.66

[a] PEG = 2-(2-hydroxyethyloxy)ethanol ($C_4 H_{10} O_3$)

3.2.2. MnaA crystal ligand soaking

Co-crystal screening of MnaA with UDP-GlcNAc substrate had not produced crystal hits to date and therefore MnaA crystals obtained from optimized PEGs II condition number 36 (0.16 M $CaCl_2$, 0.1 M tris pH 8.8, and 15% PEG 4000) were soaked in a drop of mother liquor supplemented with UDP-α-D-GlcNAc in an attempt to generate a donor-bound structure. Overnight soaks in 10 mM UDP-α-D-GlcNAc produced no change to the resultant crystal structures, with no bound ligand observed (data not shown). Overnight soaking in 50 mM UDP-α-D-GlcNAc caused all crystals to dissolve.

3.3. Discussion

3.3.1. MnaA adopts a GT-B fold with an open conformation in the unliganded state

The *P. alvei* MnaA structure contains two molecules in the crystallographic asymmetric unit with both molecules displaying near identical conformations (with an overlap Cα rmsd of 0.32 Å for 370 equivalent residues). The monomer displays an overall GT-B fold where the N- and C-terminal domains each adopt the characteristic α/β/α layered architecture of a Rossmann domain (**Figure 19**). This arrangement is consistent with the architecture of previously published structures of other bacterial non-hydrolyzing UDP-GlcNAc 2-epimerase enzymes (Campbell, Mosimann *et al.,* 2000, Badger, Sauder *et al.,* 2005, Velloso, Bhaskaran *et al.,* 2008, Chen, Huang *et al.,* 2014, Mann, Muller *et al.,* 2016, Hurlburt, Guan *et al.,* 2020), for which substrate binding is observed in the cleft located between the two domains (Campbell, Mosimann *et al.,* 2000, Velloso, Bhaskaran *et al.,* 2008, Chen, Huang *et al.,* 2014, Mann, Muller *et al.,* 2016, Hurlburt, Guan *et al.,* 2020) (**Table 4**).

In the absence of substrate, MnaA adopts an unliganded conformation that matches the "open" state of these other enzymes (Badger, Sauder *et al.,* 2005, Chen, Huang *et al.,* 2014, Hurlburt, Guan *et al.,* 2020), where the inner surfaces of the binding cleft are exposed to solvent. This open state is thought to facilitate substrate entry and binding within the substrate cleft (Campbell, Mosimann *et al.,* 2000, Mann, Muller *et al.,* 2016). In these previously published UDP-GlcNAc 2-epimerase structures, substrate binding appears to induce a significant conformation change in which the two Rossmann-like domains "clamp down" on the bound substrate, whereupon the epimerization reaction proceeds in the "closed" enzyme state (Campbell, Mosimann *et al.,* 2000, Velloso, Bhaskaran *et al.,* 2008, Mann, Muller *et al.,* 2016).

Table 4. *Summary of all bacterial non-hydrolyzing UDP-GlcNAc 2-epimerase structures in the PDB*

PDB	Identity to MnaA	Resolution	Organism	Conformation	Catalytic site substrate	Allosteric site substrate	Publication reference
3BEO	67.2 %	1.70 Å	*Bacillus anthracis*	Closed	UDP *	UDP-GlcNAc	Velloso, Bhaskaran *et al.*, (2008)
1O6C	63.0 %	2.90 Å	*Bacillus subtilis*	Open	-	-	Badger, Sauder *et al.*, (2005)
4FKZ	63.0 %	1.69 Å	*Bacillus subtilis*	Closed	UDP *	UDP-GlcNAc	Structure released in 2013; no associated publication
3OT5	59.2 %	2.20 Å	*Listeria monocytogenes*	Open	-	-	Structure released in 2010; no associated publication
5ENZ	58.0 %	1.91 Å	*Staphylococcus aureus*	Closed	UDP *	-	Mann, Muller *et al.*, (2016)
1F6D	53.8 %	2.50 Å	*Escherichia coli*	Partially open	UDP *	-	Campbell, Mosimann *et al.*, (2000)
1VGV	53.8 %	2.31 Å	*Escherichia coli*	Closed-liganded	UDP-GlcNAc		Badger, Sauder *et al.*, (2005) *
3DZC	50.7 %	2.35 Å	*Vibrio cholerae*	Open	-	-	Structure released in 2008; no associated publication
6VLB	48.1%	1.85 Å	*Neisseria meningitidis*	Open	-	-	Hurlburt, Guan *et al.*, (2020)
6VLC	48.1%	2.15 Å	*Neisseria meningitidis*	Closed	UDP-GlcNAc	-	Hurlburt, Guan *et al.*, (2020)
5DLD	47.8 %	1.45 Å	*Burkholderia vietnamiensis*	Closed	UDP *	UDP-GlcNAc	Structure released in 2015; no associated publication
1V4V	45.4 %	1.8 Å	*Thermus thermophilus*	Open	-	-	Structure released in 2003; no associated publication
4NEQ	32.1 %	2.85 Å	*Methanocaldococcus jannaschii*	Open	-	-	Chen, Huang *et al.*, (2014)
4NES	32.1 %	1.42 Å	*Methanocaldococcus jannaschii*	Closed	UDP *	UDP-GlcNAc	Chen, Huang *et al.*, (2014)
4HWG	29.5 %	2.00 Å	*Rickettsia bellii*	Open	-	-	Structure released in 2003; no associated publication

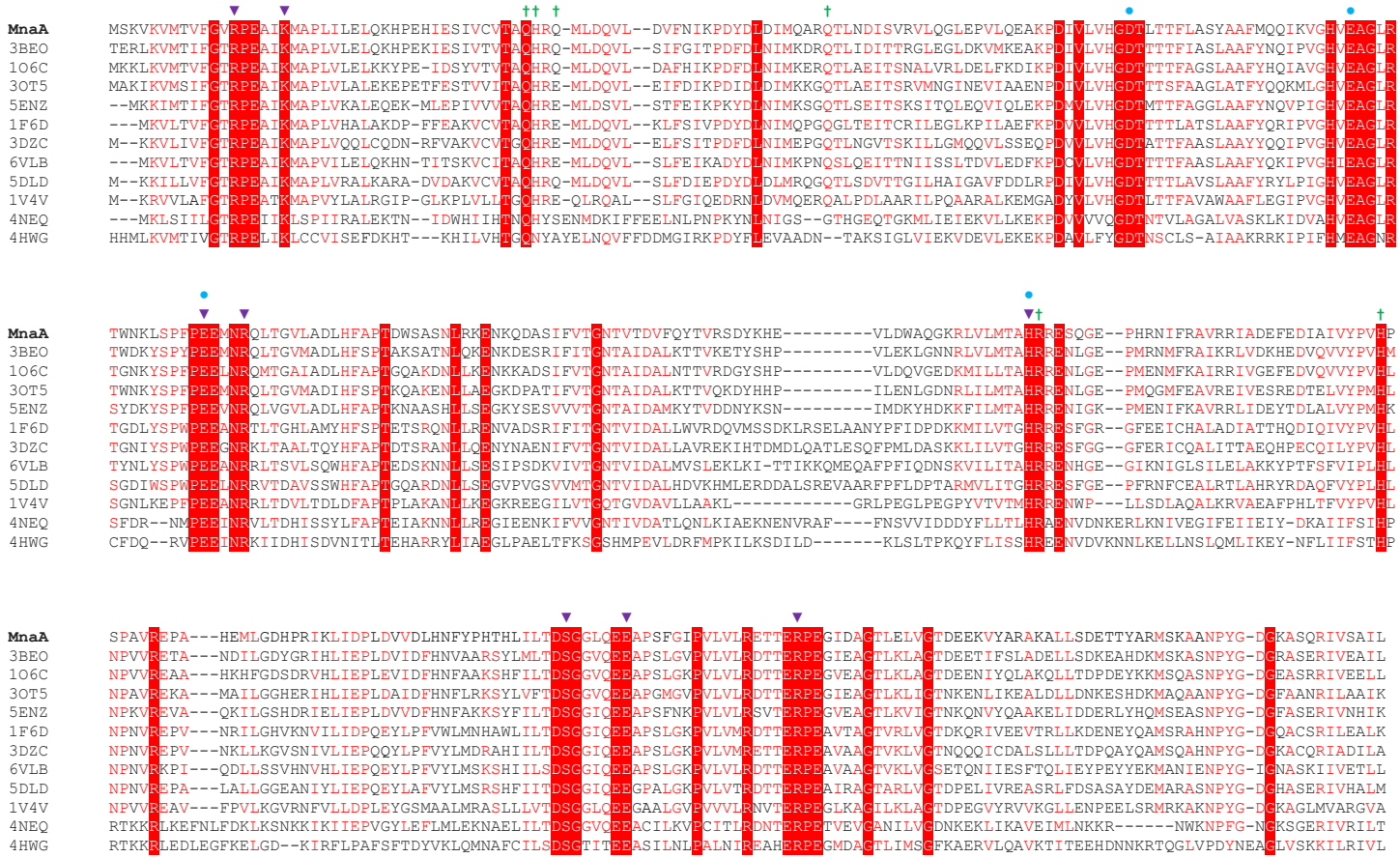
* Only the UDP component of UDP-GlcNAc substrate was modeled in the catalytic site of the indicated structures

Key residues involved in substrate binding for these other UDP-GlcNAc epimerase enzymes are conserved in *P. alvei* MnaA (Arg13, Lys18, Glu135, Arg139, His208, Ser285, and Glu291) (**Figure 20 & Table 5**). In addition, residues Asp99, Glu121, Glu135 and His208 of the putative catalytic site (**Figure 20; Table 5; & Figure 21**) are conserved in the *P. alvei* MnaA amino acid sequence. Thus, despite variance in the overall amino acid sequences of these structures (from about 66.2–29.5% sequence identity) the residues most important to catalysis (and substrate binding) appear to be conserved.

As mentioned, some bacterial non-hydrolyzing UDP-GlcNAc 2-epimerases function *via* allosteric regulation. Crystal structures of UDP-GlcNAc 2-epimerases from Gram-positive *Bacillus anthracis* (Velloso, Bhaskaran *et al.*, 2008) and *Staphylococcus aureus* (Mann, Muller *et al.*, 2016), as well as the archaeon *Methanococcus jannaschii* (Chen, Huang *et al.*, 2014), reveal that this allosteric regulation occurs *via* binding of UDP-GlcNAc substrate in a conserved site next to the active site. In these structures the active site contains only a bound UDP molecule. This is due to hydrolysis of the GlcNAc residue when enzyme is soaked or co-crystallized with its natural substrate, UDP-GlcNAc (Velloso, Bhaskaran *et al.*, 2008, Chen, Huang *et al.*, 2014).

UDP-GlcNAc binding at the allosteric site likely functions to exclude solvent from the active site and optimize the catalytic conformation of the enzyme (Campbell, Mosimann *et al.*, 2000, Velloso, Bhaskaran *et al.*, 2008, Chen, Huang *et al.*, 2014, Mann, Muller *et al.*, 2016). Additionally, the GlcNAc moiety of the allosteric effector is proximal to the active site and is poised to make extensive interactions with UDP molecule bound therein (Velloso, Bhaskaran *et al.*, 2008, Chen, Huang *et al.*, 2014, Mann, Muller *et al.*, 2016). In these enzymes, the observed allosteric site residues are not uniformly conserved, but the chemical identities are preserved in

Figure 20. *Comparison of bacterial non-hydrolyzing UDP-GlcNAc 2-epimerases.*



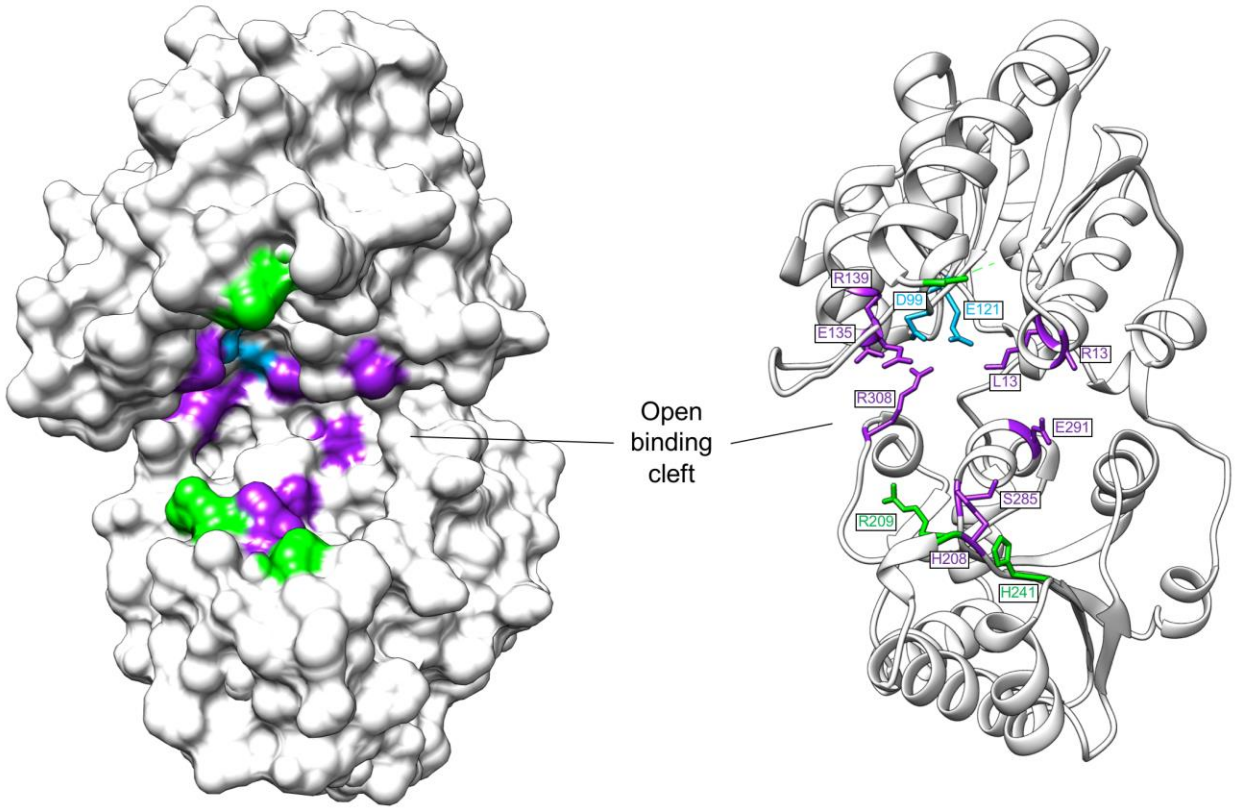
The sequence alignment of all know bacterial non-hydrolyzing UDP-GlcNAc 2-epimerase structures in the PDB (see summary in **Table 4**; if multiple structures of the same enzyme from the same organism are published, only one representative sequence is shown to avoid redundancy). Red shading indicates residues that are 100% conserved in all epimerases while red font indicates highly conserved residues or residues of similar identity. The putative catalytic residues are indicated as blue circles above the sequence, residues that bind the UDP-GlcNAc substrate are designated by purple triangles, and residues that bind the UDP-GlcNAc allosteric effector are designated by green crosses.

Table 5. *Allosteric/catalytic site residues from other UDP-GlcNAc 2-epimerases present in* P. alvei *MnaA*

Substrate binding site residues	Putative catalytic residues	Allosteric site residues
Arg13	Asp99	Gln42
Lys18	Glu121	His43*
Glu135	Glu135	Gln45*
Arg139	His208	Gln69
His208		Arg209
Ser285		His241
Glu291		

*residues indicated by asterisk were not modelled in crystal structure

Figure 21. *Positions of conserved allosteric and catalytic site residues in MnaA cleft*



Surface representation (*left*) and ribbon model (*right*) showing the locations of conserved binding site residues (purple), putative catalytic site residues (blue), and allosteric site residues (green) in the MnaA cleft. The residues that are indicated correspond to those that are listed in **Table 5**.

most cases (*e.g.*, interchange of polar residues glutamine or serine) (**Figure 20**). Notably, *P. alvei* MnaA shows the greatest level of amino acid sequence identity (67.2%) with the *B. anthracis* MnaA UDP-GlcNAc 2-epimerase (**Table 4**); they share the same five allosteric residues (**Figure 20 & Table 5**). Therefore, contact with UDP-GlcNAc at this site may similarly impact the binding and catalysis of the *P. alvei* MnaA enzyme. Velloso *et al*. (2008) hypothesize that binding of the allosteric effector prevents the more thermodynamically stable reaction intermediates, UDP and 2-acetoamidoglucal, from diffusing away from the catalytic site during the epimerization reaction (Velloso, Bhaskaran *et al.*, 2008). Further, binding at the allosteric site may shield the catalytic site from solvent in the closed catalytic state of the enzyme and thereby prevent unproductive reactions from occurring between water and bound substrate (Velloso, Bhaskaran *et al.*, 2008, Chen, Huang *et al.*, 2014, Mann, Muller *et al.*, 2016). The next step in verifying these possibilities would be to conduct kinetic and structural studies of substrate-bound enzyme. However, this is challenging, and at the time of writing, there remains no structural insight into *how* the less-stable UDP-GlcNAc or UDP-ManNAc substrates are able to depart the catalytic site, while the more-stable 2-acetoamidoglucal and UDP intermediates remain (Velloso, Bhaskaran *et al.*, 2008).

To improve understanding of the *P. alvei* MnaA epimerization reaction mechanism, we attempted structural characterization by X-ray crystallographic methods. One goal was to obtain MnaA co-crystal structures in complex with the UDP-GlcNAc substrate. This structure could verify whether *P. alvei* MnaA undergoes the same conformational change upon UDP-GlcNAc binding to the allosteric site as observed in other published UDP-GlcNAc 2-epimerase structures (Velloso, Bhaskaran *et al.*, 2008, Chen, Huang *et al.*, 2016, Mann, Muller *et al.*, 2016). Unfortunately, co-crystallization of MnaA and UDP-GlcNAc has not yet yielded crystals. The next strategy involved soaking experiments using crystals obtained from the optimized unliganded

MnaA crystallization condition [see Results, section 3.2.2. MnaA crystal ligand soaking, *supra*, for a detailed description of soaking conditions]. Most soaking trials shattered the crystal lattice and did not yield specimens amenable to X-ray diffraction. If *P. alvei* MnaA substrate- and effector- binding is at all analogous to that of *B. anthracis* MnaA, crystal dissolution could be due to UDP-GlcNAc binding of the allosteric site and a concomitant large conformational change in the protein. This theory, while likely based on known MnaA structures, remains uncertain and awaits characterization of an intact, substrate-bound *P. alvei* MnaA structure.

3.3.2. Future work & liganded structures

Obtaining liganded co-crystal structures of *P. alvei* MnaA is an important step in understanding the molecular recognition of substrate in UDP-GlcNAc 2-epimerases. Many aspects of this process remain uncertain. For example, it is unclear how the allosteric site discriminates between epimers UDP-GlcNAc and UDP-ManNAc while the catalytic site does not. In published UDP-GlcNAc 2-epimerase structures, and presumably for *P. alvei* MnaA as well, the allosteric site exclusively and specifically binds UDP-GlcNAc in order to bring about a conformational shift that activates the protein. When UDP-ManNAc alone is available, the enzyme is inactive (Velloso, Bhaskaran *et al.,* 2008). Either UDP-ManNAc does not bind to the allosteric site or, if some level of binding occurs, the substrate does not induce the conformational shift necessary for enzyme activation. By contrast, the catalytic site does not discriminate between the epimers and can interconvert UDP-GlcNAc and UDP-ManNAc (meaning the catalytic site can recognize both) (Tanner 2002, Velloso, Bhaskaran *et al.,* 2008).

Since UDP-GlcNAc ligand soaks did not yield usable MnaA crystals, future experiments should use co-crystallization techniques. Velloso *et al.* (2008) had moderate success applying this approach to a UDP-GlcNAc 2-epimerase, though their resultant structures show UDP-GlcNAc bound only in the allosteric site (Chen, Huang *et al.,* 2016, Mann, Muller *et al.,* 2016). As with many other epimerase structures, the catalytic site contains UDP rather than intact substrate (Chen, Huang *et al.,* 2016, Mann, Muller *et al.,* 2016). Obtaining such a structure could tell us whether *P. alvei* MnaA binds UDP-GlcNAc at a conserved allosteric site, precisely how UDP-GlcNAc is recognized, and yield insight as to whether this binding causes the open conformation (observed in the unliganded structure) to shift to a closed conformation.

Substrate analogs also can enable structural studies examining substrate recognition and catalytic mechanisms in these enzymes. Mann *et al.* (2016) found that the antibiotic tunicamycin, a GT inhibitor that is structurally related to UDP-GlcNAc, also inhibits MnaA activity. Although the method of inhibition in MnaA is structurally undefined, tunicamycin has been shown to bind (through non-competitive interaction) to the active sites of other bacterial wall-teichoic biosynthetic enzymes, such as TarO (Hancock, Wiseman *et al.*, 1976) and MraY (Ikeda, Wachi *et al.*, 1991, Campbell, Singh *et al.*, 2011), and could presumably function *via* a similar inhibitory mechanism in UDP-GlcNAc 2-epimerases. Therefore, as non-competitive inhibition may function to help retain the intact substrate within the enzyme catalytic site, tunicamycin (or other substrate analogs (Al-Rawi, Hinderlich *et al.*, 2004, Stolz, Reiner *et al.*, 2004, Xu, Brenning *et al.*, 2013)) could be explored to generate a co-crystal structure of MnaA with a sugar bound at the catalytic site if co-crystallization with UDP-GlcNAc, alone, results in no sugar at the catalytic site.

3.4. Conclusions

The *P. alvei* MnaA unliganded crystal structure reveals that the enzyme adopts an open conformation characterized by an accessible cleft between the N- and C-terminal domains. Conservation of residues critical in binding substrate and allosteric activator UDP-GlcNAc suggests that *P. alvei* MnaA activity may be regulated in a similar fashion. Confirming this hypothesis will require further structural studies, including the characterization of intact substrate-bound enzymes.

3.5. Methods

3.5.1. MnaA overexpression and purification

Recombinant *P. alvei* MnaA epimerase (**Figure 22**), consisting of amino acids 1-384 with a Leu-Glu linker followed by a C-terminal His_6 tag, was expressed as previously described (Hager, Lopez-Guzman *et al.*, 2018). The gene for MnaA was inserted into the pET22b(+) expression vector (Novagen) and then heat-shock-transformed into *E. coli* BL21 Star™ (DE3) (Thermo Fisher Scientific, Waltham, MA, USA) for overexpression at 37°C with 200 rpm shaking in selective LB broth medium (10 g/L tryptone, 5 g/L yeast extract, 10 g/L NaCl) supplemented with 100 μg/mL ampicillin. Protein overexpression was induced at mid-exponential growth phase (OD_{600} ~0.6) *via* the addition of isopropyl-β-D-thiogalactopyranosid (IPTG) to a final concentration of 0.6 mM. Cells were incubated 4 additional hours after IPTG induction, and subsequently harvested by centrifugation at 6,500 rpm in a Beckmann JA-14 rotor (Beckman Coulter, Brea, CA, USA) for 25 min.

Figure 22*. MnaA expression construct protein sequence*

MSKVKVMTVF GVRPEAIKMA PLILELQKHP EHIESIVCVT AQHRQMLDQV LDVFNIKPDY
DLDIMQARQT LNDISVRVLQ GLEPVLQEAK PDIVLVHGDT LTTFLASYAA FMQQIKVGHV
EAGLRTWNKL SPFPEEMNRQ LTGVLADLHF APTDWSASNL RKENKQDASI FVTGNTVTDV
FQYTVRSDYK HEVLDWAQGK RLVLMTAHRR ESQGEPHRNI FRAVRRIADE FEDIAIVYPV
HPSPAVREPA HEMLGDHPRI KLIDPLDVVD LHNFYPHTHL ILTDSGGLQE EAPSFGIPVL
VLRETTERPE GIDAGTLELV GTDEEKVYAR AKALLSDETT YARMSKAANP YGDGKASQRI
VSAILHSFGV LEERPEPFHT KFTNLEHHHH HH

Amino acid sequence of the MnaA construct used in this study. MnaA residues (1-384) are indicated in black font. The Leu-Gly linker and C-terminal His_6 tag (derived from the pET22b expression vector) are indicated in grey font.

The cell pellet obtained from 3 L of overexpression culture was resuspended in 100 mL of lysis buffer (5% v/v glycerol, 20 mM imidazole, 1 M NaCl, 0.2 M arginine, 50 mM NaH_2PO_4, pH 7.4) supplemented with a SIGMAFAST™ Protease Inhibitor Tablet (Sigma-Aldrich, St. Louis, MO, USA), 1 mL of 17 mg/mL phenylmethylsulfonyl fluoride (PMSF) in isopropanol, 2 mM tris(2-carboxyethyl)phosphine (TCEP), and 3 mg of DNase I. The resuspended cells were disrupted using a Misonix Sonicator 3000 (5 min total time, 10 s on, 10 s off, initial power output 6), and Triton® X-100 was added to a final concentration of 0.1% v/v after sonication. The cell lysate was centrifuged at 6,500 rpm for 25 min in a Beckmann JA-17 rotor (Beckman Coulter, Brea, CA, USA), and the resultant supernatant was removed and centrifuged again at 10,000 rpm for 60 min to further remove cell debris. The final supernatant fraction (cell crude extract) was incubated for 30 min at 4 °C with 3 mL of Thermo Scientific™ HisPur™ Ni-NTA resin (Thermo Fisher Scientific, Waltham, MA, USA) pre-equilibrated in lysis buffer. The lysate and Ni-NTA mixture was loaded onto a column and washed with 50 mL of wash buffer (5% v/v glycerol, 25 mM imidazole, 1 M NaCl, 50 mM NaH_2PO_4, 2 mM TCEP, pH 7.4). Recombinant MnaA protein was eluted using a gradient of wash buffer (50 mL) and elution buffer (50 mL; 5% v/v glycerol, 250 mM imidazole, 1 M NaCl, 50 mM NaH_2PO_4, 2 mM TCEP, pH 7.4) and collected at a flow rate of 1 mL/min in 1.5 mL fractions. Fraction were analyzed for purity by SDS-PAGE gel (12.5% acrylamide) stained in Coomassie Brilliant Blue. Those fractions determined to be of the highest purity were pooled and dialyzed into new buffer.

3.5.2. MnaA crystallization

The recombinant, purified and pooled MnaA sample was dialyzed into 4 L of dialysis buffer (20 mM HEPES, pH 7.5) for 4 hours at 4 °C with gentle stirring using 3.5K MWCO Thermo Scientific™ SnakeSkin™ dialysis tubing (Thermo Fisher Scientific, Waltham, MA, USA). Dialyzed sample was then concentrated to 17 mg/mL in a 3K MWCO Pall Macrosep® Advance Centrifugal Device (Pall Corporation, Port Washington, NY, USA) *via* centrifugation at 5,000 rpm in a JA-17 rotor kept at 4 °C. Protein concentration was measured by A_{280} on a NanoDrop ND-1000 spectrophotometer (Thermo Fisher Scientific, Waltham, MA, USA).

The concentrated MnaA sample was filtered through a 0.22 μm centrifugal filter minutes prior to set-up of crystallization conditions. Crystal screens of the recombinant purified MnaA (17 mg/mL) were prepared using an Art Robbins Instrument crystal gryphon robot and Hampton 96-well Intelli plates. Crystallization plates were stored at 18 °C. Initial crystals of MnaA were obtained from PEGs II condition number 36 (0.2 M $CaCl_2$, 0.1 M tris pH 8.5, 20% (w/v) PEG 4000) 1 week after plating. This condition was further optimized by raising the pH and changing the concentration of $CaCl_2$ and PEG 4000 in the condition (optimized condition: 0.16 M $CaCl_2$, 0.1 M tris pH 8.8, and 15% PEG 4000). Larger crystallization drops were also prepared (1.5 μL mother liquor to 1.5 μL 17 mg/mL MnaA) *via* hanging drop vapor diffusion at 18 °C in order to generate larger crystals of higher quality for use in subsequent X-ray diffraction experiments and data collection.

3.5.3. Data collection, structure solution, and refinement

A crystal obtained by hanging drop vapor diffusion from the optimization of PEGs II condition 36 was resuspended in a mother liquor solution containing 20% MPD for 10-20 seconds

prior to being flash-frozen at 100°K for data collection. X-ray diffraction data were collected with a Rigaku MicroMax-007HF X-ray generator coupled to a Rigaku Dectris Pilatus3R 200K-A detector. HKL2000 (Otwinowski and Minor 1997) was used to scale, average, and integrate the data. The data set was solved by molecular replacement using a model of MnaA predicted by Phyre2 (Kelley, Mezulis *et al.,* 2015) and PHASER from the CCP4 program suite (McCoy, Grosse-Kunstleve *et al.,* 2007, Winn, Ballard *et al.,* 2011). Coot (Emsley, Lohkamp *et al.,* 2010) and Refmac5 from the CCP4 (Murshudov, Skubak *et al.,* 2011, Winn, Ballard *et al.,* 2011) suite were used for all subsequent model building and refinement steps.

3.5.4. MnaA crystal ligand soaking

MnaA crystals grown by hanging drop vapor diffusion in optimized PEGs II condition 36 (0.16 M $CaCl_2$, 0.1 M tris pH 8.8, and 15% PEG 4000) were soaked overnight at 18 °C in mother liquor solution with additions of either 10 mM or 50 mM α-D-GlcNAc (Sigma-Aldrich, St. Louis, MO, USA).

3.5.5. Visualization and graphics

All figures depicting protein structures and surfaces were created using the UCSF Chimera Extensible Molecular Modeling System, developed by the Resource for Biocomputing, Visualization, and Informatics at the University of California, San Francisco, with support from NIH P41-GM103311 (Pettersen, Goddard *et al.,* 2004).

References

Al-Rawi, S., S. Hinderlich, W. Reutter and A. Giannis (2004). "Synthesis and biochemical properties of reversible inhibitors of UDP-N-acetylglucosamine 2-epimerase." Angew Chem Int Ed Engl **43**(33): 4366-4370.

Albers, S. V. and B. H. Meyer (2011). "The archaeal cell envelope." Nat Rev Microbiol **9**(6): 414-426.

Altman, E., J. R. Brisson, P. Messner and U. B. Sleytr (1990). "Chemical characterization of the regularly arranged surface layer glycoprotein of *Clostridium thermosaccharolyticum* D120-70." Eur J Biochem **188**(1): 73-82.

Altman, E., C. Schaffer, J. R. Brisson and P. Messner (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.

Antikainen, J., L. Anton, J. Sillanpaa and T. K. Korhonen (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**(2): 381-394.

Araki, Y. and E. Ito (1989). "Linkage units in cell walls of gram-positive bacteria." Crit Rev Microbiol **17**(2): 121-135.

Archibald, A. R., J. Baddiley and N. L. Blumsom (1968). "The teichoic acids." Adv Enzymol Relat Areas Mol Biol **30**: 223-253.

Ash, C., F. G. Priest and M. D. Collins (1993). "Molecular identification of rRNA group 3 bacilli (Ash, Farrow, Wallbanks and Collins) using a PCR probe test. Proposal for the creation of a new genus *Paenibacillus*." Antonie Van Leeuwenhoek **64**(3-4): 253-260.

Awram, P. and J. Smit (2001). "Identification of lipopolysaccharide O antigen synthesis genes required for attachment of the S-layer of *Caulobacter crescentus*." Microbiology **147**(6): 1451-1460.

Badger, J., J. M. Sauder, J. M. Adams, S. Antonysamy, K. Bain, M. G. Bergseid, S. G. Buchanan, M. D. Buchanan, Y. Batiyenko, J. A. Christopher, S. Emtage, A. Eroshkina, I. Feil, E. B. Furlong, K. S. Gajiwala, X. Gao, D. He, J. Hendle, A. Huber, K. Hoda, P. Kearins, C. Kissinger, B. Laubert, H. A. Lewis, J. Lin, K. Loomis, D. Lorimer, G. Louie, M. Maletic, C. D. Marsh, I. Miller, J. Molinari, H. J. Muller-Dieckmann, J. M. Newman, B. W. Noland, B. Pagarigan, F. Park, T. S. Peat, K. W. Post, S. Radojicic, A. Ramos, R. Romero, M. E. Rutter, W. E. Sanderson, K. D. Schwinn, J. Tresser, J. Winhoven, T. A. Wright, L. Wu, J. Xu and T. J. Harris (2005). "Structural analysis of a set of proteins resulting from a bacterial genomics project." Proteins **60**(4): 787-796.

Baillie, L., R. Hebdon, H. Flick-Smith and D. Williamson (2003). "Characterisation of the immune response to the UK human anthrax vaccine." FEMS Immunol Med Microbiol **36**(1-2): 83-86.

Baranova, E., R. Fronzes, A. Garcia-Pino, N. Van Gerven, D. Papapostolou, G. Pehau-Arnaudet, E. Pardon, J. Steyaert, S. Howorka and H. Remaut (2012). "SbsB structure and lattice reconstruction unveil Ca2+ triggered S-layer assembly." Nature **487**(7405): 119-122.

Bayer, E. A., J. P. Belaich, Y. Shoham and R. Lamed (2004). "The cellulosomes: multienzyme machines for degradation of plant cell wall polysaccharides." Annu Rev Microbiol **58**: 521-554.

Berman, H. M., J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov and P. E. Bourne (2000). "The Protein Data Bank." Nucleic Acids Res **28**(1): 235-242.

Beveridge, T. J. (1981). "Ultrastructure, chemistry, and function of the bacterial wall." Int Rev Cytol **72**: 229-317.

Beveridge, T. J. (1988). "The bacterial surface: general considerations towards design and function." Can J Microbiol **34**(4): 363-372.

Beveridge, T. J. and R. G. Murray (1974). "Superficial macromolecular arrays on the cell wall of *Spirillum putridiconchylium*." J Bacteriol **119**(3): 1019-1038.

Bharat, T. A. M., A. von Kugelgen and V. Alva (2021). "Molecular Logic of Prokaryotic Surface Layer Structures." Trends Microbiol **29**(5): 405-415.

Bhavsar, A. P., L. K. Erdman, J. W. Schertzer and E. D. Brown (2004). "Teichoic acid is an essential polymer in *Bacillus subtilis* that is functionally distinct from teichuronic acid." J Bacteriol **186**(23): 7865-7873.

Blackler, R. J., A. Lopez-Guzman, F. F. Hager, B. Janesch, G. Martinz, S. M. L. Gagnon, O. Haji-Ghassemi, P. Kosma, P. Messner, C. Schaffer and S. V. Evans (2018). "Structural basis of cell wall anchoring by SLH domains in *Paenibacillus alvei*." Nat Commun **9**(1): 3120.

Brinton, C., J. McNary and J. Carnaham (1969). Purification and *in vitro* assembly of a curved network of identical protein subunits from the outer surface of a Bacillus. Bact Proc **48**.

Brown, S., J. P. Santa Maria, Jr. and S. Walker (2013). "Wall teichoic acids of gram-positive bacteria." Annu Rev Microbiol **67**: 313-336.

Buckmire, F. L. and R. G. Murray (1970). "Studies on the cell wall of *Spirillum serpens*. 1. Isolation and partial purification of the outermost cell wall layer." Can J Microbiol **16**(10): 1011-1022.

Buckmire, F. L. and R. G. Murray (1973). "Studies on the cell wall of *Spirillum serpens*. II. Chemical characterization of the outer structured layer." Can J Microbiol **19**(1): 59-66.

Calabi, E., F. Calabi, A. D. Phillips and N. F. Fairweather (2002). "Binding of *Clostridium difficile* surface layer proteins to gastrointestinal tissues." Infect Immun **70**(10): 5770-5778.

Calabi, E., S. Ward, B. Wren, T. Paxton, M. Panico, H. Morris, A. Dell, G. Dougan and N. Fairweather (2001). "Molecular characterization of the surface layer proteins from *Clostridium difficile*." Mol Microbiol **40**(5): 1187-1199.

Campbell, J., A. K. Singh, J. P. Santa Maria, Jr., Y. Kim, S. Brown, J. G. Swoboda, E. Mylonakis, B. J. Wilkinson and S. Walker (2011). "Synthetic lethal compound combinations reveal a fundamental connection between wall teichoic acid and peptidoglycan biosyntheses in *Staphylococcus aureus*." ACS Chem Biol **6**(1): 106-116.

Campbell, R. E., S. C. Mosimann, M. E. Tanner and N. C. Strynadka (2000). "The structure of UDP-N-acetylglucosamine 2-epimerase reveals homology to phosphoglycosyl transferases." Biochemistry **39**(49): 14993-15001.

Carl, M. and G. A. Dasch (1989). "The importance of the crystalline surface layer protein antigens of rickettsiae in T-cell immunity." J Autoimmun **2 Suppl**: 81-91.

Cava, F., M. A. de Pedro, H. Schwarz, A. Henne and J. Berenguer (2004). "Binding to pyruvylated compounds as an ancestral mechanism to anchor the outer envelope in primitive bacteria." Mol Microbiol **52**(3): 677-690.

Chauvaux, S., M. Matuschek and P. Beguin (1999). "Distinct affinity of binding sites for S-layer homologous domains in *Clostridium thermocellum* and *Bacillus anthracis* cell envelopes." J Bacteriol **181**(8): 2455-2458.

Chen, S. C., C. H. Huang, S. J. Lai, C. S. Yang, T. H. Hsiao, C. H. Lin, P. K. Fu, T. P. Ko and Y. Chen (2016). "Mechanism and inhibition of human UDP-GlcNAc 2-epimerase, the key enzyme in sialic acid biosynthesis." Sci Rep **6**: 23274.

Chen, S. C., C. H. Huang, C. S. Yang, J. S. Liu, S. M. Kuan and Y. Chen (2014). "Crystal structures of the archaeal UDP-GlcNAc 2-epimerase from *Methanocaldococcus jannaschii* reveal a conformational change induced by UDP-GlcNAc." Proteins **82**(7): 1519-1526.

Claus, H., E. Akca, T. Debaerdemaeker, C. Evrard, J. P. Declercq, J. R. Harris, B. Schlott and H. Konig (2005). "Molecular organization of selected prokaryotic S-layer proteins." Can J Microbiol **51**(9): 731-743.

Djukic, M., D. Becker, A. Poehlein, S. Voget and R. Daniel (2012). "Genome sequence of *Paenibacillus alvei* DSM 29, a secondary invader during European foulbrood outbreaks." J Bacteriol **194**(22): 6365.

Egelseer, E. M., T. Danhorn, M. Pleschberger, C. Hotzy, U. B. Sleytr and M. Sara (2001). "Characterization of an S-layer glycoprotein produced in the course of S-layer variation of *Bacillus stearothermophilus* ATCC 12980 and sequencing and cloning of the sbsD gene encoding the protein moiety." Arch Microbiol **177**(1): 70-80.

El-Gebali, S., J. Mistry, A. Bateman, S. R. Eddy, A. Luciani, S. C. Potter, M. Qureshi, L. J. Richardson, G. A. Salazar, A. Smart, E. L. L. Sonnhammer, L. Hirsh, L. Paladin, D. Piovesan, S.

C. E. Tosatto and R. D. Finn (2019). "The Pfam protein families database in 2019." Nucleic Acids Res **47**(D1): D427-D432.

Ellwood, D. C. and D. W. Tempest (1969). "Control of teichoic acid and teichuronic acid biosyntheses in chemostat cultures of *Bacillus subtilis* var. niger." Biochem J **111**(1): 1-5.

Emsley, P., B. Lohkamp, W. G. Scott and K. Cowtan (2010). "Features and development of Coot." Acta Crystallogr D Biol Crystallogr **66**(Pt 4): 486-501.

Engelhardt, H. and J. Peters (1998). "Structural research on surface layers: a focus on stability, surface layer homology domains, and surface layer-cell wall interactions." Journal of Structural Biology **124**: 276-302.

Fagan, R. P. and N. F. Fairweather (2014). "Biogenesis and functions of bacterial S-layers." Nat Rev Microbiol **12**(3): 211-222.

Fagan, R. P. and N. F. Fairweather (2014). "Biogenesis and functions of bacterial S-layers." Nature Reviews in Microbiology **12**(3): 211-222.

Fioravanti, A., F. Van Hauwermeiren, S. E. Van der Verren, W. Jonckheere, A. Goncalves, E. Pardon, J. Steyaert, H. De Greve, M. Lamkanfi and H. Remaut (2019). "Structure of S-layer protein Sap reveals a mechanism for therapeutic intervention in anthrax." Nat Microbiol **4**(11): 1805-1814.

Forsgren, E. (2010). "European foulbrood in honey bees." J Invertebr Pathol **103 Suppl 1**: S5-9.

Fujimoto, S., A. Takade, K. Amako and M. J. Blaser (1991). "Correlation between molecular size of the surface array protein and morphology and antigenicity of the *Campylobacter fetus* S layer." Infect Immun **59**(6): 2017-2022.

Funfhaus, A. and E. Genersch (2012). "Proteome analysis of *Paenibacillus larvae* reveals the existence of a putative S-layer protein." Environ Microbiol Rep **4**(2): 194-202.

Genersch, E., E. Forsgren, J. Pentikainen, A. Ashiralieva, S. Rauch, J. Kilwinski and I. Fries (2006). "Reclassification of *Paenibacillus larvae* subsp. *pulvifaciens* and *Paenibacillus larvae* subsp. *larvae* as *Paenibacillus larvae* without subspecies differentiation." Int J Syst Evol Microbiol **56**(Pt 3): 501-511.

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

Glaubert, A. M. and U. B. Sleytr (1975). "Analysis of regular arrays of subunits on bacterial surfaces: evidence for a dynamic process of assembly." J Ultrastruct Res **50**(1): 103-116.

Goel, A. K. (2015). "Anthrax: A disease of biowarfare and public health importance." World J Clin Cases **3**(1): 20-33.

Grossar, D., V. Kilchenmann, E. Forsgren, J. D. Charriere, L. Gauthier, M. Chapuisat and V. Dietemann (2020). "Putative determinants of virulence in *Melissococcus plutonius*, the bacterial agent causing European foulbrood in honey bees." Virulence **11**(1): 554-567.

Hager, F. F., A. Lopez-Guzman, S. Krauter, M. Blaukopf, M. Polter, I. Brockhausen, P. Kosma and C. Schaffer (2018). "Functional Characterization of Enzymatic Steps Involved in Pyruvylation of Bacterial Secondary Cell Wall Polymer Fragments." Front Microbiol **9**: 1356.

Hancock, I. C., G. Wiseman and J. Baddiley (1976). "Biosynthesis of the unit that links teichoic acid to the bacterial wall: inhibition by tunicamycin." FEBS Lett **69**(1): 75-80.

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

Houwink, A. L. (1953). "A macromolecular mono-layer in the cell wall of *Spirillum spec*." Biochim Biophys Acta **10**(3): 360-366.

Houwink, A. L. (1956). "Flagella, gas vacuoles and cell-wall structure in *Halobacterium halobium*; an electron microscope study." J Gen Microbiol **15**(1): 146-150.

Huber, C., N. Ilk, D. Runzler, E. M. Egelseer, S. Weigert, U. B. Sleytr and M. Sara (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**(1): 197-205.

Hurlburt, N. K., J. Guan, H. Ong, H. Yu, X. Chen and A. J. Fisher (2020). "Structural characterization of a nonhydrolyzing UDP-GlcNAc 2-epimerase from *Neisseria meningitidis* serogroup A." Acta Crystallogr F Struct Biol Commun **76**(Pt 11): 557-567.

Hynonen, U. and A. Palva (2013). "*Lactobacillus* surface layer proteins: structure, function and applications." Appl Microbiol Biotechnol **97**(12): 5225-5243.

Ikeda, M., M. Wachi, H. K. Jung, F. Ishino and M. Matsuhashi (1991). "The *Escherichia coli* mraY gene encoding UDP-N-acetylmuramoyl-pentapeptide: undecaprenyl-phosphate phospho-N-acetylmuramoyl-pentapeptide transferase." J Bacteriol **173**(3): 1021-1026.

Ilk, N., P. Kosma, M. Puchberger, E. M. Egelseer, H. F. Mayer, U. B. Sleytr and M. Sara (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**(24): 7643-7646.

Janesch, B., A. Koerdt, P. Messner and C. Schaffer (2013). "The S-layer homology domain-containing protein SlhA from *Paenibacillus alvei* CCM 2051(T) is important for swarming and biofilm formation." PLoS One **8**(9): e76566.

Janesch, B., P. Messner and C. Schäffer (2013). "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**(3): 565-575.

Jernigan, J. A., D. S. Stephens, D. A. Ashford, C. Omenaca, M. S. Topiel, M. Galbraith, M. Tapper, T. L. Fisk, S. Zaki, T. Popovic, R. F. Meyer, C. P. Quinn, S. A. Harper, S. K. Fridkin, J. J. Sejvar, C. W. Shepard, M. McConnell, J. Guarner, W. J. Shieh, J. M. Malecki, J. L. Gerberding, J. M. Hughes, B. A. Perkins and T. Anthrax Bioterrorism Investigation (2001). "Bioterrorism-related inhalational anthrax: the first 10 cases reported in the United States." Emerg Infect Dis **7**(6): 933-944.

Kattke, M. D., J. E. Gosschalk, O. E. Martinez, G. Kumar, R. T. Gale, D. Cascio, M. R. Sawaya, M. Philips, E. D. Brown and R. T. Clubb (2019). "Structure and mechanism of TagA, a novel membrane-associated glycosyltransferase that produces wall teichoic acids in pathogenic bacteria." PLoS Pathog **15**(4): e1007723.

Kawai, E., H. Akatsuka, A. Idei, T. Shibatani and K. Omori (1998). "*Serratia marcescens* S-layer protein is secreted extracellularly via an ATP-binding cassette exporter, the Lip system." Mol Microbiol **27**(5): 941-952.

Kelley, L. A., S. Mezulis, C. M. Yates, M. N. Wass and M. J. Sternberg (2015). "The Phyre2 web portal for protein modeling, prediction and analysis." Nat Protoc **10**(6): 845-858.

Kern, J., R. Wilton, R. Zhang, T. A. Binkowski, A. Joachimiak and O. Schneewind (2011). "Structure of surface layer homology (SLH) domains from *Bacillus anthracis* surface array protein." J Biol Chem **286**(29): 26042-26049.

Kessel, M., I. Wildhaber, S. Cohen and W. Baumeister (1988). "Three-dimensional structure of the regular surface glycoprotein layer of *Halobacterium volcanii* from the Dead Sea." EMBO J **7**(5): 1549-1554.

Kirk, J. A., D. Gebhart, A. M. Buckley, S. Lok, D. Scholl, G. R. Douce, G. R. Govoni and R. P. Fagan (2017). "New class of precision antimicrobials redefines role of *Clostridium difficile* S-layer in virulence and viability." Sci Transl Med **9**(406).

König, H., H. Claus and A. Varma (2010). "Prokaryotic cell wall compounds – structure and Biochemistry." Springer, Berlin.

Krauter, S., C. Schäffer and P. Kosma (2021). "Synthesis of a pyruvylated *N*-acetyl-β-D-mannosamine containing disaccharide repeating unit of a cell wall glycopolymer from *Paenibacillus alvei* " Arkivoc: 137-151.

Lahooti, M. and C. R. Harwood (1999). "Transcriptional analysis of the *Bacillus subtilis* teichuronic acid operon." Microbiology (Reading) **145 (Pt 12)**: 3409-3417.

Lang, W. K., K. Glassey and A. R. Archibald (1982). "Influence of phosphate supply on teichoic acid and teichuronic acid content of *Bacillus subtilis* cell walls." J Bacteriol **151**(1): 367-375.

Lemaire, M., I. Miras, P. Gounon and P. Beguin (1998). "Identification of a region responsible for binding to the cell wall within the S-layer protein of *Clostridium thermocellum*." Microbiology (Reading) **144 (Pt 1)**: 211-217.

Liu, S. Y., F. C. Gherardini, M. Matuschek, H. Bahl and J. Wiegel (1996). "Cloning, sequencing, and expression of the gene encoding a large S-layer-associated endoxylanase from *Thermoanaerobacterium* sp. strain JW/SL-YS 485 in *Escherichia coli*." J Bacteriol **178**(6): 1539-1547.

Lupas, A., H. Engelhardt, J. Peters, U. Santarius, S. Volker and W. Baumeister (1994). "Domain structure of the *Acetogenium kivui* surface layer revealed by electron crystallography and sequence analysis." J Bacteriol **176**(5): 1224-1233.

Mader, C., C. Huber, D. Moll, U. B. Sleytr and M. Sara (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**(6): 1758-1768.

Mann, P. A., A. Muller, K. A. Wolff, T. Fischmann, H. Wang, P. Reed, Y. Hou, W. Li, C. E. Muller, J. Xiao, N. Murgolo, X. Sher, T. Mayhood, P. R. Sheth, A. Mirza, M. Labroli, L. Xiao, M. McCoy, C. J. Gill, M. G. Pinho, T. Schneider and T. Roemer (2016). "Chemical Genetic Analysis and Functional Characterization of Staphylococcal Wall Teichoic Acid 2-Epimerases Reveals Unconventional Antibiotic Drug Targets." PLoS Pathog **12**(5): e1005585.

Martinez, O. E., B. J. Mahoney, A. K. Goring, S. W. Yi, D. P. Tran, D. Cascio, M. L. Phillips, M. M. Muthana, X. Chen, M. E. Jung, J. A. Loo and R. T. Clubb (2021). "Insight into the molecular basis of substrate recognition by the wall teichoic acid glycosyltransferase TagA." J Biol Chem **298**(2): 101464.

McCoy, A. J., R. W. Grosse-Kunstleve, P. D. Adams, M. D. Winn, L. C. Storoni and R. J. Read (2007). "Phaser crystallographic software." J Appl Crystallogr **40**(Pt 4): 658-674.

Mesnage, S., T. Fontaine, T. Mignot, M. Delepierre, M. Mock and A. Fouet (2000). "Bacterial SLH domain proteins are non-covalently anchored to the cell surface via a conserved mechanism involving wall polysaccharide pyruvylation." Embo Journal **19**(17): 4473-4484.

Messner, P., C. Schaffer and P. Kosma (2013). "Bacterial cell-envelope glycoconjugates." Adv Carbohydr Chem Biochem **69**: 209-272.

Messner, P. and U. B. Sleytr (1992). "Crystalline bacterial cell-surface layers." Adv Microb Physiol **33**: 213-275.

Messner, P., U. B. Sleytr, R. Christian, G. Schulz and F. M. Unger (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**(2): 211-218.

Mignot, T., B. Denis, E. Couture-Tosi, A. B. Kolsto, M. Mock and A. Fouet (2001). "Distribution of S-layers on the surface of *Bacillus cereus* strains: phylogenetic origin and ecological pressure." Environ Microbiol **3**(8): 493-501.

Murshudov, G. N., P. Skubak, A. A. Lebedev, N. S. Pannu, R. A. Steiner, R. A. Nicholls, M. D. Winn, F. Long and A. A. Vagin (2011). "REFMAC5 for the refinement of macromolecular crystal structures." Acta Crystallogr D Biol Crystallogr **67**(Pt 4): 355-367.

Naumova, I. B. and A. S. Shashkov (1997). "Anionic polymers in cell walls of gram-positive bacteria." Biochemistry (Mosc) **62**(8): 809-840.

Oatley, P., J. A. Kirk, S. Ma, S. Jones and R. P. Fagan (2020). "Spatial organization of *Clostridium difficile* S-layer biogenesis." Sci Rep **10**(1): 14089.

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

Peng, Z., A. Addisu, S. Alrabaa and X. Sun (2017). "Antibiotic Resistance and Toxin Production of *Clostridium difficile* Isolates from the Hospitalized Patients in a Large Hospital in Florida." Front Microbiol **8**: 2584.

Pettersen, E. F., T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng and T. E. Ferrin (2004). "UCSF Chimera--a visualization system for exploratory research and analysis." J Comput Chem **25**(13): 1605-1612.

Poppinga, L., B. Janesch, A. Funfhaus, G. Sekot, E. Garcia-Gonzalez, G. Hertlein, K. Hedtke, C. Schaffer and E. Genersch (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**(5): e1002716.

Rasmussen-Ivey, C. R., M. J. Figueras, D. McGarey and M. R. Liles (2016). "Virulence Factors of Aeromonas hydrophila: In the Wake of Reclassification." Front Microbiol **7**: 1337.

Ravi, J. and A. Fioravanti (2021). "S-layers: The Proteinaceous Multifunctional Armors of Gram-Positive Pathogens." Front Microbiol **12**: 663468.

Ries, W., C. Hotzy, I. Schocher, U. B. Sleytr and M. Sara (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**(12): 3892-3898.

Rodrigues-Oliveira, T., A. Belmok, D. Vasconcellos, B. Schuster and C. M. Kyaw (2017). "Archaeal S-Layers: Overview and Current State of the Art." Front Microbiol **8**: 2597.

Runzler, D., C. Huber, D. Moll, G. Kohler and M. Sara (2004). "Biophysical characterization of the entire bacterial surface layer protein SbsB and its two distinct functional domains." J Biol Chem **279**(7): 5207-5215.

Sapp, J. (2005). "The prokaryote-eukaryote dichotomy: meanings and mythology." Microbiol Mol Biol Rev **69**(2): 292-305.

Sara, M. (2001). "Conserved anchoring mechanisms between crystalline cell surface S-layer proteins and secondary cell wall polymers in Gram-positive bacteria?" Trends Microbiol **9**(2): 47-49; discussion 49-50.

Sara, M. and U. B. Sleytr (1994). "Comparative studies of S-layer proteins from *Bacillus stearothermophilus* strains expressed during growth in continuous culture under oxygen-limited and non-oxygen-limited conditions." J Bacteriol **176**(23): 7182-7189.

Sara, M. and U. B. Sleytr (2000). "S-Layer proteins." J Bacteriol **182**(4): 859-868.

Schaffer, C., H. Kahlig, R. Christian, G. Schulz, S. Zayni and P. Messner (1999). "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 (Reading) **145 (Pt 7)**: 1575-1583.

Schaffer, C. and P. Messner (2004). "Surface-layer glycoproteins: an example for the diversity of bacterial glycosylation with promising impacts on nanobiotechnology." Glycobiology **14**(8): 31R-42R.

Schaffer, C. and P. Messner (2005). "The structure of secondary cell wall polymers: how Gram-positive bacteria stick their cell walls together." Microbiology (Reading) **151**(Pt 3): 643-651.

Schäffer, C. and P. Messner (2017). "Emerging facets of prokaryotic glycosylation." FEMS Microbiology Reviews **41**(1): 49-91.

Schaffer, C., N. Muller, P. K. Mandal, R. Christian, S. Zayni and P. Messner (2000). "A pyrophosphate bridge links the pyruvate-containing secondary cell wall polymer of *Paenibacillus alvei* CCM 2051 to muramic acid." Glycoconj J **17**(10): 681-690.

Schäffer, C., N. Müller, P. K. Mandal, R. Christian, S. Zayni and P. Messner (2000). "A pyrophosphate bridge links the pyruvate-containing secondary cell wall polymer of *Paenibacillus alvei* CCM 2051 to muramic acid." Glycoconj J **17**(10): 681-690.

Schaffer, C., T. Wugeditsch, H. Kahlig, A. Scheberl, S. Zayni and P. Messner (2002). "The surface layer (S-layer) glycoprotein of *Geobacillus stearothermophilus* NRS 2004/3a. Analysis of its glycosylation." J Biol Chem **277**(8): 6230-6239.

Schüttelkopf, A. W. and D. M. van Aalten (2004). "PRODRG: a tool for high-throughput crystallography of protein-ligand complexes." Acta Crystallogr D Biol Crystallogr **60**(Pt 8): 1355-1363.

Shimotahira, N., Y. Oogai, M. Kawada-Matsuo, S. Yamada, K. Fukutsuji, K. Nagano, F. Yoshimura, K. Noguchi and H. Komatsuzawa (2013). "The surface layer of *Tannerella forsythia* contributes to serum resistance and oral bacterial coaggregation." Infect Immun **81**(4): 1198-1206.

Silhavy, T. J., D. Kahne and S. Walker (2010). "The bacterial cell envelope." Cold Spring Harb Perspect Biol **2**(5): a000414.

Sleytr, U., M. Sára, D. Pum, B. Schuster, P. Messner and C. Schäffer (2002). Self-assembly protein systems: microbial S-layers. Biopolymers: polyamides and complex proteinaceous materials I, Weinheim, Germany: Wiley-VCH.

Sleytr, U. B. (1975). "Heterologous reattachment of regular arrays of glycoproteins on bacterial surfaces." Nature **257**(5525): 400-402.

Sleytr, U. B. (1976). "Self-assembly of the hexagonally and tetragonally arranged subunits of bacterial surface layers and their reattachment to cell walls." J Ultrastruct Res **55**(3): 360-377.

Sleytr, U. B. (2016). "Curiosity and passion for science and art." World Scientific **7**.

Sleytr, U. B., E.-M. Egelseer, N. Ilk, P. Messner, C. Schäffer, D. Pum and B. Schuster (2010). "Nanobiotechnological applications of S-layers *Prokaryotic Cell Wall Components – Structure and Biochemistry*." (König H, Claus H. & Varma A, eds), Springer, Heidelberg, Germany**:** 459-481.

Sleytr, U. B. and P. Messner (1983). "Crystalline surface layers on bacteria." Annu Rev Microbiol **37**: 311-339.

Sleytr, U. B. and P. Messner (1988). "Crystalline surface layers in procaryotes." J Bacteriol **170**(7): 2891-2897.

Sleytr, U. B., P. Messner, D. Pum and M. Sara (1996). "Crystalline bacterial cell surface proteins." Elsevier.

Sleytr, U. B., P. Messner, D. Pum and M. Sara (1999). "Crystalline bacterial cell surface layers (s layers): from supramolecular cell structure to biomimetics and nanotechnology." Angew Chem Int Ed Engl **38**(8): 1034-1054.

Sleytr, U. B., B. Schuster, E. M. Egelseer and D. Pum (2014). "S-layers: principles and applications." FEMS Microbiol Rev **38**(5): 823-864.

Smit, E. and P. H. Pouwels (2002). "One repeat of the cell wall binding domain is sufficient for anchoring the *Lactobacillus acidophilus* surface layer protein." J Bacteriol **184**(16): 4617-4619.

Steindl, C., C. Schaffer, T. Wugeditsch, M. Graninger, I. Matecko, N. Muller and P. Messner (2002). "The first biantennary bacterial secondary cell wall polymer and its influence on S-layer glycoprotein assembly." Biochem J **368**(Pt 2): 483-494.

Stolz, F., M. Reiner, A. Blume, W. Reutter and R. R. Schmidt (2004). "Novel UDP-glycal derivatives as transition state analogue inhibitors of UDP-GlcNAc 2-epimerase." J Org Chem **69**(3): 665-679.

Suhr, M., F. L. Lederer, T. J. Gunther, J. Raff and K. Pollmann (2016). "Characterization of Three Different Unusual S-Layer Proteins from *Viridibacillus arvi* JG-B58 That Exhibits Two Super-Imposed S-Layer Proteins." PLoS One **11**(6): e0156785.

Sychantha, D., R. N. Chapman, N. C. Bamford, G. J. Boons, P. L. Howell and A. J. Clarke (2018). "Molecular Basis for the Attachment of S-Layer Proteins to the Cell Wall of *Bacillus anthracis*." Biochemistry **57**(13): 1949-1953.

Tanner, M. E. (2002). "Understanding nature's strategies for enzyme-catalyzed racemization and epimerization." Acc Chem Res **35**(4): 237-246.

Thompson, S. A. (2002). "Campylobacter surface-layers (S-layers) and immune evasion." Ann Periodontol **7**(1): 43-53.

Usenik, A., M. Renko, M. Mihelic, N. Lindic, J. Borisek, A. Perdih, G. Pretnar, U. Muller and D. Turk (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**(3): 514-521.

Velloso, L. M., S. S. Bhaskaran, R. Schuch, V. A. Fischetti and C. E. Stebbins (2008). "A structural basis for the allosteric regulation of non-hydrolysing UDP-GlcNAc 2-epimerases." EMBO Rep **9**(2): 199-205.

von Kugelgen, A., H. Tang, G. G. Hardy, D. Kureisaite-Ciziene, Y. V. Brun, P. J. Stansfeld, C. V. Robinson and T. A. M. Bharat (2020). "In Situ Structure of an Intact Lipopolysaccharide-Bound Bacterial Surface Layer." Cell **180**(2): 348-358 e315.

Wagenaar, J. A., M. A. van Bergen, M. J. Blaser, R. V. Tauxe, D. G. Newell and J. P. van Putten (2014). "Campylobacter fetus infections in humans: exposure and disease." Clin Infect Dis **58**(11): 1579-1586.

Ward, J. B. (1981). "Teichoic and teichuronic acids: biosynthesis, assembly, and location." Microbiol Rev **45**(2): 211-243.

Watson, S. W. and C. C. Remsen (1970). "Cell envelope of *Nitrosocystis oceanus*." J Ultrastruct Res **33**(1): 148-160.

Willing, S. E., T. Candela, H. A. Shaw, Z. Seager, S. Mesnage, R. P. Fagan and N. F. Fairweather (2015). "*Clostridium difficile* surface proteins are anchored to the cell wall using CWB2 motifs that recognise the anionic polymer PSII." Mol Microbiol **96**(3): 596-608.

Winn, M. D., C. C. Ballard, K. D. Cowtan, E. J. Dodson, P. Emsley, P. R. Evans, R. M. Keegan, E. B. Krissinel, A. G. Leslie, A. McCoy, S. J. McNicholas, G. N. Murshudov, N. S. Pannu, E. A. Potterton, H. R. Powell, R. J. Read, A. Vagin and K. S. Wilson (2011). "Overview of the CCP4 suite and current developments." Acta Crystallogr D Biol Crystallogr **67**(Pt 4): 235-242.

Woese, C. R. and G. E. Fox (1977). "Phylogenetic structure of the prokaryotic domain: the primary kingdoms." Proc Natl Acad Sci U S A **74**(11): 5088-5090.

Woese, C. R., O. Kandler and M. L. Wheelis (1990). "Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya." Proc Natl Acad Sci U S A **87**(12): 4576-4579.

Xu, Q., M. G. Resch, K. Podkaminer, S. H. Yang, J. O. Baker, B. S. Donohoe, C. Wilson, D. M. Klingeman, D. G. Olson, S. R. Decker, R. J. Giannone, R. L. Hettich, S. D. Brown, L. R. Lynd, E. A. Bayer, M. E. Himmel and Y. J. Bomble (2016). "Dramatic performance of *Clostridium thermocellum* explained by its wide range of cellulase modalities." Science Advances **2**(2): e1501254.

Xu, Y., B. Brenning, A. Clifford, D. Vollmer, J. Bearss, C. Jones, V. McCarthy, C. Shi, B. Wolfe, B. Aavula, S. Warner, D. J. Bearss, M. V. McCullar, R. Schuch, A. Pelzek, S. S. Bhaskaran, C. E. Stebbins, A. R. Goldberg, V. A. Fischetti and H. Vankayalapati (2013). "Discovery of Novel Putative Inhibitors of UDP-GlcNAc 2-Epimerase as Potent Antibacterial Agents." ACS Med Chem Lett **4**(12): 1142-1147.

Yue, D., M. Nordhoff, L. H. Wieler and E. Genersch (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**(6): 1612-1620.

Zarschler, K., B. Janesch, B. Kainz, R. Ristl, P. Messner and C. Schaffer (2010). "Cell surface display of chimeric glycoproteins via the S-layer of *Paenibacillus alvei*." Carbohydrate Research **345**(10): 1422-1431.

Zarschler, K., B. Janesch, B. Kainz, R. Ristl, P. Messner and C. Schäffer (2010). "Cell surface display of chimeric glycoproteins via the S-layer of *Paenibacillus alvei*." Carbohydr Res **345**(10): 1422-1431.

Appendix A – Journal Permissions

Journal Permission for use of materials presented in Chapters 1 & 2

ELSEVIER LICENSE TERMS AND CONDITIONS

Apr 18, 2022

This Agreement between Dr. Max Legg ("You") and Elsevier ("Elsevier") consists of your license details and the terms and conditions provided by Elsevier and Copyright Clearance Center.

The publisher has provided special terms related to this request that can be found at the end of the Publisher's Terms and Conditions.

License Number	5292010741103
License date	Apr 18, 2022
Licensed Content Publisher	Elsevier
Licensed Content Publication	Journal of Biological Chemistry
Licensed Content Title	The S-layer homology domains of Paenibacillus alvei surface protein SpaA bind to cell wall polysaccharide through the terminal monosaccharide residue
Licensed Content Author	Max S.G. Legg,Fiona F. Hager-Mair,Simon Krauter,Susannah M.L. Gagnon,Arturo Lòpez-Guzmán,Charlie Lim,Markus Blaukopf,Paul Kosma,Christina Schäffer,Stephen V. Evans
Licensed Content Date	Apr 1, 2022
Licensed Content Volume	298
Licensed Content Issue	4
Licensed Content Pages	1
Start Page	101745
End Page	
Type of Use	reuse in a thesis/dissertation
Portion	full article
Circulation	1
Format	both print and electronic
Are you the author of this Elsevier article?	Yes
Will you be translating?	No
Title	Advancing understanding of secondary cell wall polymer binding and synthesis in S-layers of Gram-Positive bacteria
Institution name	University of Victoria
Expected presentation date	Apr 2022
Order reference number	123456789
Requestor Location	Dr. Max Legg University of Victoria Victoria, BC V8P 5C2 Canada Attn: Dr. Max Legg
Publisher Tax ID	GB 494 6272 12
Billing Type	Invoice
Billing Address	Dr. Max Legg University of Victoria Victoria, BC V8P 5C2 Canada Attn: Dr. Max Legg
Total	**0.00 CAD**
Terms and Conditions	

Appendix B – Summary of Expression Studies for TagA & CsaB

Table 6. *Summary of expression constructs evaluated for TagA*

Construct	Max yield from 3L expression	Attempted crystal screen	Crystal hits	Comments
MBP-TagA	0 mg	No	-	Construct did not express
TagAΔ(K202-Q252)	91 mg	Yes	No	Construct is soluble but has not crystallized
TagAΔ(M1-Y13,K202-Q252) HDX-1	37 mg	Yes	No	Construct is stable but has not crystallized
TagAΔ(G191-Q252) HDX-2	15 mg	No	-	Purified sample precipitated within one day
TagAΔ(M1-Y13,G191-Q252) HDX-3	0 mg	No	-	Construct did not express
TagAΔ(M1-S18,K202-Q252) -5 Nterm	42 mg	Yes	No	Purified sample precipitated within one day
TagAΔ(M1-E23,K202-Q252) -10 Nterm	0 mg	No	-	Construct did not express

Table 7. *TagA expression construct identities*

Construct name	Description & amino acid sequence of construct
MBP-TagA	*Amino acid sequence of maltose-binding protein fused to full length TagA protein*: MKIEEGKLVIWINGDKGYNGLAEVGKKFEKDTGIKVTVEHPDKLEEKFPQVAATGDGPDIIFWAHDRFGGYAQSGLLAEITPDKAFQDKLYPFTWDAVRYNGKLIAYPIAVEALSL IYNKDLLPNPPKTWEEIPALDKELKAKGKSALMFNLQEPYFTWPLIAADGGYAFKYENGKYDIKDVGVDNAGAKAGLTFLVDLIKNKHMNADTDYSIAEAAFNKGETAMTINGPWA WSNIDTSKVNYGVTVLPTFKGQPSKPFVGVLSAGINAASPNKELAKEFLENYLLTDEGLEAVNKDKPLGAVALKSYEEELAKDPRIAATMENAQKGEIMPNIPQMSAFWYAVRTAV INAASGRQTVDEALKDAQTNSSSNNNNNNNNNNLGDDDDKVPEFMLEMKDVPKVSIYGIPFSKLNMEDTVKLLVDHIDNGRSVQVITANPIMVMAALENDKNMAVMKQADLIVPDG TGIVWAANKFGDPVVERVTGFDLLHELMKMGELRKWKVYLLGAAPDVVQAAAERLQEQYPSISIVGYRDGFFKSDEDEKVVAQIREANPDLLFVARGADTQEPWIAKYRHQLQVPF MMGVGGSFDVISGKAKRAPMWVQKLRIEWLYRLAKEPKRFMRMLALPKFAWKVMRDKDKPTKVQ
TagAΔ(K202-Q252)	*Amino acid sequence of the TagA protein construct with amino acids K202-Q252 deleted from C-terminus (construct design was based on the homologous TarA crystal structurePDB: 5WB4)*: MGSSHHHHHHSSGENLYFQGHMLEMKDVPKVSIYGIPFSKLNMEDTVKLLVDHIDNGRSVQVITANPIMVMAALENDKNMAVMKQADLIVPDGTGIVWAANKFGDPVVERVTGFDL LHELMKMGELRKWKVYLLGAAPDVVQAAAERLQEQYPSISIVGYRDGFFKSDEDEKVVAQIREANPDLLFVARGADTQEPWIAKYRHQLQVPFMMGVGGSFDVISG
TagAΔ(M1-Y13,K202-Q252) HDX-1	*Amino acid sequence of the TagA protein construct with amino acids K202-Q252 deleted from C-terminus, and amino acids M1-Y13 deleted from the N-terminus (HDX was performed on the TagAΔ(K202-Q252) construct to determine that amino acids M1-Y13 may come from a highly disordered protein region)*: MGSSHHHHHHSSGENLYFQGHGIPFSKLNMEDTVKLLVDHIDNGRSVQVITANPIMVMAALENDKNMAVMKQADLIVPDGTGIVWAANKFGDPVVERVTGFDLLHELMKMGELRKW KVYLLGAAPDVVQAAAERLQEQYPSISIVGYRDGFFKSDEDEKVVAQIREANPDLLFVARGADTQEPWIAKYRHQLQVPFMMGVGGSFDVISG
TagAΔ(G191-Q252) HDX-2	*Amino acid sequence of the TagA protein construct with amino acids G191-Q252 deleted from C-terminus (HDX was performed on the TagAΔ(K202-Q252) construct to determine that amino acids G191-K202 may come from a highly disordered protein region)*: MGSSHHHHHHSSGENLYFQGHMLEMKDVPKVSIYGIPFSKLNMEDTVKLLVDHIDNGRSVQVITANPIMVMAALENDKNMAVMKQADLIVPDGTGIVWAANKFGDPVVERVTGFDL LHELMKMGELRKWKVYLLGAAPDVVQAAAERLQEQYPSISIVGYRDGFFKSDEDEKVVAQIREANPDLLFVARGADTQEPWIAKYRHQLQVPFMM
TagAΔ(M1-Y13,G191-Q252) HDX-3	*Amino acid sequence of the TagA protein construct with amino acids G191-Q252 deleted from C-terminus, and amino acids M1-Y13 deleted from the N-terminus (HDX was performed on the TagAΔ(K202-Q252) construct to determine that amino acids G191-K202 and M1-Y13 may come from a highly disordered protein region)*: MGSSHHHHHHSSGENLYFQGHGIPFSKLNMEDTVKLLVDHIDNGRSVQVITANPIMVMAALENDKNMAVMKQADLIVPDGTGIVWAANKFGDPVVERVTGFDLLHELMKMGELRKW KVYLLGAAPDVVQAAAERLQEQYPSISIVGYRDGFFKSDEDEKVVAQIREANPDLLFVARGADTQEPWIAKYRHQLQVPFMM
TagAΔ(M1-S18,K202-Q252) -5 Nterm	*Amino acid sequence of the TagAΔ(M1-Y13,K202-Q252) HDX-1 protein construct with a further 5 amino acids deleted from the N-terminus*: MGSSHHHHHHSSGENLYFQGHKLNMEDTVKLLVDHIDNGRSVQVITANPIMVMAALENDKNMAVMKQADLIVPDGTGIVWAANKFGDPVVERVTGFDLLHELMKMGELRKWKVYLL GAAPDVVQAAAERLQEQYPSISIVGYRDGFFKSDEDEKVVAQIREANPDLLFVARGADTQEPWIAKYRHQLQVPFMMGVGGSFDVISG
TagAΔ(M1-E23,K202-Q252) -10 Nterm	*Amino acid sequence of the TagAΔ(M1-Y13,K202-Q252) HDX-1 protein construct with a further 10 amino acids deleted from the N-terminus*: MGSSHHHHHHSSGENLYFQGHDTVKLLVDHIDNGRSVQVITANPIMVMAALENDKNMAVMKQADLIVPDGTGIVWAANKFGDPVVERVTGFDLLHELMKMGELRKWKVYLLGAAPD VVQAAAERLQEQYPSISIVGYRDGFFKSDEDEKVVAQIREANPDLLFVARGADTQEPWIAKYRHQLQVPFMMGVGGSFDVISG

Table 8. *Summary of expression constructs evaluated for CsaB*

Construct	Max yield from 3L expression	Attempted crystal screen	Crystal hits	Comments
CsaB	125 mg	Yes	No	Purified sample precipitated within days
CsaB -5 ΔCterm	48 mg	No	-	Purified sample precipitated within one day
CsaB -10 ΔCterm	4 mg	No	-	Purified sample precipitated within one day
CsaB -15 ΔCterm	0 mg	No	-	Construct did not express
CsaB -20 ΔCterm	0 mg	No	-	Construct did not express
CsaB -5 ΔNterm	80 mg	Yes	No	Purified sample precipitated within one day
CsaB -10 ΔNterm	0 mg	No	-	Construct did not express

Table 9. *CsaB expression construct identities*

Construct name	Description & sequence
CsaB	*Amino acid sequence of full length CsaB protein construct:* MASKATRIVLSGYYGFNNSGDEAVLLSILTALERAGNEAGTAIEPVVLSGDPETTSRLYGVQAVHRMKPGALLGAIRSSDALISGGGSLLQDATSSKTIPYYLGVLKLAQWFRK PTFIYSQGVGPVNRESFYPYIRHVFSRAAYVSVRDRESAELLMRMGIGQDNIHVVPDPVMGLRLPASQGEKSELDDGKGFDEAGRPYVGVSLRFWNQDRSDMDSIVDMLLQLSQ MQNVHLRFLPFHGASDEEASRYVMKKLETMVPEADGLCGGQSANATEAGSVMSLCAPLEHPQTMLREVSQCRVLVGMRLHSLIYAASQEVPLAGISYDPKIDQFLHRLHEKAIG TTAKLDAVRAASYMCELLDNVDGWRKDHKEAITVLKQEAEQPAQQIVKWLRHKALEHHHHHH
CsaB -5 ΔCterm	*Amino acid sequence of CsaB protein construct with 5 amino acids deleted from the C-terminus:* MASKATRIVLSGYYGFNNSGDEAVLLSILTALERAGNEAGTAIEPVVLSGDPETTSRLYGVQAVHRMKPGALLGAIRSSDALISGGGSLLQDATSSKTIPYYLGVLKLAQWFRK PTFIYSQGVGPVNRESFYPYIRHVFSRAAYVSVRDRESAELLMRMGIGQDNIHVVPDPVMGLRLPASQGEKSELDDGKGFDEAGRPYVGVSLRFWNQDRSDMDSIVDMLLQLSQ MQNVHLRFLPFHGASDEEASRYVMKKLETMVPEADGLCGGQSANATEAGSVMSLCAPLEHPQTMLREVSQCRVLVGMRLHSLIYAASQEVPLAGISYDPKIDQFLHRLHEKAIG TTAKLDAVRAASYMCELLDNVDGWRKDHKEAITVLKQEAEQPAQQIVKWLEHHHHHH
CsaB -10 ΔCterm	*Amino acid sequence of CsaB protein construct with 10 amino acids deleted from the C-terminus:* MASKATRIVLSGYYGFNNSGDEAVLLSILTALERAGNEAGTAIEPVVLSGDPETTSRLYGVQAVHRMKPGALLGAIRSSDALISGGGSLLQDATSSKTIPYYLGVLKLAQWFRK PTFIYSQGVGPVNRESFYPYIRHVFSRAAYVSVRDRESAELLMRMGIGQDNIHVVPDPVMGLRLPASQGEKSELDDGKGFDEAGRPYVGVSLRFWNQDRSDMDSIVDMLLQLSQ MQNVHLRFLPFHGASDEEASRYVMKKLETMVPEADGLCGGQSANATEAGSVMSLCAPLEHPQTMLREVSQCRVLVGMRLHSLIYAASQEVPLAGISYDPKIDQFLHRLHEKAIG TTAKLDAVRAASYMCELLDNVDGWRKDHKEAITVLKQEAEQPAQLEHHHHHH
CsaB -15 ΔCterm	*Amino acid sequence of CsaB protein construct with 15 amino acids deleted from the C-terminus:* MASKATRIVLSGYYGFNNSGDEAVLLSILTALERAGNEAGTAIEPVVLSGDPETTSRLYGVQAVHRMKPGALLGAIRSSDALISGGGSLLQDATSSKTIPYYLGVLKLAQWFRK PTFIYSQGVGPVNRESFYPYIRHVFSRAAYVSVRDRESAELLMRMGIGQDNIHVVPDPVMGLRLPASQGEKSELDDGKGFDEAGRPYVGVSLRFWNQDRSDMDSIVDMLLQLSQ MQNVHLRFLPFHGASDEEASRYVMKKLETMVPEADGLCGGQSANATEAGSVMSLCAPLEHPQTMLREVSQCRVLVGMRLHSLIYAASQEVPLAGISYDPKIDQFLHRLHEKAIG TTAKLDAVRAASYMCELLDNVDGWRKDHKEAITVLKQEALEHHHHHH
CsaB -20 ΔCterm	*Amino acid sequence of CsaB protein construct with 20 amino acids deleted from the C-terminus:* MASKATRIVLSGYYGFNNSGDEAVLLSILTALERAGNEAGTAIEPVVLSGDPETTSRLYGVQAVHRMKPGALLGAIRSSDALISGGGSLLQDATSSKTIPYYLGVLKLAQWFRK PTFIYSQGVGPVNRESFYPYIRHVFSRAAYVSVRDRESAELLMRMGIGQDNIHVVPDPVMGLRLPASQGEKSELDDGKGFDEAGRPYVGVSLRFWNQDRSDMDSIVDMLLQLSQ MQNVHLRFLPFHGASDEEASRYVMKKLETMVPEADGLCGGQSANATEAGSVMSLCAPLEHPQTMLREVSQCRVLVGMRLHSLIYAASQEVPLAGISYDPKIDQFLHRLHEKAIG TTAKLDAVRAASYMCELLDNVDGWRKDHKEAITVLEHHHHHH
CsaB -5 ΔNterm (M1-A5)	*Amino acid sequence of CsaB protein construct with 5 amino acids deleted from the N-terminus:* MTRIVLSGYYGFNNSGDEAVLLSILTALERAGNEAGTAIEPVVLSGDPETTSRLYGVQAVHRMKPGALLGAIRSSDALISGGGSLLQDATSSKTIPYYLGVLKLAQWFRKPTFI YSQGVGPVNRESFYPYIRHVFSRAAYVSVRDRESAELLMRMGIGQDNIHVVPDPVMGLRLPASQGEKSELDDGKGFDEAGRPYVGVSLRFWNQDRSDMDSIVDMLLQLSQMQNV HLRFLPFHGASDEEASRYVMKKLETMVPEADGLCGGQSANATEAGSVMSLCAPLEHPQTMLREVSQCRVLVGMRLHSLIYAASQEVPLAGISYDPKIDQFLHRLHEKAIGTTAK LDAVRAASYMCELLDNVDGWRKDHKEAITVLKQEAEQPAQQIVKWLRHKALEHHHHHH
CsaB -10 ΔNterm (M1-L10)	*Amino acid sequence of CsaB protein construct with 10 amino acids deleted from the N-terminus:* MSGYYGFNNSGDEAVLLSILTALERAGNEAGTAIEPVVLSGDPETTSRLYGVQAVHRMKPGALLGAIRSSDALISGGGSLLQDATSSKTIPYYLGVLKLAQWFRKPTFIYSQGV GPVNRESFYPYIRHVFSRAAYVSVRDRESAELLMRMGIGQDNIHVVPDPVMGLRLPASQGEKSELDDGKGFDEAGRPYVGVSLRFWNQDRSDMDSIVDMLLQLSQMQNVHLRFL PFHGASDEEASRYVMKKLETMVPEADGLCGGQSANATEAGSVMSLCAPLEHPQTMLREVSQCRVLVGMRLHSLIYAASQEVPLAGISYDPKIDQFLHRLHEKAIGTTAKLDAVR AASYMCELLDNVDGWRKDHKEAITVLKQEAEQPAQQIVKWLRHKALEHHHHHH