Taylor & Francis
Taylor & Francis Group

ADDENDUM

🔓 OPEN ACCESS ⟳ Check for updates

Serine-rich repeat proteins from gut microbes

Dimitrios Latousakis ⓘ, Donald A. MacKenzie, Andrea Telatin ⓘ, and Nathalie Juge

The Gut Microbes and Health Institute Strategic Programme, Quadram Institute Bioscience, Norwich Research Park, Norwich, UK

ABSTRACT

Serine-rich repeat proteins (SRRPs) have emerged as an important group of cell surface adhesins found in a growing number of Gram-positive bacteria. Studies focused on SRRPs from streptococci and staphylococci demonstrated that these proteins are *O*-glycosylated on serine or threonine residues and exported via an accessory secretion (aSec) system. In pathogens, these adhesins contribute to disease pathogenesis and represent therapeutic targets. Recently, the non-canonical aSec system has been identified in the genomes of gut microbes and characterization of their associated SRRPs is beginning to unfold, showing their role in mediating attachment and biofilm formation. Here we provide an update of the occurrence, structure, and function of SRRPs across bacteria, with emphasis on the molecular and biochemical properties of SRRPs from gut symbionts, particularly Lactobacilli. These emerging studies underscore the range of ligands recognized by these adhesins and the importance of SRRP glycosylation in the interaction of gut microbes with the host.

Introduction

SRRPs cover a functionally and structurally diverse family of glycoproteins found in many Gram-positive bacteria[1]. These proteins were originally identified in oral bacteria, such as streptococci and later in staphylococci,[2–11] where their expression has been linked to virulence.[11–13] Recently, SRRPs have also been reported in gut commensal microbes[14-16] including *Lactobacillus reuteri*, a Gram-positive bacterial species inhabiting the gastrointestinal (GI) tract of vertebrates[17,18] and *Streptococcus salivarius*, a pioneer colonizer and commensal bacterium of the human GI tract.[14] SRRPs are composed of distinct subdomains: a cleavable and unusually long signal peptide which is followed by an accessory Sec transport (AST) domain, a short serine-rich repeat region (SRR-1), a binding region (BR) (also known originally as a 'basic region' due to its unusual composition of basic amino acids), a second and much larger SRR-2, and a short non-repeat region that includes a LPTXG cell wall anchoring motif[1,19,20] (Figure 1). SRRP-BRs encompass a range of 3D structures, highlighting a relationship between their structural folds and binding ligands.

Export of SRRPs onto the bacterial surface occurs through a dedicated non-canonical Sec translocase, SecY2A2.[21] This accessory secretion (aSec) system is encoded by genes that are normally co-located with the *srrp* gene(s) within a gene cluster and is composed of the motor protein SecA2, the translocon channel SecY2 and three to five accessory Sec proteins (Asp1-5). In addition, this gene cluster also contains genes encoding a variable number of glycosyltransferases (GTs), ranging between two to ten.[22]

Recent studies identified SecA2/Y2 clusters in the genomes of various *Lactobacillus* species,[16,18,23] suggesting a conserved role of SRRPs among gut symbionts that possess the SecA2/Y2 cluster (Figure 2). One of the characteristics of SRRPs is that they undergo glycosylation, resulting in a range of glycan structures, reflecting the unique glycosylation pathways of the bacterial strains. It is expected that the combination of fold and glycosylation pattern will dictate the specificity of SRRPs in mediating the interaction of the bacteria with their respective host or niche within a host.

This article has been republished with minor changes. These changes do not impact the academic content of the article.

ⓘ Supplemental data for this article can be accessed on the publisher's website.
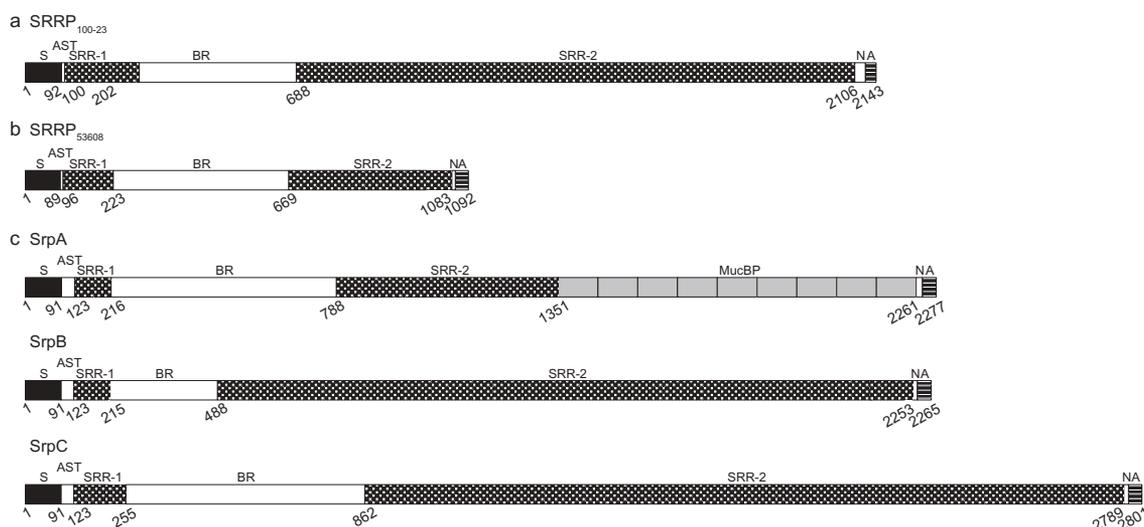
Figure 1. Structural domain organization of characterized serine-rich repeat proteins (SRRPs) from gut commensal bacteria. (a) *L. reuteri* strain 100–23, (b) *L. reuteri* strain ATCC 53608, (c) *Strep. salivarius* strain JIM8777. SRRPs are composed of an N-terminal signal peptide (S; black boxes); an AST domain; a non-repeat domain that mediates adhesion (binding region; BR); two serine-rich repeat domains (SRR) flanking the BR (SRR-1 and SRR-2; checkered boxes); and a second non-repeat domain (N), followed by a C-terminal cell wall anchor domain (A; striped boxes). In contrast to the majority of SRRPs, SrpA from *Strep. salivarius* strains, including JIM8777, also harbors a number of MucBP (mucin-binding protein) domains before and/or after SRR-2 (nine after SRR-2 in the case of JIM8777; grey boxes). The SRRs are composed of serine residues alternating with, most frequently, either an alanine, valine, or threonine residue. Numbers represent the starting amino acid positions of each domain. White boxes represent non-repeat regions.
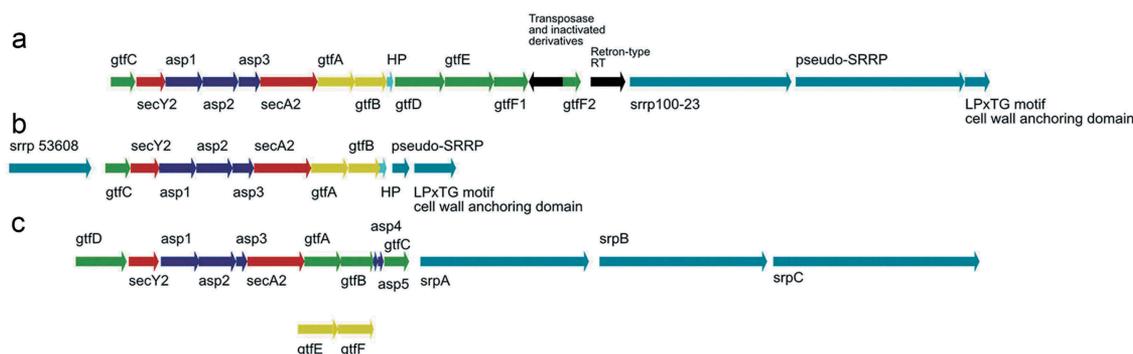
Figure 2. Organization of functionally characterized *secA2/Y2* clusters from gut commensal bacteria. (a) *L. reuteri* 100–23, (b) *L. reuteri* ATCC 53608 and (c) *Strep. salivarius* JIM8777. The genes encoding the SecA2/Y2 translocation machinery are shown in red, the accessory secretion proteins Asp1-5 in blue and the priming GTs, GtfA, and GtfB in *L. reuteri* and GtfE and GtfF in *Strep. salivarius*, in yellow. Genes encoding additional GTs are shown in green and the genes encoding SRRPs are illustrated in teal. Black arrows represent genes that are not part of the aSec machinery. HP: hypothetical protein.

Occurrence of SRRPs in bacterial genomes

A detailed bioinformatics analysis of all genome-sequenced bacteria classed as "non-pathogenic" or commensal revealed genes encoding full-length SRRPs and SecA2/Y2 secretion systems in a number of species from the order Lactobacillales, including strains of *L. reuteri*, *Lactobacillus oris*, *Lactobacillus salivarius*, *Lactobacillus johnsonii*, *Lactobacillus fructivorans*, *Lactococcus lactis*, *Streptococcus salivarius*, *Streptococcus vestibularis*, *Streptococcus mitis*, *Streptococcus oralis*, *Streptococcus cristatus*, and *Streptococcus* sp. strain DD12, with none found so far in other major lactobacilli species such as *Lactobacillus plantarum* (Table S1). Of particular interest are some strains of the major commensal of the oral mucosa and oropharyngeal tract, *Strep. salivarius*,[14,24] which possess three distinct SRRPs – SrpA, SrpB and SrpC (Table S1). It should be noted that although many strains of *Strep. salivarius*, *Strep. vestibularis*, *Strep. mitis*,

Strep. oralis and Strep. cristatus are classed as commensals, some have pathogenic potential, having been associated with diseases such as endocarditis or isolated clinically in cases of meningitis, ureolytic bacteremia, spinal operation infections and in immunocompromised situations.[25–31] In addition, recent bioinformatics analysis of 58 genome-sequenced L. reuteri strains showed that homologues of SRRP (and the corresponding linked aSec gene cluster) were mostly found in rodent and pig isolates of L. reuteri and absent from strains of human origin.[18]

Our detailed bioinformatics analysis also identified strains that appeared to possess only an incomplete aSec gene cluster, a SRRP that lacked a C-terminal cell wall anchor (possibly the result of a pseudogene, although still capable of exporting a SRRP extracellularly) or an obvious pseudo-SRRP whose domains were encoded by at least two adjacent ORFs (Table S1). These include strains of Streptococcus thoraltensis, Lactobacillus gasseri, Lactobacillus rhamnosus, Lactobacillus murinus, Lactobacillus nagelii and Lactobacillus mucosae, a species closely related to L. reuteri. Interestingly, L. mucosae LM1 appeared to have two SecA2/Y2 accessory gene clusters (LBLM1_RS03980-_RS04070 and LBLM1_RS11555-_RS04660), both with associated pseudo-SRRPs, although the latter cluster does not possess any glycosyltransferase (gtf) genes. A similar organisation occurs in Streptococcus sp. strain DD12[32] having two SecA2/Y2 gene clusters STRDD12_00537-_00544 and STRDD12_00642-_00656, with associated SRRP and/or pseudo-SRRP genes with the former cluster lacking any gtf genes (Table S1).

Domain organization of SRRPs

Overall organization

SRRPs are composed of distinct subdomains: an extended atypical signal sequence peptide of around 90 aa at the N-terminus, followed by an AST domain, a short SRR-1 region, a BR, a second and much larger SRR-2, and a LPTXG cell wall anchoring motif.[1,19,20] The AST domain, required for efficient targeting of the SRRP to the aSec machinery,[20] is rich in alanine, serine and threonine residues (30–50%) and is typically 30–40 aa

long although much shorter AST sequences occur in some SRRPs (Table S2). The SRR domains are typically composed of alternating serine residues, most frequently separated by an alanine, valine or threonine, but other residues such as glutamate, methionine, and leucine can also be found. SRR-2 shows high variability in terms of the nature and number of repeats between species and strains, resulting in the large diversity of sizes occurring in the SRRP family. It has been suggested that the number of repeats in the SRR-2 domains may have evolved to enable the binding region to reach out across the cell surface to mediate attachment.[1]

Table S2 illustrates the domain organization of a range of SRRPs from commensal and pathogenic bacteria. Although not an exhaustive list, examples are included from most of the species that have an aSec and contain at least one SRRP or pseudo-SRRP. Pseudogenes were included in this analysis because many appear capable of expressing a secreted protein that lacks only the LPXTG cell wall anchor motif so that the protein produced may still have a biological function although it would not be covalently attached to the cell surface. The aSec-specific secretion signal ranged in length from 55 aa of the SRRP from L. fructivorans to 101 aa of the SRRPs from L. salivarius. SRRPs from one or more species can have a predominance of the same SRR-2 motif but this does not apply to all SRRPs from the same species or even SRRPs from the same strain if multiple SRRPs are present, suggesting that these SRR-2 sequences evolved independently or had been acquired via horizontal transfer from other species and then been subject to evolutionary drift (Table S2). For example, the SRRPs from L. reuteri rodent strain 100–23C (SRRP$_{100-23}$) and of pig strain ATCC 53608 (SRRP$_{53608}$) display distinct SRR-2 repeating patterns (Table S2). The 10-aa serine-rich motif or repeat (SRR) 'SLSNSVSMSE' occurs 91 times in the SRR-2 domain of SRRP$_{100-23}$ and is interspersed with another 10 repeats with slight sequence variations (Table S2). The 'SLSNSVSMSE' repeat motif also occurs 10 times in the SRR-2 domain of SRRP$_{53608}$ along with another 15 repeats of differing sequence and is found in several other SRRPs from pig, rodent and sourdough isolates of L. reuteri and from strains of L. johnsonii. Similar repeating patterns ranging in size from 10 to 20 aa of alternating serine residues but with different sequences are also

found in the SRR-2 domains of the other SecA2/Y2-secreted proteins, including those from pathogenic streptococci and staphylococci (Table S2). In contrast to other Gram-positive bacteria which have a unique SRR glycoprotein-encoding gene (see above), *Strep. salivarius* expresses three large and glycosylated surface-exposed proteins – SrpA, SrpB, and SrpC – that show characteristics of SRR glycoproteins and are secreted through an aSec system.

Binding domains

The sequence analysis of SRRP-BRs in lactobacilli genomes was reported previously.[18] Here the analysis was extended to 140 SRRP-BR domains (covering 108 commensal-associated SRRPs and 32 pathogen/clinical-associated SRRPs) using a MUSCLE multiple sequence alignment[33,34] and generating a Maximum Likelihood tree displayed as a circular phylogram[35] (Figure 3). In most cases, SRRPs from commensal bacteria shared low homology in the circular diagram with the well-characterized SRRPs of pathogens – Fap1, GspB, Hsa, PsrP, SraP, SrpA, Srr-1, and Srr-2. This was confirmed in pairwise global alignments of the "pathogenic" SRRP-BRs with selected BR sequences from SRRPs of *L. reuteri* and SrpA/SrpC from *Strep. salivarius* JIM8777 giving aa % identity and aa % similarity values generally <20% and <30%, respectively (Table S3A&B). Specifically, $SRRP_{53608}$-BR shared only 16.8% aa identity and 27.5% aa similarity with *Strep. gordonii* GspB-BR and 13.0% aa identity and 25.0% aa similarity with *Strep. parasanguinis* Fap1-NRα (Table S3A&B). The *Strep. agalactiae* Srr-1 adhesin that includes the K4 sub-domain implicated in the binding to cytokeratin 4 (CK4) showed 16.7% aa identity and 30.3% aa similarity to $SRRP_{53608}$-BR (Table S3A&B). In a few instances, BRs from pathogenic or clinical isolates showed homology to SRRP-BR domains from commensal strains as shown for the BRs from PsrP of *Strep. pneumoniae* and SraP from *Staph. haemolyticus* with six *Strep. salivarius* SrpBs (Figure 3). This was further confirmed by pairwise global alignment with aa % identity values of 46.5% and 35.2% and aa % similarity values of 59.9% and 42.5% for PsrP and SraP, respectively, with SrpB from *Strep. salivarius* JIM8777 (Table S3A&B). Closer examination of the region of homology between the PsrP-BR and SrpB-BR (Figure S1A)

revealed that it included the PsrP binding region to keratin 10 (KRT10),[36] a 71 aa region with 52.1% aa identity and 64.8% aa similarity to the corresponding region in SrpB (Figure S1B) that was also present in the five other members of the SrpB cluster (Figure S1C). This suggests that *Strep. salivarius* SrpB could bind to keratin expressed by the different epithelial cell lines to which this bacterial species is known to bind.[14]

The circular phylogram further showed some host-specific clades for *L. reuteri* BRs (Figure 3). Three major clusters were evident – one group consisting mainly of SRRPs from porcine isolates that included strain ATCC 53608 but also rodent isolate 100–23 (with $SRRP_{53608}$-BR and $SRRP_{100-23}$-BR sharing 44.8% aa identity and 61.5% aa similarity in a pairwise global alignment) (Table S3A&B), a second group mainly of pseudo-SRRPs from rodent and sourdough strains and a third group of pseudo-SRRPs of porcine origin. Interestingly, BRs from *Strep. salivarius* were divided into three main groups – the first comprising SrpA and SrpB from strains JIM8777, JF, ATCC 27945, NCTC 8618, FDAARGOS 259 and HSISS4, a second containing SrpC from the same group of six strains and a third group of SrpA, SrpB and SrpC from strains ATCC 25975, CCHSS3 and 57.I. In *Strep. salivarius* genomes, SrpC is often annotated as the flagellin-specific chaperone, FliS, and in this analysis three FliS-BRs from two other streptococcal species clustered with the group of six SrpCs from *Strep. salivarius* (Figure 3).

Although originally termed the "basic region" in SRRPs first studied from pathogenic bacteria because of the high isoelectric point (pI)[20] of the binding region, it has become evident in recent years that the corresponding region in many other SRRPs has a net acidic pI.[1,11] In the present bioinformatics analysis of SRRP amino acid sequences, the majority of BRs (79%) showed a net "strong acidic" pI (arbitrarily set as ranging between pH 4.00–5.55) (Table S4). A few BRs (17%), although still net acidic, showed a pI value between 5.55 and 7.00 (classed as "weak acidic"); these include mainly BRs from SrpC and pseudo-SrpC from *Strep. salivarius* strains along with FliS from *Strep. parasanguinis* and *Strep. vestibularis* strains and SRRPs from *Strep. cristatus*.
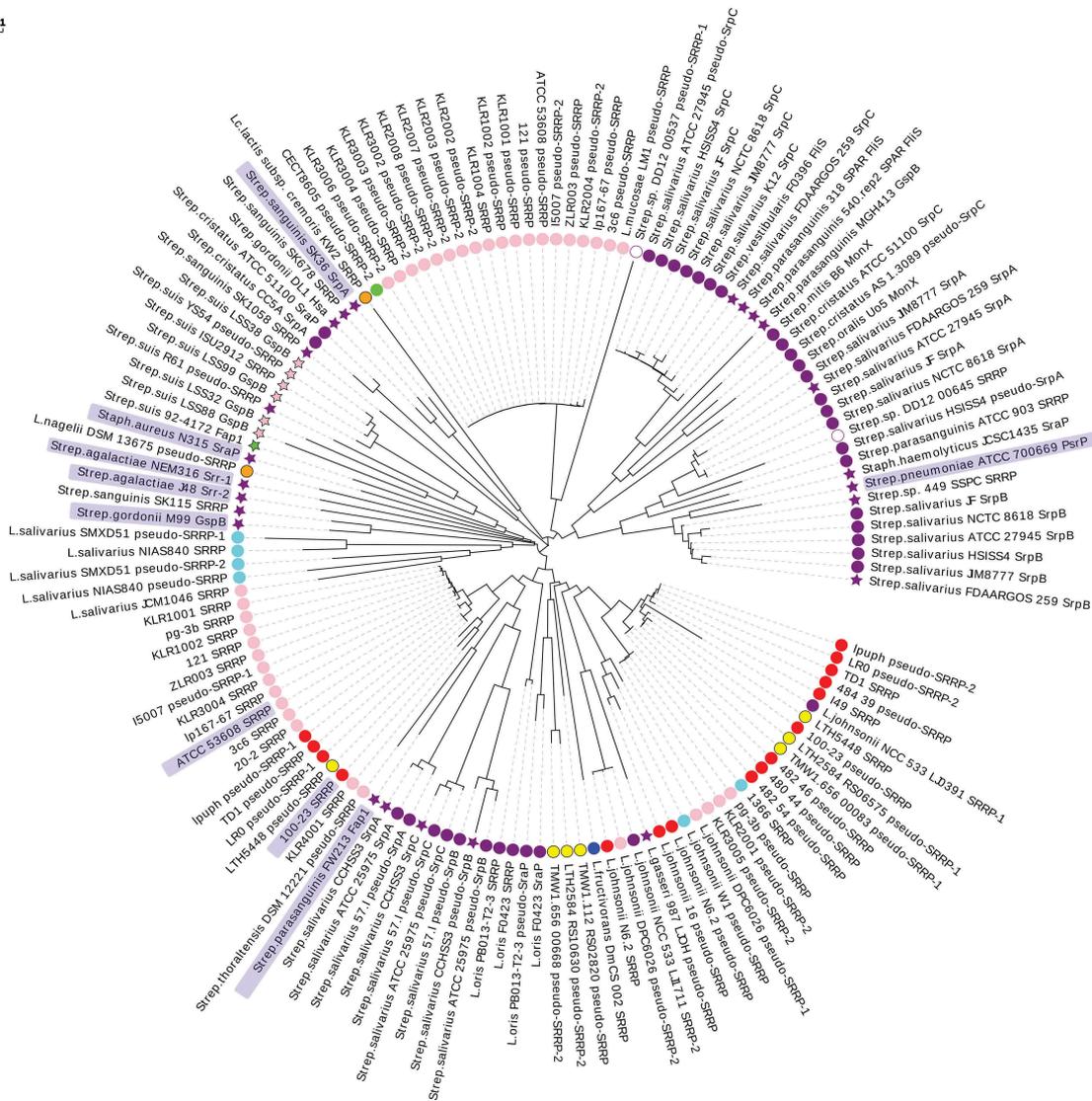
Figure 3. Circular phylogram representation of SRRP-BR domains from commensal and pathogenic bacteria. A MUSCLE multiple sequence alignment[33] was carried out in MEGA-X[34] using BR domain aa sequences from a total of 108 commensal-associated SRRPs and 32 pathogen/clinical-associated SRRPs. A guide tree was generated from the second iteration and a Maximum Likelihood phylogenetic tree displayed as a circular phylogram with EvolView software (http://www.evolgenius.info/evolview/).[35] SRRP-BRs are displayed as follows: circle, from a commensal or non-pathogenic strain; star, from a pathogenic or clinical isolate. Host or source origins of strains are indicated as follows: pink, porcine; red, rodent; purple, human; aqua, avian; lime, bovine; blue, insect; yellow, sourdough; orange, other fermented food or drink; white, chimpanzee fruit residues. BRs shown with only a strain name are all from *L. reuteri*. The entries for which a SRRP-BR crystal structure is available are shaded in a light blue box. The scale bar represents the branch length expressed as the number of aa substitutions per site.

Only a minority of BRs examined (4%) had a net basic pI (>8.00) with five out of six being from SRRPs of pathogenic strains (such as GspB and Hsa from *Strep. gordonii*, PsrP from *Strep. pneumoniae*, a SRRP from *Strep. sanguinis* and a SraP from *Strep. haemolyticus*). The calculated pI is an estimation of the average electrostatic charge over the entire BR domain and is therefore not predictive of BR function. However, it is worth noting

that pH can influence BR structural conformation and binding specificity, as observed for *L. reuteri* SRRPs[18] and *Strep. parasanguinis* Fap1,[37] so that information on BR pI, especially around the binding site(s), may inform on the types of ligands recognized by SRRPs under particular pH conditions.

In addition to the typical BRs, some SRRPs from commensal and pathogenic bacteria contain

other putative binding regions that could modulate or enhance binding. For example, the gene encoding SrpA in *Strep. salivarius* genomes contains a varying number of repeat motifs annotated as mucin-binding protein (MucBP) located on either side of the SRR-2 domain (Figure S2A). Some SRRPs from other streptococci, including the *Strep. salivarius* clinical isolate CCHSS3[38] and commensal isolate 57.I,[39] contain unusually long BR domains that include an L-type lectin region,[40] a single Rib/alpha/Esp surface antigen repeat[41] and a hyperosmolarity resistance protein Ebh (N-terminal domain) region[42] (Table S2). The SRRPs found in many strains of the pathogenic species *Streptococcus suis* also have an exceptionally large BR domain that can contain motifs such as a bacterial group 3 immunoglobulin (Ig)-like domain (Figure S2B) or a FlgD Tudor-like domain (Table S2). These putative binding domains in addition to the well-characterized BR domains may contribute to the binding of SRRPs to a variety of ligands.

Glycosylation of SRRPs

SRRP glycosylation pathways

The aSec system contains the necessary genes encoding proteins that facilitate the expression, glycosylation and subsequent secretion of SRRPs.[1] Interestingly, glycosylation of SRRP appears to promote secretion of the adhesin through the SecA2/Y2 system, and inhibit its export through the canonical SecA system.[43]

The best-studied examples of SecA2/SecY2-mediated glycosylation systems are from pathogenic *Strep. parasanguinis*, *Strep. pneumoniae*, *Strep. gordonii*, *Strep. agalactiae*, and *Staphylococcus aureus*.[44–46] In all cases, the glycosylation process is initiated by a two-protein glycosyltransferase complex, consisting of GtfA and GtfB that mediate the addition of *N*-acetylglucosamine (GlcNAc) to serine and threonine residues within the SRR domains of the adhesins. These enzymes interact with the acceptor SRRP and with each other through a conserved domain DUF1975 and mediate the addition of the reducing GlcNAc. GtfA acts as a glycosyltransferase (GT), whereas GtfB interacts with the acceptor protein as a chaperone.[47,48] In certain species, such as

Staph. aureus, these are the only GTs encoded in the aSecA cluster, and the glycan is not further extended.[8,10] In bacteria that express additional GTs, GtfC, formerly annotated as nucleotide sugar synthase (Nss), catalyzes the second glycosylation step, in most cases by adding a glucose moiety.[44,45,49,50] The third glycosylation step is mediated by a domain of unknown function (DUF), DUF1792, that is present in a bifunctional enzyme (dGT1)[45,50,51] and has been shown to adopt a GT-D fold.[51] The dGT1 enzyme also harbors a GT-A domain which has been shown to either generate a branching point as in Fap1 with the addition of GlcNAc,[44] or linearly extend the glycan as in PsrP with the addition of either a Glc or Gal residue.[45] GTs involved in the glycosylation of PsrP exhibit functional redundancy, as more than one GT can mediate the third and fourth glycosylation steps.[45] Additional GTs are in some cases involved in the further elongation of the glycan structures.[44,45] While the genes encoding GtfA and GtfB are conserved in all aSec loci identified to date, the number of genes encoding the additional GTs involved in the glycosylation of SRRPs varies between species, resulting in a range of glycan structures.

Furthermore, previous work on the glycosylation of GspB from *Strep. gordonii* M99 showed that additional heterogeneity is introduced by the *O*-acetylation of GlcNAc residues by Asp2.[52,53] This modification has also been identified in Srr-1 from *Strep. agalactiae* H36B.[7] The catalytic residues of Asp2 identified by Seepersaud et al. (2012)[53] are conserved in *Strep. gordonii* M99, *L. reuteri* 100–23, *L. reuteri* ATCC 53608 and *Strep. salivarius* JIM8777, suggesting that Asp2 may also perform the same reaction in these organisms.

In *L. reuteri*, the intact aSec cluster has mostly been found in strains of murine or porcine origin, and it appears to be absent from strains of human origin (see above). In addition to the SecA2 and SecY2 translocases and the accessory secretion-associated proteins Asp1-3, the *L. reuteri* ATCC 53608 SecA2/Y2 glycosylation system contains genes encoding the priming GtfA$_{53608}$ and GtfB$_{53608}$, and a gene encoding GtfC$_{53608}$ whereas in *L. reuteri* 100-23C, the SecA2/Y2 cluster includes eight genes encoding predicted GTs, including GtfA$_{100-23}$, GtfB$_{100-23}$, and GtfC$_{100-23}$ (Figure 2).[16,18,23] A recent study demonstrated that GtfA/B are involved in GlcNAc attachment to

SRRP$_{100-23}$ and SRRP$_{53608}$ while GtfC$_{53608}$ extends the chain with a GlcNAc residue and GtfC$_{100-23}$ with Glc[15] (Figure 4). To date, all characterized GtfCs have been shown to add a Glc residue onto the GlcNAc core, therefore this was the first report of a GtfC from the SecA2/Y2 system showing ligand specificity to UDP-GlcNAc.[15] The specificity of *L. reuteri* GtfC$_{53608}$ was further supported by Differential Scanning Fluorimetry (DSF) and Saturation Transfer Difference (STD) NMR analyses, showing a preference for UDP-GlcNAc, in line with the MS/GC-MS analyses.[15]

Glycosylation of the three SRRPs of *Strep. salivarius* JIM8777 follows a slightly distinct process. Despite the presence of genes encoding putative GtfA and GtfB, the first glycosylation appears to be unusually carried out by two genetically linked GTs, GtfE and GtfF, encoded outside of the *secA2/Y2* locus[14] (Figure 4). In addition, genes encoding a putative GtfC and a putative dGT1 are present in the aSec cluster, suggesting a subsequent glycosylation pathway similar to that of SRRP$_{100-23}$ from *L. reuteri* 100-23.[15]
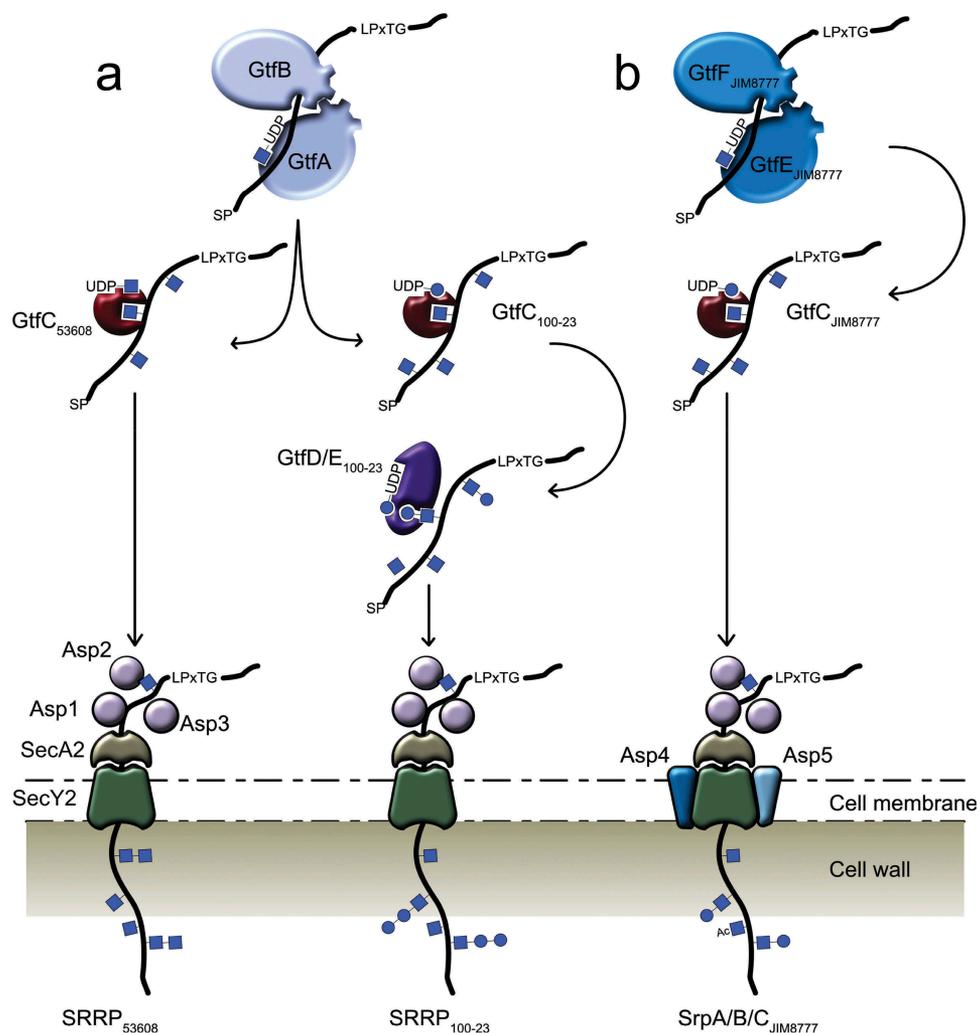
Figure 4. SRRP glycosylation mechanisms in gut commensal bacteria. (a) SRRP glycosylation in *L. reuteri* strains ATCC 53608 and 100–23. The GtfA/B complex initiates the glycosylation of the *L. reuteri* SRRP with GlcNAc residues, while GtfC extends the glycans with either GlcNAc (GtfC$_{53608}$) or Glc (GtfC$_{100-23}$). The glycosylated SRRP$_{53608}$ is then secreted through the aSec system, whereas the SRRP$_{100-23}$ is further extended by GtfD and/or GtfE before secretion. (b) SRRP glycosylation in *Strep. salivarius* strain JIM8777. SrpA, SrpB, and SrpC glycosylation are mediated by the GtfE/F complex, GTs encoded by a genetic locus unlinked to the aSec genomic island. After extension and acetylation of the glycans by other enzymes in the aSec system, the adhesins are secreted through the SecA2/Y2 channel. Blue circle: glucose, blue square: GlcNAc, SP: signal peptide.

Glycosylation profile of SRRPs

SRRPs from pathogenic *Streptococcus* and *Staphylococcus* species carry a range of different glycan structures (Table 1), reflecting differences in the organization of the SecA2/Y2 accessory cluster of these strains.

In the commensal *Strep. salivarius* JIM8777 strain, all three SRRP substrates of the SecA2/Y2 system, SrpA, SrpB, and SrpC were found to be glycosylated.[14] LC-MS/MS analysis revealed the presence of hexose (Hex), *N*-acetylhexosamine (HexNAc) and *O*-acetyl-*N*-acetylhexosamine (*O*-AcHexNAc) linked to peptides in both SRR-1 and SRR-2 domains (Table 1). The SRR-1 domain showed high glycosylation diversity with 24 different glycan combinations whereas HexNAc or *O*-AcHexNAc modified peptides were found in the SRR-2 domain, with all serine and threonine residues of the C-terminal SRR-2 domain being glycosylated. Hex residues were only present on peptides carrying at least three HexNAc residues, suggesting that HexNAc modification is a prerequisite for Hex glycosylation (Figure 4). Sequence homology suggested glycosylation of these regions in all three *Strep. salivarius* JIM8777 SRRPs. Furthermore, MS analysis of *Strep. salivarius* SrpA, SrpB, and SrpC identified *O*-acetylated GlcNAc residues[14], supporting Asp2 activity (see above).

The glycosylation profile of SRRPs from *L. reuteri* was recently determined using a combination of bioinformatics analysis, lectin screening, LC-MS-based sugar nucleotide profiling, MALDI-ToF, and GC-MS analyses. This study showed that the *L. reuteri* ATCC 53608 and 100-23C strains were capable of performing protein glycosylation and that SRRP$_{100-23}$ and SRRP$_{53608}$ were glycosylated with Hex-Hex-HexNAc and di-HexNAc moieties, respectively. Following *in vivo* glycoengineering in *E. coli*, NMR analysis and enzymatic treatment further showed that SRRP$_{53608}$ was glycosylated with GlcNAcβ(1→6)-GlcNAcα moieties. Together, it was suggested that SRRP$_{100-23}$ is glycosylated with

Table 1. Glycan structures identified on SRRPs from different microorganisms.

SRRP	Origin	Glycans	
Fap1	*Strep. parasanguinis* FW213	(1) Glc1-3GlcNAc1-(Rha1-3Glc1)-2,6Glc1-6GlcNAc-[51]	
GspB	*Strep. gordonii* M99	(1) Glc-Glc-GlcNAc-[50]	
		(2) *O*-AcGlcNAc-[50]	
PsrP	*Strep. pneumoniae* TIGR4	(1) Gal-Glc-Glc-GlcNAc-[45]	
		(2) Glc-Glc-Glc-GlcNAc-[45]	
		(3) Gal-Gal-Glc-GlcNAc-[45]	
		(4) Glc-Gal-Glc-GlcNAc-[45]	
SraP	*Staph. aureus* COL	(1) GlcNAc-[10]	
Srr-1	*Strep. agalactiae* H36B	Contains HexNAc, *O*-AcHexNAc, Hex[7]	ND*
Srr-2	*Strep. agalactiae* COH1	Contains GlcNAc, Glc[49]	ND*
SrpA	*Strep. salivarius* JIM8777	(1) Hex-GlcNAc-[14]	
SrpB		(2) *O*-AcHexNAc-[14]	
SrpC		(3) GlcNAc-[14]	
SRRP$_{100-23}$	*L. reuteri* 100-23	(1) Glc-Glc-GlcNAc-[15]	
		(2) Gal-Glc-GlcNAc-[15]	
		(3) GlcNAc-[15]	
SRRP$_{53608}$	*L. reuteri* ATCC 53608	(1) GlcNAcα-[15]	
		(2) GlcNAcβ1-6GlcNAcα-[15]	

*ND: Not determined; PR: Protein; Ac; *O*-acetyl-; Monosaccharide symbols follow the Symbol Nomenclature for Glycans system.[54] Key: glucose (blue circle), GlcNAc (blue square), rhamnose (green triangle), galactose (yellow circle), hexose (white circle).

GlcNAc and Hex-Glc-GlcNAc whereas SRRP$_{53608}$ is glycosylated with GlcNAc and di-GlcNAc moieties[15] (Figure 4) (Table 1). Although both strains encode a predicted Asp2, *O*-acetylation could not be confirmed biochemically due to the conditions used in the MS analysis.[15] The number of GTs in the *L. reuteri* 100-23C SecA2/Y2 cluster exceeds the number of sugars on SRRP$_{100-23}$, as also reported for some streptococcal SecA2/Y2 systems.[14,45] To date, there is no generic explanation for the presence of additional genes encoding GTs in the genomes of these strains. In some cases, gene duplication is observed, which may lead to functional redundancy, whereas insertion of genetic elements into genes encoding GTs may lead to gene inactivation. A defective glycosylation of SRRPs in pathogenic bacteria led to impaired binding of the respective bacteria onto model substrates and reduced virulence in mouse models.[1,55,56] The glycosylation of SRRPs in *Lactobacillus* species, as demonstrated for *L. reuteri* strains, is likely to impact on the adhesion capacity of these strains.

Functional and structural properties of SRRPs

Bacterial attachment to host surfaces is a pivotal event in the biological and infectious processes of both commensal and pathogenic bacteria, respectively. The role of SRRPs in Gram-positive bacterial pathogenesis has been investigated extensively (see Lizcano et al., 2012[1] for a review). *Streptococcus* and *Staphylococcus* SRRPs contribute towards a wide range of diseases including sub-acute bacterial endocarditis, community-acquired pneumonia, and meningitis.[57,58] It is believed that the main function of the SRR domains is to form a rigid, *O*-glycosylated, rod-like structure to extend the N-terminal BR domain out from the cell surface to facilitate efficient binding.[1,70] Crystal structures of seven SRRP-BRs have been reported for Gram-positive pathogens to date, highlighting a relationship between their structural folds and binding ligands (Table 2). These include *Strep. parasanguinis* Fap1 (2KUB and 2X12),[70] *Strep. gordonii* GspB (3QC5/6),[71] *Strep. sanguinis* SrpA (5EQ2),[67] Srr-1 and Srr-2 paralogues of *Strep. agalactiae* (4MBO/R),[68,72] *Staph. aureus* SraP (4M0(0–3))[65] and *Strep. pneumoniae* PsrP (3ZGH/I).[36] For

example, Srr-1, Srr-2, and PsrP each adopt variations of the DEv-IgG fold[36,68,73] and mediate binding to proteins with long α-helical folds, with Srr-1 binding to cytokeratin-4;[69,74,75] Srr-1 and Srr-2 binding to fibrinogen Aα,[2,68,76–78] and PsrP adhering to cytokeratin-10,[13] as well as DNA.[12,36] Other SRRP-BRs are composed of two or more domain folds and include – from N- to C-terminus – the helical and CnaA folds for Fap1,[70] CnaA, siglec and 'unique' subdomains for GspB,[71] siglec and 'unique' subdomains for SrpA,[67] and a legume lectin-like fold, a β-grasp fold and two eukaryotic cadherin-like modules for SraP.[65] The GspB, Hsa, and SrpA homologs from *Strep. gordonii* and *Strep. sanguinis* species bind to sialoglycan ligands on microarrays and platelets.[63,79] Hsa binds to both α(2–3) sialyllactosamine and sialyl-T antigen, whereas GspB binds only to the latter.[59,71,79–85] SraP specifically recognizes N-acetylneuraminic acid.[79,86] Recently, a novel SRRP, SssP1, associated with a SecA2/Y2 gene cluster has been identified in strain CZ130302 of *Strep. suis*, an important Gram-positive pathogen in the swine industry and emerging zoonotic pathogen for humans. SssP1 participates in adhesion to host cells including HEp-2 and human brain microvessel endothelial cells (HBMECs) and contributes to *Strep. suis* virulence. Although SssP1 lacks an LPXTG cell wall anchor which may classify it as a pseudo-SRRP, it contains additional N3 and SRR-3 domains at its C-terminus which may play a role in tethering it to the bacterial cell surface. The SssP1-BR, like those from several other *Strep. suis* SRRPs, contains two bacterial group 3 Ig-like domains (Figure S1B). Recombinant SssP1 proteins that contained these Ig-like-1 and Ig-like-2 domains, respectively, were shown to be involved in the binding to sialic acids[11] (see Table 2).

In the gut commensal bacterium *Strep. salivarius*, SrpA, SrpB and SrpC are the main factors underlying its multifaceted adhesion and, their glycosylation plays a major role in host colonization.[14,65] SrpB and SrpC play complementary adhesive roles involved in several steps of the colonization process: auto-aggregation, biofilm formation and adhesion to a variety of host epithelial cells and components. At least one of the *Strep. salivarius* SRR glycoproteins is important for colonization in mice. A preliminary analysis of the binding targets for these three SRRPs indicated that SrpC bound to mucin MUC2 and to

Table 2. Binding substrates of SRRPs from different microorganisms.

SRRP name	Origin	Binding Substrate
Fap1	*Strep. parasanguinis FW213*	Undetermined saliva components[3]
GspB	*Strep. gordonii M99*	Human salivary proteins, sialyl-T antigen, glycoprotein Ibα[59-61]
Hsa	*Strep. gordonii DL1*	Sialyl-T antigen, α(2,3) sialyllactosamine, sialic acid-containing MUC7, human salivary proteins, glycoprotein Ibα[59,61,62]
SRRP$_{SK1}$	*Strep. sanguinis SK1*	Sialyl-T antigen, sialyl-Lewis X, α(2,3) sialyllactosamine[63]
SRRP$_{NCTC10712}$	*Strep. sanguinis NCTC10712*	Sialyl-T antigen, sialyl-Lewis X, α(2,3) sialyllactosamine[63]
SRRP$_{SK678}$	*Strep. sanguinis SK678*	Sialyl-Lewis X, α(2,3) sialyllactosamine[63]
SRRP$_{SF100}$	*Strep. mitis SF100*	Sialyl-T antigen[63]
PsrP	*Strep. pneumoniae TIGR4*	Keratin-10, BR$_{PsrP}$, extracellular DNA[12,13,64]
SraP	*Staph. aureus COL*	Surface molecules of blood platelets, sialylated structures[65,66]
SrpA	*Strep. sanguinis SK36*	Platelets, sialic acid residues[67]
Srr-1	*Strep. agalactiae CNCTC 10/84*	Platelets, keratin-4, fibrinogen[2,68,69]
Srr-2	*Strep. agalactiae COH1*	Fibrinogen[68]
SssP1	*Strep. suis CZ130302*	HEp-2, human microvessel endothelial cells, sialylated structures[11]
SRRP$_{100-23}$	*L. reuteri 100-23*	Mouse forestomach stratified squamous epithelium[17]
SRRP$_{53608}$	*L. reuteri ATCC 53608*	Porcine gastric mucin, DNA, intestinal epithelium, polygalacturonic acid, rhamnogalacturonan I and chondroitin sulfate A[18]
SrpB	*Strep. salivarius JIM8777*	Epithelial cells (HT-29, A549, FaDu, HEp-2)[14]
SrpC	*Strep. salivarius JIM8777*	Extracellular matrix components (MUC2, fibronectin, MaxGel ECM, laminin, Elastin)[14]

various extracellular matrix proteins whereas SrpB participated in bacterial autoaggregation and was involved in adhesion to various human epithelial cell lines.[14] The ligands for SrpA have yet to be identified, despite being annotated as having MucBP repeat motifs (Figure S1A), but it was suggested that SrpA could be an adhesin involved in interspecies co-aggregation.[14] No crystal structure of *Strep. salivarius* SRRP-BRs is currently available (see Table 2).

In *L. reuteri*, the SecA2/Y2 cluster and SRRP in the murine isolate *L. reuteri* 100-23 is crucial for adhesion of the bacteria to the forestomach epithelium of the murine GI tract, as shown by colonization experiments in germ-free mice with *L. reuteri* 100-23 wild-type and mutants.[17] Mutants lacking the s*ecA2* gene showed defective adhesion, whereas mutants lacking *srrp* showed the most reduced biofilm formation, compared to other putative adhesins tested.[17] In contrast to all structurally characterized SRRP-BRs reported to date, *L. reuteri* SRRP-BR displays a fold typically adopted by extracellular pectate lyase PelC-like proteins.[18] The BR crystal structures of SRRP$_{100-23}$ and SRRP$_{53608}$ from *L. reuteri* ATCC 53608, revealed a "β-solenoid" fold comprising β-strands coiled in a repetitive pattern to form a right-handed helix with three parallel β-sheets, which is unique in the SRRP family.[18]

SRRP$_{53608}$-BR bound to host epithelial cells, mucins, and DNA at neutral pH and recognized polygalacturonic acid (PGA), rhamnogalacturonan I or chondroitin sulfate A at acidic pH. Mutagenesis confirmed the role of the BR putative binding site in the interaction of SRRP$_{53608}$-BR with PGA. Long molecular dynamics (MD) simulations showed that SRRP$_{53608}$ undergoes a pH-dependent conformational change that would mediate the different binding profiles observed. A Dali search revealed that the structure of *L. reuteri* SRRP-BRs also shows similarity to some extracellular adhesive proteins with PelC-like folds in pathogens, such as pertactins or bacteriophage tail-spike proteins.[18]

A summary of the structural organization and fold of SRRP-BRs structurally characterized to date is provided in Figure 5, where we also report the CATH[87] family of each protein. All the proteins, excluding the *L. reuteri* ones, belong to the "Immunoglobulin-like" family. The structures of Srr-1 and Srr-2 (*Strep. agalactiae*), while structurally in the same CATH category, do not share specific domains with the other SRRPs of the list, and share a higher structural similarity with clumping factor A of *Staph. aureus* (Newman strain). The structure of *L. reuteri* SRRP-BRs is radically different (CATH 2.160.20, Pectate Lyase C-like) to all other SRRP-BRs from pathogenic bacteria for which a crystal structure is available, perhaps reflecting a different role in gut-commensal interactions. Interestingly, a sequence similarity search revealed an uncharacterized non-
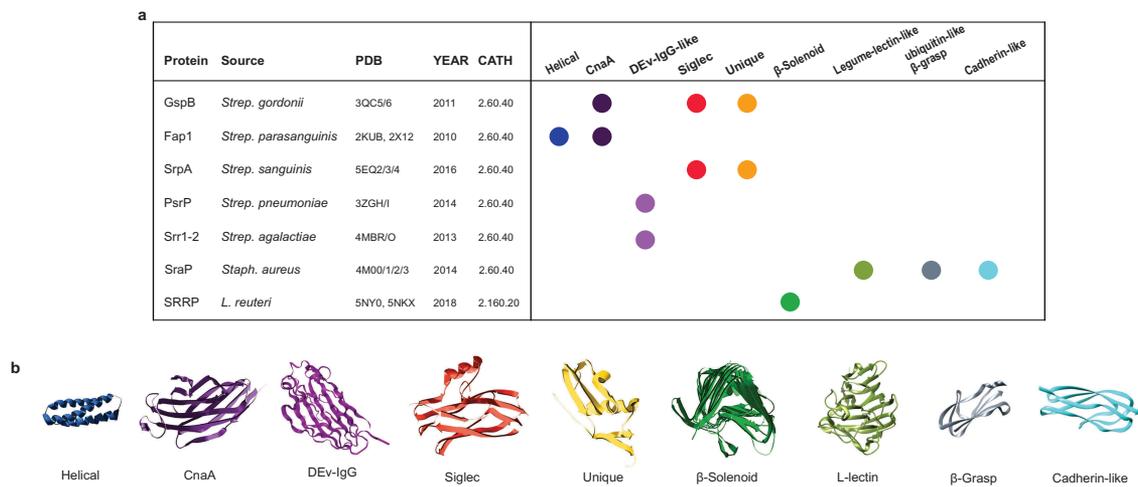
Figure 5. Structural repertoire of the SRRP binding regions. (a) Summary of the domains found in the available structures of SRRPs. For each protein, we report the name, the source organism, the PDB code of the available structures, the year of the publication of the structure and the CATH code. On the right, a schematic representation is included of the structural domains present in the proteins. (b) Example of three-dimensional structure of each domain found in the SRRP-BRs.

SRR protein (WP_003669640.1) in *L. reuteri* human strain DSM 20016 (lacking the aSec pathway), that is likely to share the same β-solenoid organization (according to homology modelling) and could be evolutionary related to the SRRPs found in *L. reuteri* strains 100-23 and ATCC 53608, isolated from rat and pig, respectively. The variety of domains found in the binding region of proteins from different organisms, suggests that the evolutionary adaptation to different niches could require domain shuffling, eventually acquiring existing domains from other proteins. The diversity in SRRP-BR structures is reflected in their primary amino acid sequences, as illustrated in Figure 3.

SRR glycoproteins from *L. reuteri* strains may contribute to the mechanisms underpinning *L. reuteri* adaptation to their vertebrate hosts. In addition, the presence of complete SecA2/Y2 clusters with an intact SRRP in the genomes of other *Lactobacillus* species (see above), suggest a common role of SRR glycoproteins in adhesion to host epithelia, which may be related to the ecological context of these strains (see Duar et al., 2017[88] for a review).

Conclusions and future directions

Bacterial adhesion is a critical step for colonization of the host. Cell surface proteins (adhesins or lectins) mediate attachment to host cells, initiate colonization and define a bacterial cell and tissue tropism by interacting with host proteins or glycoconjugates. SRRPs expressed by Gram-positive bacteria are important mediators of the interaction between streptococci and staphylococci with the host tissues. SRRPs show a high sequence diversity, which is reflected by the modularity and structural folds of their binding domains. The finding that SRRPs occur in non-pathogenic bacteria expands the role played by this important family of glycosylated adhesins in mediating gut microbiota–host interactions. Knowledge of the structure and function of SRRPs from gut microbes may be used for the selection of probiotic strains targeting different vertebrate hosts. Another potential application is to explore these adhesion/biofilm formation properties to compete with clinically important pathogens, reduce infection and combat antimicrobial resistance. However, more work is needed to ascertain the exact nature of the receptors of these adhesins *in vivo*.

Protein glycosylation is a major feature of SRRPs. These adhesins harbor different glycan structures reflecting differences in the organization of the SecA2/Y2 accessory cluster of these strains, and in their binding capacity. Investigating the co-evolution of the SRR glycoproteins with their specialized glycan modifying and export systems is warranted to gain novel evolutionary insights into host/niche-

specific adaptation. This potential requirement for the coupling of glycosylation and secretion has been proposed as a mechanism to ensure that the adhesin is optimally modified for binding. Whether the post-translational modification and transport of SRRPs follow distinct processes and the impact of protein glycosylation on the biological role of SRRP adhesins are important research questions which remain to be addressed. The discovery that glycosylation is strain-specific provides a molecular track to explore the impact of protein glycosylation in biofilm formation, adhesion, and colonization of gut symbionts. Furthermore, novel information on the cellular pathways leading to the glycosylation of SRRPs expands the range of glycosyltransferase specificities and potential glycoengineering applications for the recombinant production of glycoprotein conjugates in different cell types.

Disclosure of interest

Funding

ORCID

Dimitrios Latousakis http://orcid.org/0000-0002-2472-9328
Andrea Telatin http://orcid.org/0000-0001-7619-281X

References

1. Lizcano A, Sanchez CJ, Orihuela CJ. A role for glycosylated serine-rich repeat proteins in Gram-positive bacterial pathogenesis. Mol Oral Microbiol. 2012;27:257–269. doi:10.1111/j.2041-1014.2012.00653.x.
2. Seo HS, Xiong YQ, Sullam PM. Role of the serine-rich surface glycoprotein Srr1 of *Streptococcus agalactiae* in the pathogenesis of infective endocarditis. PLoS One. 2013;8:e64204. doi:10.1371/journal.pone.0064204.
3. Wu H, Mintz KP, Ladha M, Fives-Taylor PM. Isolation and characterization of Fap1, a fimbriae-associated adhesin of *Streptococcus parasanguis* FW213. Mol Microbiol. 1998;28:487–500.
4. Zhou M, Wu H. Glycosylation and biogenesis of a family of serine-rich bacterial adhesins. Microbiol (Reading, Engl). 2009;155:317–327. doi:10.1099/mic.0.025221-0.
5. Bensing BA, Sullam PM. An accessory sec locus of *Streptococcus gordonii* is required for export of the surface protein GspB and for normal levels of binding to human platelets. Mol Microbiol. 2002;44:1081–1094.
6. Bensing BA, Gibson BW, Sullam PM. The *Streptococcus gordonii* platelet binding protein GspB undergoes glycosylation independently of export. J Bacteriol. 2004;186:638–645.
7. Chaze T, Guillot A, Valot B, Langella O, Chamot-Rooke J, Di Guilmi A-M, Trieu-Cuot P, Dramsi S, Mistou M-Y. O-Glycosylation of the N-terminal region of the serine-rich adhesin Srr1 of *Streptococcus agalactiae* explored by mass spectrometry. Mol Cell Proteomics. 2014;13:2168–2182. doi:10.1074/mcp.M114.038075.
8. Siboo IR, Chaffin DO, Rubens CE, Sullam PM. Characterization of the accessory Sec system of *Staphylococcus aureus*. J Bacteriol. 2008;190:6188–6196. doi:10.1128/JB.00300-08.
9. Takamatsu D, Bensing BA, Sullam PM. Genes in the accessory sec locus of *Streptococcus gordonii* have three functionally distinct effects on the expression of the platelet-binding protein GspB. Mol Microbiol. 2004;52:189–203. doi:10.1111/j.1365-2958.2004.03978.x.
10. Li Y, Huang X, Li J, Zeng J, Zhu F, Fan W, Hu L. Both GtfA and GtfB are required for SraP glycosylation in *Staphylococcus aureus*. Curr Microbiol. 2014;69:121–126. doi:10.1007/s00284-014-0563-2.
11. Zhang Y, Lu P, Pan Z, Zhu Y, Ma J, Zhong X, Dong W, Lu C, Yao H. SssP1, a *Streptococcus suis* fimbria-like protein transported by the SecY2/A2 system, contributes to bacterial virulence. Appl Environ Microbiol. 2018;84:e01385–18. doi:10.1128/AEM.01385-18.
12. Sanchez CJ, Shivshankar P, Stol K, Trakhtenbroit S, Sullam PM, Sauer K, Hermans PW, Orihuela CJ. The pneumococcal serine-rich repeat protein is an intra-species bacterial adhesin that promotes bacterial aggregation in vivo and in biofilms. PLoS Pathog. 2010;6:e1001044. doi:10.1371/journal.ppat.1000975.
13. Shivshankar P, Sanchez C, Rose LF, Orihuela CJ. The *Streptococcus pneumoniae* adhesin PsrP binds to Keratin 10 on lung cells. Mol Microbiol. 2009;73:663–679. doi:10.1111/j.1365-2958.2009.06796.x.
14. Couvigny B, Lapaque N, Rigottier-Gois L, Guillot A, Chat S, Meylheuc T, Kulakauskas S, Rohde M, Mistou M-Y, Renault P, et al. Three glycosylated serine-rich repeat proteins play a pivotal role in adhesion and colonization of the pioneer commensal bacterium, *Streptococcus salivarius*. Environ Microbiol. 2017;19:3579–3594. doi:10.1111/1462-2920.13853.
15. Latousakis D, Nepravishta R, Rejzek M, Wegmann U, Le Gall G, Kavanaugh D, Colquhoun IJ, Frese S, MacKenzie DA, Walter J, et al. Serine-rich repeat protein adhesins from *Lactobacillus reuteri* display strain specific glycosylation profiles. Glycobiology. 2019;29:45–58. doi:10.1093/glycob/cwy100.

16. Latousakis D, Juge N. How sweet are our gut beneficial bacteria? A focus on protein glycosylation in *Lactobacillus*. Int J Mol Sci. 2018;19:136. doi:10.3390/ijms19010136.

17. Frese SA, MacKenzie DA, Peterson DA, Schmaltz R, Fangman T, Zhou Y, Zhang C, Benson AK, Cody LA, Mulholland F, et al. Molecular characterization of host-specific biofilm formation in a vertebrate gut symbiont. PLoS Genet. 2013;9:e1004057. doi:10.1371/journal.pgen.1004057.

18. Sequeira S, Kavanaugh D, MacKenzie DA, Šuligoj T, Walpole S, Leclaire C, Gunning AP, Latousakis D, Willats WGT, Angulo J, et al. Structural basis for the role of serine-rich repeat proteins from *Lactobacillus reuteri* in gut microbe–host interactions. Proc Natl Acad Sci 2018; 115:E2706–2715. doi:10.1073/pnas.1715016115

19. Rigel NW, Braunstein M. A new twist on an old pathway - accessory secretion systems. Mol Microbiol. 2008;69:291–302. doi:10.1111/j.1365-2958.2008.06294.x.

20. Feltcher ME, Braunstein M. Emerging themes in SecA2-mediated protein export. Nat Rev Microbiol. 2012;10:779–789. doi:10.1038/nrmicro2874.

21. Prabudiansyah I, Driessen AJM. The canonical and accessory Sec system of gram-positive bacteria. Curr Top Microbiol Immunol. 2017;404:45–67. doi:10.1007/82_2016_9.

22. Bensing BA, Seepersaud R, Yen YT, Sullam PM. Selective transport by SecA2: an expanding family of customized motor proteins. Biochim Biophys Acta. 2014;1843:1674–1686. doi:10.1016/j.bbamcr.2013.10.019.

23. Wegmann U, MacKenzie DA, Zheng J, Goesmann A, Roos S, Swarbreck D, Walter J, Crossman LC, Juge N. The pan-genome of *Lactobacillus reuteri* strains originating from the pig gastrointestinal tract. BMC Genomics. 2015;16:1023. doi:10.1186/s12864-015-2216-7.

24. Delorme C, Abraham AL, Renault P, Guedon E. Genomics of *Streptococcus salivarius*, a major human commensal. Infect Genet Evol. 2015;33:381–392. doi:10.1016/j.meegid.2014.10.001.

25. Gautam M, Chopra KB, Douglas DD, Stewart RA, Kusne S. *Streptococcus salivarius* Bacteremia and spontaneous bacterial peritonitis in liver transplantation candidates. Liver Transplant. 2007;13:1582–1588. doi:10.1002/lt.21277.

26. Doyuk E, Ormerod OJ, Bowler I. Native valve endocarditis due to *Streptococcus vestibularis* and *Streptococcus oralis*. J Infect. 2002;45:39–41.

27. Cunliffe NA, Jacob AJ. *Streptococcus vestibularis* bacteraemia. J Infect. 1997;34:85. doi:10.1016/S0163-4453(97)80017-5.

28. Sehu MM, Heney C, Chandra S, Bergh H, Nimmo G. *Streptococcus salivarius* meningitis post spinal procedure: diagnosis by 16S and a call to better aseptic practices. Int J Infect Dis. 2014;21:405. doi:10.1016/j.ijid.2014.03.1256.

29. Gupta SS, Sarasam R, Wartak S, Namana V. Splenic and Kidney Infarct: sequelae of Subacute *Streptococcus mitis* Bacterial Endocarditis. J Glob Insect Dis. 2017;9:123–124. doi:10.4103/jgid.jgid_181_16.

30. Byrd VS, Nemeth AS. A case of infective endocarditis and spinal epidural abscess caused by *Streptococcus mitis* bacteremia. Case Rep Infect Dis. 2017 Article ID 7289032. 3.

31. Matthys C, Claeys G, Verschraegen G, Wauters G, Vogelaers D, De Baere T, Verhelst R, Van Nooten G, Van Coster R, Vaneechoutte M. *Streptococcus cristatus* isolated from a resected heart valve and blood cultures: case reports and application of phenotypic and genotypic techniques for identification. Acta Clin Belg. 2006;61:196–200. doi:10.1179/acb.2006.034.

32. Denapaite D, Rieger M, Köndgen S, Brückner R, Ochigava I, Kappeler P, Mätz-Rensing K, Leendertz F, Hakenbeck R. Highly variable *streptococcus oralis* strains are common among viridans streptococci isolated from primates. mSphere. 2016;1:e00041–15. doi:10.1128/mSphere.00041-15.

33. Edgar RC. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. BMC Bioinformatics. 2004;5:113. doi:10.1186/1471-2105-5-113.

34. Knyaz C, Stecher G, Li M, Kumar S, Tamura K. MEGA X: molecular evolutionary genetics analysis across computing platforms. Mol Biol Evol. 2018;35:1547–1549. doi:10.1093/molbev/msy096.

35. He ZL, Zhang HK, Gao SH, Lercher MJ, Chen WH, Hu SN. Evolview v2: an online visualization and management tool for customized and annotated phylogenetic trees. Nucleic Acids Res. 2016;44:W236–241. doi:10.1093/nar/gkw370.

36. Schulte T, Löfling J, Mikaelsson C, Kikhney A, Hentrich K, Diamante A, Ebel C, Normark S, Svergun D, Henriques-Normak B, et al. The basic keratin 10-binding domain of the virulence-associated pneumococcal serine-rich protein PsrP adopts a novel MSCRAMM fold. Open Biol. 2014;4:130090. doi:10.1098/rsob.140120.

37. Garnett JA, Simpson PJ, Taylor J, Benjamin SV, Tagliaferri C, Cota E, Chen Y-YM, Wu H, Matthews S. Structural insight into the role of Streptococcus parasanguinis Fap1 within oral biofilm formation. Biochem Biophys Res Commun. 2012;417:421–426. doi:10.1016/j.bbrc.2011.11.131.

38. Delorme C, Guédon E, Pons N, Cruaud C, Couloux A, Loux V, Chiapello H, Poyart C, Gautier C, Sanchez N, et al. Complete genome sequence of the clinical *Streptococcus salivarius* strain CCHSS3. J Bacteriol. 2011;193:5041–5042. doi:10.1128/JB.05416-11.

39. Geng JN, Huang SC, Li SL, Hu SN, Chen YYM. Complete genome sequence of the ureolytic

Streptococcus salivarius strain 57.I. J Bacteriol. 2011;193:5596–5597. doi:10.1128/JB.05670-11.

40. Peumans WJ, Van Damme EJM, Barre A, Rouge P. Classification of plant lectins in families of structurally and evolutionary related proteins. In: Wu AM, editor. Molecular immunology of complex carbohydrates-2. New York (NY): Kluwer Academic/Plenum Publ; 2001. p. 27–54.

41. Wastfelt M, StalhammarCarlemalm M, Delisse AM, Cabezon T, Lindahl G. Identification of a family of streptococcal surface proteins with extremely repetitive structure. J Biol Chem. 1996;271:18892–18897.

42. Kuroda M, Tanaka Y, Aoki R, Shu D, Tsumoto K, Ohta T. *Staphylococcus aureus* giant protein Ebh is involved in tolerance to transient hyperosmotic pressure. Biochem Biophys Res Commun. 2008;374:237–241. doi:10.1016/j.bbrc.2008.07.037.

43. Chen Q, Sun B, Wu H, Peng Z, Fives-Taylor PM. Differential roles of individual domains in selection of secretion route of a *Streptococcus parasanguinis* serine-rich adhesin, Fap1. J Bacteriol. 2007;189:7610–7617. doi:10.1128/JB.00748-07.

44. Zhu F, Zhang H, Yang T, Haslam SM, Dell A, Wu H. Engineering and dissecting the glycosylation pathway of a streptococcal serine-rich repeat adhesin. J Biol Chem. 2016;291:27354–27363. doi:10.1074/jbc. M116.752998.

45. Jiang Y-L, Jin H, Yang H-B, Zhao R-L, Wang S, Chen Y, Zhou C-Z. Defining the enzymatic pathway for polymorphic *O*-glycosylation of the pneumococcal serine-rich repeat protein PsrP. J Biol Chem. 2017;292:6213–6224. doi:10.1074/jbc.M116.770446.

46. Takamatsu D, Bensing BA, Sullam PM. Four proteins encoded in the gspB-secY2A2 operon of *Streptococcus gordonii* mediate the intracellular glycosylation of the platelet-binding protein GspB. J Bacteriol. 2004;186:7100–7111. doi:10.1128/JB.186.21.7100-7111.2004.

47. Shi -W-W, Jiang Y-L, Zhu F, Yang Y-H, Shao Q-Y, Yang H-B, Ren Y-M, Wu H, Chen Y, Zhou C-Z. Structure of a novel O-linked N-acetyl-D-glucosamine (O-GlcNAc) transferase, GtfA, reveals insights into the glycosylation of pneumococcal serine-rich repeat adhesins. J Biol Chem. 2014;289:20898–20907. doi:10.1074/jbc.M114.581934.

48. Chen Y, Seepersaud R, Bensing BA, Sullam PM, Rapoport TA. Mechanism of a cytosolic O-glycosyltransferase essential for the synthesis of a bacterial adhesion protein. Proc Natl Acad Sci USA. 2016;113:E1190– 1199. doi:10.1073/pnas.1600494113.

49. Zhu F, Zhang H, Wu H. A conserved domain is crucial for acceptor substrate binding in a family of glucosyltransferases. J Bacteriol. 2015;197:510–517. doi:10.1128/JB.02267-14.

50. Chen Y, Bensing BA, Seepersaud R, Mi W, Liao M, Jeffrey PD, Shajahan A, Sonon RN, Azadi P, Sullam PM, et al. Unraveling the sequence of cytosolic

reactions in the export of GspB adhesin from *Streptococcus gordonii*. J Biol Chem. 2018;293:5360–5373. doi:10.1074/jbc.RA117.000963.

51. Zhang H, Zhu F, Yang T, Ding L, Zhou M, Li J, Haslam SM, Dell A, Erlandsen H, Wu H. The highly conserved domain of unknown function 1792 has a distinct glycosyltransferase fold. Nat Commun. 2014;5:4339. doi:10.1038/ncomms5972.

52. Seepersaud R, Sychantha D, Bensing BA, Clarke AJ, Sullam PM. O-acetylation of the serine-rich repeat glycoprotein GspB is coordinated with accessory Sec transport. PLoS Pathog. 2017;13:e1006558. doi:10.1371/journal.ppat.1006558.

53. Seepersaud R, Bensing BA, Yen YT, Sullam PM. The accessory Sec protein Asp2 modulates GlcNAc deposition onto the serine-rich repeat glycoprotein GspB. J Bacteriol. 2012;194:5564–5575. doi:10.1128/JB.01000-12.

54. Varki A, Cummings RD, Aebi M, Packer NH, Seeberger PH, Esko JD, Stanley P, Hart G, Darvill A, Kinoshita T, et al. Symbol nomenclature for graphical representations of glycans. Glycobiology. 2015;25:1323–1324. doi:10.1093/glycob/cwv091.

55. Mistou M-Y, Dramsi S, Brega S, Poyart C, Trieu-Cuot P. Molecular dissection of the secA2 locus of group B *Streptococcus* reveals that glycosylation of the Srr1 LPXTG protein is required for full virulence. J Bacteriol. 2009;191:4195–4206. doi:10.1128/JB.01673-08.

56. Lizcano A, Akula Suresh Babu R, Shenoy AT, Saville AM, Kumar N, D'Mello A, Hinojosa CA, Gilley RP, Segovia J, Mitchell TJ, et al. Transcriptional organization of pneumococcal psrP-secY2A2 and impact of GtfA and GtfB deletion on PsrP-associated virulence properties. Microbes Infect. 2017;19:323–333. doi:10.1016/j.micinf.2017.04.001.

57. Mitchell TJ. The pathogenesis of streptococcal infections: from tooth decay to meningitis. Nat Rev Microbiol. 2003;1:219–230. doi:10.1038/nrmicro771.

58. Tong SYC, Davis JS, Eichenberger E, Holland TL, Fowler VG. Staphylococcus aureus infections: epidemiology, pathophysiology, clinical manifestations, and management. Clin Microbiol Rev. 2015;28:603–661. doi:10.1128/CMR.00134-14.

59. Takamatsu D, Bensing BA, Prakobphol A, Fisher SJ, Sullam PM. Binding of the streptococcal surface glycoproteins GspB and Hsa to human salivary proteins. Infect Immun. 2006;74:1933–1940. doi:10.1128/IAI.74.3.1933-1940.2006.

60. Xiong YQ, Bensing BA, Bayer AS, Chambers HF, Sullam PM. Role of the serine-rich surface glycoprotein GspB of Streptococcus gordonii in the pathogenesis of infective endocarditis. Microb Pathog. 2008;45:297–301. doi:10.1016/j.micpath.2008.06.004.

61. Takamatsu D, Bensing BA, Cheng H, Jarvis GA, Siboo IR, López JA, Griffiss JM, Sullam PM. Binding of the Streptococcus gordonii surface glycoproteins GspB and Hsa to specific carbohydrate structures on platelet

membrane glycoprotein Ibα. Mol Microbiol. 2005;58:380–392. doi:10.1111/j.1365-2958.2005.04830.x.

62. Takahashi Y, Yajima A, Cisar JO, Konishi K. Functional analysis of the Streptococcus gordonii DL1 sialic acid-binding adhesin and its essential role in bacterial binding to platelets. Infect Immun. 2004;72:3876–3882. doi:10.1128/IAI.72.7.3876-3882.2004.

63. Bensing BA, Khedri Z, Deng L, Yu H, Prakobphol A, Fisher SJ, Chen X, Iverson TM, Varki A, Sullam PM. Novel aspects of sialoglycan recognition by the Siglec-like domains of streptococcal SRR glycoproteins. Glycobiology. 2016;26:1222–1234. doi:10.1093/glycob/cww042.

64. Schulte T, Mikaelsson C, Beaussart A, Kikhney A, Deshmukh M, Wolniak S, Pathak A, Ebel C, Löfling J, Fogolari F, et al. The BR domain of PsrP interacts with extracellular DNA to promote bacterial aggregation; structural insights into pneumococcal biofilm formation. Sci Rep. 2016;6:32371. doi:10.1038/srep32371.

65. Yang Y-H, Jiang Y-L, Zhang J, Wang L, Bai X-H, Zhang S-J, Ren Y-M, Li N, Zhang Y-H, Zhang Z, et al. Structural insights into srap-mediated Staphylococcus aureus adhesion to host cells. PLoS Pathog. 2014;10: e1004169. doi:10.1371/journal.ppat.1004169.

66. Siboo IR, Chambers HF, Sullam PM. Role of SraP, a serine-rich surface protein of staphylococcus aureus, in binding to human platelets. Infect Immun. 2005;73:2273–2280. doi:10.1128/IAI.73.4.2273-2280.2005.

67. Bensing BA, Loukachevitch LV, McCulloch KM, Yu H, Vann KR, Wawrzak Z, Anderson S, Chen X, Sullam PM, Iverson TM. Structural basis for sialoglycan binding by the Streptococcus sanguinis SrpA adhesin. J Biol Chem. 2016;291:7230–7240. doi:10.1074/jbc.M115.701425.

68. Seo HS, Minasov G, Seepersaud R, Doran KS, Dubrovska I, Shuvalova L, Anderson WF, Iverson TM, Sullam PM. Characterization of fibrinogen binding by glycoproteins Srr1 and Srr2 of Streptococcus agalactiae. J Biol Chem. 2013;288:35982–35996. doi:10.1074/jbc.M113.513358.

69. Samen U, Reinscheid DJ, Reinscheid DJ, Borges F. The surface protein srr-1 of Streptococcus agalactiae binds human keratin 4 and promotes adherence to epithelial HEp-2 cells. Infect Immun. 2007;75:5405–5414. doi:10.1128/IAI.00717-07.

70. Ramboarina S, Garnett JA, Zhou M, Li Y, Peng Z, Taylor JD, Lee W-C, Bodey A, Murray JW, Alguel Y, et al. Structural insights into serine-rich fimbriae from Gram-positive bacteria. J Biol Chem. 2010;285:32446–32457. doi:10.1074/jbc.M110.128165.

71. Pyburn TM, Bensing BA, Xiong YQ, Melancon BJ, Tomasiak TM, Ward NJ, Yankovskaya V, Oliver KM, Cecchini G, Sulikowski GA, et al. A structural model for binding of the serine-rich repeat adhesin GspB to host carbohydrate receptors. PLoS Pathog. 2011;7: e1002112. doi:10.1371/journal.ppat.1002112.

72. Sundaresan R, Samen U, Ponnuraj K. Expression, purification, crystallization and preliminary X-ray diffraction studies of the human keratin 4-binding domain of serine-rich repeat protein 1 from Streptococcus agalactiae. Acta Crystallogr F-Struct Biol Cryst Commun. 2011;67:1582–1585. doi:10.1107/S1744309111040413.

73. Sundaresan R, Samen U, Ponnuraj K. Structure of KRT4 binding domain of Srr-1 from Streptococcus agalactiae reveals a novel beta-sheet complementation. Int J Biol Macromol. 2015;75:97–105. doi:10.1016/j.ijbiomac.2014.12.048.

74. Sheen TR, Jimenez A, Wang NY, Banerjee A, van Sorge NM, Doran KS. Serine-rich repeat proteins and pili promote Streptococcus agalactiae colonization of the vaginal tract. J Bacteriol. 2011;193:6834–6842. doi:10.1128/JB.00094-11.

75. Wang N-Y, Patras KA, Seo HS, Cavaco CK, Rösler B, Neely MN, Sullam PM, Doran KS. Group B streptococcal serine-rich repeat proteins promote interaction with fibrinogen and vaginal colonization. J Infect Dis. 2014;210:982–991. doi:10.1093/infdis/jiu151.

76. Seo HS, Mu R, Kim BJ, Doran KS, Sullam PM. Binding of glycoprotein srr1 of Streptococcus agalactiae to fibrinogen promotes attachment to brain endothelium and the development of meningitis. PLoS Path. 2012;8: e64204. doi:10.1371/journal.ppat.1002947.

77. Six A, Bellais S, Bouaboud A, Fouet A, Gabriel C, Tazi A, Dramsi S, Trieu-Cuot P, Poyart C. Srr2, a multifaceted adhesin expressed by ST-17 hypervirulent Group B Streptococcus involved in binding to both fibrinogen and plasminogen. Mol Microbiol. 2015;97:1209–1222. doi:10.1111/mmi.13097.

78. van Sorge NM, Quach D, Gurney MA, Sullam PM, Nizet V, Doran KS. The group B streptococcal serine-rich repeat 1 glycoprotein mediates penetration of the blood-brain barrier. J Infect Dis. 2009;199:1479–1487. doi:10.1086/598217.

79. Deng L, Bensing BA, Thamadilok S, Yu H, Lau K, Chen X, Ruhl S, Sullam PM, Varki A, Orihuela CJ. Oral streptococci utilize a Siglec-like domain of serine-rich repeat adhesins to preferentially target platelet sialoglycans in human blood. PLoS Pathog. 2014;10:e1004540. doi:10.1371/journal.ppat.1004540.

80. Kim AR, Ahn B, Kim HY, Seo HS, Yun CH, Han SH. Serine-rich repeat adhesin gordonii surface protein B is important for Streptococcus gordonii biofilm formation. J Endodont. 2016;42:1767–1772. doi:10.1016/j.joen.2016.08.016.

81. Oguchi R, Takahashi Y, Shimazu K, Urano-Tashiro Y, Kawarai T, Konishi K, Karibe H. Contribution of Streptococcus gordonii Hsa adhesin to biofilm formation. Jpn J Infect Dis. 2017;70:399–404. doi:10.7883/yoken.JJID.2017.E001.

82. Takahashi Y, Takashima E, Shimazu K, Yagishita H, Aoba T, Konishi K. Contribution of sialic acid-binding adhesin to pathogenesis of experimental endocarditis

caused by *Streptococcus gordonii* DL1. Infect Immun. 2006;74:740–743. doi:10.1128/IAI.74.1.740-743.2006.

83. Takamatsu D, Bensing BA, Cheng H, Jarvis GA, Siboo IR, López JA, Griffiss JM, Sullam PM. Binding of the Streptococcus gordonii surface glycoproteins GspB and Hsa to specific carbohydrate structures on platelet membrane glycoprotein Ibalpha. Mol Microbiol. 2005;58:380–392. doi:10.1111/j.1365-2958.2005.04830.x.

84. Urano-Tashiro Y, Takahashi Y, Oguchi R, Konishi K. Two arginine residues of *Streptococcus gordonii* sialic acid-binding adhesin Hsa are essential for interaction to host cell receptors. PLoS One. 2016;11:e0154098. doi:10.1371/journal.pone.0154098.

85. Yajima A, Urano-Tashiro Y, Shimazu K, Takashima E, Takahashi Y, Konishi K. Hsa, an adhesin of *Streptococcus gordonii* DL1, binds to alpha 2-3-linked sialic acid on glycophorin A of the erythrocyte membrane. Microbiol

Immunol. 2008;52:69–77. doi:10.1111/j.1348-0421.2008.00015.x.

86. Loukachevitch LV, Bensing BA, Yu H, Zeng J, Chen X, Sullam PM, Iverson TM. Structures of the *Streptococcus sanguinis* SrpA binding region with human sialoglycans suggest features of the physiological ligand. Biochemistry. 2016;55:5927–5937. doi:10.1021/acs.biochem.6b00704.

87. Dawson NL, Lewis TE, Das S, Lees JG, Lee D, Ashford P, Orengo CA, Sillitoe I. CATH: an expanded resource to predict protein function through structure and sequence. Nucleic Acids Res. 2017;45:D289–295. doi:10.1093/nar/gkw1098.

88. Duar RM, Lin XB, Zheng J, Martino ME, Grenier T, Pérez-Muñoz ME, Leulier F, Gänzle M, Walter J. Lifestyles in transition: evolution and natural history of the genus *Lactobacillus*. FEMS Microbiol Rev. 2017;41:S27–48. doi:10.1093/femsre/fux030.